

Article

The Impact and Consequences of Correcting for Acquiescence When Correlated Residuals are Present

Ana Hernández-Dorado^{id}, Pere Joan Ferrando^{id} and Andreu Vigil-Colet^{id}

Universitat Rovira i Virgili (Spain)

ARTICLE INFO

Received: January 9, 2024

Accepted: June 25, 2024

Keywords:

Acquiescence
Correlated residuals
Response biases

Palabras clave:

Aquiescencia
Residuales correlacionados
Sesgos de respuesta

ABSTRACT

Background: In spite of the interest generated by controlling variance unrelated to the content in personality measures, few studies have assessed the combined impact of acquiescence (ACQ) and correlated residuals in the factor-analytic structural solutions that serve as a basis for item calibration. **Method:** This article compares three control procedures in databases that simultaneously exhibit acquiescence (ACQ) and correlated residuals: the SIREN method (Navarro-Gonzalez, et al., 2024; ACQ control), the MORGANA method (Ferrando et al., 2022; 2023; correlated residuals control), and a combined double control method. A ‘control’ procedure was also examined in which the presence of both determinants was ignored. **Results:** The findings reveal significant differences between the three control methods, with the ACQ control method and the combined method proving more effective. Moreover, when the residual variance shifts to the factor loadings, it appears to be captured by the ACQ correction method. **Conclusions:** The findings suggest using a mixed correction procedure in order to obtain less biased item parameter estimates when both sources of unwanted variance are operating.

El Impacto y las Consecuencias de Corregir por Aquiescencia Cuando hay Residuales Correlacionados

RESUMEN

Antecedentes: A pesar del interés que suscita el control de la varianza no relacionada con el contenido en medidas de personalidad, pocos estudios han evaluado los impactos combinados de la aquiescencia (ACQ) y los residuales correlacionados en la estimación de las estructuras factoriales que sirven de base para la calibración de los ítems. **Método:** Este artículo compara tres procedimientos de control en bases de datos que presentan simultáneamente aquiescencia y residuales correlacionados: método SIREN (Navarro-Gonzalez, et al., 2024; control de ACQ), método MORGANA (Ferrando et al., 2022; 2023; control de residuales correlacionados) y un método doble de control combinado. También se examinó un procedimiento ‘control’ en el que se ignoraba la presencia los dos determinantes. **Resultados:** Los resultados muestran diferencias significativas entre los tres métodos de control, siendo más eficientes el método de control de ACQ y el combinado. Además, cuando la varianza residual se desplaza a las cargas factoriales, esta parece ser captada por el método de corrección de ACQ. **Conclusiones:** Los hallazgos sugieren utilizar un procedimiento de corrección mixto para obtener estimaciones menos sesgadas de los parámetros de los ítems cuando ambas fuentes de varianza no deseada están presentes.

Cite as: Hernández-Dorado, A., Ferrando, P. J., & Vigil-Colet, A. (2025). The impact and consequences of correcting for acquiescence when correlated residuals are present.

Psicothema, 37(1), 11-20. <https://doi.org/10.70478/psicothema.2025.37.02>

Corresponding author: Ana Hernández-Dorado, ahernandezd@professor.universidadviiu.com

This article is published under Creative Commons License 4.0 CC-BY-NC

The measurement of personality variables via questionnaires presents a series of problems that have been extensively debated in the literature since the 1930's. The general issue is that the response to a personality item can potentially be impacted by a series of determinants other than the 'content' intended to be measured (Cronbach, 1946). From a factor analytic (FA) view, when this is the case, two problems can be expected to occur. First, a clear structure cannot be attained. Second, individual factor scores estimate derived from the structural solution cannot be univocally interpreted (Messick, 1995).

While there is an enormous amount of literature about the unwanted, non-content response determinants to personality items (e.g. Arias et al., 2024; Vigil-Colet et al., 2020), studies focused on the joint impact of different sources are much scarce (e.g. Ferrando & Anguiano-Carrasco 2010). This is somewhat surprising if we consider that, when responding to a personality item, several types of unwanted determinants are probably jointly operating, and that the overall impact can possibly reflect complex relations among them. Furthermore, research on the topic and derived procedures (e.g. Ferrando et al., 2003; Navarro-Gonzalez et al., 2016) has considered the same general type of variables (mainly, the joint occurrence of several response styles: acquiescence, social desirability or extreme response).

The present paper also focuses on the joint impact of two non-content response determinants. However, in contrast to previous research, they are now from two different types: Response biases on the one hand, and Method effects on the other. Specifically, we shall focus on acquiescence as the determinant in the first group, and local dependence (redundancy, correlated residuals) in the second.

Acquiescence

Acquiescence (ACQ) is one of the most studied response biases, either as the only distorting determinant (Bentler et al., 1971) or accompanied by others, mainly social desirability (SD) (Ferrando et al., 2009; Navarro-Gonzalez et al., 2016). It is generally viewed as a respondent-dependent response style, and, with regards to its impact, when operating tends to increase the correlations between items that are worded in the same direction but are not conceptually related (Podsakoff, 2003). Evidence suggests that about 6% of participants respond in a pronounced acquiescent or disacquiescent (DACQ) way (Hinz et al., 2007), and that about 4% of the variance of personality items is due to ACQ (Danner et al., 2015). In both cases, the percentages are far from being trivial, and therefore, the effects, detection and correction of ACQ have been extensively researched (Baumgartner & Steenkamp, 2001; Billiet & McClendon, 2000; de la Fuente & Abad, 2020; Savalei & Falk, 2014). In psychometric terms, ACQ may affect the structural estimates at the calibration stage, distort the individual score estimates at the scoring stage, and bias the model-data fit assessment results. Focusing more specifically on calibration, the impact of ACQ can (a) vary intra-individually depending on the measures or trait being measured (Ray, 1983); (b) cause biases in the invariance of the content loadings in comparative studies; and (c) generate substantial bias in the estimation of stability coefficients and cross-lagged effects between variables over time in panel models (Billiet & Davidov, 2008).

The results above justify the interest in measuring or controlling ACQ, and, so far, most of the existing procedures fall within two

broad categories: First, 'A priori' methods linked to test design, mainly the use of balanced scales (Ray, 1979) or the inclusion of specific ACQ scales (e.g. Watson 1992; Krosnick, 1999; Saris et al., 2010). In general, the use of this type of procedures can be considered when the user is in a position to fully design the test. However, they increase the complexity of the analyses, and there is no convincing evidence of their advantages in terms of ACQ control. The second category, the 'Ex post facto' methods, are mainly based on statistical control of the data. These are procedures that can be applied when we are not in position to design the test. Table 1 summarizes the main control methods that have been developed in recent decades.

Several studies have compared the pros and cons of the procedures in table 1 are in terms of performance (e.g. de la Fuente & Abad, 2020; Primi et al., 2019; Savalei & Falk, 2014). In general, the random intercept factor analysis (RIFA) method usually emerges as the winner in terms of the acceptable overall performance-simplicity trade-off. However, this first place is due more to easiness of implementation than to real differences in effectiveness with the rest of the methods.

We briefly revise, finally, more complex studies in which the joint impact of ACQ and another response style has been assessed. So far, two 'secondary' response styles have been considered: Extreme Response (ER) (Cheung & Rensvold, 2000; Weijters et al., 2010; Park & Wu, 2019) and social desirability (SD) (e.g. Hand & Brazzell, 1965; Ferrando & Anguiano-Carrasco, 2010). At the practical level, Ferrando et al. (2009) proposed a procedure for controlling the bias caused by SD and ACQ simultaneously which has been used in the construction of certain personality scales as OPERAS (Vigil-Colet et al., 2013) or INCA (Morales-Vives et al., 2019).

Local Dependence-Correlated Residuals

The terms 'Local dependencies', 'Correlated residuals', 'Doublets' or 'Shared specificities' refer to a common phenomenon which, can be defined as follows. First, a pair (or a small group) of items continues to be related after the influence of the common content they measure has been partialled out. Second, this residual relation is due to causes different from additional shared common contents, such as context effects, redundancies in the evoked situation, or wording similarities (Ferrando et al., 2022; Ferrando et al., 2023). A main point here is that the causes just described are not linked to individual response tendencies (as in ACQ) but to specific properties of the items. So, their existence is mostly related to the design of the measurement instrument.

Research on residual correlations has focused above all on the convenience or not of allowing residuals to be modeled in factor analytic (FA) solutions. The effects of not controlling them, however, have been far less assessed (perhaps because researchers view this as a problem of test construction). As a result, residual analysis is rarely undertaken. Furthermore, these effects, have been mostly assessed at the calibration stage, and are: (a) biased item parameter estimates, and (b) distorted model-data fit assessment (Montoya & Edwards, 2021).

Within a FA framework, the detection of doublets is mainly based on the inspection of the residual covariance matrix (Ferrando et al., 2022). In the case of a traditional exploratory factor analysis (EFA), where the residual covariances are forced to be zero, an un-modeled doublet can result on an increase of the corresponding fitted residual,

Table 1
Summary of Some ACQ-Control Methods

Procedure	Author	Description	Pros	Cons
Ipsative method	Chan & Bentler (1993)	Developed for fitting factor models to data that have an ipsative structure,	Only feasible alternative if one wants to factor analyze measures with this structure	Weak functioning if there is a violation of the assumption of fully balanced scales and homogeneous ACQ loadings
General factor style Model	Billiet & McClendon (2000)	They developed a model that included a style factor that affected all items.	Easy to implement and improvement in goodness-of-fit indices compared to the uncorrected ACQ model.	The scale is required to be balanced and loadings on the style factor are required to be tau-equivalent
Unrestricted FA with Target Rotation	Ferrando et al. (2003)	Unrestricted FA model in which ACQ is explicitly modeled as a secondary factor orthogonal to content.	Allows for differential ACQ loadings to be estimated.	Fully balanced scales are required.
RIFA	Maydeu-Olivares & Coffman (2006)	CFA model in which the additional ACQ factor is restricted to have equal loadings. Correction method derived from an adaptation of the rotation method (Lorenzo-Seva & Rodriguez-Fornells, 2006) that allows for the removal of variance due to ACQ in partially balanced scales.	Easy to implement.	Tau-equivalence in ACQ factor loadings
Partially Balanced EFA	Lorenzo-Seva & Ferrando (2009)	A hybrid model that combines an EFA part where item-factor loadings are freely estimated and a restricted CFA part where item-factor loadings on the RI/ARS factor α are restricted to follow a predefined pattern	Works also with partially balanced scales. Robust.	A minimal number of reversed items is required.
RI-EFA model	Aichholzer (2014)	Multi-stage procedure designed for fitting restricted FA solution in data matrices that have been cleaned from ACQ bias.	Can be extended to testing measurement invariance over subgroups or over time as well as to testing covariates of the RI/ARS factor and, hence, causes of such bias	Implementation is complex
Hybrid CFA-EFA (Siren)	Navarro-Gonzalez et al. (2023)	Multi-stage procedure designed for fitting restricted FA solution in data matrices that have been cleaned from ACQ bias.	Allows restricted solutions to be fitted with the standard linear FA model or the non-linear graded-response model.	Sequential and conditional ad lib procedure that necessarily entails a loss of efficiency.

an overall increase in the residual covariances, or a propagation ‘shift’ leading to an overestimation of the factor loadings involved in the doublet. This propagation effect can well make the doublet undetectable, so fitted residual inspection, despite being the fastest and simplest approach, is not always the most appropriate. The partial-correlations method or the MORGANA method (Ferrando et al., 2022), despite being more complex, are expected to attain better results. MORGANA is derived from the concept of Expected Parameter Change (EPC; Saris et al., 1987) and is able to minimize the propagation effects of substantial doublets to other residuals or to the factor loadings. MORGANA would include two indices: EREC and ENIDE. The first quantifies the amount of misspecification in the residual correlation and is that considered in this study.

Beyond the (strict) FA framework, there has been recently an increased interest in new (or adapted) methods for detecting local dependence, such as the Bayesian Lasso method (Pan et al., 2017), which, according to the authors, is able to achieve both model parsimony and an identifiable model. On the other hand, Christensen et al. (2023), use Weighted Topological Overlap (WTO) to assess the similarity between variables in a network, and EBICglasso to adjust the network model. By comparing the similarity between variables to the fit of the network model, this approach is expected to be able to detect the presence of violations of local independence.

Regardless of the used detection method, once detected, the user must decide what to do with these residuals, and there are two obvious options: (a) remove one of the redundant items or (b) include them in the model. This second choice would increase the number of additional parameters leading to better model-data fit results. However, this improvement in fit is likely to imply capitalization of chance and a loss of replicability or reproducibility of the results.

In summary, the research to date on unintended response determinants to personality items has gaps that need to be filled, and

one of them is to determine the possible impact of two unwanted elements of different origin: ACQ on the one hand and the presence of correlated residuals on the other. The presence of each one of them separately is known to cause distortions at several levels. So, it seems relevant to ask which effects can be expected if they occur jointly. More specifically, the main goal of this study is to assess the combined impact of ACQ and correlated residuals on the item structural estimates and model-data-fit results of personality measures. And, in order to derive general predictions, we shall consider the usual FA framework with residual correlations restricted to be zero.

The remaining of the article is structured as follows. First, we shall derive some basic algebraic predictions, which, although not necessary to understand the general purposes of the article, provide a basis for a better understanding of the next steps. Next, two simulation studies, based on certain independent variables that are known to impact the structure and the goodness of fit of the model will be undertaken. The simulations will be complemented with an empirical study based on personality data. Finally, we shall discuss the implications of the obtained results.

Basic Predictions

Consider a test that measures a personality trait (θ) and that is made up of n continuous-response items. All of them have a response scale oriented in the same direction (e.g. 0: strongly disagree vs. 5: strongly agree), but half are positively oriented and the other half are reverted. First, suppose that the scale is ACQ-free, but the residuals for items x_1 and x_2 are correlated (Figure 1). The basis model is:

$$x_{ij} = \alpha_j \theta_i + \varepsilon_{ij} \quad (1)$$

where x_{ij} is the standardized score of person i on item j , α is the factor loading, and ε is the measurement error. The factor (θ) is the trait that is scaled in a z-score metric (mean 0 and variance 1). Once the α values are determined, it is necessary to obtain the residual correlations. For items j and k , the residual covariance between these two items is:

$$res_{jk} = \sigma_{jk} - \alpha_j \alpha_k \tag{2}$$

and, under the assumption that the residuals are uncorrelated, the model-implied correlation between a pair of items is given by (Harman, 1962; pp 120-1):

$$\sigma_{jk} = \alpha_j \alpha_k + res_{jk} = \alpha_j \alpha_k + \theta = \alpha_j \alpha_k \tag{3}$$

where σ_{jk} is the inter-item correlation. However, if correlated residuals exist, but have been forced to be zero, the resulting omitted covariances will tend to be re-assigned and may (a) spread throughout the elements of the residual matrix or (b) bias the content loading estimates; specifically, this second effect tends to focus on one of the two items involved in the doublet.

Taking now a step forward, let us consider an expanded unrestricted model, that was already proposed by Ferrando et al. (2003), and that includes two uncorrelated common factors: (a) a content factor (θ_1) and (b) an acquiescence factor (θ_2) (Figure 2).

$$x_{ij} = \alpha_{j1} \theta_{1i} + \alpha_{j2} \theta_{2i} + \varepsilon_{ij} \tag{4}$$

The loadings on the content factor (α_{j1}) are balanced (half of the items will be positive and the other half negative) and larger than those on the ACQ factor which are all positive. If so, the sum of the content loadings will be null ($\sum_j^n \alpha_{j1} = 0$). The sum of the standardized scores becomes:

$$X_i = \sum_j^n x_{ij} = \theta_{1i} \sum_j^n \alpha_{j1} + \theta_{2i} \sum_j^n \alpha_{j2} + \sum_j^n \varepsilon_{ij} \tag{5}$$

and the model-based residual covariance is

$$res_{jk} = \alpha_{j2} \alpha_{k2} \tag{6}$$

that is, fitting a one-dimensional model like (1) is expected to identify the ACQ factor. Furthermore, if, (a) the scale is well balanced, and (b) as assumed the residual correlations are all zero, then the loadings on this factor will be unbiased estimates of the item proneness to elicit ACQ (Ferrando & Lorenzo-Seva, 2009). However, if correlated residuals exist, they are expected to be absorbed in the estimated ACQ loadings that will then become biased. Overall, if the bidimensional model (4) with a content factor and an ACQ factor is fitted under balanced conditions but with non-zero correlated residuals, it is assumed that, in order to keep res_{jk} as close to zero as possible, the ACQ factor would absorb both the true ACQ and (to certain extent) any possible doublet that exists.

In light of these predictions, the aims of our research are to assess: (1) the impact of correcting acquiescence in the estimated correlated residuals, and (2) how the use or omission of (a) ACQ estimation and (b) residual correction methods affect the estimation of content

Figure 1
Unidimensional Model

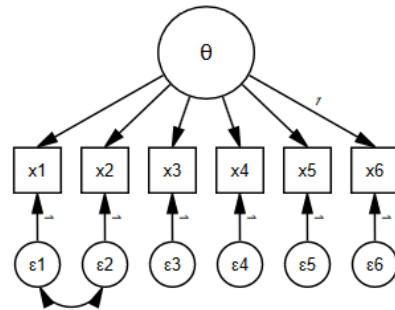
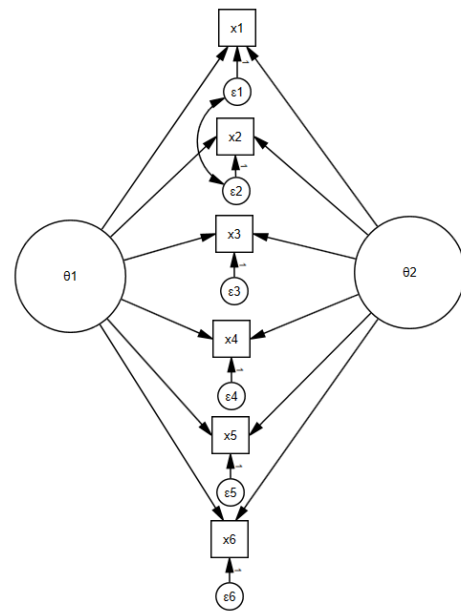


Figure 2
The Bidimensional for ACQ Bias



Note. The bidimensional for ACQ bias (Ferrando & Lorenzo-Seva, 2010; Ferrando et al., 2003) that includes a residual correlation between item x1 and item x2. θ_1 represent the content factor, θ_2 represent the acquiescence factor and residual terms are shown as ε .

factor loadings and model-data fit results. This information will enable us to propose a tentative procedural guide for scenarios containing a combined presence of variance unrelated to content.

Study 1: Monte Carlo Simulation Study

The aim of the reported simulation study was to assess the predictions above. More specifically, the aim was to assess the impact on factor loadings, number, and magnitude of the detected residuals when controlling and ‘eliminating’ variance due to acquiescence and detected correlated residuals. Four conditions were compared: control without any correction, acquiescence bias correction, residuals correction, and combined or mixed correction.

Method

Instruments

All samples were simulated under the two-dimensional model with one and/or two content factors (F1 and F2) and an acquiescence factor (ACQ) (Figure 2). Depending on the number of factors, the number of items varied, with 6 continuous items in the case of single-content factor models and 8 continuous items in two-content factor models. In all cases, the factor loadings were completely balanced. The number of items per factor was selected based on theoretical considerations and empirical evidence (at least three items per factor) from the literature. Both the simulated data and the analyses; were carried out with R.

In the ACQ correction, the ‘*acqhybrid*’ function from the ‘*siren*’ package (Navarro-Gonzalez et al., 2023) is used. This function fits a restricted solution factor analysis through a two-step procedure: In the first step an ACQ factor is estimated and its effects are partially excluded from the inter-item correlation matrix. In the second step, a restricted confirmatory factor analysis (CFA) solution is fitted to the reduced or ‘cleaned’ matrix.

In the control procedure, an EFA was conducted using the ‘*fa*’ function from the ‘*psych*’ package (Revelle, 2015).

In this study, we chose to use 500 replicas because previous research indicated that a larger number did neither significantly affect power nor produce substantial changes in the results (Ferrando et al., 2016).

Procedure

The study design was a full factorial $4 \times 2 \times 2 \times 2$, and the following variables were manipulated: (1) type of analysis: controlling only ACQ, controlling only residuals, controlling both residuals and ACQ, and no controlling any bias; (2) number of factors; (3) location of items that exhibit residual correlation (within the same factor or in different factors); and (4) the sign of the item pair exhibiting residual correlation (items written in the same direction, the same sign, vs. items written in different directions, opposite signs). Additionally, the size of both the content and the ACQ loadings were controlled.

Of the 32 simulated datasets, eight were analyzed by controlling for acquiescence (see above). To analyze the effect of correcting only the correlated residuals, the MORGANA method was used (see above). The combined procedure used the two previous methods: firstly, the variance due to ACQ was detected and eliminated, and next, the residual correlations were estimated from the ‘cleaned’ correlation matrix.

Data Analysis

The dependent variables were: (a) the number of doublets detected (whether false or true positives), (b) the estimated value of the EREC index for the true positives, (c) the difference between simulated and estimated factor loadings, and (d) goodness-of-fit indices. In order to examine the absorption effect, a contingency table and two analyses of variance (ANOVA) were conducted to determine whether the number of detected doublets and the EREC values depended on the prior correction of acquiescence. Because EREC has proven to be a highly sensitive index, only

values greater than .20 were considered (Ferrando et al., 2022). The difference between the estimated and simulated factor loadings was assessed by computing the root mean square error (RMSE), which is defined as

$$RMSE = \frac{1}{p} \sqrt{\sum_{i=1}^c (\hat{\alpha}_{ip} - \alpha_i)^2} \quad (7)$$

Where p is the total number of replicas, α_i is the true value of the loadings on the factor, and $\hat{\alpha}_{ip}$ is the estimated loading value on item i in replica p .

Results

Contingency Table

The contingency (Table 2) shows the number of doublets detected by EREC under the conditions of RES (only Residuals are corrected) and combined (ACQ and Residuals are corrected), and the results support the previous algebraically-derived predictions. Under the combined condition, EREC does not detect any doublets, that is, absorption occurs in 100% of cases, regardless of whether the items have the same sign or opposite signs. However, the probability of absorption decreases by approximately 25% when the items are in different factors (recall that the simulation considers two orthogonal factors). When there is no prior ACQ correction, the results vary depending on whether the items have the same sign or opposite signs. When the two items that form the doublet are positive, there is a tendency to overestimate the magnitude of the correlated residuals. It is interesting to note that, in this group, in only 0.73% of cases does the EREC index not detect the simulated doublet. This fact is especially noteworthy when compared to the condition in which the doublet falls on items with opposite signs, in which the number of false negatives reaches 23.5%. In the no-correction condition (of ACQ) in uncorrelated two-factor models, overestimation occurs in 100% of the cases. This contrasts significantly with the no-correction condition of ACQ in one-factor models, where overestimation is slightly lower and affects less than 65% of the sample (when the doublet involves items of the same sign and when it involves items of opposite signs).

ANOVAs

The results of the ANOVAs mentioned above are now summarized. Regarding the number of detected doublets, the results are consistent with the trend observed in the contingency table concerning the number of factors ($F(1) = 27,190.0$; $p < .05$; $\eta_p^2 = .15$) and the correction method used ($F(1) = 100,500.0$; $p < .05$; $\eta_p^2 = .58$). Additionally, EREC demonstrated robustness when the residual was situated between positive items or items with opposite signs ($F(1) = 132.0$; $p < .05$; $\eta_p^2 < .001$) and when the ACQ factor exhibited higher factorial loadings.

In relation to the size of the EREC values, only those counted as true positives were considered, and values below .2 were omitted. The results showed significant differences with a large effect size in the variable ‘number of factors’ ($F(1) = 44,953.6$; $p < .05$; $\eta_p^2 = .54$) and residual location ($F(1) = 192,662.0$; $p < .05$; $\eta_p^2 = .84$), as well as in the interaction of both factors ($F(1) = 7,438.1$; $p < .05$; $\eta_p^2 = .16$).

RMSE

Table 3 shows the obtained RMSE values when the residuals are between items with the same sign. The results obtained with residuals between items of opposite signs follow the same trend, although with slightly higher RMSE index values. Furthermore, the RMSE values show greater variability when the number of factors increases and when the residual is between different factors. There

are no significant differences between the four procedures when analyzing one-factor models, with the combined procedure being the most accurate, compared to the ACQ procedure.

In bidimensional models, when the residual is between two items of the same factor, the ACQ procedure is clearly superior to the others, a trend that repeats when the residual is between items of different factors. In the rest of the methods tested, it is observed that when the residual is between different items there is greater

Table 2
Contingency Table. Number of Doublets Detected by the EREC Index

N° D	1 factor				2 factors			
	Same Directions		Opposite Directions		Same Directions		Opposite Directions	
	Combined	RES	Combined	RES	Combined	RES	Combined	RES
0	4000 (100%)	29 (.73%)	4000 (100%)	941 (23.52%)	2942 (73.55%)		2973 (74.32%)	
1		1434 (35.85%)		726 (18.15%)	365 (9.13%)		329 (8.23%)	
2		2537 (63.42%)		2333 (58.33%)	244 (6.1%)		253 (6.33%)	
3					449 (11.22%)	4000 (100%)	445 (11.12%)	4000 (100%)

Note. N°D = Number of Doublets; RES=without prior acquiescence correction; Same directions = they are items that measure the trait from the same direction; Opposite Directions = The residual is found among items that measure the trait in opposite ways.

Table 3
RMSE Values With the Residual Simulated Between Items With the Same Orientation.

			I1	I2	I3	I4	I5	I6	I7	I8	
CONTROL	1F		.044	.036	.043	.034	.033	.035			
		RES (1-3)	F1	.071	.065	.077	.063	.043	.044	.047	.045
	2F		.044	.047	.045	.049	.046	.048	.048	.048	.046
		RES (1-5)	F1	.088	.109	.110	.108	.109	.101	.101	.101
	ACQ	1F		.049	.046	.050	.045	.049	.047		
			RES (1-3)	F1	.061	.053	.060	.057	<.0001	<.0001	<.0001
2F			<.0001	<.0001	<.0001	<.0001	.045	.047	.045	.048	
		RES (1-5)	F1	.053	.048	.048	.047	<.0001	<.0001	<.0001	<.0001
RES		1F		.045	.041	.045	.042	.040	.038		
			RES (1-3)	F1	.087	.081	.088	.081	.047	.048	.047
	2F		.042	.043	.041	.044	.050	.050	.050	.048	
		RES (1-5)	F1	.103	.125	.124	.126	.099	.092	.091	.090
	COMBINED	1F		.031	.027	.030	.029	.028	.027		
			RES (1-3)	F1	.086	.088	.090	.086	.048	.046	.047
2F			.039	.045	.040	.038	.046	.047	.049	.046	
		RES (1-5)	F1	.095	.095	.100	.088	.056	.055	.054	.054
			.052	.050	.048	.052	.052	.054	.051	.051	

Note. N°D = Number of Doublets; ACQ = only acquiescence correction; RES=without prior acquiescence correction; RES (1-3) = the residual is located between items 1 and 3; both within the same factor; RES (1-5) = The residual is located between items 1 and 5 in different factors.

variability in the RMSE values. The Combined, RES, and Control procedures show moderate errors, with RES appearing to have the least measurement accuracy. However, in these bidimensional models, when observing the second content factor, it is noted that in general the RMSE values are clearly lower than in the first factor in the ACQ, RES, and Combined procedures. This trend, however, is not observed in the Control procedure.

Goodness of Fit

Table 4 shows the main goodness-of-fit indices of the four procedures analyzed. Overall, the ACQ and Combined procedures show good fit results, as does the RES procedure. However, in the one-factor model, the root mean square residual (RMSR) values are slightly better in the combined model. Results in Table 4 are limited to simulations where the residual was between items of the same sign but similar trend is observed for opposite signs.

Table 4
Goodness of Fit Indices

		TLI	RMSEA	RMSR	
Control	F1	.932	.083	.050	
	F2	RES (1-3)	.925	.070	.030
		RES (1-5)	.883	.080	.040
ACQ	F1	.999	<.0001	.006	
	F2	RES (1-3)	.997	<.0001	.028
		RES (1-5)	.999	<.0001	.031
RES	F1	.999	<.0001	.073	
	F2	RES (1-3)	.999	.009	.051
		RES (1-5)	.997	.010	.063
Combined	F1	.998	.0001	.022	
	F2	RES (1-3)	.999	<.0001	.021
		RES (1-5)	.999	<.0001	.028

Note. ACQ = only acquiescence correction; RES=without prior acquiescence correction; RES (1-3) = the residual is located between items 1 and 3; both within the same factor; RES (1-5) = The residual is located between items 1 and 5 in different factors.

Study 2: Empirical Example

So as to illustrate with real data the results obtained via simulation, we shall re-analyze an existing dataset that presumably contains correlated residuals.

Method

Participants

Respondents were 2,429 adults, with an age range between 18 and 60 years ($M = 29.15$; $SD = 14.65$) and of which 38.37% were men.

Instruments

We shall re-analyze an existing dataset used in the calibration of the Overall Personality Assessment Scales (OPERAS; Vigil-Colet et al., 2013). Specifically, we shall re-analyze the data corresponding to the Extraversion subscale (EX), which comprises seven items that are almost fully balanced (four measuring extraversion and three

introversion), and all of them positively worded. The scale scores show high reliability and very low levels of social desirability bias.

Procedure

Four EFA's were conducted: a content-only EFA without any correction, an EFA with ACQ correction, an EFA with correlated-residuals correction, and a mixed exploratory analysis that corrected for ACQ and removed correlated residuals. The used procedures were those described in the simulation study. Given the ordered-categorical nature of the data, the EFA's were based on polychoric inter-item correlation matrices and fitted with the Unweighted Least Squares (ULS) criterion.

Data Analysis

All analyses were carried out using R, utilizing the same packages as in the previous simulation studies.

Results

The most apparent result in Table 5 is that the EREC values decrease dramatically when the data is pre-corrected for acquiescence. Without correction, residual correlation is detected between pairs 2 - 4 and 5 - 6 (see Table 5). The detected pairs exhibit clear semantic redundancy; so, the results are submitted to be correct. However, once the variance due to acquiescence is partialled-out, no substantial correlated residuals are longer detected (Table 6).

Table 5
Detected Doublets According to EREC Index

Doublets	EREC Index	
	RES	Combined
2 - 4	.578	
5 - 6	.470	.203
2 - 5		.120

Note. Values less than .2 will be considered trivial. RES=without prior acquiescence correction

Table 6
Main Detected Doublets

Doublets	Items
2 - 4	2. Me desenvuelvo bien en situaciones sociales 4. Hago amigos con facilidad
5 - 6	5. Prefiero que otros sean el centro de atención 6. Permanezco en segundo plano

Note. 2. I handle social situations well; 4. I make friends easily; 5. I prefer others to be the center of attention; 6. I stay in the background

Table 7
Estimated Loadings in Each of the Procedures.

Control	ACQ	RES	Combined
.660	.624	.685	.644
.766	.722	.696	.686
-.686	-.684	-.715	-.647
.746	.718	.674	.702
-.554	-.659	-.495	-.680
-.625	-.702	-.580	-.691
.664	.625	.688	.635

Table 7 compares the loading estimates when EFA's are performed (a) without correction, (b) correcting only acquiescence, (c) correcting only residuals, and (d) performing a complete

correction. Factor loadings for items 5 and 6 exhibit the greatest variability across different correction types, the difference being maximal between the acquiescence correction option and the residuals correction option.

Table 8
Goodness of Fit Indices

	GFI	TLI	RMSEA	RMSR
Control	---	.781	.17	.084
ACQ	.999	.998	.023	.021
RES	.987	.987	.047	---
Combined	.999	.999	.00	---

Table 8 displays the goodness of fit indices estimated in each of the procedures. The control procedure is the only one that does not reach an acceptable fit. The ACQ correction and residual correction procedures exhibit good fit in terms of goodness of fit index (GFI) and Tucker and Lewis index (TLI) and moderate fit in root mean square error of approximation (RMSEA) terms. As expected, the combined procedure yields the best results.

Discussion

The current research has attempted to explore the potential impact of two non-content sources of error or unwanted determinants of different origins: ACQ and correlated residuals. Previous studies found that, separately, both, ACQ and correlated residuals can distort structural item estimates and goodness-of-fit assessment at the calibration stage. However, the combined effect of their joint occurrence does not appear to have been addressed until now.

Through two studies, we have attempted to determine what occurs when we correct for ACQ in a dataset that includes more than one of the unwanted determinants. The predictions made above provide analytical evidence that part of the correlated residual variance may be absorbed by the ACQ factor when ACQ corrections are applied, a prediction that has been supported by the simulation results: It was found that MORGANA, even when being a very sensitive procedure, was unable to detect almost any simulated residual doublet (true positive) when there was prior ACQ correction, regardless of the items' location and sign. This result also holds in cases where the doublet is located in two different factors with no prior correction.

At the same time, however, the results suggest that, even though there is indeed a clear absorption effect by the ACQ factor, this effect does not seem to have a negative impact on the model fit results and the accuracy of the content factor loading estimates. In general, the trend when using the ACQ correction method and the combined method is that the accuracy in the estimation of content loadings is very good. Furthermore, the estimation of the second element of the pair is slightly more accurate than that of the first element; however, the overall trend is that the greater the bias (difference between simulated and estimated loading) in the first element of the pair, the greater the bias will be in the second element of the pair.

The simulation study only considered fully and essentially balanced item sets, and one-factor, and two-uncorrelated-factor models with high loadings on the content factors; a very simple an 'ideal' set of conditions indeed. So, results cannot be naively

generalized to more complex models, and further intensive research is needed. However, even when acknowledging its preliminary nature, we believe that the results obtained here provide useful information that can be considered for practical applications.

Based on the obtained results, it can be preliminarily concluded that, when correcting for acquiescence in a dataset that contains correlated residuals, we are absorbing not only the portion of variance attributed to this response bias but also part of the variance that is due to the existing correlated residuals. However, when both sources are jointly corrected, the absorption effect is expected to be much weaker, goodness of model-data fit is expected to slightly improve, and the structural estimates for the content factors are expected to be essentially unbiased. So, what we tentatively suggest is that, when fitting a balanced measure that already aims to correct for ACQ in a dataset in which correlated residuals are also suspected, the best approach is to perform a dual correction procedure. This suggestion, however, must be qualified. In principle, it is expected to be appropriate in scenarios in which the number of common factors (content and ACQ) is reasonably well known. And, even in this case, the dual correction needs not be sequential (as done here) but could be also simultaneous (we are exploring this issue at present). On the other hand, if the number of common factors cannot be reasonably well specified 'a priori', then it could be envisaged to undertake first a residual correlation assessment using a procedure aimed at detecting residuals with no need to specify the number of common factors in advance, and next estimating the ACQ factor. In this respect, potentially useful residual detection methods that do not require the number of common factors to be specified are those by Christensen, Garrido, and Golino (2023), and the partial correlation (image) approach (Ferrando et al., 2022) mentioned above.

Author Contributions

Ana Hernández-Dorado: Writing – Original Draft, Investigation and Software. **Pere J. Ferrando:** Conceptualization, Formal Analysis, Writing – Review & Editing. **Andreu Vigil-Colet:** Data Curation, Writing – Review & Editing.

Funding

This project has been made possible by the support of the Ministerio de Ciencia e Innovación, the Agencia Estatal de Investigación (AEI) and the European Regional Development Fund (ERDF) (PID2020-112894GB-I00). This funding source had no role in the design of this study, data collection, management, analysis, and interpretation of data, writing of the manuscript, and the decision to submit the manuscript for publication.

Declaration of Interests

The authors declare that there is no conflict of interest.

Data Availability Statement

All the codes used in the simulation are deposited on GitHub and interested users can access them via [<https://bit.ly/githubRESAC>]. The data used in the experimental part will be available to interested users upon reasonable request.

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>
- Arias, V. B., Ponce, F. P., Garrido, L. E., Nieto-Cañaveras, M. D., Martínez-Molina, A., & Arias, B. (2024). Detecting non-content-based response styles in survey data: An application of mixture factor analysis. *Behavior Research Methods*, 56, 3242–3258. <https://doi.org/10.3758/s13428-023-02308-w>
- Baumgartner, H., & Steenkamp, J. B. 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>
- 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. K. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(4), 608-628. http://doi.org/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>
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-cultural Psychology*, 31(2), 187-212. <https://doi.org/10.1177/0022022100031002003>
- Christensen, A. P., Garrido, L. E., & Golino, H. (2023). Unique variable analysis: A network psychometrics method to detect local dependence. *Multivariate Behavioral Research*, 58(6), 1165-1182. <https://doi.org/10.1080/00273171.2023.2194606>
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6(4), 475-494. <https://doi.org/10.1177/001316444600600405>
- 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>
- de la Fuente, J., & Abad, F. J. (2020). Comparing methods for modeling acquiescence in multidimensional partially balanced scales. *Psicothema* 32(4), 590-597. <http://doi.org/10.7334/psicothema2020.96>
- Ferrando, P. J., & Anguiano-Carrasco, C. (2010). Acquiescence and social desirability as item response determinants: An IRT-based study with the Marlowe-Crowne and the EPQ Lie scales. *Personality and Individual Differences*, 48(5), 596-600. <https://doi.org/10.1016/j.paid.2009.12.013>
- Ferrando, P. J., Hernández-Dorado, A., & Lorenzo-Seva, U. (2022). Detecting correlated residuals in exploratory factor analysis: New proposals and a comparison of procedures. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(4), 630-638. <https://doi.org/10.1080/10705511.2021.2004543>
- Ferrando, P. J., Hernández-Dorado, A., & Lorenzo-Seva, U. (2023). A simple two-step procedure for fitting fully unrestricted exploratory factor analytic solutions with correlated residuals. *Structural Equation Modeling: A Multidisciplinary Journal*, 31(3), 420–428. <https://doi.org/10.1080/10705511.2023.2267181>
- Ferrando, P. J., & Lorenzo-Seva, U. (2010). Acquiescence as a source of bias and model and person misfit: A theoretical and empirical analysis. *British Journal of Mathematical and Statistical Psychology*, 63(2), 427-448. <https://doi.org/10.1348/000711009X470740>
- 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>
- Hand, J., & Brazzell, C. O. (1965). Contamination in measures of acquiescence and social desirability. *Psychological Reports*, 16(3), 759-760. <https://doi.org/10.2466/pr0.1965.16.3.759>
- Harman, H. H. (1962). *Modern Factor Analysis* (2a ed.). University of Chicago Press.
- Hinz, A., Michalski, D., Schwarz, R., & Herzberg, P. Y. (2007). The acquiescence effect in responding to a questionnaire. *GMS Psycho-Social Medicine*, 4, 1-9. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2736523/pdf/PSM-04-07.pdf>
- Krosnick, J. A. (1999). Maximizing questionnaire quality. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of Political Attitudes* (pp. 37–57). Academic Press.
- 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>
- Lorenzo-Seva, U., & Rodríguez-Fornells, A. (2006). Acquiescent responding in balanced multidimensional scales and exploratory factor analysis. *Psychometrika*, 71(4), 769-777. <https://doi.org/10.1007/s11336-004-1207-4>
- 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>
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational measurement: Issues and Practice*, 14(4), 5-8. <https://doi.org/10.1111/j.1745-3992.1995.tb00881.x>
- Montoya, A. K., & Edwards, M. C. (2021). The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. *Educational and Psychological Measurement*, 81(3), 413–440. <https://doi.org/10.1177/0013164420942899>

- Morales-Vives, F., Cosí, S., Lorenzo-Seva, U., & Vigil-Colet, A. (2019). The inventory of callous-unemotional traits and antisocial behavior (inca) for young people: Development and validation in a community sample. *Frontiers in Psychology, 10*, Article 713. <https://doi.org/10.3389/fpsyg.2019.00713>
- Navarro-González, D., Lorenzo-Seva, U., & Vigil-Colet, A. (2016). How response bias affects the factorial structure of personality self-reports. *Psicothema, 28*(4), 465-470. <http://www.doi.org/10.7334/psicothema2016.113>
- Navarro-Gonzalez, D., Ferrando, P. J., Morales-Vives, F., Hernandez-Dorado, A., & Navarro-Gonzalez, M. D. (2023). Package 'siren' (version 1.0.4) [Software package]. <https://cran.r-project.org/web/packages/siren/index.html>
- Navarro-Gonzalez, D., Ferrando, P. J., Morales-Vives, F., & Hernandez-Dorado, A. (2024). SIREN: A hybrid CFA-EFA R package for controlling acquiescence in restricted factorial [Manuscript submitted for publication]. Department of Psychology, University of Rovira i Virgili.
- Pan, J., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso. *Psychological methods, 22*(4), 687-704. <https://doi.org/10.1037/met0000112>
- Park, M., & Wu, A. D. (2019). Item Response Tree Models to Investigate Acquiescence and Extreme Response Styles in Likert-Type Rating Scales. *Educational and Psychological Measurement, 79*(5), 911-930. <https://doi.org/10.1177/0013164419829855>
- Podsakoff, P.M., MacKenzie, S.B., Lee, J., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(879), 1010-1037. <https://doi.org/10.1037/0021-9010.88.5.879>
- Primi, R., Santos, D., De Fruyt, F., & John, O. P. (2019). Comparison of classical and modern methods for measuring and correcting for acquiescence. *British Journal of Mathematical and Statistical Psychology, 72*(3), 447-465. <https://doi.org/10.1111/bmsp.12168>
- Ray, J. J. (1979). Is the acquiescent response style not so mythical after all? Some results from a successful balanced F scale. *Journal of Personality Assessment, 43*(6), 638-643. https://doi.org/10.1207/s15327752jpa4306_14
- Ray, J. J. (1983). Reviving the problem of acquiescent response bias. *Journal of Social Psychology, 121*(1), 81-96. <https://doi.org/10.1080/00224545.1983.9924470>
- Revelle, M. W. (2015). Package 'psych'. *The Comprehensive R Archive Network, 337*, Article 338. <https://mirror.ibcp.fr/pub/CRAN/web/packages/psych/psych.pdf>
- Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology, 105*-129. <https://doi.org/10.2307/271030>
- Saris, W. E., Satorra, A., & Van der Veld, W. M. (2010). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling, 16*, 561-582. <https://doi.org/https://doi.org/10.1080/10705510903203433>
- 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>
- Vigil-Colet, A., Morales-Vives, F., Camps, E., Tous, J., & Lorenzo-Seva, U. (2013). Development and validation of the overall personality assessment scale (OPERAS). *Psicothema, 25*(1), 100-106. <https://www.redalyc.org/pdf/727/72725690017.pdf>
- 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>
- Watson, D. (1992). Correcting for acquiescent response bias in the absence of a balanced scale: An application to class consciousness. *Sociological Methods & Research, 21*(1), 52-88. <https://doi.org/10.1177/0049124192021001003>
- Weijters, B.; Geuens, M.; Schillewaert, N. (2010). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement, 34*(2), 105 -121. <https://doi.org/10.1177/0146621609338593>