Psicothema

Psicothema 2020, Vol. 32, No. 4, 590-597
doi: 10.7334/psicothema2020.96

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Comparing Methods for Modeling Acquiescence in Multidimensional Partially Balanced Scales

Javier de la Fuente[1] and Francisco J. Abad[2]

[1] University of Texas at Austin and [2] Universidad Autónoma de Madrid

Abstract

Background: The inclusion of direct and reversed items in scales is a commonly-used strategy to control acquiescence bias. However, this is not enough to avoid the distortions produced by this response style in the structure of covariances and means of the scale in question. This simulation study provides evidence on the performance of two different procedures for modelling the influence of acquiescence bias on partially balanced multidimensional scales: a method based on exploratory factor analysis (EFA) with target rotation, and a method based on random intercept factor analysis (RIFA). **Method:** The independent variables analyzed in a simulation study were sample size, number of items per factor, balance of substantive loadings of direct and reversed items, size and heterogeneity of acquiescence loadings, and inter-factor correlation. **Results:** The RIFA method had better performance over most of the conditions, especially for the balanced conditions, although the variance of acquiescence factor loadings had a certain impact. In relation to the EFA method, it was severely affected by a low degree of balance. **Conclusions:** RIFA seems the most robust approach, but EFA also remains a good alternative for medium and fully balanced scales.

Keywords: Acquiescence bias, response style, multidimensional scales, simulation study, factor analysis.

Resumen

Comparación de Métodos Para Modelar la Aquiescencia en Escalas Multidimensionales Parcialmente Balanceadas. **Antecedentes:** la inclusión de ítems directos e inversos en escalas es una estrategia comúnmente utilizada para controlar el sesgo de aquiescencia. No obstante, esto es insuficiente para evitar las distorsiones producidas por este estilo de respuesta en la estructura de covarianzas y medias de la escala. El presente estudio de simulación aporta evidencia sobre el rendimiento de dos procedimientos para controlar la influencia del sesgo de aquiescencia en escalas multidimensionales parcialmente balanceadas: un método basado en análisis factorial exploratorio con rotación target (EFA), y un método basado en el análisis factorial confirmatorio con intercepto aleatorio (RIFA). **Método:** las variables independientes del estudio de simulación fueron: tamaño muestral, número de ítems por factor, balanceo de los pesos sustantivos de los ítems directos e inversos, tamaño y heterogeneidad de los pesos en aquiescencia, y correlación entre factores. **Resultados:** el método RIFA tiene mejor funcionamiento en general, especialmente para las condiciones balanceadas, aunque la varianza de los pesos de aquiescencia tuvo impacto en su rendimiento. El método EFA se ve principalmente afectado en la situación de bajo balanceo. **Conclusiones:** el RIFA parece la aproximación más robusta, aunque el EFA se mantiene como una alternativa a considerar para escalas con balanceo medio o completo.

Palabras clave: sesgo de aquiescencia, estilo de respuesta, escalas multidimensionales, estudio de simulación, análisis factorial.

Acquiescence bias is a response style characterized by a systematic trend of showing a high degree of agreement with items, irrespective of their content, and mainly using the highest response categories (Baumgartner & Steenkamp, 2001; Paulhus, 1991; Van Vaerenbergh & Thomas, 2012). From a psychometric standpoint, acquiescence style is often considered a source of error variance, distorting self-reported measures (Hofstee et al., 1998). Although acquiescence responding has some domain-specific behavior, it also shows some similarities with personality traits (Bentler et al., 1971; DiStefano & Motl, 2006) and, in several studies, consistency across domains (e.g., Danner et al., 2015). In that regard, medium to high correlations have been found between acquiescence factors obtained from personality and attitude scale items (Danner et al., 2015). In terms of stability, test-retest correlations for acquiescence bias of around .6 have been reported for both short (e.g., two months; Danner et al., 2015) and long-term (e.g., four years, Billiet & Davidov, 2008) time intervals.

The inclusion of positively (PK) and negatively (NK) keyed items is the traditional approach for controlling acquiescence bias. PK and NK items have opposite semantic meanings, thus measuring opposite poles of the same theoretical construct. Nonetheless, the inclusion of NK items on a scale is controversial (Suárez-Alvarez et al., 2018), and does not ensure that the means and covariance structure are free from acquiescence bias (Vigil-Colet et al., 2020). In that regard, several studies have demonstrated

Received: March 23, 2020 • Accepted: June 28, 2020
Corresponding author: Javier de la Fuente
University of Texas at Austin
108 E Dean Keeton St
78712 Austin, Texas (Estados Unidos)
e-mail: j.delafuente@utexas.edu

the negative impact of acquiescence bias on the psychometric properties of scales, distorting items' means and variances, as well as covariances across items and other variables (Baumgartner & Steenkamp, 2001; Greenleaf, 1992; Weijters et al., 2010). These distortions produce biased reliability and convergent validity estimations, as well as correlated error terms that reduce data fit to theoretical models (Danner et al., 2015; Plieninger, 2017).

Different statistical methods based on the common factor model have been proposed for controlling acquiescence bias, aiming at explicitly modeling an acquiescence factor controlling that source of systematic variance (Billiet & McClendon, 2000; Ferrando et al., 2003; Lorenzo-Seva & Ferrando, 2009; Maydeu-Olivares & Coffman, 2006). Savalei and Falk (2014) conducted a simulation study to assess the performance of three different procedures in recovering the substantive factor loadings from a unidimensional balanced scale (same number of PK and NK items) in the presence of acquiescence bias. The three methods analyzed were: an exploratory factor analysis (EFA) method (Ferrando et al., 2003), the random intercept factor analysis (RIFA) model (Maydeu-Olivares & Coffman, 2006), and the ipsative method proposed by Chan and Bentler (1993). Results from the simulation indicated that acquiescence bias only had a remarkable impact on factor loadings recovery when its influence was strong. All three models presented an adequate performance when their assumptions were met.

The present study aims at assessing the performance of EFA and RIFA models proposed by Lorenzo-Seva and Ferrando (2009) and Maydeu-Olivares and Coffman (2006), respectively. The performance of both methods has already been compared, but only with unidimensional and totally balanced scales, unusual features in common scales and questionnaires. Moreover, previous simulation studies have been mainly focused on the recovery of items loadings on substantive factors, overshadowing the recovery of acquiescence loadings. Finally, only the confirmatory version of the RIFA model has been compared against the EFA approach, which might be an unfair comparison, since results are not generalizable to situations in which the factor structure is unknown. However, the RIFA can be also used in the exploratory structural equation modelling framework (e.g., Aichholzer, 2014) and this is the approach followed here, which provides a better comparison with the EFA method.

The present study aims at providing evidence on the performance of both procedures in a multidimensional context, focusing on the quality of the recovery of 1) substantive factor loadings, 2) acquiescence factor loadings, and 3) the correlations between substantive factors.

In order to assess the performance of both methods, a Monte Carlo simulation study was conducted, manipulating a wide array of variables directly related to the assumptions of both models.

Regarding the EFA method, it is expected a better parameter recovery when PK and NK have similar loadings in the substantive factors. In relation to the RIFA procedure, it is mainly expected a negative impact of heterogeneity of the acquiescence factor loadings, considering this model assumes tau-equivalence in this factor (Maydeu-Olivares & Coffman, 2006).

Method

Procedure

The simulated factorial model for assessing the performance of the EFA and RIFA methods is displayed in Figure 1 (for the 10 items condition). This model comprises two substantive factors (F1 and F2), and an acquiescence factor AQ (see Figure 1). The proportion of PK was maintained equal (i.e., 66%) in all conditions. Loadings of the PK items on the substantive factors was fixed to .70, for all the items in all the conditions.

A factorial design was used, manipulating the following variables: 1) sample size (N = 200, 500, or 1000), 2) number of items per factor (Jf = 5, 8, or 12), 3) average of the PK and NK item loadings on the acquiescence factor (AQL = .15 or .3), 4) heterogeneity between PK and NK loadings on the AQ factor (AQH = Null, Medium, or High, corresponding to PK and NK loadings differences of 0, .1, and .2, respectively), 5) Balance of PK and NK loadings on the substantive factors (Balance = Low, Medium, or Full, corresponding to loadings for the NK items of .5, .6 and .7, respectively), and 6) correlation between substantive factors (Phi = 0, .3, or .5). Note that the item AQ loadings on each condition depended on the levels on the AQL and AQH independent variables. For instance, for AQL = .30 and AQH = High, AQ loadings were fixed to .40 for the PK items and to .20 for the NK items.

Thus, the simulation study comprised a total of 486 experimental conditions (3 × 3 × 2 × 3 × 3 × 3). All conditions were chosen according to their impact on the models' assumptions, as well as the degree of realism in applied contexts. For each condition, 20 samples matrices of standardized continuous variables were simulated following the common factor model.

Two approaches for modeling acquiescence were considered. In both cases, datasets are analyzed without recoding the reversed items. In the EFA procedure, a set of items balanced in direction is first selected (i.e., if there was *K* NK items, the first *K* PK items were considered). Then, the loading on acquiescence for a standardized item in the balanced subset is approached as (Lorenzo et al., 2009):

$$\hat{\lambda}_{jaq} = \left[\frac{\sum_{g=1}^{nb} r_{jg}}{\sqrt{\sum_{g=1}^{nb} \sum_{h=1}^{nb} r_{gh}}} \right], \quad (1)$$

where nb is the number of items in the balanced subtest, and r_{jg} is the corresponding element of the item correlation matrix. Then, three factors are extracted and rotated to a partially specified orthogonal target matrix, taking the approached loadings as target values. From that rotation, a transformation matrix **T** is obtained,

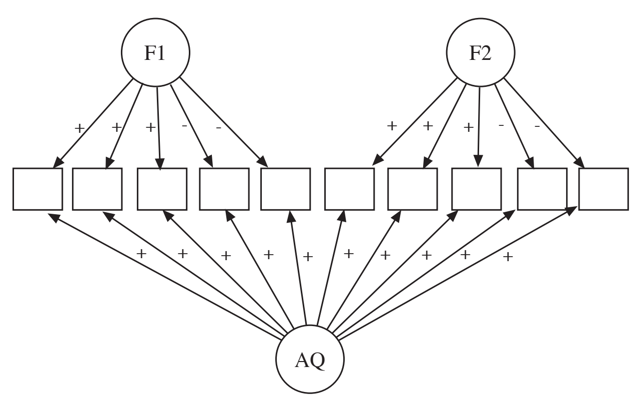

Figure 1. Multidimensional model for generating simulated responses. F1 = Substantive Factor 1; F2 = Substantive Factor 2, AQ = Acquiescence Factor

and used to rotate the solution for the complete scale. Finally, the part of the pattern matrix corresponding with the substantive factors is rotated with partially specified Target rotation, according to the theoretical model, for simplifying the interpretation.

The approximation $\hat{\lambda}_{jaq}$ works best when the number of items increases, the loadings on acquiescence factor are high, and the residual variances are low. Note also that this approach is based on the assumption that, for the balanced set, the mean loadings of PK and NK on the substantive factors are equal. In case this condition is not met, the approximation is expected to work worse for oblique substantive factor structures. For instance, for a two-substantive factor simple structure, if assumptions are violated, the equation for a standardized item *j* measuring the first factor can be expanded as:

$$\left[\frac{\sum_{g=1}^{nb} r_{jg}}{\sqrt{\sum_{g=1}^{nb}\sum_{h=1}^{nb} r_{gh}}}\right] = \lambda_{jaq}\left[\frac{\left(\sum_{g=1}^{nb}\lambda_{gaq}\right)}{\sqrt{\mathrm{var}(X)}}\right] + \frac{\lambda_{j1}\left(\sum_{g=1}^{nb}\lambda_{g1}\right)}{\sqrt{\mathrm{var}(X)}} + \left[\frac{\lambda_{j1}\left(\sum_{g=1}^{nb}\lambda_{g2}\right)\phi 12}{\sqrt{\mathrm{var}(X)}}\right] + \left[\frac{\mathrm{var}(e_j)}{\sqrt{\mathrm{var}(X)}}\right],$$

Where $(\sum_{g=1}^{nb} \lambda_{g1})$ and $(\sum_{g=1}^{nb} \lambda_{g2})$ are the sum of loadings on the first and second substantive factors, ϕ_{12} is the correlation between the substantive factors, $var(e_j)$ is the item uniqueness, and $var(X)$ is the variance of the balanced test score (i.e., the sum of the standardized item scores). Thus, the equation 1 approach is based in the assumption of $(\sum_{g=1}^{nb} \lambda_{g1})$, $(\sum_{g=1}^{nb} \lambda_{g2})$, ϕ_{12}, and $var(e_j)$ being close to zero, and the ratio $\left[\frac{\left(\sum_{g=1}^{nb}\lambda_{gaq}\right)}{\sqrt{\mathrm{var}(X)}}\right]$ being close to 1.

In the RIFA model, an ESEM with two factors and a random intercept is considered. The variance of the random intercept is set free to be estimated, with non-null values reflecting the presence of systematic variance associated with acquiescence or individual differences in response style. Including a random intercept is mathematically equivalent to adding an additional orthogonal acquiescence factor in which all the loadings on the factor are constrained to be equal and positive. This approach was the followed here, in order to compare the EFA and the RIFA models. Note that for applying the ESEM version of the RIFA model, we followed a two-step procedure. First, a SEM two factor-model, with the random intercept, was estimated. For the two-factors, an unrestricted factor model was specified in which the minimum constraints for identification were imposed (i.e., factor variances were set to 1, factor covariances were constrained to be zero, and factor loadings were constrained to follow an echelon pattern; see Rosseel, 2020). Second, the part of the pattern matrix corresponding with the substantive factors was rotated with partially specified orthogonal Target rotation, as in the EFA method.

The RIFA approximation is expected to work best when the loadings on acquiescence are homogeneous. Furthermore, in the case of unbalanced scales some confounding between the substantive and acquiescence factors can occur (e.g., for the extreme case in which all the items are direct, the acquiescence and content factors should not be separable). Thus, the study of the RIFA model in the unbalanced case deserves attention.

Data analysis

The mean bias error for the loadings (MBE) and the root mean squared error (RMSE) were used to assess the accuracy on the recovery of parameters (i.e., inter-factor correlation and factor loadings). Note that for computing MBE the sign of estimated and population substantive loadings of reversed items was changed (i.e., a positive bias always indicate average overestimation of the loading size). Additionaly, the Tucker congruency coefficients were also obtained (Tucker, 1951), as a measure of similarity between the population factors and its corresponding estimation. The limits of the congruency coefficient range from -1 to 1, where values between .85 and .94 indicate an acceptable similarity between factors, and values over .95 represent high levels of similarity between factors factores (Lorenzo-Seva & ten Berge, 2006). For all solutions, factors were aligned with their corresponding population factor-structure following the least-squared criterion using the *faAlign* function (Waller, 2019). Finally, for each model, the Ten Berge estimated factor scores (ten Berge et al., 1999) were computed, and correlations with true factor scores were obtained.

Additionally, a series of univariate analysis of variance (ANOVAs) were conducted in order to quantify the effect sizes, using partial omega squared (η^2_p) as effect size measure. Cohen's guidelines were used for interpreting it, with .01, .06 and .14 representing small, medium, and large effect sizes, respectively (Cohen, 1988). For all ANOVAs, RMSE was used as dependent variable, since the sign of mean bias could vary across conditions, thus challenging its interpretation.

All factor analysis included in the simulation were implemented in the R environment. For the EFA, maximum likelihood estimation was used using the R Package *Psych* (Revelle, 2019). For applying rotation, the *GPArotation* package was used (Bernaards & Jennrich, 2005). Finally, the RIFA was implemented using the R package *lavaan* (Rosseel, 2020).

Results

Results contained in tables 1 and 2 indicate the RIFA method presented a better performance and less variability across conditions, for both the recovery of the factor loadings and the correlation between the substantive factors, obtaining a mean congruency coefficient over .95 in 83% conditions (mean c.c. = .978). In 78.8% of conditions, the RIFA method provided a higher congruency coefficient than the EFA method. On the other hand, the EFA method obtained an appropriate mean congruency coefficient (c.c. > .95) in 62.8% of the conditions (mean c.c. = .948).

Recovery of loadings on the substantive factors

As shown in Table 3, sample size and balance presented a strong positive in the estimation of the substantive factor loadings on both the EFA (η^2_p[N] = .417; η^2_p [Balance] = .661) and RIFA (η^2_p[N] = .612; η^2_p[Balance] = .217) procedures, although the negative effect of low balance was larger in the EFA procedure (e.g., RMSE[EFA: low balance] = .077 > RMSE[RIFA: low balance] = .047). On the other hand, the RIFA method was negatively affected by the acquiescence factor loadings heterogeneity (η^2_p[AQH] = .239).

Regarding the average bias for the EFA method, it was slightly negative across all conditions, producing an average factor loadings underestimation of -.011, with larger underestimations in the most adverse conditions (e.g., mean bias [EFA; low balance] = -.026). For the RIFA method, the mean bias was very small, always below .004 in absolute value.

Recovery of items loadings on the acquiescence factor

Both EFA and RIFA methods were negatively affected by the heterogeneity of the items loadings on the acquiescence factor, and, although it presented a considerable effect on the EFA method (η^2_p[AQH] = .177), its effect was larger on the RIFA method (η^2_p[AQH] = .954), as expected. In fact, the lowest congruency coefficient for the RIFA method was found in conditions of high heterogeneity [AQH = High] = .951. Additionally, the EFA was more affected by the reduction of sample size (η^2_p[EFA; N] = .334), the higher inter-factor correlation (η^2_p[EFA; Phi] = .173), the lower mean loading on the acquiescence factor (η^2_p[EFA; AQL] = .558), and, specially, the low balance of PK and NK substantive loadings (η^2_p[EFA; balance] = .831). Unexpectedly, the number of items did not present a large effect size on the EFA method (η^2_p[Jf] = .004). In the EFA method, two large two-way interactions were found, both involving balance (η^2_p[Balance x Phi] = .147; η^2_p[AQL x Balance] = .166). In general the negative effect of having a non-full balanced set was higher the larger the inter-factor correlation, and the lower the acquiescence loading. To sum up, the RIFA method was more precise and consistent across all conditions, especially in conditions of low balance (RMSE[EFA; balance = low] =. 174 > RMSE[RIFA; balance = low] = .062).

Table 1

	EFA							
	MBE			RMSE			C.C.	Cor.
Variable level	Subs.	AQ	Phi	Subs.	AQ	Phi		
N								
200	-.013	.029	-.039	.066	.131	.073	.931	.904
500	-.010	.031	-.032	.047	.105	.053	**<u>.952</u>**	.909
1000	-.010	.031	-.031	.038	.094	.045	**<u>.960</u>**	.911
Jf								
5	-.007	.025	-.026	.053	.112	.054	.945	.880
8	-.011	.029	-.033	.049	.108	.056	.948	.913
12	-.015	.037	-.042	.049	.110	.062	**<u>.950</u>**	.931
AQL								
0.15	-.012	.036	-.036	.052	.134	.059	.913	.899
0.3	-.010	.025	-.032	.048	.086	.055	**<u>.983</u>**	.917
AQH								
Null	-.014	.025	-.042	.050	.098	.063	**<u>.966</u>**	.906
Medium	-.012	.030	-.036	.050	.109	.058	.949	.909
High	-.007	.036	-.024	.051	.123	.051	.929	.909
Balance								.
Full	-.001	.004	-.003	.032	.059	.039	**<u>.977</u>**	940
Medium	-.006	.021	-.014	.042	.096	.042	**<u>.953</u>**	.921
Low	-.026	.065	-.084	.077	.174	.090	.913	.862
Phi								
0	-.006	.024	-.025	.046	.097	.055	**<u>.954</u>**	.914
0.3	-.011	.031	-.037	.049	.112	.060	.947	.907
0.5	-.016	.036	-.040	.056	.121	.056	.942	.903
Total	-.011	.030	-.034	.050	.110	.057	.948	.908

Note: MBE = Mean Bias Error; RMSE = Root of Mean Squared Error; N = sample size; Jf = number of indicators per factor; AQL = average loading on the acquiescence factor; AQH = heterogeneity between PK and NK loadings on the AQ factor; Balance = Balance of PK and NK loadings on the substantive factors in absolute value; Subs = Substantive factors; AQ = Acquiescence factors; Phi = correlation between substantive factors; C.C. = congruence coefficients; C.C. ≥ .95 appears in bold and underlined. Cor. = average correlation between true and estimated substantive factor score

Table 2

	RIFA							
	MBE			RMSE			C.C.	Cor.
Variable level	Subs.	AQ	Phi	Subs.	AQ	Phi		
N								
200	-.004	.015	-.016	.056	.061	.059	**<u>.975</u>**	.921
500	-.002	.016	-.009	.038	.057	.039	**<u>.979</u>**	.922
1000	-.002	.017	-.009	.030	.055	.029	**<u>.980</u>**	.923
Jf								
5	-.002	.012	-.010	.045	.059	.045	**<u>.978</u>**	.893
8	-.003	.015	-.011	.041	.058	.041	**<u>.978</u>**	.927
12	-.004	.020	-.014	.039	.057	.041	**<u>.978</u>**	.947
AQL								
0.15	-.001	.016	-.008	.039	.058	.041	**<u>.967</u>**	.920
0.3	-.002	.015	-.015	.044	.058	.043	**<u>.988</u>**	.924
AQH								
Null	-.001	.001	-.005	.036	.016	.041	**<u>.997</u>**	.925
Medium	-.004	.017	-.015	.040	.054	.043	**<u>.985</u>**	.924
High	-.003	.032	-.015	.048	.104	.044	**<u>.951</u>**	.918
Balance								
Full	-.002	.011	-.006	.036	.055	.040	**<u>.979</u>**	.939
Medium	-.003	.014	-.009	.040	.057	.041	**<u>.978</u>**	.925
Low	-.004	.023	-.020	.047	.062	.047	**<u>.977</u>**	.902
Phi								
0	-.002	.015	-.007	.040	.058	.046	**<u>.978</u>**	.922
0.3	-.003	.016	-.013	.040	.058	.043	**<u>.978</u>**	.922
0.5	-.004	.016	-.015	.045	.058	.038	**<u>.977</u>**	.922
Total	-.003	.016	-.012	.041	.058	.042	**<u>.978</u>**	.922

Note: MBE = Mean Bias Error; RMSE = Root of Mean Squared Error; N = sample size; Jf = number of indicators per factor; AQL = average loading on the acquiescence factor; AQH = heterogeneity between PK and NK loadings on the AQ factor; Balance = Balance of PK and NK loadings on the substantive factors in absolute value; Subs = Substantive factors; AQ = Acquiescence factors; Phi = correlation between substantive factors; C.C. = congruence coefficients; C.C. ≥ .95 appears in bold and underlined; Cor. = average correlation between true and estimated substantive factor score

Table 3
Size Effects (η^2_p) of the Univariate Analysis of variance (ANOVAs) of RMSE

	EFA			RIFA		
Effect type variables	**Subs.**	**AQ**	**Phi**	**Subs.**	**AQ**	**Phi**
Main effects						
N	**<u>.417</u>**	**<u>.334</u>**	.079	**<u>.612</u>**	.096	.124
Jf	.026	.004	.008	.067	.014	.002
AQL	.019	**<u>.558</u>**	.003	.064	.000	.001
AQH	.000	**<u>.177</u>**	.014	**<u>.239</u>**	**<u>.954</u>**	.001
Balance	**<u>.661</u>**	**<u>.831</u>**	**<u>.253</u>**	**<u>.217</u>**	.101	.008
Phi	.085	**<u>.173</u>**	.002	.065	.000	.010
Two-Way Interactions						
AQL x Balance	.004	**<u>.166</u>**	.004	.002	.004	.001
Balance x Phi	.011	**<u>.147</u>**	.017	.000	.000	.000

Note: N = sample size; Jf = number of indicators per factor; AQL = average loading on the acquiescence factor; AQH = heterogeneity between PK and NK loadings on the AQ factor; Balance = Balance of PK and NK loadings on the substantive factors in absolute value; Subs = Substantive factors; AQ = Acquiescence factors; Phi = correlation between substantive factors; Large effects appear underlined in bold. Only interactions with a large effect (i.e., $\eta^2_p \geq .14$) in some of the methods are shown

In terms of mean bias, it was positive in practically all conditions and methods, but larger for the EFA method (mean bias [EFA] = .030; mean bias[RIFA] = .016). The largest overestimation was shown in conditions of low balance in the EFA method (mean bias [EFA; balance =low] = .065).

Recovery of the correlations between substantive factors

The performance of the EFA procedure was affected by the lack of balance (η^2_p[Balance = .253). In general, the RIFA procedure was also more robust and consistent, showing lower RMSE in all conditions, but especially in low balance conditions (RMSE [EFA; Balance = low] = .090 > RMSE [RIFA; Balance = low] = .047). In relation to bias, it was negative for both the methods, although was smaller for the RIFA.

Recovery of factor scores

The correlation between true and estimated factor scores were higher in average for the RIFA method (Cor. [RIFA] = .922 > Cor. [EFA] = .908). The largest difference was found in the condition of Low Balance (Cor. [RIFA; balance = Low] = .902 > Cor. [EFA; balance = Low] = .862).

Discussion

The present simulation study offers, for the first time, evidence on the efficacy of two procedures based on the common factor model to control acquiescence bias on partially balanced multidimensional scales. For this, a set of key variables related to the assumptions of each of the methods have been manipulated, such as the balancing of loadings in each substantive factor of the direct and reverse items, the size and variance of the loadings on the acquiescence factor, and the number of items per factor. In addition, the magnitude of the correlation between the substantive factors has been manipulated, since its impact on the effectiveness of both procedures is currently unknown. The study focuses on the quality of the recovery of the different simulated factor structures, considering the loadings on acquiescence and both substantive factors, as the correlation between the latter.

As stated in the introduction, the inclusion of reverse items (i.e., negatively keyed items) for the control of acquiescence has been one traditional approach, although it has shown to be controversial and led to psychometric problems (Suárez et al., 2018; Vigil et al., 2020), revealing the presence of unintended systematic variance, and the emergence of method factors associated with the direction, direct or reversed, of the items (DiStefano, & Motl, 2006; Weijters & Baumgartner, 2012; Weijters et al., 2013). Several approaches have been considered for the statistical control of these artifacts, including a method factor for the subset of direct or inverse items (Tomás & Oliver, 1999), correlated uniqueness within the items of the same direction (Marsh, 1996), or, the approaches described here, the inclusion of a general method factor affecting all items (Billiet & McClendon, 2000). Here we take into account the specific EFA approach proposed by Lorenzo-Seva and Ferrando (2009) and the RIFA approach of Maydeu-Olivares and Coffman (2006).

Regarding the recovery of parameters, the simulation results indicate that the performance of the EFA procedure was mainly positively affected by the degree of balancing between direct and reverse items, and the average loading of the items on the acquiescence factor. These effects are consistent with the hypotheses of the study, since the degree of balance and the average acquiescence loading of the items are directly related to the assumptions of the EFA method (Lorenzo-Seva et al., 2003). The impact of balance and the absence of effects of the number of indicators in the recovery of substantive factors loadings is also consistent with the results of the study by Savalei and Falk (2014), although they analyzed a more restricted set of conditions.

In relation to the RIFA method, it presented a superior performance in the parameter recovery across most of the conditions, especially in those with a low degree of balancing. In line with the results of the simulation study by Savalei and Falk (2014), this method was robust to the violation of the assumption of tau-equivalence in the acquiescence loadings. In spite of this, the estimation was negatively affected by the heterogeneity, and, for recovering the substantive loadings, also by the degree of unbalance. This last effect was not detected in the Savalei and Falk (2013), which can be due to their limited set of conditions. The difficulty of the method may be due to the lower average substantive loadings in the low balance condition. The simulated situation is plausible and realistic, since inverse items tend to have lower loadings in the substantive factors (Weijters & Baumgartner, 2012) and higher in the acquiescence factor, since they are usually longer and more complex syntactically (Condon et al., 2006). Finally, the estimation of the correlation between the substantive factors was only affected by the sample size, showing a lower accuracy with smaller sample sizes.

The evidence provided by the present study indicates that the RIFA procedure has a number of advantages over the EFA method for controlling acquiescence bias on partially balanced scales. First, the RIFA method is easier to implement, requiring only to add to the model an additional orthogonal factor in which all items (without recoding) have weights equal to 1. Therefore, this procedure allows for controlling the acquiescence bias by adding a single parameter, also allowing to test the hypothesis of variance nullity associated with the response style. Secondly, the RIFA method is robust against the violation of the assumption of tau-equivalence in the acquiescence factor loadings. This robustness, together with the precision of the estimates, make this method a precise tool to control acquiescence in confirmatory analyzes of the latent structure of questionnaires and scales. However, the results of the present study indicate that the effectiveness in loading recovery can be compromised when there are high heterogeneous acquiescence loadings. Figure 2 shows the average performance of both EFA and RIFA methods, depending on the balance of substantive loadings and the heterogeneity of acquiescence loadings. Although RIFA is generally more robust to the violation of its assumptions, it can be shown that EFA shows a reasonable performance for Medium and Full balance conditions and, in conditions with High heterogeneity of acquiescence loadings, the EFA overcomes the RIFA. The same trend is found for the recovery of factor scores when full balance is achieved (see Figure 3). Note that Figure 3 shows that estimated factor scores are usually better correlated with true factor scores when acquiescence is modelled except for the presence of a strong violation of the assumptions of each technique (i.e., Low balance for the EFA, and High heterogeneity for the RIFA).

Therefore, despite the robustness exhibited by the RIFA method, the EFA procedure remains as an acceptable alternative when the scale is balanced, and the interest of the study focuses on the degree in which each item elicits acquiescence. Furthermore, it is expected

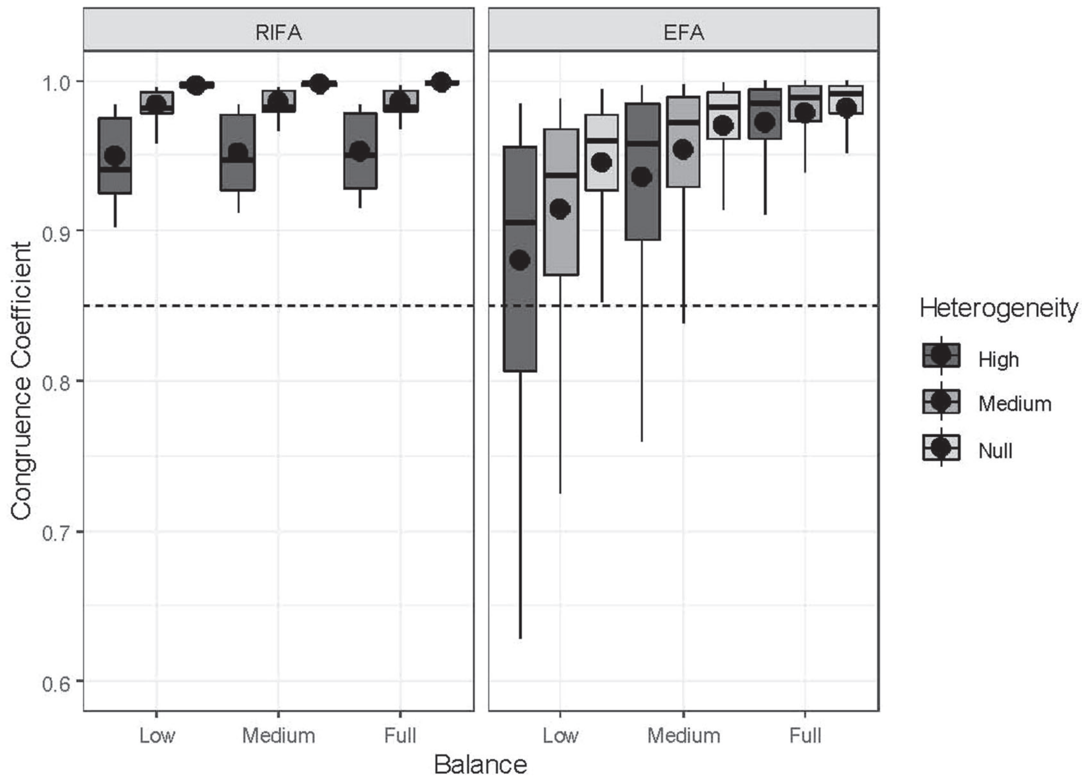

Figure 2. Average congruence coefficient as function of balance, heterogeneity, and method

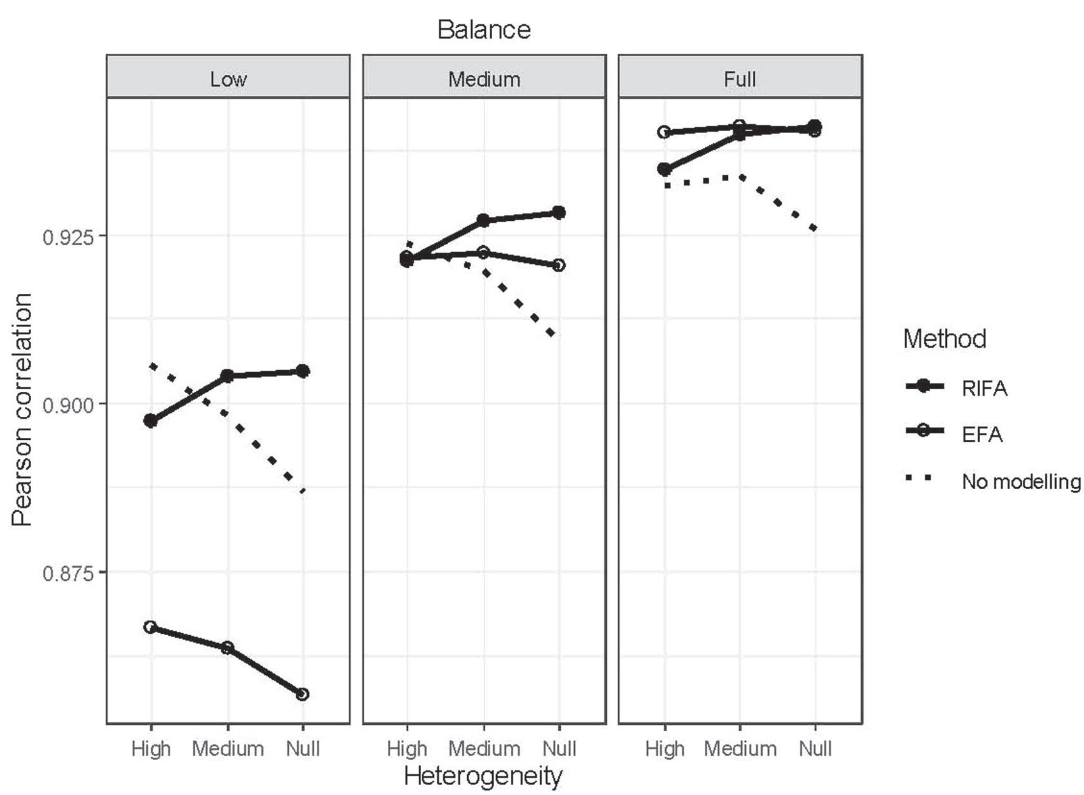

Figure 3. Average correlation between true and estimated factor scores as function of balance, heterogeneity, and method (No modelling - EFA model without acquiescence – is added as baseline)

that a larger unbalance of the number of direct and reverse items might have more detrimental effects on the RIFA model, than on the EFA. Here we have focused in a fixed percentage of direct items (i.e., 66%).

Finally, it is important to highlight that there are other more recent statistical models for the control of acquiescence bias than those analyzed in this study. For example, the procedure proposed by Ferrando et al. (2016) allows controlling the acquiescence bias even when it correlates with the substantive factors. Also, Ferrando et al. (2009) have proposed a different procedure that can be used for correcting simultaneously for acquiescence and social desirability. As future directions, it is considered relevant to carry out new simulation studies in which the performance of other newer procedures is evaluated.

Acknowledgements

This research was partially supported by Ministerio de Ciencia, Innovación y Universidades, Spain (Grant PSI2017-85022-P), European Social Fund, and Cátedra de Modelos y Aplicaciones Psicométricas (Instituto de Ingeniería del Conocimiento and Autonomous University of Madrid).

References

Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality, 53*, 1-4. https://doi.org/10.1016/j.jrp.2014.07.001

Baumgartner, H., & Steenkamp, J. E. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research, 38*(2), 143-156. https://doi.org/10.1509/jmkr.38.2.143.18840

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin, 76*(3), 186-204. https://doi.org/10.1037/h0031474

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

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*(5), 676-696. https://doi.org/10.1177/0013164404272507

Billiet, J. B., & Davidov, E. (2008). Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design. *Sociological Methods & Research, 36*(4), 542-562. https://doi.org/10.1177/0049124107313901

Billiet, J. B., & Mcclendon, M. J. (2000). Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items. *Structural Equation Modeling: A Multidisciplinary Journal, 7*(4), 608-628. https://doi.org/10.1207/S15328007SEM0704_5

Chan, W., & Bentler, P. M. (1993). The Covariance Structure Analysis of Ipsative Data. *Sociological Methods & Research, 22*(2), 214-247. https://doi.org/10.1177/0049124193022002003

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.

Condon, L., Ferrando, P. J., & Demestre, J. (2006). A note on some item characteristics related to acquiescent responding. *Personality and Individual Differences, 40*(3), 403-407. https://doi.org/10.1016/j.paid.2005.07.019

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

Distefano, C., & Motl, R. W. (2006). Further Investigating Method Effects Associated With Negatively Worded Items on Self-Report Surveys. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(3), 440-464. https://doi.org/10.1207/s15328007sem1303_6

Ferrando, P. J., Lorenzo-Seva, U., & Chico, E. (2003). Unrestricted Factor Analytic Procedures for Assessing Acquiescent Responding in Balanced, Theoretically Unidimensional Personality Scales. *Multivariate Behavioral Research, 38*(3), 353-374. https://doi.org/10.1207/S15327906MBR3803_04

Ferrando, P. J., Lorenzo-Seva, U., & Chico, E. (2009). A general factor-analytic procedure for assessing response bias in questionnaire measures. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(2), 364-381. https://doi.org/10.1080/10705510902751374

Ferrando, P. J., Morales-Vives, F., & Lorenzo-Seva, U. (2016). Assessing and Controlling Acquiescent Responding When Acquiescence and Content Are Related: A Comprehensive Factor-Analytic Approach. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(5), 713-725. https://doi.org/10.1080/10705511.2016.1185723

Greenleaf, E. A. (1992). Measuring Extreme Response Style. *Public Opinion Quarterly, 56*(3), 328. https://doi.org/10.1086/269326

Hofstee, W.K.B., Ten Berge, J.M.F., & Hendriks, A.A. (1998). How to score questionnaires. *Personality and Individual Differences, 25*(5), 897-909. https://doi.org/10.1016/S0191-8869(98)00086-5

Lorenzo-Seva, U., & Ten Berge, J. M.F. (2006). Tucker's Congruence Coefficient as a Meaningful Index of Factor Similarity. *Methodology, 2*(2), 57-64. https://doi.org/10.1027/1614-2241.2.2.57

Lorenzo-Seva, U., & Ferrando, P. J. (2009). Acquiescent responding in partially balanced multidimensional scales. *British Journal of Mathematical and Statistical Psychology, 62*(2), 319-326. https://doi.org/10.1348/000711007X265164

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*(4), 810-819. https://doi.org/10.1037//0022-3514.70.4.810

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods, 11*(4), 344-362. https://doi.org/10.1037/1082-989X.11.4.344

Paulhus, D. L. (1991). Measurement and Control of Response Bias. *Measures of Personality and Social Psychological Attitudes,* 17-59. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Plieninger, H. (2017). Mountain or Molehill? A Simulation Study on the Impact of Response Styles. *Educational and Psychological Measurement*, *77*(1), 32-53. https://doi.org/10.1177/0013164416636655

Revelle, W. (2019). *psych: Procedures for psychological, psychometric, and personality research* (Version 1.9.12) [Computer software]. Northwestern University, Evanston, Illinois, R package. https://CRAN.Rproject.org/package=psych

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Suárez-Alvarez, J., Pedrosa, I., Lozano Fernández, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema, 30*(2), 149-158. https://doi.org/10.7334/psicothema2018.33

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. https://doi.org/10.1080/00273171.2014.931800

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies*. Educational Testing Service.

Van Vaerenbergh, Y., & Thomas, T. D. (2012). Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and

Remedies. *International Journal of Public Opinion Research, 25*(2), 195-217. https://doi.org/10.1093/ijpor/eds021

Vigil-Colet, A., Navarro-González, D., & Morales-Vives, F. (2020). To reverse or to not reverse Likert-type items: That is the question. *Psicothema, 32*(1), 108-114. https//doi.org/10.7334/psicothema2019.286

Waller, N.G. (2019). *Fungible: Psychometric functions from the Waller lab* (Version 1.86) [Computer software]. http://cran.r-project.org/package=fungible

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods, 15*(1), 96-110. https://doi.org/10.1037/a0018721

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236-247. https://doi.org/10.1016/j.ijresmar.2010.02.004

Weijters, B., & Baumgartner, H. (2012). Misresponse to Reversed and Negated Items in Surveys: A Review. *Journal of Marketing Research, 49*(5), 737-747. https://doi.org/10.1509/jmr.11.0368

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods, 18*(3), 320-334. https://doi.org/10.1037/a0032121