ECONOMICS & Sociology

Berumen, S. A., & Rabadán-Pérez, F. (2025). Methodology for the reconstruction of citizen surveys in different time periods. *Economics and Sociology*, *18*(2), 271-288. doi:10.14254/2071-789X.2025/18-2/14

METHODOLOGY FOR THE RECONSTRUCTION OF CITIZEN SURVEYS IN DIFFERENT TIME PERIODS

Sergio A. Berumen
King Juan Carlos University, Madrid, Spain
E-mail: sergio.berumen@urjc.es
ORCID 0000-0003-1773-9955

Francisco Rabadán-Pérez
King Juan Carlos University, Madrid, Spain
E-mail: francisco.rabadanl@urjc.es
ORCID 0000-0002-4324-7244

Received: March, 2024
1st Revision: May, 2025
Accepted: June, 2025

DOI: 10.14254/2071-789X.2025/18-2/14

ABSTRACT. In the evaluation of the impact generated by public policies, it is common practice to collect data in surveys carried out at different points in time. The study presents a comparison of two main methodological problems: (i) that questions asked are not always identical, and (ii) that data may not always be interpreted with the same methodology. To assess the extent to which the factorial structure is maintained, this paper proposes an original application of the principle of parsimony, like the well-known elbow rule in sedimentation plotting. The statistical contribution is validated through two publications of the Spanish Sociological Research Centre (CIS), published in different years. The results of the analysis show that, despite the variations mentioned above, the factorial structure of the responses remains the same, suggesting that the changes may be attributable to policy intervention.

JEL Classification: C1, H1, P5, Z1.

Keywords: public opinion, public policies, surveys, evaluation methods.

Introduction

In governance, the implementation of new policies is just as important as the citizenry's assessment of the effectiveness of the policies that are in force. The evaluation of policies that have been in place for some time is equally crucial, even if it is conducted less frequently due to either changes in the staffing of government bodies or changes in political parties in power, which can result in a lack of interest in recognising the successes of the previous government. Indeed, conducting evaluations between periods/years is invaluable, as it makes it possible to identify the progress made and the objectives to be achieved. In this context, this research is a methodological contribution to the evaluation of the impact generated by public policies based on the responses of citizens in surveys carried out at different times.

In the evaluation of the impact of public policies, the comparison of surveys faces recurrent methodological problems, such as: (i) difficulties associated with the lack of continuity in the questions; (ii) difficulties inherent in the selection of the variables to be

compared; (iii) the need for recoding between non-coincident questionnaires; and (iv) the discretionary nature of the units of measurement of the variables.

The present research formulates a methodological proposal focusing on the differences in the factorial structure of two surveys carried out by the Spanish Sociological Research Centre (CIS) on the policy against tax fraud in Spain. To check factorial consistency and assess the extent to which the structure is maintained, goodness-of-fit measures and contrasts from confirmatory factorial analysis and structural equations (Structural Equation Modelling, SEM) have been used. Further, a novel interpretation of the principle of parsimony has been added to the statistics methods of RMSEA (Root Mean Error of Approximation), FIT (Factor Fit of the Complete Model) and BIC (Bayesian Information Criterion). The main statistical utility is that the factors extracted from the first survey can be reconstructed, in whole or in part, in subsequent studies with a partially different question structure.

1. Literature review

Spearman (1927) set the guidelines for determining the intelligence factor in order to identify a latent variable through the observed variables. The so-called latent factor or latent variable is of special relevance in social science due to the principle of interpretability, which is particularly important in the comparison of citizen surveys conducted in different periods/years. To this end, certain principles must be complied with (Rabadan-Pérez, et al., 2022): (i) justify why it is necessary to compare both periods/years; (ii) know the dimensional structure of each survey, and from it, analyse to what extent they can be reproduced/compared with each other; (iii) know the methods that enable to assess the differences in the dimensional structure; (iv) make contrasts with respect to common magnitudes in both periods/years; (v) in case it is not possible to fully reconstruct the dimensional structure, identify variables that partially explain the factor, recode them if necessary, and reconstruct a new factor that is common to both surveys; and (vi) if the dimensional structure cannot be fully reconstructed, the variables that partially explain the factor have to be identified, recoded if necessary, and a new factor common to both surveys have to be reconstructed.

2. Methodological proposal

When a factor cannot be extracted in both surveys, due to missing or modified variables which are at least partially comparable, a factorial structure as faithful as possible should be maintained between the two periods/years, for which the following is proposed: (i) identify outliers according to the Mahalanobis distance criterion (1936) and their exclusion in subsequent dimension reduction analyses when $sig \leq 10^{-3}$ (Rencher 1995: 83-89); (ii) whenever possible, perform *Maximum Likelihood Estimation* (MLE) extraction; the fundamental assumption for maximum likelihood extraction is multivariate normality, whereby, because in citizen surveys most variables are polytomous or dichotomous in nature, the skewness and kurtosis coefficients should be in the range $[-1, 1]$, or between $[-2, 2]$; (iii) the ideal form of comparison is to maintain an extraction with the same variables in both questionnaires (*Exploratory Factor Analysis*, EFA); (iv) when the previous point is not possible, proceed to the comparison of close variables (those with the highest factor loadings in the reference period with respect to the factor) because they best describe the behaviour of the factor; (v) when it is not possible to work with close variables but there is a certain theoretical parallelism between them, recode them into common categories, and (vi) when none of the above is possible, as a last resort it is proposed to resort to variables that express an idea of complementarity with respect to the variables of the reference period/year.

As far as the unit of measurement is concerned, even if one only wishes to compare the impact of different variables of the policy in question in one period/year, the different ranges in the Likert scales (1932) and the heterogeneity of variables makes it necessary to resort to typing (to compare the effects of variables in terms of standard deviations), corrected by the scale, inverse or direct, in which the questionnaire question is asked so that the relationship between variables can be interpreted more clearly. To achieve this, this research proposes a corrected z^*typification that transforms all variables to a direct scale, as follows:

$$z_i^* = (-1)^a * \left(\frac{x_i - E(x_i)}{\sqrt{V(x_i)}}\right); \quad \begin{cases} inverse\ relation \Rightarrow a = 1 \\ direct\ relation \Rightarrow a = 0 \end{cases}$$

Strictly concerning the comparison method, it is necessary to select a set of target variables of a polytomous or dichotomous nature that are capable of directly measuring the object of study, although it is not possible to measure said object of study directly: (i) when the polytomous variable is ordinal in nature, yet does not differentiate its categories significantly in terms of contrast, proceed to dichotomise it; (ii) when the objective is to assess whether the differences in sample means (Welch 1938) are statistically significant, proceed to the Monte Carlo test of equality of means and asymptotics, and (iii) as a result, include those differences between both periods/years that are significant and eliminate irrelevant or redundant information.

On the basis of the above, in the present investigation: (i) the interpretation of the factors was carried out by conceptually associating the variables with the highest factorial loadings with the factor, which was considered to be the causal variable of the observed variables; (ii) the reverse or direct scale in the questionnaire was taken into account in relation to the increase or decrease of the factor with the variables that allowed it to be extracted; (iii) when the extraction did not allow a correct interpretation of the factor, the elimination of those variables whose conceptual relationship with the factor was not justified was accepted; and (iv) once the number of factors had been selected, the observed variables that participated in the extraction were replaced by the factorial scores for each case.

Finally, regarding factorial reduction: (i) for the interpretation of the factors, the variables with the highest factorial loadings were conceptually associated with the factor that was considered to be the causal variable of the observed variables; (ii) the reverse or direct scale in the questionnaire was taken into account in relation to the increase or decrease of the factor with the variables that allowed its extraction; (iii) when the extraction did not allow a correct interpretation of the factor, the removal of those variables whose conceptual relationship with the factor was not justified was accepted; and (iv) once the number of factors had been selected, the observed variables involved in the extraction were replaced by the factorial scores for each case.

3. Empirical validation

This methodological proposal has been validated by means of two surveys of the CIS from different years, both focusing on tax fraud policy in Spain: i) for the reference period/year, the *CIS3259* survey, targeting the population of both sexes aged 18 and over: of the 2500 interviews projected in the stratified cluster sampling design, 2464 were carried out (CIS 2019); and (ii) for the period/year to be compared, the *CIS2910*, with similar characteristics regarding the sampling design: of the 2500 interviews planned, 2468 were carried out (CIS 2011). The variables identified that reflect the respondents' perception of the level of tax fraud are P15, P17, P18, P19 and P20 in the reference period questionnaire. The result of the selection of complete cases for extraction yielded a sample size of 761, which

verifies Hogarty's criteria (2005). Furthermore, almost all variables showed a skewness and kurtosis coefficient in the [-2.2] range.

3.1. Maximum likelihood factorial analysis for the CIS3259 survey

In this section, the full EFA has been carried out since, being the reference period, it cannot be accepted that there are missing variables in the exploratory analysis.

Exploratory analysis

The determinant of the correlation matrix is close to zero ($|R| = 1'5738 * 10^{-4}$). The statistic B (Bartlett 1951) follows a distribution χ_k^2, asymptotically with sample size N (being k the number of independent variables, if the population correlation matrix R is equal to the identity matrix I). In order for there to be a sufficient correlation between variables to undertake factorial analysis, this hypothesis must be rejected. This null hypothesis is rejected in this study at a significance level of $\alpha = 0'05$. For an effect size, $w = 0'3$, the power of the contrast $(1 - \beta)$ calculated with *G*Power*, reaches its maximum value $(1 - \beta) = 1$ for a critical value of Bartlett's statistic B* = 52,1923 at a χ_{37}^2. The Kaiser-Meyer-Olkin sample adequacy statistic (Kaiser 1970) has been found to be adequate ($KMO = 0'818$), as well as the sample adequacy measures for each of the variables they check (in this section MSA most of them are greater than 0.8):

$$MSA_i \geq 0'7 \quad \forall i = 1, \dots, p$$

All communalities h_i improved after maximum-likelihood extraction. *Table 1* shows how 57.3% of the total variance explained (*TVE*) is explained with six factors and, after Varimax rotation (Kaiser 1958), 46%.

Table 1. Total variance explained (VTE) by the factorial model

	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
Factor (i)	λ_i	VTE	$\sum_i VTE$	λ_i	VTE	$\sum_i VTE$	λ_i	VTE	$\sum_i VTE$
MLE1	4.27	15.3%	15.3%	3.74	13.4%	13.4%	2.48	8.9%	8.9%
MLE2	3.42	12.2%	27.4%	0.00	10.0%	23.4%	2.28	8.1%	17.0%
MLE3	2.84	10.1%	37.6%	2.33	8.3%	31.7%	2.15	0.0%	24.7%
MLE4	2.08	7.4%	45.0%	1.55	5.5%	37.2%	2.11	7.5%	32.2%
MLE5	1.90	6.8%	51.8%	1.40	5.0%	0.0%	2.03	7.3%	39.5%
MLE6	1.55	5.5%	57.3%	1.07	3.8%	46.0%	1.84	6.6%	46.0%

Source: *own compilation*

The factors obtained are identified (according to measures of form, factorial loadings and questionnaire scale) as follows: (i) happiness: (MLE1); (ii) intolerance towards tax fraud: (MLE2); (iii) the need for long-term public policy financing (MLE3); (iv) dissatisfaction with the functioning of welfare state public services (MLE4); (v) social mobility (MLE5); and (vi) citizen morale (MLE6).

Confirmatory analysis

In the parallel analysis (Velicer 1976a), the common components or factors that have higher eigenvalues than those that would be obtained randomly were selected.

Table 2. Factorial loadings and variance ratio

	MLE2	MLE1	MLE3	MLE6	MLE5	MLE4
Factorial loadings	2.55	2.45	2.09	2.03	1.99	1.84
Proportion of variance	0.09	0.09	0.07	0.07	0.07	0.07
Cumulative variance	0.09	0.18	0.25	0.33	0.40	0.46
Proportion of variance explained	0.20	0.19	0.16	0.16	0.15	0.14
Proportion of cumulative variance	0.20	0.39	0.55	0.70	0.86	1.00

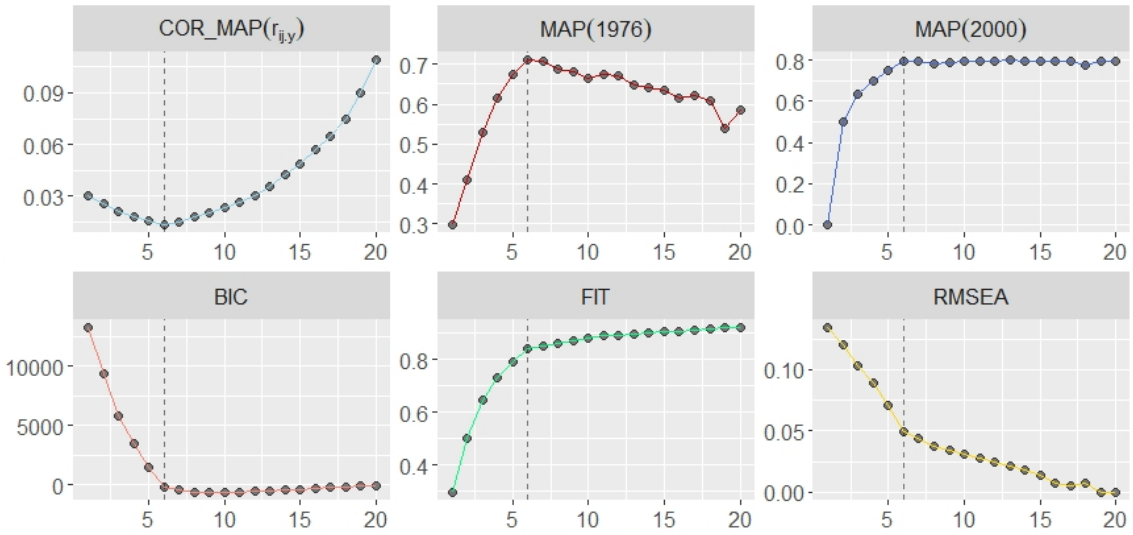
Source: *own compilation*

The Hofmann (1978) complexity index for each item, being $a_i$the factorial loading for the factor i, is defined as:

$$c_i = \frac{(\sum_i a_i^2)^2}{\sum_i a_i^4}$$

This represents the average number of latent variables needed to explain the observed variables. While a perfect simple structure solution has a complexity of 1, in the sense that each element would only load on one factor, a solution with uniformly distributed elements has a complexity greater than 1. Thus, it is important that the structure is not simple and perfect, because then the reduction of dimensions would not make sense, and so: (i) the higher c_i the better the quality of the variable to participate in the factorial extraction; and (ii) the average complexity index should be higher than unity.

To confirm the factorial consistency, further statistics related to the confirmatory factor analysis are analysed.



Graph 1. Criteria for the confirmation of the number of factor for CIS3259
Source: *own compilation*

The graph $COR\_MAP(r_{ij\cdot y})$ shows the partial correlations of the factorial model as the extraction progresses:

$$r_{ij\cdot y} = \frac{r_{ij} - r_{iy}r_{jy}}{\left[\left(1 - r_{iy}^2\right)\left(1 - r_{jy}^2\right)\right]^2} \in [-1,1] \quad (equation\ 11;\ Velicer\ 1976)$$

The numerator contains the covariance between each pair of variables(i, j), which will decrease as factors are extracted because less and less variance is left unexplained by the

model. The denominator contains the two correlation terms for each variable as well as the component that is being removed. In the proposal, both numerator and denominator have been reduced with the number of factors extracted, which has progressively explained the variance of the model.

At a certain point, the denominator has started to decrease more than the numerator, because more of each variable's own variance is being removed than of the joint variance explained by the factorial model, thereby increasing the partial correlation. The MAP 1976 chart (Velicer 1976b) indicates the optimal number of factors for different levels of complexity. Velicer's MAP criterion computes the mean of the partial correlations as the extraction progresses, inviting stopping when more variance is removed from the variables themselves than from the common variance explained by the model. In this case, the value is reached in six factors where the common variance is proportionally higher than the variance of the variables themselves.

The goodness-of-fit measures of the factorial model are expressed in the RMSEA(Smith *et al* 1998), FIT (Clark & Bowles 2018) and BIC (Price 2013)plots, as follows:

$$RMSEA = \sqrt{max\left[\frac{\frac{\chi^2}{df} - 1}{N - 1}, 0\right]};$$

For a correct fit of the model, a value of the $RMSEA$ is required that is less than or equal to $0{,}05$, and if this is not met, then at least less than $0{,}09$ (Tennant & Pallant 2012). *Graph 1* shows the three proposed goodness-of-fit indicators (RMSEA, BIC and FIT) and shows a clear similarity to the elbow rule that is usually applied to the sedimentation plot for the selection of the number of factors when there is a sharp change in slope.

Interpretation of the factors

Simultaneously, the factors are named according to the behaviour of the variables saturating in the extraction, while the multivariate normality of the distribution is assessed. In the research, across all the contrasts for the variables, the Kolmogorov-Smirnov test (Stephens 1979) rejected the null hypothesis that the observed variables are univariate normal with a significance level $\alpha = 0'05$, in fact, only two factors passed this test (MLE1 and MLE4). This one-dimensional test, however, is very restrictive from the multivariate point of view, which is why we use the criterion of skewness and kurtosis, taking advantage of the sample size, and the result shows that all the shape statistics of the variables that participate in the model belong to the interval [-1,2], with the exception of the variable p4_3 in factor 6. Therefore, in general, the joint distribution of the factors is not rejected as multivariate normal.

The *Tables 3 to 8* show the shape statistics, together with the variables ordered from highest to lowest factorial loadings in absolute value; therefore, the first variable will always be the one that best explains the factor (proximate variable). The scale refers to the order in which the question is asked in the survey. Thus, the variable closest to the MLE1 factor is *P1_3* "satisfaction with respondent's social life". From the relationship between the variables, their interpretation is inferred and is referred to as factor 1.

Table 3. Factor 1. Happiness due to social integration and personal well-being: measures of form, factorial loadings, and scale (MLE1).

Label	Asymmetry	Kurtosis	Loading Factorial	Scale
P1_3. Their social life	- 0.886	1.088	0.760	(+)
P1_4. Their standard of living	- 0.439	0.08	0.702	(+)
P2. Personal happiness scale (0-10)	-0.665	0.598	0.670	(+)
P1_1. Their family life	-1.233	2.534	0.654	(+)
P1_2. Their health	-0.987	1.271	0.630	(+)

Source: *own compilation*

Factor 2 saturates questions from group P27* (*Table 4*) which reflects the respondent's level of intolerance towards actions related to tax fraud and which are allegedly carried out by third parties (they are formulated in an impersonal way). The variables that saturate in the factor have the following taxes in common: Spanish Personal Income Tax (IRPF), Value Added Tax (IVA) and Corporate Income Tax (IS), but tax evasion (P27_9), one-off offences such as being paid "under the table" (P27_2) or receiving a payment one is not entitled to (P27_4) are not covered. In summary, P27 focuses on intolerance towards those who are obliged to declare taxes, in the terms described in *Law 58/2003 of 17 December 2003*.

Table 4. Factor 2. Taxpayer intolerance of tax evasion: measures of form, factorial loadings, and scale (MLE2)

Label	Asymmetry	Kurtosis	Loading Factorial	Scale
P27_1. Failure to declare all income for income tax purposes (IRPF)	-1.367	1.826	0.685	(−)
P27_5. Being self-employed and not charging VAT	-1.248	0.881	0.684	(−)
P27_3. Taking a deduction that you are not entitled to when paying taxes (VAT or personal income tax return)	-1.422	2.253	0.677	(−)
P27_8. For a small business to avoid or avoid paying corporation tax	-1.723	2.95	0.642	(−)
P27_7. For a large company to avoid or evade paying corporate income tax	-1.984	3.177	0.565	(−)

Source: *own compilation*

Factor 3 saturates variables from group P8* (*Table 5*) which captures the respondent's judgement on whether the tax revenue they pay is adequately administered in relation to policy needs. This factor does not include welfare state policies, the maintenance of state security forces or the creation or maintenance of infrastructure and is limited to long-term public policies.

Table 5. Factor 3. Perception of the need for long-term public policy funding: measures of form, factorial loadings, and scale (MLE3)

Label	Asymmetry	Kurtosis	Loading Factorial	Scale
P8_14. Science and technology research	-1.626	1.756	0.678	(−)
P8_13. Development cooperation	-0.862	-0.246	0.669	(−)
P8_12. Environmental protection	-1.230	0.502	0.657	(−)
P8_7. Culture	-0.880	-0.256	0.579	(−)
P8_8. Housing	-0.996	-0.028	0.569	(−)

Source: *own compilation*

Factor 4 saturates variables from group P6* (*Table 6*), which reflects the level of satisfaction on a reverse scale with the functioning of various public services. Those covered by the factor are directly related to the welfare state.

Table 6. Factor 4. Dissatisfaction with the functioning of welfare state public services: measures of form, factorial loadings, and scale (MLE4)

Label	Asymmetry	Kurtosis	Loading Factorial	Scale
P6_6. Social Services	0.194	-0.413	0.708	(−)
P6_9. Aid to dependent persons	-0.197	-0.413	0.649	(−)
P6_4. Administration of Justice	-0.309	-0.365	0.624	(−)
P6_5. Public Safety	0.324	-0.35	0.603	(−)
P6_2. Health care	0.123	-0.309	0.553	(−)

Source: *own compilation*

In factor 5 (*Table 7*), the variable with the highest factorial loading is "household income" (P51), followed by "personal income" (P52), and then the respondent's level of educational qualification and status as estimated by the CIS.

Table 7. Factor 5. Social lift: measures of form, factorial loadings, and scale (MLE5)

Label	Asymmetry	Kurtosis	Loading Factorial	Scale
P51. Household income	-0.043	-1.182	0.794	(+)
P52. Personal income	0.605	-0.494	0.698	(+)
Respondent's education [recoded].	-0.317	-1.093	0.651	(+)
STATUS Socio-economic status of the respondent [recoded].	0.122	-1.378	-0.568	(−)

Source: *own compilation*

Factor 6 saturates questions from group P4* (*Table 8*), which reflects the respondent's demand for the qualification of "good citizen" due to their moral and civic behaviour. In this factor, the variable with the highest factorial loadings reflects tolerance towards "the opinions of others, even if they are different from one's own" (P4_5), followed by variables referring to one's own responsibilities: "responsibility and honesty" (P4_6), " compliance with laws and regulations" (P4_4) and "no tax evasion" (P4_3).

Table 8. Factor 6. Civic morality: measures of form, factorial loadings, and scale (MLE6)

Label	Asymmetry	Kurtosis	Loading Factorial	Scale
P4_5. Respect the opinions of others, even if they differ from one's own.	-1.118	0.845	0.784	(+)
P4_6. Be a responsible and honest person	-1.352	1.287	0.709	(+)
P4_4. Always comply with laws and regulations	-1.208	1.147	0.626	(+)
P4_3. Do not evade tax	-1.94	4.982	0.462	(+)

Source: *own compilation*

One-dimensional significance tests for the reference period

The target questions of the *CIS3259* survey, which consist of four categories, are recoded into two categories for simple hypothesis testing. In addition, the variables in the questionnaire appear in reverse scale, so it is necessary to recode them directly. The hypothesis of equality of means is then tested, provided that the univariate normality test has

been passed, or that the factors or variables come from a multivariate normal distribution following the criterion of skewness and kurtosis.

In order to find differences between dichotomous categories, we proceeded to carry out contrasts of mean differences of the factors regarding these, and only the contrasts where the mean difference was significant are shown, as indicated in the *Tables 9 to 13*. The selection of the contrast method is as follows: i) when the Levene's test (Levene 1960) for equality of variances has been passed, the Student's t-test is performed, and ii) otherwise, the Welch's asymptotic test is performed(1938).

According to the results shown in *Table 9*, the most tax-conscious citizens: (i) are happier, by 0.21 Standard Deviations (SD), and (ii) have a lower perceived need for long-term public policy funding, by 0.19 SD, and (iii) are lower on the social ladder, by 0.17 SD.

Table 9. Student's t-test for equality of means of the factors on RP_16. Do you think that Spaniards, when it comes to paying taxes, are very aware and responsible, quite aware and responsible, not very aware or not at all aware and responsible? (0: slightly or not at all aware, 1: fairly or very aware)

Variable	Sig.	Mean Difference	IC (95%)	
			Lower	Upper
F1. Happiness	0.02	-0.21	-0.34	-0.08
F3. The need for long-term public policy financing	0.004	0.19	0.06	0.32
F5. Social elevator	0.014	0.17	0.03	0.30

Source: *own compilation*

Respondents who consider themselves very aware and responsible when paying taxes (*Table 10*) are less tolerant of tax fraud (0.42 SD) and are more satisfied with the functioning of public services of the welfare state (0.35 SD).

Table 10. Student's t-test for equality of means of the factors on RP_17. Do you personally consider yourself to be very aware and responsible, quite aware and responsible, not very aware and responsible or not very aware and responsible? (0: slightly or not at all aware, 1: fairly or very aware)

Variable	Sig.	Mean Difference	IC (95%)	
			Lower	Upper
F2. Intolerance towards tax fraud	0	-0.42	-0.63	-0.21
F4. Dissatisfaction with the functioning of welfare state public services	0.001	0.35	0.14	0.56

Source: *own compilation*

Respondents who considered that there is a lot of tax evasion (*Table 11*) were more likely to value the financing of long-term public policies (0.48 SD) and were less satisfied with the services provided by the welfare state (0.49 SD), making them morally committed citizens with tax compliance.

Table 11. Student's t-test for equality of means of the factors on RP_18. In your opinion, do you think that in Spain there is a lot of tax fraud, quite a lot, a little or very little tax fraud? (0 = a little or very little, 1 = a lot or quite a lot).

Variable	Sig.	Mean Difference	IC (95%)	
			Lower	Upper
F3. The need for long-term public policy financing	0.005	-0.48	-0.82	-0.15
F4. Dissatisfaction with the functioning of welfare state public services	0.004	-0.49	-0.83	-0.15

Source: *own compilation*

Respondents who felt that a lot of people report all their tax income (*Table 12*) appreciated that there is a greater need for public policy funding in the long term (0.24 SD), and they were also slightly higher on the social scale (0.16 SD).

Table 12. Welch's contrast for equality of means of the factors on RP_19. Among the people you know, how many people do you actually declare all their income when filing their personal income tax return (IRPF): all or almost all, quite a lot, some or none?" (0 = some or none, 1 = almost all or quite a lot).

Variable	Sig.	Mean Difference	IC (95%)	
			Lower	Upper
F3. The need for long-term public policy financing	0.004	-0.24	-0.41	-0.08
F5. Social elevator	0.024	-0.16	-0.31	-0.02

Source: *own compilation*

Finally, respondents who thought that those who have more pay more (*Table 13*) were more dissatisfied with the functioning of the welfare state's public services (0.44 SD).

Table 13. Student's t-test for equality of means of the factors on RP_15. And do you think that, in general, taxes are levied fairly, that is, that those who have more pay more, or do you think that this is not the case?

Variable	Sig.	Mean Difference	IC (95%)	
			Lower	Upper
F4. Dissatisfaction with the functioning of welfare state public services	0	-0.44	-0.65	-0.23

Source: *own compilation*

3.2. Simultaneous Factorial Analysis for Surveys CIS3259 and CIS2910

In this second stage, we face the problem of missing or differently coded variables.

A comparison of the two surveys shows that there are missing variables to reproduce the *CIS2910* factorial model, and only factors F3 and F4 can be reconstructed. Thus, we proceeded to recode the 2019 survey and the 2011 variables that are comparable in a similar way, leaving a total of 4,932 cases, and then proceeded to the maximum likelihood extraction in the same way as in the previous section.

After selecting complete cases for the extraction of both factors, a sample size of 2447 was left, which verifies the criteria of Hogarty (2005), due to the fact that practically all the variables have a skewness and kurtosis coefficient belonging to the interval [-2,2].

The *Table 14* shows the main validation statistics for the following factorial analyses: (i) EFA MLE6 is the model for the reference period; (ii) EFA MLE 2 is the reconstructed factorial analysis with the variables that participated in EFA MLE 6, including all cases from both years; (iii) EFA *CIS3259* is the reconstructed model, but only with the cases from the *CIS3259* questionnaire; and (iv) EFA *CIS2910* is the reconstructed model but only with the cases from the *CIS2910* questionnaire.

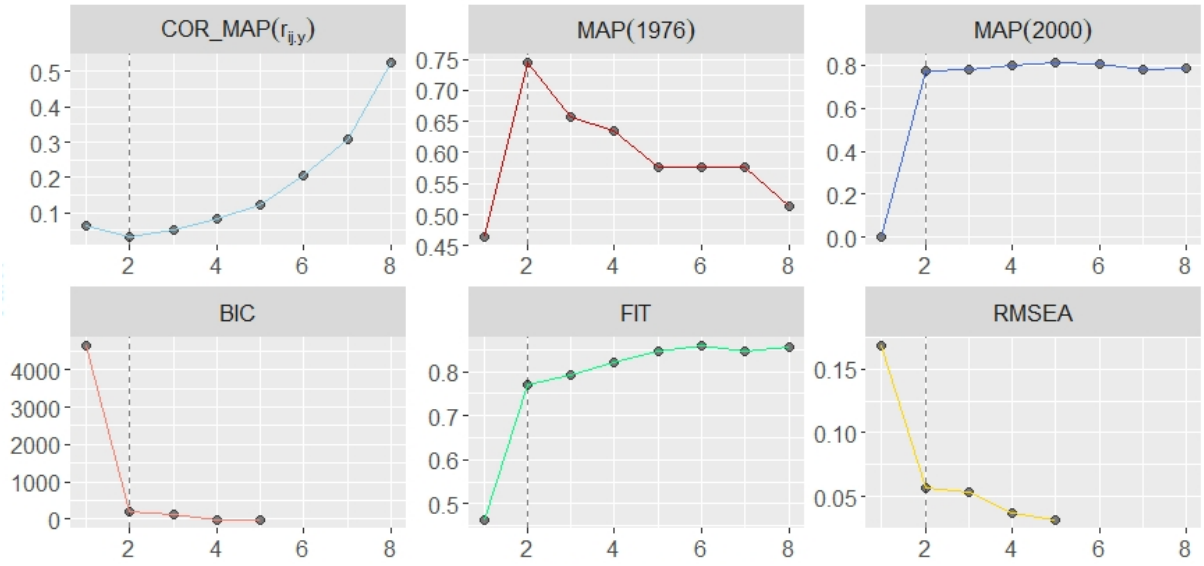
Considering the small number of variables involved in the extraction of the replicated factorial model (EFA MLE 2), the determinant of the correlation matrix is close to zero. Thus, all models are validated (in this section, most *MSA* are above 0.8), although there is a certain decrease in the goodness of fit of the model referring to the *CIS2910* survey (from 2011), compared to the model for *CIS3259* (from 2019), which is directly related to the first model (EFA MLE 6) that only considers the structure for 2019.

Table 14. Statistics and validation contrasts of the factorial models of both periods and of the joint factorial model

	EFA MLE 6 CIS3259	EFA MLE 2 CIS3259 CIS2010	CIS3259	CIS2910
N = complete cases	719	2474	1297	1177
$\|\boldsymbol{R}\|$	0.000	0.099	0.07	0.153
$\boldsymbol{KMO}$	0.818	0.802	0.829	0.757
$\boldsymbol{Sig.Bartlett}$	0	0	0	0
Mean Item pomplexity	1.1	1	1	1
Number of sufficient factors (Velicer test)	6	2	2	2
$\boldsymbol{RMSEA}$	0.046	0.053	0.052	0.058
$\boldsymbol{IC_{\alpha=0'1}(RMSEA)}$	(0′043 ,0′048)	(0′048 , 0′058)	(0′046, 0′059)	(0′052, 0′065)
RSMR	0.02	0.03	0.03	0.03
RSMR Corrected	0.03	0.04	0.03	0.04
BIC	-383.16	166.36	-1.23	39.51
FIT (*based off upon diagonal values*)	0.99	0.99	0.99	0.98
NNFI	0.909	0.940	0.947	0.915

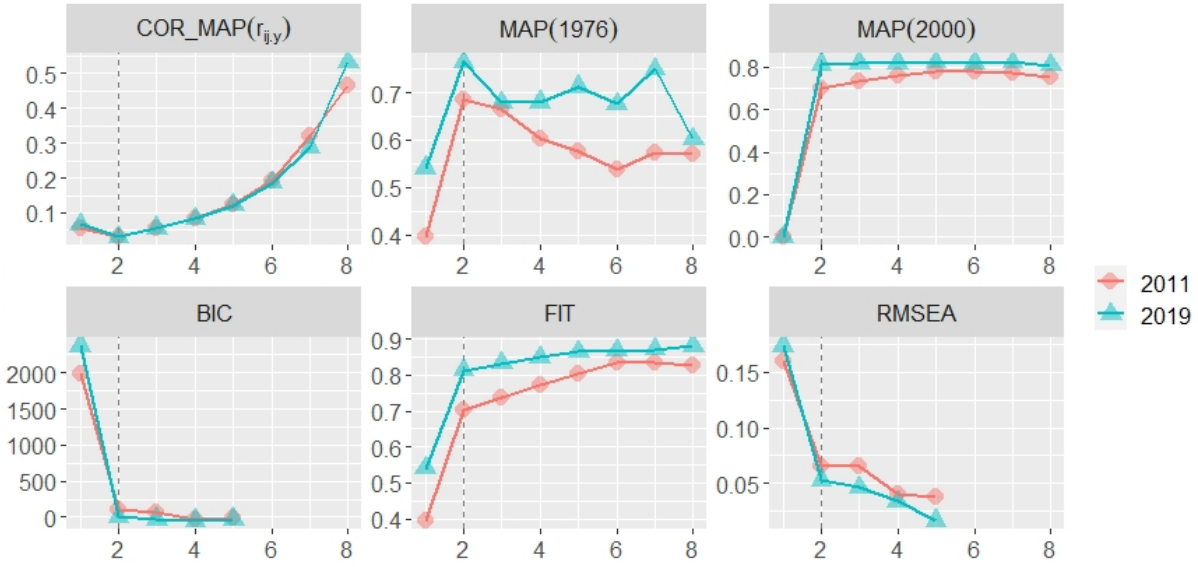
Source: *own compilation*

Regarding the selection of the number of factors, *Graph 2* is interpreted in the same way as for the factorial model for the reference period in *Graph 1*. It can also be seen how the principle of parsimony is reflected in the elbow of the goodness-of-fit plots. On this occasion, Velicer's MAP criterion indicates the selection of two factors in both the original and revised versions.



Graph 2. Criteria for confirmation of the number of factors for the EFA MLE 2 joint factorial model.
Source: *own compilation*

Finally, the *Graphs 2 and 3* confirm the extraction of two factors, so that the extraction of both factors is consistent with respect to the two years and the common variables.



Graph 3. Criteria for confirmation of the number of factors for the EFA MLE 2 factorial models CIS3259 and CIS2910 separately
Source: *own compilation*

4. Results

This third stage deals with the comparison of the questionnaires when the factorial structure and the methodology applied for this purpose cannot be reproduced. In this section, only those variables whose difference is significant in terms of hypothesis testing are shown, with special relevance to those representing the 2019 factors (*CIS3259*).

Table 15. Correspondence table for the comparison of variables in 2011 and 2019

2019 Factor	Relation to EFA model MLE 6 CIS3259 (from 2019)	
MLE1	No information Not reconstructible	
MLE2	Proxy	P27_1. Failure to declare all income for income tax purposes (IRPF)
MLE3	Reconstructed	F3. The need for long-term public policy financing
MLE4	Reconstructed	F4. Dissatisfaction with the functioning of welfare state public services
	Conceptual	P44. Assessment of current personal economic situation
MLE5	3rd in load (recoded)	MIX
MLE6	No information Not reconstructible	

Source: *own compilation*

The best possible comparison is obtained by retrieving the 2019 factorial structure: this is the case for MLE3 and MLE4. The second best alternative is to compare proximate variables: this is the case with MLE2. The third best option is to compare those variables that, having participated in the factor, have the highest possible factorial loadings with respect to the 2019 model: this is the case with F5, represented by MIX. The fourth option is the conceptual comparison; although the variable does not appear in the original factor, there is a parallelism of theoretical content between the latent variable and the observed variable in both

periods: in this case, P44 (*CIS2910*) is conceptually close to the variables P51(*CIS3259*) "household income" and P52(*CIS3259*) "personal income".

In accordance with the methodology formulated above, *Table 16* shows the respondents' self-identification variables relevant for the analysis of tax fraud, and other important aspects.

Table 16. Table of target variables in both periods and other variables of interest

	Variable
Target variables	P10. Consideration received by the company for the payment of taxes and contributions to public authorities
	P18. Extent of tax fraud that exists
	P24. Assessment of the government's efforts to combat tax fraud
Curiosity	P35. Age of respondent
Complementary to F4	P8_1. Education, P8_2. Public works, P8_3. Unemployment protection P8_4. Defence, P8_5. Public security, P8_6. Health, P8_9. Justice P8_10. Social Security/Pensions, P8_11. Transport and communications P8_15. Aid to dependent persons

Source: *own compilation*

The most relevant differences for the years 2011 and 2019 are reflected in *Table 17*, which shows the heterogeneity of the unit of measurement, from ordinal variables to variables such as age with unit of measurement in years, and even in direct and inverse scales. It is also observed that: It is also observed that: (i) regarding the perceived need for long-term public policies, the average age of the surveyed population has increased, and (ii) there is a correlation between higher educational attainment and better personal financial status, with a decrease in the perceived need for public works financing.

Table 17. Student's t-test or Welch's asymptotic test for significant mean differences (MD) for both periods

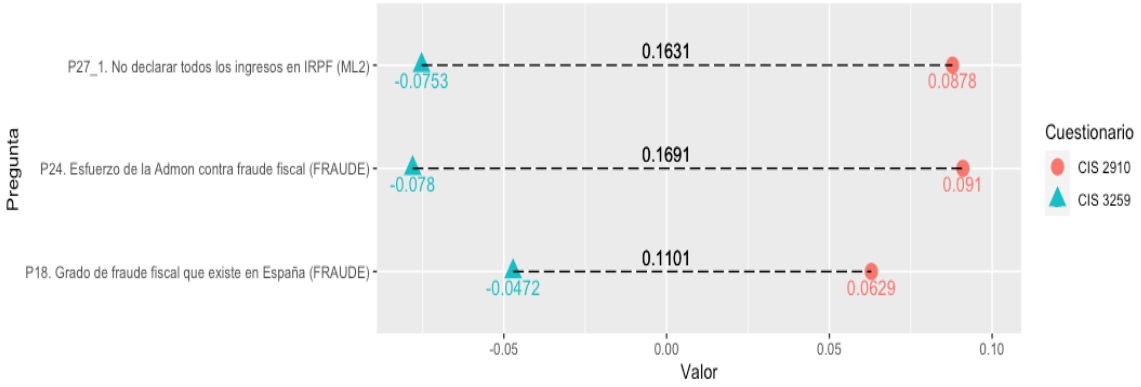
	P8_2. Public works	Survey *CIS3359* on *CIS2910* scale	F3. The need for funding in long-term public policies	P35. Age of respondent	P44. Situación económica personal (INV)
Unit of measurement	Likert: (3 categories)	Likert (5 categories)	SD	Years	Likert (5 categories)
Mean *CIS2910*	1.98	3.0772	-0.2717347	47.16	2.94
Mean *CIS3259*	2.25	3.5713	0.2465935	50.62	2.16
Difference	0.27	0.4941	0.5183282	3.46	-0.78

Source: *own compilation*

The *Graphs 4 to 7* show the positions of the mean of each variable in both questionnaires, taking as the unit of measurement the standard deviations with respect to the mean of the variables typified with scale sign correction (z^*) in the set of cases that make up both years. This measure makes it possible to assess which have been the most relevant changes in the average behaviour of the variables.

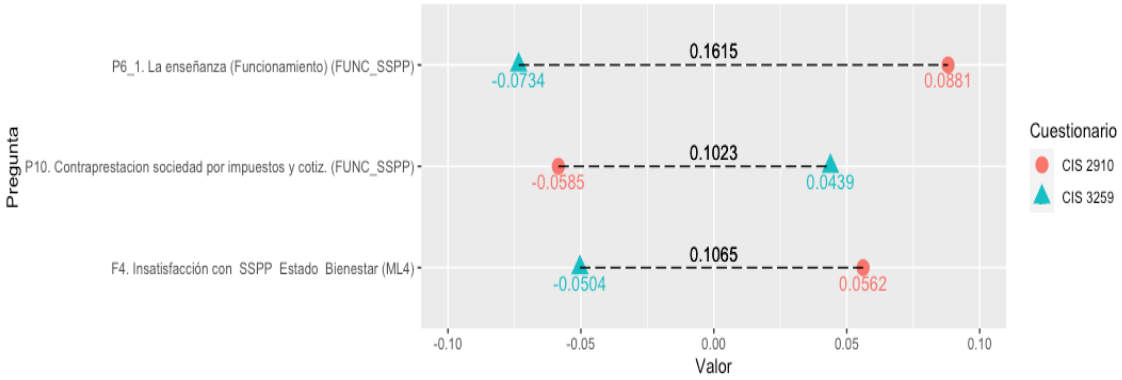
The *Graph 4* shows insignificant differences in the tax fraud perception variables (a maximum of 0.17 SD). The scores *Z* must be interpreted taking into account that *P27_1, P24* and *P18* are on a reverse scale. This means that: (i) tolerance towards not declaring all income in personal income tax has increased (*P27_1*); (ii) the perception of the Administration's

effort to fight tax fraud has increased (*P24*), and (iii) the perception of the degree of tax fraud that exists in Spain has increased (*P18*).



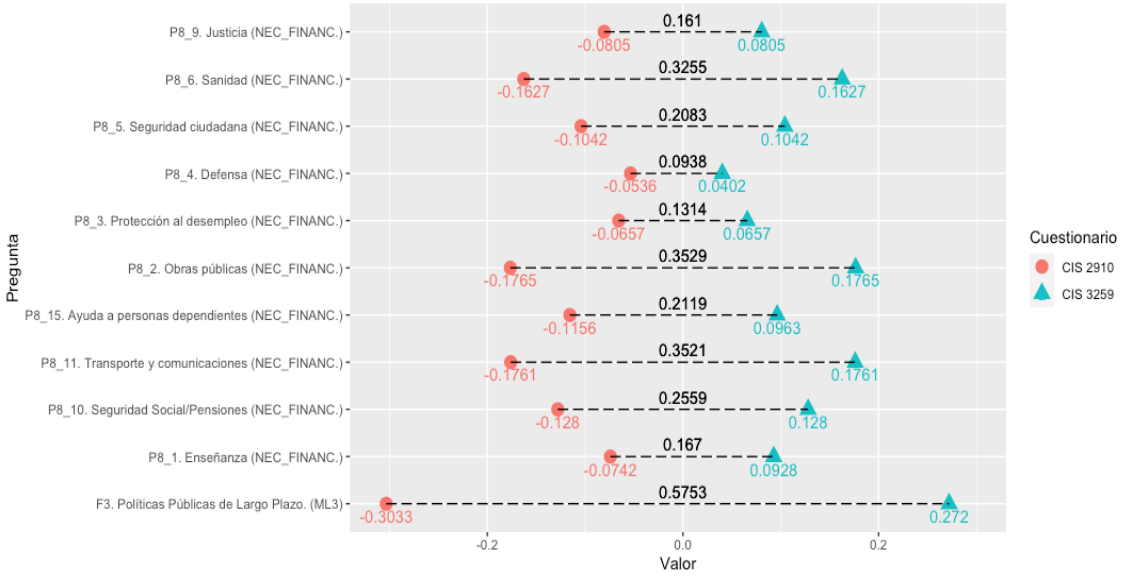
Graph 4. Statistically significant differences: tax fraud CIS2910 and CIS3259 in Z-scores*
Source: *own compilation*

P6_1 and P10 are also on a reverse scale. *Graph 5* therefore shows a slight improvement in satisfaction with education (P6_1) and a slight decrease in the feeling of being compensated for the payment of taxes and social security contributions. Regarding factor 4, satisfaction with public welfare state services has increased slightly. Like the interpretation of *Graph 4*, the differences are not particularly significant.



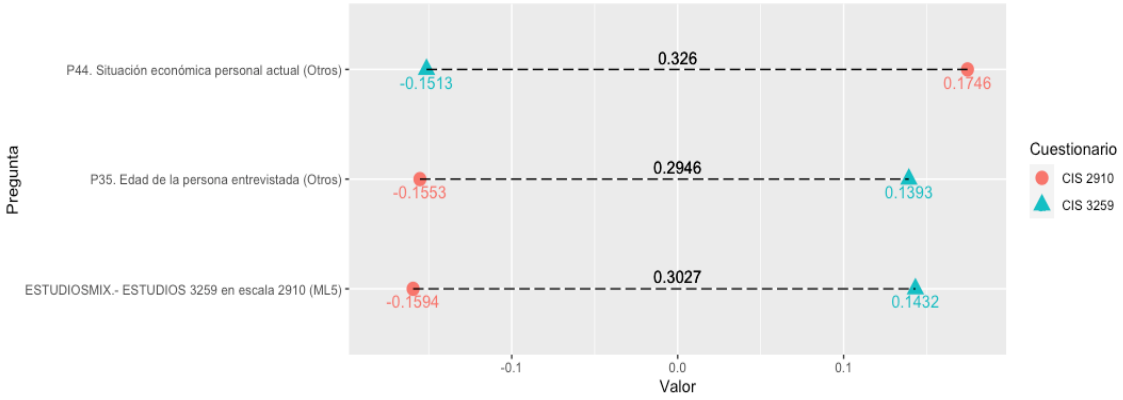
Graph 5. Significant differences: performance of public services CIS2910 and CIS3259 in Z-scores*
Source: *own compilation*

The *Graph 6* shows the questions *P8_**, which reflect the opinion on whether the public resources allocated are insufficient (on a reverse scale) for the maintenance of the public services mentioned above. In general terms, the perception of the need for funding of these public services has increased. Factor 3 has increased, reflecting a significant increase in the perception of the need to allocate resources to long-term planning, indicative of greater state intervention; this difference is very notable because it represents a shift in the centre of gravity of the distribution of factorial scores by almost 0.6 SD.



Graph 6. Significant differences: need for public service funding CIS2910 and CIS3259 in Z-scores*
Source: *own compilation*

Given the impossibility of reconstructing factor 5 (social elevator) from the baseline study, two nearby variables, P44 (reverse coded) and Studies-Mix, are used. Regarding *Graph 7*, the personal economic situation has improved (*P44* is on the reverse scale), the population has aged, and the level of education attained has increased, on average by about 0.3 SD.



Graph 7. Significant differences: personal profile CIS2910 and CIS3259 in Z* scores
Source: *own compilation*

Conclusion

Throughout the research, the methodological difficulties described in the introduction have been confirmed. With these considerations in mind, the paper explains the process of transition from exploratory analysis (Conway & Huffcutt 2003) to confirmatory analysis. The methodological proposal is a procedure which, in effect, makes it possible to recover the factorial structure, even if only partially, to make comparisons between variables, and to recode the units of measurement and scale, all of this in surveys carried out on a public policy in two periods/years. To verify this relationship, the usual requirements for the selection of variables participating in the exploratory model have been increased, both in the value of the

factorial loading (to consider that the variables are saturated by the factor, the loading should be of at least 0.7), and in the improvement of the communalities in the maximum-likelihood extraction, which together has allowed us to compare the exploratory model with contrasts and criteria of the appropriate number of factors, typical of confirmatory factor analysis and SEMs.

The above results confirm the predominance of the principle of interpretability in survey reconstruction over exploratory criteria (Kaiser, sedimentation plot and TVE). The number of factors selected for the interpretability criterion is supported by the results of the number of factors according to the criteria of the confirmatory factorial analysis through the tests (COR_MAP, MAP76 and MAP2000) and the goodness-of-fit statistics (RMSEA, FIT and BIC).

In the selection of the number of factors it has been shown that the elbow rule of the sedimentation plot can be extrapolated to the plots of the confirmatory statistics (RMSEA, FIT and BIC) depending on the number of factors. However, two widely accepted criteria in SEM modelling need to be taken into consideration: (i) strictness with the separation in the factorial loadings, where a criterion of a difference of 0.3 between the maximum loading and the previous one is proposed, and (ii) improvement in the communalities in the maximum likelihood extraction.

The results achieved show that the elbow rule applied to the confirmatory statistics for the number of factors (RMSEA, FIT and BIC – *graphs 1, 2 and 3*), the number of factors recommended by the application of the principle of parsimony is in line with the correlation selection criteria of Velicer $\left[min\left(r_{ij\cdot y}\right)\right]$, MAP (1976), MAP (2000). The Hofmann complexity index is particularly interesting for detecting the least suitable variables to be excluded from the factorial model from a quantitative point of view, irrespective of the interpretability principle.

Finally, the proposed methodology tests the consistency of the factorial structure for the reconstruction of an earlier period (*CIS2910*) with respect to a reference period (*CIS3259*), by allowing: (i) partial reconstruction of the factorial model, (ii) in the absence of the proximate variable, replace it with the common variable with the highest factorial loading; (iii) use the saturations of the variables to weight their level of importance in each period, and (iv) check whether there is a weakening or improvement in the quality of the factorial model (RMSEA, FIT and BIC) in the period in which it is reconstructed. By virtue of the above, the validity of the methodological approach described in section 2 for the selection of those differences in means that are significant is confirmed, and also that the transformations of the variables in the form z_i^*considerably facilitate the interpretation and comparison between these differences.

References

Bartlett, M. S. (1951). A Further Note on Tests of Significance in Factor Analysis. *British Journal of Statistical Psychology*, *4*(1), 1-2. https://doi.org/10.1111/j.2044-8317.1951.tb00299.x

Centro de Investigaciones Sociológicas (2011). 2910| OPINIÓN PÚBLICA Y POLÍTICA FISCAL (XXVIII). Madrid: CIS. https://www.cis.es/cis/export/sites/default/-Archivos/Marginales/2900_2919/2910/es2910mar.pdf

Centro de Investigaciones Sociológicas (2019). Estudio Nº 3259 Opinión Pública y Política Fiscal (XXXVI). Madrid: CIS. https://www.cis.es/cis/opencm/ES/1_encuestas/estudios/ver.jsp?estudio=14464

Clark, D. A., & Bowles, R. P. (2018). Model Fit and Item Factor Analysis: Overfactoring,

Underfactoring, and a Program to Guide Interpretation. *Multivariate Behavioral Research*, *53*(4), 544-558. https://www.tandfonline.com/doi/full/10.1080/00273171.2018.1461058

Conway, J. M., & Huffcutt, A.I. (2003). A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research. *Organizational Research Methods*, *6*(2), 147-168. https://doi.org/10.1177/1094428103251541

Hofmann, R. J. (1978). Complexity And Simplicity As Objective Indices Descriptive Of Factor Solutions. *Multivariate Behavioral Research*, *13*(2), 247-250. https://doi.org/10.1207/s15327906mbr1302_9

Hogarty, K. Y., Hines, C.V., Kromrey, J.D., Ferron, J.M., & Mumford, K.R. (2005). The Quality of Factor Solutions in Exploratory Factor Analysis: The Influence of Sample Size, Communality, and Overdetermination. *Educational and Psychological Measurement*, *65*(2), 202-226. https://doi.org/10.1177/0013164404267287

Kaiser, H.F. (1958). The Varimax Criterion for Analytic Rotation in Factor Analysis. *Psychometrika*, *23*(3), 187-200. https://doi.org/10.1007/BF02289233

Kaiser, H.F. (1970). A Second Generation Little Jiffy. *Psychometrica*, *35*(4), 401-415. https://doi.org/10.1007/BF02291817

Levene, H. (1960). Robust for Equality of Variances. pp. 278-292. In: Ingram Olkin, Harold Hotelling et al. (eds). *Contributions to Probability and Statistics in Honor of Harold Hotelling*. Stanford: Stanford University Press. https://www.semanticscholar.org/paper/Contributions-to-Probability-and-Statistics%3A-Essays-Burkholder-Olkin/fd2d9dfc0226f1ae63f18e16a3489a33dd538525

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, *22*(140), 55. https://psycnet.apa.org/record/1933-01885-001

Mahalanobis, P.Ch. (1936). On the Generalized Distance in Statistics. *Proceedings of the National Institute of Science of India*, *2*(1), 49-55. http://bayes.acs.unt.edu.8083/BayesContent/class/Jon/MiscDocs/1936_Mahalanobis.pdf

Rabadán-Pérez, F., Berumen, S. A., Guiance-Lapido, J., & Hernández-Mora, C. (2022). Reconstrucción y consistencia factorial: la regla del codo aplicada al RMSEA, análisis paralelo y otras pruebas confirmatorias. *Revista De Métodos Cuantitativos Para La Economía Y La Empresa*, *33*, 353-385. https://doi.org/10.46661/revmetodoscuanteconempresa.5464

Rencher, A.C. (1995). *Methods of Multivariate Analysis*. New York: John Wiley & Sons. https://www.semanticscholar.org/paper/Methods-of-multivariate-analysis-(second-edition)-Rencher/a83cfec9c23390a10e5c215c375480b8cd3a1565

Smith, R.M., Schumacker, R.E., & Bush, M.J. (1998). Using Item Mean Squares to Evaluate Fit to the Rasch Model. *Journal of Outcome Measurement*, *2*(1), 66-78. PMID: 9661732. https://pubmed.ncbi.nlm.nih.gov/9661732/

Spearman, C. (1927). The Abilities of Man their Nature and Measurement. *Nature*, *120*(3014), 181-183. https://doi.org/10.1038/120181a0

Stephens, M.A. (1979). Test of Fit for the Logistic Distribution Based on the Empirical Distribution Function. *Biometrika*, *66*(3), 591-595. https://doi.org/10.2307/2335180

Tennant, A., & Pallant, J.F. (2012). The Root Mean Square Error of Approximation (RMSEA) as a Supplementary Statistic to Determine Fit to the Rasch Model with Large Sample Sizes. *Rasch Measurement Transactions*, *25*(4), 1348-1349. https://www.rasch.org/rmt/rmt254d.htm

Velicer, W.F. (1976a). Determining the Number of Components from the Matrix of Partial Correlations. *Psychometrika*, *41*(3), 321-327. https://doi.org/10.1007/BF02293557

Velicer, W.F. (1976b). The Relation Between Factor Score Estimates, Image Scores, and Principal Components Scores. *Educational and Psychological Measurement*, *36*(1), 149-159. https://doi.org/10.1177%2F001316447603600114

Welch, B.L. (1938). The Significance of the Difference Between two Means When the Population Variances are Unequal. *Biometrika*, *29*(3-4), 350–362. https://doi.org/10.1093/biomet/29.3-4.350