Rethinking the exploration of dichotomous data: Mokken Scale Analysis vs. Factorial Analysis

Abstract

The need to determine the correct dimensionality of theoretical constructs and generate valid measurement instruments when underlying items are categorical has generated a significant volume of research in the social sciences. This paper presents two studies contrasting different categorical exploratory techniques. The first study compares Mokken Scale Analysis (MSA) and two factor-based exploratory techniques for non-continuous variables: Item Factor Analysis (IFA) and Normal Ogive Harmonic Analysis Robust Method (NOHARM). Comparisons are conducted across techniques and in reference to the common Principal Component Analysis (PCA) model using simulated data under conditions of two-dimensionality with different degrees of correlation (r = .0 – r =.6). The second study shows the theoretical and practical results of using MSA and NOHARM (the factorial technique which functioned best in the first study) on two non-simulated data sets. The non-simulated data is particularly interesting because MSA was used to solve a theoretical debate. Based on the results from both studies, we show that the ability of NOHARM to detect dimensionality and scalability is similar to MSA when the data comprise two uncorrelated latent dimensions; however, NOHARM is preferable when data are drawn from instruments containing latent dimensions weakly or moderately correlated. The paper discusses the theoretical and practical implications of these findings.

Keywords: *Mokken Scale Analysis, Principal Component Analysis, NOHARM, factor analysis, dimensionality.*

Introduction

Phenomena studied in social sciences are complex, so theoretical controversy often abounds with regard to the internal structure (dimensionality) of different concepts and constructs[i]. Exploratory techniques are frequently employed by researchers to study latent structures and their relation with the theories concerned; meanwhile, empirical evidence for the dimensionality of the instruments feed back into the theory, providing evidence to corroborate or invalidate theoretical propositions. In social sciences field research, several alternative approaches have been suggested for factor analyzing dichotomous data. In this article, we use simulations and re-analyses of existing data to delineate the relative strengths and weaknesses of alternative factor analytic approaches designed for use when the underlying measured items are dichotomous.

We also contrast alternative factor analytic approaches to Principal Component Analysis (PCA) – a widely used default option in numerous statistical packages (Conway and Huffcutt 2003). The methodological literature clearly documents how PCA tends to identify spurious factors when used with categorical data with non-centered distributions – an issue that is especially relevant in the case of dichotomous variables with a large number of items (Bernstein and Teng 1989). Note, however, that while we briefly review known issues associated with PCA and use PCA as a comparison approach to help illustrate our ideas, our focus centers on comparing other proposed analytic approaches.

Dichotomous variables are widely employed in different sociological and political domains. For example, they have been used to study the construct of political efficacy (Mokken 1971) and the construct of attitudes toward welfare and social spending (Jacoby 1994; these two examples will be used as practical cases in study two of the present paper, as explained below). Dichotomous variables are also quite common in other domains. For instance, they have been employed in different work and organizational literatures to evaluate the presence or absence of certain organizational opportunities (O'Reilly and Chatman 1986) and workplace accommodations (Konrad, Moore, Doherty, and Breward 2013). In human resources research, dichotomous variables have been used to study organizational performance (Delaney and Huselid 1996) and technological mechanisms for inter-organizational coordination (Bensaou and Venkatraman 1995). In occupational health literature, dichotomous measures have been used for assessing work intensification (Bamberger et al. 2016), work safety climate (Arcury et al. 2015) and need for recovery (De Vries, Michielsen, and van Heck 2003). Finally, in the military contexts dichotomous measures have been used to study self-reported symptoms (Kelton et al. 2010) and combat experiences (Wilk et al. 2010). What is common across all these examples is that items were measured dichotomously, yet theory suggests that items likely cluster in meaningful ways.

To explore the internal structure of dichotomous item scales, several authors have proposed using the exploratory Mokken Scale Analysis (MSA) solution (e.g. Hemker, Sijtsma, and Molenaar 1995; Mokken 1971; van der Eijk and Rose 2015; van Schurr 2003), placing the MSA at the center of a methodological debate about the comparative advantages of different exploratory techniques used to identify the dimensionality of data (i.e. Finch 2010, 2011; Kuijpers, van der Ark and Croon 2013; Tate 2003; van Abswoude, van der Ark and Sijtsma 2004; Wismeijer, Sijtsma, van Assen and Vingerhoets, 2008).

This paper extends previous literature comparing MSA with factorial analysis models designed to treat ordinal and dichotomous data. Specifically, we examine an Item Factor Analysis or IFA (Wirth and Edwards, 2007) which uses the tetrachoric or polychoric correlation matrix between the items as the main source of information to develop the factor analysis, and a Non-linear Factor Analysis approach, the Normal Ogive Harmonic Analysis Robust Method (NOHARM), which applies a harmonic analysis based on the bivariate proportion of item successes to perform the factor analysis (McDonald 1997).

As a second contribution, we compare MSA and factor analysis techniques under conditions of multi-dimensionality, specifically consisting of instruments with two latent dimensions which are either uncorrelated (r = .00), or correlated (*r* ranging from 0.1 to 0.6). We believe both extensions are important because multi-dimensionality is common in different social sciences constructs (Robert, Keohane, and Moravcsik 2011; Sieberer 2011) and MSA has been employed under various circumstances in sociological and political literature (see examples below) as well as other social sciences literatures (for example: Gonzalez Roma, Schaufeli, Bakker and Lloret 2006).

We address these objectives through two studies. The first compares MSA to IFA and NOHARM (and to PCA as a counterpoint) in a simulation study. The second study examines the practical relevance of the results obtained from the first study by applying MSA and NOHARM (the factor analysis technique which functioned best) to two applied cases using non-simulated data. As previously mentioned, the first case consists of the original example that Mokken used to present the model in 1971, when he studied the political efficacy construct (Mokken 1971). The second case consists of an open debate on the construct attitudes toward welfare and social spending (Jacoby 1994), where different authors have debated the dimensionality of the construct based on different exploratory studies. The examples were chosen primary because a) the dimensionality of both constructs is a matter of theoretical debate presumably resolvable by applying MSA, b) data are publicly available, and c) the data are of historic and current interest.

In sum, the two studies contained in the paper are intended to provide social scientists with recommendations and guidelines on the potentialities and weaknesses of

MSA compared to factor solutions designed to explore the dimensionality and structure of categorical data.

Theoretical Background

Default exploratory analysis: Traditional Principal Component Analysis (PCA)

An exploratory analysis of the internal structure of a data set is a process to identify the latent factors present and was originally based on the covariance matrix or Pearson's correlations. Although PCA is not strictly speaking a latent factor detection technique (Fabrigar Wegener MacCallum and Frin 1999), it is commonly used in many applied situations because it often provides results similar to those from linear factor analysis (Velicer and Jackson 1990). PCA on a Pearson's correlation matrix is routinely used in combination with the Kaiser's rule, which consists of retaining those factors or components with eigenvalues greater than 1. Finally, the relationships between each item and the factors (also called the item's factor/component loading) is identified using rotation procedures, which are applied to obtain a more easily interpretable solution and can also be employed to detect independent factors (orthogonal rotation) and associated factors (oblique rotation). Normally, applied researchers use default orthogonal rotations like varimax (Conway and Huffcutt 2003). We refer to this specific type of PCA (person's correlation, Kaiser's rule and varimax rotation) as the "traditional PCA", which is also known as "Little Jiffy" (Kaiser 1970; Gorsuch 1990), and in the next paragraph we briefly review the known issues with this approach.

In traditional PCA, the underlying variables must be continuous for the technique to operate optimally. When PCA is used on ordinal or dichotomous data the correlation magnitude (Pearson coefficient or *phi* in the case of dichotomous data) between two or more items depends not only on the actual magnitude of the relationship, but also on the response rate for each of the alternatives in each item. This implies that the correlations obtained for items with very different response rates will tend to be understated, while items with similar response rates will tend to display higher levels of association (McDonald 1999). Overall, differences in prevalence will produce sub-groups of items which are erroneously associated more closely with each other than with other variables potentially producing extra factors (Bernstein and Teng 1989). Spurious factors can lead investigators to believe that theoretical constructs are overly complex; to order subjects on the basis of non-existent latent dimensions, and to eliminate from the instrument certain items which may be informative for the theoretical construct measured merely because they appear to be associated with latent dimensions other than those to which they actually belong. Additionally, PCA detects principal components rather than latent factors and it overstates explained variance (Fabrigar et al. 1999) which can lead applied researchers to overestimate the quality of their instruments.

More recent variants of the PCA have been developed as a response to these limitations to include replacing the matrix of Pearson’s correlations with the tetrachoric or polychoric correlations matrix (Kolenikov and Angeles 2004; 2009) and the non-linear version of the PCA (Mori, Kuroda, and Makino 2016). Similarly, other researchers recommend using PCA in combination with more sophisticated techniques instead of the Kaiser’s rule to decide the number of components to be retained (see Hayton, Allen and Scarpello 2004). Nevertheless, these alternatives are not often employed by applied researchers who continue using what is called “Little Jiffy” (Kaiser 1970; Gorsuch 1990). This is due to several reasons: first, PCA in its Little Jiffy version is the default option in several statistical packages (Conway and Huffcutt 2003). Second, applied researchers may be unlikely to use analytical techniques if the statistical packages are not userfriendly (Allison 2012). For these reasons, we include traditional PCA (aka, the "Little Jiffy") as a baseline comparative technique.

IFA and NOHARM

In an effort to respond to the limitations of classical factor analysis techniques and PCA, factor-based alternatives have been designed and proposed for applications to non-continuous data since at least the mid 1970s (Christofferson 1975; Muthen 1978). These include IFA, which replaces the matrix of Pearson covariances and correlations between items with a polychoric or a tetrachoric correlation matrix, thereby allowing the application of adequate factor extraction techniques (Flora and Curran 2004; Forero, Maydeu-Olivares and Gallardo-Pujol 2009; Muthen 1984), including the DWLS (Diagonally Weighted Least Squares) estimation which is one of the most widely recommended (Flora and Curran 2004) and which we use in this study.

IFA provides precise results and largely avoids the problem of detecting extra factors in many circumstances (Forero et al. 2009). However, the literature warns that IFA is not entirely free of problems. For example, Flora and Curran (2004) note that: 1) the tetrachoric and polychoric correlation matrices cannot always be inverted; 2) small sample sizes can result in unreliable estimates of the degree of association between variables; and 3) the tetrachoric and polychoric correlations appear to be poor indicators of the association between items when they are highly asymmetrical.

An alternative solution to IFA was proposed by McDonald (1967; 1999) under the name NOHARM which was developed from a general non-linear factor analysis model and allows estimation of the parameters of the multidimensional normal ogive model in Item Response Theory for binary data (equivalent to a non-linear common factor model, McDonald 1982; 1999).

NOHARM, (as IFA) is a partial information procedure (bivariate information methods) that approximates the IRT normal ogive model with a polynomial function that expresses the probability of the correct answer to the dichotomous item in terms of a linear, quadratic and product term of (multiple) abilities. It estimates the common factor model assuming that the analyzed (dichotomous) n - items arise by dichotomizing a n-variate normal density. The estimation of the parameters occurs through a double step procedure that employs the first and second order marginal of the contingency table, applying an iterative process based on Unweighted Least Squares to allow convergence towards the values that best explain the observed matrix of conjoint success between the items. (Additional technical details of this model can be found in McDonald 1967; 1999; as well as in Maydeu-Olivares 2001). The iterative procedure produces final parameters which can easily be transformed into the number of factors and factor loadings of a conventional factor model, including alternative solutions for orthogonal and oblique rotations.

Exploratory application of NOHARM entails obtaining the factor solution for models with an increasing number of factors, and assessing each one using goodness-of-fit indices based on the size of the unexplained residuals left by the model. According to the authors of the model, this empirical criterion must always be cross-checked against the theoretical sustainability of the latent factors identified when using applied data (McDonald and Mok 1995; McDonald 2000).

Various simulations (e.g. Maydeu-Olivares 2001; Tate 2003, Finch 2010; 2011) have demonstrated the capacity of NOHARM to detect dataset dimensionality. These studies also show that the quality of the solutions obtained by NOHARM are usually similar to or better than IFA, and the model has the additional advantage of providing good estimates even in small samples with as little as 200 cases without suffering from the convergence and estimation problems which affect the IFA procedure (McDonald 1999).

The Mokken Scale Analysis

As a response to traditional PCA limitations, several authors propose using MSA (Hemker, et al. 1995; van der Eijk and Rose 2015; van Schuur 2003). MSA is a unidimensional procedure developed by Mokken (1971) as a probabilistic version of Guttman scalogram analysis (Guttman 1950). Given its characteristics and properties, it is treated in the literature as a Non-Parametric Item Response Theory (Sijtsma and Molenaar 1987). van Schuur (2003) provides a detailed description of the technical details and mathematical bases for the functioning of the MSA model; however, we briefly highlight technical features essential for the application of MSA in this paper.

The main index used to assess the dimensionality of the data is the Coefficient of Homogeneity (H_{ij} index), which can be defined as the ratio of the covariance between any pair of items and the maximum value which that covariance can take considering the response frequency distribution for the two items (van Schuur 2003). This index shows the degree to which two items are associated and expresses the degree to which they may belong to the same latent dimension.

Based on the H_{ij} values for each pair of items, an index H can be obtained for the association of each item with the rest of the scale (H_i index) and for the total scale (H_t index), which would express the mean level of the relationship between the items and can therefore serve as an indicator of unidimensionality for a set of items (Kuijpers et al. 2013). This index reaches a value of 1 when the scale is perfectly unidimensional, but it has conventionally been held that H_t-values equal to or above a lower boundary, c, (usually set at .30) indicate a scale of adequate quality (Mokken and Lewis 1982). Hence, if the H_t-value for a set of variables is slightly lower than the cut-off point, it is common to eliminate items with low H_i scores until H_t is equal to or higher than .30.

This procedure is not problem-free because in the case of multiple latent variables models, the values of the H_{ij} indexes will depend on two types of relationship: a) on the one hand, if the two items pertain to the same latent variable, the H_{ij} index will reflect the strength of the factorial loading between each item and the (common) latent variable; b) on the other hand, if the two items pertain to two different latent variables, the H_{ij} index will reflect the strength of the factorial loading of each item with its respective latent variable multiplied by the correlation between the two latent variables. Paradoxically, a problematic situation appears when the correlation between latent variables is high and when the items are of high quality (high factorial loadings). In this case, two or more items belonging to different latent dimensions may obtain a sufficiently high H_{ij} among themselves (i.e. above .30) as the product of correlations between the different latent dimensions with which they are associated. As a consequence, MSA would erroneously indicate the existence of a unidimensional structure in multidimensional situations.

Present study

We present two studies in the next sections to extend the scope of the comparisons between MSA and other exploratory techniques described in the literature to situations of multidimensionality. In the first, we analyze simulated data, while in the second we use real data from two applied cases where dimensionality was disputed. We end the paper with a general discussion highlighting the practical implications of our findings and offering some guidelines for social scientists. We also describe the limitations inherent in the present study and our recommendations for future research.

Study 1

Method

Data generation.

A Monte Carlo study was conducted. Data were generated using the R software (R Development Core Team, 2012), according to the following factor model:

$$X_{ij} = \sum_{k=1}^{k} \lambda_{jk} \times F_k + \left(1 - \sum_{k=1}^{k} \lambda_{jk}\right)^{1/2} \times \delta_{\mathrm{j}}$$

where X_{ij} corresponds to the simulated response of the subject *i* in the ítem *j*, λ_{jk} is the factor loading of the ítem *j* in the factor *k* (*k* = 1, 2). F_k corresponds to latent factor with distribution N(0,1) y δ_j is the error of measurement with distribution N(0,1)[1].

The continuous simulated responses were dichotomized using different threshold to each item for generate a wide range of difficulties. See details below.

Simulated conditions.

Bi-dimensional tests[2] of 22 items were generated (11 items by factor), considering two sample sizes (n=500, n=1000) and seven levels of correlation between the factors, since ρ = 0.0 to 0.6. We decided to work with 11 items in each factor so that we could simulate 8 items with heterogeneous and asymmetrical difficulty (from p = 0.10 to 0.40 and p = 0.60 to 0.90), with 3 additional symmetric items (p = .5), obtaining a symmetrically balanced subscale with a number of items for each factor that is common in social sciences (Marsh Morin Parker and Kaur 2014; Tabachnick and Fidell 2001). Regarding the correlations between factors, according to Gursoy, Chi and Karadag (2013) we simulated up to 0.6, since higher values (i.e. 0.7 or more) imply that the two factors

[1] The proposed factorial method is equivalent to the IRT two parameter normal ogiva model, (McDonald 1999)
[2] In the current study we only included results related to instruments containing two-dimensional latent structures. Nevertheless, we found our results to be consistent also in the case of three-dimensional latent structure.

share more than the half of their variance, so they could be modelled as a single factor[3]. Finally, regarding the sample size, we simulated samples with 500 because the literature suggests that different factorial techniques start to present recovery problems with smaller samples. We used 1000 to provide a large sample example noting that the advantages of additional sample size beyond 1000 is quite low (Asún, RDZ-Navarro and Alvarado, 2016). Note that samples within the range of 500 to 1000 are reasonably common in the social sciences (Conway and Huffcutt 2003).

The items of each test were simulated with homogeneous ability to represent the latent variable, according to three factor loadings ($\lambda = 0.5, 0.7$ y 0.9), corresponding to items of low ($h^2 = 0.25$), mean ($h^2 = 0.49$) and high ($h^2 = 0.81$) quality. Meanwhile, the items were simulated with heterogeneous difficulty (from $p = 0.10$ to 0.90), so that there is an item at each level of difficulty in each test, except the intermediate level ($p = 0.5$) where three items were created. In each condition, 500 replicates were performed, so the results represent averages and standard deviations obtained from these conditions.

Data analysis.

The PCA and IFA procedures were run using the package Psych version 1.5.8 for R (Revelle 2015); MSA was estimated using the package Mokken version 2.7.7 for R (van der Ark 2007); and the NOHARM model was run on version 1.8-9 of package Sirt for R (Robitzsch 2014).

To compare the functioning of the different techniques we looked at the capacity of each one to obtain a solution which fits the data (according to the goodness-of-fit criteria for each technique) and adequately reproduce the simulated structure based on

[3] Beside the reported results, we simulated additional conditions for correlations above 0.6 and found that a coherent patterns of results,(with the ones we report in the paper) so that the techniques that presented recovery problems (traditional PCA, MSA) with lower correlations didn't work, while NOHARM reached a good functioning till correlation of 0.9 and IFA reached good functioning till correlation of 0.9, but only for samples of 1000.

two criteria: a) the latent structure identified, which must coincide with the two latent dimensions actually simulated; b) the quality of recovery of the structure simulated, based on the simulated lambda.

The empirical acceptability of the solutions obtained was assessed on the following basis: a solution was considered acceptable in the case of traditional PCA if the factors were retained employing the Kaiser rule (selected because it is the most commonly used criterion in practice). For MSA, according to Mokken and Lewis (1982), the threshold for acceptability was a H_i-value equal to or greater than .30. In the case of IFA and NOHARM we used the criteria suggested by Schreiber et al. 2006, who indicated that both RMSEA and SRMR should present values less than or equal to .08. This double criterion is widely accepted in the methodological literature and widely applied by social researchers (Hu and Bentler 1999).

Results

Highly discriminating items.

For test compounds for items in the highly discriminating condition ($\lambda = .9$) the PCA analysis detected (mistakenly) the presence of 4 components regardless of the relationship between the factors. We observed that when $\rho = 0$, the variance of the first component was 25.7% and 26.1% for $n = 1000$ and $n = 500$ respectively, while it increased to 38.4% when $\rho = .6$.

In contrast, the MSA clearly detected the two-dimensional nature of the data when factors were independent (detecting two 11 items scales, $H_t = 0.8$); however, we observed that from $\rho \geq .3$ and especially from $\rho \geq .4$, items tended to cluster in a single scale regardless of the size of sample.

The IFA procedure retrieved two factors irrespective of the size of the sample and the correlation between factors, with the only difference that lambdas tended to be slightly higher with $n = 1000$. In all simulated conditions, the SRMR adjustment and RMSEA indices indicated a proper fit.

Finally, NOHARM recovered two factors in all conditions, with the lambdas closest to the simulated values (λ between 0.89 and 0.90). In all conditions SRMR indices and adjustment RMSEA indicated an excellent fit.

Moderately discriminating items.

For test compounds for items moderately discriminating ($\lambda = .7$) the PCA analysis detected (mistakenly) the presence of 3 or 4 components. We observed that when $\rho = 0$, the variance of the first component was 17.2% and 17.6% for $n = 1000$ and $n = 500$ respectively, which increased to approximately 24.7% when $\rho = .6$. The MSA clearly detected the two-dimensional nature of the data with ρ between 0 and .4 independent of sample size. However, with $\rho > .4$ the items tended to be grouped into a single scale.

For its part, regardless of the sample size, the IFA procedure retrieved two factors in all conditions of ρ, estimating lambdas from 0.67 to 0.71. In all simulated conditions, SRMR and RMSEA indicated a good fit, taking maximum values of between 0.042 and 0.057 respectively in the case of $n = 1000$, and 0.057 and 0.079 respectively for $n = 500$.

Finally, NOHARM recovered two factors in all conditions regardless of the sample size and ρ, and the estimated lambda values were closest to the simulated values (λ between .68 and 0.7). In all simulated conditions, SRMR and RMSEA indicated an excellent fit, taking values of the 0.004 and 0.006 respectively for $n = 1000$, and values of the 0.005 and 0.008 respectively for $n = 500$.

Poor discriminating items.

With data composed of items with λ = 0.5, the PCA erroneously detected the presence of between 5-6 components. The performance of MSA was not much better than PCA in that MSA indicated the presence of about 5 factors in all simulated conditions.

For its part, the IFA procedure recovered two factors independently of ρ and the sample size, although tending to underestimate the lambdas of the second factor, especially with n = 500 (a condition in which IFA obtained lambdas of between .46 and .49). When n = 1000, the indices of RMSEA SRMR indicated a proper fit. However, with n = 500 RMSEA reached average values of between 0.097 and 0.098, indicating poor fit.

Finally, NOHARM recovered two factors in all conditions regardless of the sample size and ρ and estimated the lambda values closest to the simulated values (λ between .48 and .5). In all simulated conditions, SRMR and RMSEA indicate an excellent fit, taking values of the 0.005 and 0.022 respectively for n = 1000, and values of the 0.007 for both indices with 500 cases.

Detailed results for all the exploratory techniques are offered in Appendix, while table 1 offers a comparative review for the dimensionality recovery.

--- Insert Table 1 ---

Study 2

This section describes two applied examples of the measurement of constructs with dimensional structures that are, or have been, the subject of controversy. The examples show that the choice of one or other analytical technique can have practical and theoretical repercussions for applied scientists. In the examples, we contrast results a social scientist would obtain applying MSA and NOHARM (the factor procedure which worked best in study 1). In light of the results obtained from study 1, we specifically question the empirical contributions made to the literature based on the application of MSA, insofar as this technique could incorrectly have identified a unidimensional solution in the presence of a potential correlation between two latent dimensions. We then briefly discuss the possible implications for the debate over the constructs examined in the literature.

Case 1: Measurement of Political Efficacy

The dimensional structure of the "political efficacy" construct has been the subject of an interesting debate since the construct was first introduced over fifty years ago (Campbell, Gurin and Miller 1954). The construct is conceived as a latent expression of passivity in general responses to politics (Campbell, Converse, Miller and Donald 1966). Gamson formally defined the term as: "the efficacy dimension of political alienation refers to people's perception of their ability to influence" (Gamson 1968: 42). This one-dimensional conceptualization is empirically supported by the analysis performed by Mokken (1971) using MSA in the classical measurement handbook presenting the model, in which the author examined the dimensionality of an instrument designed to measure political efficacy in the Dutch political system (Mokken 1971: 255).

A different view is found, however, in the pioneering work of Lane (1959), who argued that the concept can in fact be split into at least two basic components, comprising internal and external efficacy. Internal efficacy refers to people's views of their own competences and capacities to understand and participate effectively in the political forum, while external efficacy refers to beliefs about the responsiveness of governmental authorities and institutions to citizen's demands. This multidimensional vision has received empirical support in recent decades (see, Niemi, Craig and Mattei 1991; Dyck and Lascher 2008).

Data

We obtained the original data from the investigation of political efficacy in the Dutch political system (van der Maesen 1967), as did Mokken (1971). The original instrument contained 9 items designed for dichotomous responses, to which van der Maesen added a further 8 items of a similar nature focusing specifically on the political system in the city of Amsterdam. The complete instrument comprising 17 items was applied to a sample of 1,513 Amsterdam voters in 1966. We used the matrix of item clusters proposed by Mokken (1971: 282-283).

Results

When Mokken (1971) applied MSA to the dataset he obtained an H_t scale coefficient of .39 for the 17-item instrument (see Table 5), which led him to conclude that the theoretical construct was unidimensional. In light of the results shown in table 5, the unidimensionality of the structure would have been confirmed if later recommendations to raise the cut-off point c when using exploratory MSA (Hemker, Sijtsma and Molenaar 1995) had been followed. This would have eliminated items relating to both internal and external efficacy, merging the remaining items in a single scale.

Our replication using NOHARM, found satisfactory SRMR values (.019); however, we found that the fit of the unidimensional model was inadequate[ii], as the RMSEA values were over the cut-off point (RMSEA = .09). In contrast, the two-dimensional solution displays satisfactory fit (SRMR = .012; RMSEA = .059) suggesting the presence of two moderately correlated but separately distinguishable factors (r = .63).

Additionally, chi squared difference between one-dimensional and two-dimensional solutions was significant ($\chi^2(16) = 870$; $p < .000$) confirming the substantial improvement of fit between both alternatives. Finally, following the recommendation to judge the interpretability of the factors obtained in theoretical terms (McDonald and Mok 1995; McDonald 2000), we went on to examine the content of the items. As shown in table 5, the first factor groups all of the items related with external political efficacy, while the second brings together the internal political efficacy items, providing further evidence supporting the two-dimensional structure.

--- Insert Table 5 ---

Case 2: Measurement of attitudes toward Welfare and Social Spending

An interesting debate has emerged over the last two decades with regard to the dimensional structure of a construct assessing "attitudes toward welfare and social spending". Jacoby (1994) originally posited and then found empirical evidence for the unidimensional structure of this construct, but Goren (2008) has recently called this conclusion into question using data from three National Election Studies Surveys (1992; 1996 and 2000), showing that various factor analyses for categorical data pointed to the presence of a two-dimensional structure with factors correlated at .67. Specifically, the author differentiated between two facets of the construct: a) attitudes towards Social Spending and b) attitudes toward Welfare. In the same year, Jacoby (2008) responded by re-analyzing the data from the 1996 National Election Study Survey employed by Goren (2008), showing that the relevant variables formed a clear one-dimensional structure, having obtained an H_t-value of .43 for the scale using MSA.

Data

We carried out our analysis using the same data as Jacoby (2008). Like Jacoby, we transformed the original 6 polytomous questions (which allowed three response options: reduce spending, maintain spending, or increase spending) into 12 new dichotomous items using a procedure based on the cumulative logic of the response categories for the original questions. Thus, we assumed that an individual who is in favor of raising spending (response = 3) will necessarily also be in agreement with maintaining it (response = 2 for the same question). Hence, each question can be split into two: agreement or disagreement with maintaining spending (with a response value of 0 or 1), and agreement or disagreement with increasing spending (with a response value 0 or 1). The data consisted of 1,385 cases.

Results

The application of MSA confirmed the results reported by Jacoby, consisting of a total H-value of .43 forming a single scale of 12 items, all of which had an H_i-value greater than or equal to .3. In principle, this should lead us to accept the presence of a unidimensional structure. Following the proposal made by Hemker et al. (1995) to raise the value of c, the scale loses two items between c = .30 and .40. When c = .45, however, the scale splits in two (one scale with 9 items, one with 2 items and another unscaled item), and at higher values for c the instrument fragments into 4 or 5 scales of differing sizes. Accordingly, this is the pattern to be expected in a one-dimensional instrument, as the items do not form a consistent two-dimensional structure.

Applying NOHARM to the same data (see table 6), we found an inacceptable fit with the one-dimensional structure, as the RMSEA value are over the cut-off point (SRMR = .012, RMSEA = .09). However, the bi-dimensional structure (with a moderate correlation between factors, *r* =.69) presented a clearly superior fit to the data (SRMR =

.006, RMSEA = 0.067). As in the first example, the chi squared difference between one-dimensional and two-dimensional solutions was significant ($\chi^2(11) = 426.2$; $p < .000$) confirming the substantial improvement of fit between both alternatives. Finally, following McDonald and Mok (1995) and McDonald (2000), we checked for the theoretical sustainability of the latent factors identified, and found that the first factor was composed of items related to welfare spending, while the second one for items related to social spending, coherent with Goren's (2008) proposition.

--- Insert Table 6 ---

Discussion

In line with previous literature on the subject, this study shows that traditional PCA (based on Pearson correlations) is not an adequate procedure to establish the dimensionality of instruments formed by categorical variables, especially dichotomous variables. In our simulated data, PCA frequently identified numerous spurious factors depending on the difficulty of the items (Bernstein and Teng 1989).

On the other hand, MSA correctly identified the dimensionality of the data in the presence of uncorrelated latent dimensions. This result is consistent with the findings reported by van Schuur (2003) for unidimensional structures, and it was expected in a multidimensional scenario with uncorrelated latent dimensions because any correlation between the items (measured by the H_{ij} index) will reveal the measurement of the respective latent dimension.

The situation changes, however, where correlation is present in a latent two-dimensional structure. In this scenario, MSA tends erroneously to group all of the items in a single scale, especially when the factor correlation is greater than .3. Our results are in line with the findings reported by van Abswoude, et al. (2004), who observed a tendency to lump items together in a single scale as the correlation between latent dimensions increased. We may note here that correlation between latent dimensions results in relatively high H_{ij}-values for pairs of items belonging to different dimensions, so that the model tends incorrectly to include them in the same scale.

The erroneous grouping effect tends to occur frequently wherever intermediate or high loadings items are found together with moderately correlated latent dimensions. Such situations are common in practice; therefore, MSA does not appear to be an adequate technique to explore the dimensionality of instruments with a latent structure that is potentially multidimensional.

Regarding the factorial analysis, our results show that the use of IFA substantially improves on traditional PCA. In our first study, IFA detected the bidimensional structure of the data under all the simulated condition, with low error in lambda estimation. IFA only offered an unacceptable solution in the case of n = 500 and low quality items, where the RMSEA index indicated weak fit to the data. Finally, NOHARM was able to correctly detect the structure of the data with a practically perfect empirical fit under all of the conditions analyzed. A possible explanation of this pattern of results is the following. A possible explanation of this pattern of results is the following. Even if both techniques uses partial information, the functioning of IFA depends on the (correct) estimation of the tetrachoric correlation which can become unstable with small samples and in the presence of asymmetric items (Flora and Curran, 2004), while NOHARM uses a polynomial approximation that is computationally faster and not affected by the non-convergence and estimations problems of tetrachoric correlations (McDonald, 1967, 1997).

This study makes a number of contributions from the methodological standpoint. In the first place, to the best of our knowledge, our study is the first to compare MSA with IFA and NOHARM. On the one hand, our results with simulated data are congruent with the findings reported by Finch (2011) and Tate (2003) who showed that NOHARM was better able than IFA to determine the dimensionality of data. More in general, it adds to the actual debate in the literature on the capacity of NOHARM to determine the dimensionality, where several authors tested it against other techniques like TESTFACT (Stone & Yeh 2006; Van der Linden 2016). On the other hand, we build on existing work by showing the superiority of the factorial techniques compared to MSA in both the detection of the dimensionality and the estimation of the item quality in conditions of multidimensionality.

We also add to the existing MSA literature (Mokken 1971) by expanding the available evidence concerning the functioning of the model in the presence of uncorrelated and correlated two-dimensional structures. In line with other studies, we demonstrate that Mokken should be used as a tool to weed out items only after the unidimensionality of the dataset analyzed has been established (Hemker et al. 1995; van Abswoude, et al. 2004).

Above and beyond the contribution made by this study to the methodological literature, in the next paragraphs we focus on the practical implications of our results for social scientists. In the first place, we wish to underscore the consequences of one or other analytical technique, which may be far-reaching at least as regards the theoretical implications of potentially fictitious empirical evidence and the use of subject scores.

In order to demonstrate the applied relevance of the results obtained from the first study, we analyzed data obtained from two real cases. In the first, an applied researcher would conclude that the dataset structure was uni-dimensional after applying MSA to a pool of items from the "political efficacy" construct (Mokken 1971). In light of this first study, we find that MSA could bring together items from two different but correlated latent dimensions to create a fictitious uni-dimensional scale. This would lead a social scientist to take a stand in the theoretical debate by providing empirical evidence for the unidimensionality of the construct analyzed, in line with Campbell et al. (1954) and Gamson (1968). Application of NOHARM, however, revealed the presence of two latent dimensions, providing empirical evidence that is consistent with the arguments supporting the bi-dimensionality of the political efficacy construct (Lane 1959) and discriminating between the internal and external latent dimensions defined in the literature (Niemi, Craig and Mattei 1991; Dyck and Lascher 2008).

In the second case study, the application of NOHARM detected the presence of a bi-dimensional structure of the analyzed construct, dividing attitudes towards Welfare from attitudes towards Social Spending, in line with work presented by Goren (2008). The relevance of this bi-dimensional structure is discussed by the author that offers empirical evidence and a theoretical discussion in this line.

These results clearly show that the application of certain techniques under inadequate conditions could have led social scientist to erroneously claim empirical evidence supporting one versus other theory. This is the case where MSA was applied in conditions where there may be a correlation between two or more latent dimensions.

To update and rework actual the recommendations in the literature made by van Schuur (2003) and van der Eijk and Rose, (2015), we would recommend social scientists use MSA only when the structure to be refined is clearly unidimensional. However, if researchers suspect that a multidimensional structure with a high degree of correlation between latent dimensions could exist (a likely situation in the social sciences), we would advise against using MSA, or at least that MSA should be used only in combination with other more powerful techniques to detect dimensionality (see, for example, van Abswoude et al. 2004).

NOHARM not only functioned optimally in the situations tested (in terms of both the detection of dimensionality and retention of relevant items), but also offered a series of advantages over MSA in applied terms. First, NOHARM provides more than one goodness-of-fit index, which facilitates assessment of the model's fit with the data, while MSA only provides the H-index, which is not in itself easy to interpret (as it requires modifying c and analyzing the structures obtained). Second, NOHARM allows the correlation between factors to be estimated, providing useful information to study relations with other variables, especially in complex analytical models. On the other hand, since NOHARM only works with dichotomous data, if polytomic information is available, IFA may be a good alternative.

Limitations and future research

Despite the theoretical and practical contributions made by this study, it is affected by certain limitations, which we discuss here together with the related avenues for future research. First, the latent dimensions simulated have the same make-up in terms of the number and distribution of items, so there are no principal and minor latent dimensions. Future research will need to continue comparing the exploratory techniques in the face of changes in these parameters, which may affect their capacity correctly to recognize the dimensional structure.

Second, we worked with only 500 and 1,000 cases in the first study. Additional analyses not described in this paper showed that NOHARM correctly recovers the bidimensional structure with samples as few as 200 subjects, as predicted by McDonald (1999). Nevertheless, with such small samples, we found an increasing presence of Heywood cases associated with higher factorial loading and high correlations between factors. For all these reasons, we consider that future research should deeper investigate the functioning of NOHARM under these conditions.

Third, with regard to generalization of the instrument to other lengths, we understand that the problems of fit are related with the degree of asymmetry between the items, and therefore our results could be sustained to other test lengths providing asymmetric items exist. Empirical support for these claims could be obtained in future research comparing the functioning of the techniques examined in the face of changes in sample and instrument size.

Fourth, the data considered in both studies are dichotomous. However, social scientists very often use measurement instruments based on Likert response formats. Though our study employed a way of examining these data using analytic models designed for dichotomous data (following the recoding procedure proposed in Jacoby 2008), this is an almost unexplored avenue which would benefit from further research.

Fifth, as previously mentioned, in this paper we wanted to test the traditional (and widely employed by applied researchers), PCA, so our results cannot be generalized to the whole family of PCA techniques. In this line, future research should compare NOHARM, MSA and IFA with more recent evolution of the PCA (Kolenikov, and Angeles 2004; 2009; Mori, Kuroda and Makino 2016).

Finally, with regard to MSA and its exploratory use, we followed the well-established procedure of Mokken and Lewis (1982) to study the dimensionality. Future research should explore and test alternative procedures, like the one proposed by Hemkel et al. (1995) which is an exploratory procedure that uses progressively higher thresholds of c to see the impact on the scale configuration. While this procedure does not present a clear rule to establish dimensionality, its effect under the condition studied in the present paper are unknown and (potentially) worthy being studied.

In conclusion, despite the limitations mentioned above, we trust that this study throws light on the pros and cons of applying factor approaches (and specifically NOHARM) to dichotomous data. We believe our results also have the potential to contribute to updating existing recommendations about best practices with respect to examining the dimensionality and factor structure of dichotomous data.

Arcury, T. A., Summers, P., Rushing, J., Grzywacz, J. G., Mora, D. C., Quandt, S. A., ... & Mills, T. H. (2015). Work safety climate, personal protection use, and injuries among Latino residential roofers. *American journal of industrial medicine*, *58(1)*, 69-76.

Allison, P. D. (2012). *Missing data*. Thousand Oaks, CA: Sage.

Asún, R. A., Rdz-Navarro, K., & Alvarado, J. M. (2016). Developing multidimensional Likert scales using item factor analysis: The case of four-point items. Sociological Methods & Research, 45(1), 109-133.

Bamberger, S. G., Larsen, A., Vinding, A. L., Nielsen, P., Fonager, K., Nielsen, R. N., ... & Omland, Ø. (2016). Assessment of work intensification by managers and psychological distressed and non-distressed employees: a multilevel comparison. *Industrial health*, (0).

Bensaou, M., & Venkatraman, N. (1995). Configurations of interorganizational relationships: a comparison between US and Japanese automakers. *Management science*, *41(9)*, 1471-1492.

Bernstein, H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105(3)*, 467-482.

Campbell, A., Gurin, G., & Miller, E.W. (1954). *The voter decides*. Evanston, III. Row, Peterson.

Campbell, A., Converse, P. E., Miller, W. E., & Donald, E. (1966). *The American Voter*. New York: Wiley.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40(1)*, 5-32.

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.

De Vries, J., Michielsen, H. J., & Van Heck, G. L. (2003). Assessment of fatigue among working people: a comparison of six questionnaires. *Occupational and Environmental Medicine*, *60* (suppl 1), i10-i15.

Delaney, J. T., & Huselid, M. A. (1996). The impact of human resource management practices on perceptions of organizational performance. *Academy of Management journal*, *39(4)*, 949-969.

Dyck, J. J., & Lascher E. L. (2009). Direct democracy and political efficacy reconsidered. *Political Behavior, 31(3),* 401-427.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Erin J. S. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods, 4(3)*, 272-295.

Finch, H. (2010). Item Parameter Estimation for the MIRT Model Bias and Precision of Confirmatory Factor Analysis—Based Models. *Applied Psychological Measurement, 34(1)*, 10-26.

Finch, H. (2011) Multidimensional item response theory parameter estimation with nonsimple structure items. *Applied Psychological Measurement, 35(1),* 67-82.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods, 9(4),* 466-483.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16(4),* 625-641.

Gamson, W. A. (1968). *Power and Discontent*. Homewood. IL: Dorsey.

González-Romá, V., Schaufeli, W. B., Bakker, A. B., & Lloret, S. (2006). Burnout and work engagement: Independent factors or opposite poles? *Journal of Vocational Behavior*, *68(1)*, 165-174.

Goren, P. (2008). The two faces of government spending. *Political Research Quarterly, 61(1),* 147-157.

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research, 25(1)*, 33-39.

Gursoy, D., Chi, C. G. Q., & Karadag, E. (2013). Generational differences in work values and attitudes among frontline and service contact employees. *International Journal of Hospitality Management, 32*, 40-48.

Guttman, L. (1950). The principal components of scale analysis. In *Measurement and prediction*. Studies in social psychology in World War II, vol.4, eds. Stouffer, Samuel A.; Louis Guttman; Edward A. Suchman; Paul F. Lazarsfeld; Shirley A. Star and John A. Clausen, 60-90. Princeton: Princeton University Press.

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of Unidimensional Scales From a Multidimensional Item Bank in the Polytomous Mokken IRT Model. *Applied Psychological Measurement, 19(4),* 337-352.

Hayton, J. C., Allen, D. G., & Scarpello, V. 2004. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods, 7(2),* 191-205.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Jacoby, W. G. (1994). Public attitudes toward government spending. *American Journal of Political Science*, 336-361.

Jacoby, W. G. (2008). Comment: The dimensionality of public attitudes toward government spending. *Political Research Quarterly, 61(1),* 158-161.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika, 35(4)*, 401-415

Kelton, M. L., LeardMann, C. A., Smith, B., Boyko, E. J., Hooper, T. I., Gackstetter, G. D., ... & Smith, T. C. (2010). Exploratory factor analysis of self-reported symptoms in a large, population-based military cohort. *BMC medical research methodology*, *10(1)*, 94.

Kolenikov, S., & Angeles, G. (2004). *The use of discrete data in PCA: theory, simulations, and applications to socioeconomic indices*. Chapel Hill: Carolina Population Center, University of North Carolina.

Kolenikov, S., & Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer? *Review of Income and Wealth, 55(1)*, 128-165.

Konrad, A. M., Moore, M. E., Ng, E. S., Doherty, A. J., & Breward, K. (2013). Temporary Work, Underemployment and Workplace Accommodations: Relationship to Well-being for Workers with Disabilities. *British Journal of Management*, *24(3),* 367-382.

Kuijpers, R. E., Van der Ark, L. A., & Croon, M. A. (2013) Standard errors and confidence intervals for scalability coefficients in Mokken scale analysis using marginal models. *Sociological Methodology, 43(1),* 42-69.

Lane, R. E. (1959). *Political life: Why people get involved in politics*. New York: Free Press.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual review of clinical psychology, 10*, 85-110.

Maydeu-Olivares, A. (2001). Multidimensional item response theory modeling of binary data: Large sample properties of NOHARM estimates. *Journal of Educational and Behavioral Statistics, 26(1),* 51-71.

McDonald, R. P. (1967). *Nonlinear factor analysis. Psychometric monographs*, 15.

McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement, 6(4)*, 379-396.

McDonald, R. P. (1997). Normal-ogive multidimensional model. In *Handbook of modern item response theory*, eds. Van der Linden, Wim J., and Ronald K. Hambleton, 257-269. New York: Springer.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement, 24(2),* 99-114.

McDonald, R. P., & Mok, M. C. (1995). Goodness of fit in item response models. *Multivariate Behavioral Research, 30(1),* 23-40.

Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research*. Vol. 1. Berlín: Walter de Gruyter.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 6*, 417-430.Mori, Y., Kuroda, M., & Makino, N. (2016). *Nonlinear Principal Component Analysis and Its Applications*. Singapur: Springer.

Mori, Y., Kuroda, M., & Makino, N. (2016). *Nonlinear Principal Component Analysis and Its Applications*. Springer.

Muthen, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika 43(4),* 551-560.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49(1),* 115-132.

Niemi, R. G., Craig, S. C., & Mattei, F. (1991). Measuring internal political efficacy in the 1988 National Election Study. *The American Political Science Review, 85,* 1407-1413.

O'Reilly, C. A., & Chatman, J. (1986). Organizational commitment and psychological attachment: The effects of compliance, identification, and internalization on prosocial behavior. *Journal of applied psychology*, *71(3)*, 492.

R Development Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Revelle, W. (2015). *Package Psych version 1.5.8. R package*. Retrieved from http://personality-project.org/r/psych.

Robert, O., Keohane, S. M., & Moravcsik, A. (2011). Constitutional Democracy and World Politics: A Response to Gartzke and Naoi. *International Organization, 65*, 599-604.

Robitzsch, A. (2014). *Sirt: Supplementary item response theory models. R package*. Retrieved from https://sites.google.com/site/alexanderrobitzsch/.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99(6),* 323-338.

Sieberer, U. (2011). The institutional power of Western European parliaments: A multidimensional analysis. *West European Politics, 34(4),* 731-754.

Sijtsma, K., & Molenaar, I. W. (1987). Reliability of test scores in nonparametric item response theory. *Psychometrika, 52(1),* 79-97.

Tabachnick, B. G. and Fidell, L. S. (2001). *Using multivariate statistics*. Edinburg: Pearson Education.

Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement, 27,(3),* 159-203.

van Abswoude, A. A., van der Ark, L. A., & Sijtsma, K. (2004). A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement, 28(1),* 3-24.

van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, *20(11)*, 1-19.

van der Eijk, C., & Rose, J. (2015). Risky business: factor analysis of survey data–assessing the probability of incorrect dimensionalisation. *PloS one, 10(3)*, e0118900.

van der Maesen, C. E. (1967). The angry voter in Dutch**.** *Acta Politica, II*. 169-200.

van Schuur, W. H. (2003). Mokken scale analysis: between the Guttman scale and parametric item response theory. *Political Analysis 11(2),* 139-163.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate behavioral research, 25(1),* 1-28.

Wilk, J. E., Bliese, P. D., Kim, P. Y., Thomas, J. L., McGurk, D., & Hoge, C. W. (2010). Relationship of combat experiences to alcohol misuse among US soldiers returning from the Iraq war. *Drug and alcohol dependence, 108(1*), 115-121.

Wismeijer, A. A., Sijtsma, K., van Assen, M. & Vingerhoets, A. J. (2008). A comparative study of the dimensionality of the self-concealment scale using principal components analysis and Mokken scale analysis. *Journal of Personality Assessment, 90(4),* 323-334.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: current approaches and future directions. *Psychological methods, 12(1),* 58.

Table 1: *Comparative review of results for dimensionality recovery.*

Exploratory technique	*Simulated Lambda*	*Simulated Sample size*	*ρ = .0*	*ρ = .1*	*ρ = .2*	*ρ = .3*	*ρ = .4*	*ρ = .5*	*ρ = .6*
PCA	.90	1000	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0.2)
		500	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	3.9 (0.2)
	.70	1000	3.8 (0.4)	3.7 (0.5)	3.7 (0.5)	3.6 (0.5)	3.5 (0.5)	3.3 (0.5)	3.2 (0.4)
		500	3.9 (0.4)	4.0 (0.3)	3.9 (0.4)	3.9 (0.4)	3.9 (0.4)	3.8 (0.5)	3.6 (0.5)
	.50	1000	5.4 (0.6)	5.4 (0.6)	5.4 (0.6)	5.4 (0.6)	5.4 (0.6)	5.4 (0.6)	5.3 (0.6)
		500	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)
MSA	.90	1000	2.0 (0.0)	2.0 (0.0)	2.0 (0.1)	1.6 (0.5)	1.0 (0.1)	1.0 (0.0)	1.0 (0.0)
		500	2.0 (0.0)	2.0 (0.0)	2.0 (0.2)	1.5 (0.5)	1.0 (0.2)	1.0 (0.0)	1.0 (0.0)
	.70	1000	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	1.9 (0.2)	1.1 (0.3)
		500	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	2.0 (0.1)	1.8 (0.4)	1.3 (0.4)
	.50	1000	4.9 (1.0)	5.0 (1.0)	4.9 (1.1)	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.1 (1.0)
		500	4.9 (1.1)	4.9 (1.1)	4.8 (1.1)	4.9 (1.1)	4.8 (1.2)	4.9 (1.1)	4.9 (1.1)
IFA	.90	1000	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		500	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	.70	1000	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		500	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	.50	1000	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		500	a	a	a	a	a	a	a
NOHARM	.90	1000	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		500	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	.70	1000	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		500	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	.50	1000	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		500	2.0	2.0	2.0	2.0	2.0	2.0	2.0

a = unacceptable solution for the two factor model.
Highlighted text: wrong dimensionality recovery solution

Table 2. *Political efficacy scale: results for MSA and NOHARM*

Items	MSA		NOHARM	
	difficulty	H_i index	Loads Factor 1	Loads Factor 2
1.- Members of Parliament don't care much about the opinions of people like me	.25	.40	**.65**	.04
2.- Cabinet Ministers don´t care much about the opinions of people like me	.27	.41	**.77**	-.03
3.- The political parties are only interested in my vote and not in my opinion	.27	.40	**.54**	.22
4.- People like me don't have any say about what the government does	.31	.36	**.56**	.15
5.- If I communicate my views to the municipals authorities they will be taken into account*	.33	.31	**.80**	-.20
6.- The municipal authorities don't care much about the opinions of people like me*	.34	.42	**.93**	-.10
7.- Members of the City Council don't care much about the opinions of people like me*	.35	.42	**.86**	-.04
8.- People like me don't have any say about what the City government does*	.37	.39	**.67**	.09
9.- If I communicate my views to members of the City Council they will be taken into account*	.37	.34	**.84**	-.20
10.- In the determination of city politics , the votes of people like me are not taken into account*	.62	.42	**.53**	.22
11.- In the determination of government policy , the votes of people like me are taken into account	.66	.34	**.53**	.05
12.- Sometimes politics and government seem so complicated that a person like me can't really understand what's going on	.35	.30	-.05	**.70**
13.- Sometimes city politics and government in Amsterdam seem so complicated that a person like me can't really understand what's going on*	.41	.36	.07	**.69**
14.- Because I know so little about politics, I shouldn't really vote	.63	.44	-.17	**1.02****
15.- I wouldn't go to polls, if I weren't obligated to do so	.66	.42	-.02	**.81**
16.- Because I know so little about city politics, I shouldn't really vote in municipal elections*	.68	.50	-.08	**.99**
17.- So many other people vote in the national elections that it doesn't matter much to me whether I vote or not	.79	.50	.22	**.54**

Notes: * local ítems. **This Heywood case is caused by the high correlation (r = .88) of this item with item 16. The difficulty and Hi index are taken from Mokken (1971, p. 273). The loadings for the two-factors solution were calculated from the data provided in Mokken (1971, pp. 282-283).

Table 3. *Attitudes toward the Welfare and Social Spending: results for MSA and NOHARM.*

	MSA		NOHARM, unidimensional model	NOHARM, bi-dimensional model	
Items	difficulty	H_i index	Loads Factor	Loads Factor 1	Loads Factor 2
Maintain spending on food stamps	.50	.37	.62	**0.92**	-0.15
Maintain spending on welfare programs	.40	.42	.69	**0.99**	-0.13
Maintain spending on the homeless	.87	.52	.79	-0.01	**0.82**
Maintain spending on social security	.93	.43	.54	0.12	**0.45**
Maintain spending on child care	.87	.50	.75	0.25	**0.54**
Maintain spending on poor people	.86	.57	.83	0.30	**0.58**
Increase spending on food stamps	.08	.62	.81	**0.70**	0.16
Increase spending on welfare programs	.09	.64	.83	**0.69**	0.20
Increase spending on the homeless	.53	.42	.68	-0.22	**0.93**
Increase spending on social security	.42	.30	.49	-0.05	**0.55**
Increase spending on child care	.46	.38	.63	0.03	**0.62**
Increase spending on poor people	.40	.44	.78	-0.06	**0.88**

Note: The 6 items analyzed from the 1966 CPS National Election Study (1,385 Caucasian individuals) were transformed into 12 dichotomous items following the procedure described in Jacoby (2008).

APPENDIX Table 1. *Means (and standard deviations) for PCA estimations.*

Lambda	*Sample size*		*ρ = .0*	*ρ = .1*	*ρ = .2*	*ρ = .3*	*ρ = .4*	*ρ = .5*	*ρ = .6*
.90	1000	Scales	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0.2)
		% EV F1	25.7 (0.57)	27.2 (0.89)	29.4 (0.94)	31.6 (0.97)	33.8 (0.98)	36.1 (0.98)	38.4 (0.97)
		% EV F2	24.3 (0.54)	22.8 (0.79)	20.6 (0.79)	18.4 (0.75)	16.2 (0.70)	14.0 (0.62)	11.7 (0.53)
	500	Scales	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	4.0 (0)	3.9 (0.2)
		% EV F1	26.1 (0.81)	27.3 (1.24)	29.5 (1.32)	31.7 (1.34)	33.9 (1.35)	36.1 (1.35)	38.4 (1.33)
		% EV F2	24.0 (0.79)	22.7 (1.11)	20.6 (1.09)	18.4 (1.02)	16.2 (0.95)	14.0 (0.85)	11.7 (0.77)
.70	1000	Scales	3.8 (0.4)	3.7 (0.5)	3.7 (0.5)	3.6 (0.5)	3.5 (0.5)	3.3 (0.5)	3.2 (0.4)
		% EV F1	17.2 (0.46)	18.1 (0.63)	19.8 (0.68)	20.7 (0.72)	22.0 (0.74)	23.5 (1.63)	24.7 (0.79)
		% EV F2	16.2 (0.43)	15.3 (0.55)	14.0 (0.54)	12.7 (0.50)	11.4 (0.45)	10.1 (0.65)	8.7 (0.34)
	500	Scales	3.9 (0.4)	4.0 (0.3)	3.9 (0.4)	3.9 (0.4)	3.9 (0.4)	3.8 (0.5)	3.6 (0.5)
		% EV F1	17.6 (0.66)	18.2 (0.90)	19.4 (0.98)	20.8 (1.03)	22.1 (1.09)	23.4 (1.10)	24.7 (1.12)
		% EV F2	16.0 (0.63)	15.3 (0.75)	14.1 (0.75)	12.8 (0.70)	11.4 (0.63)	10.1 (0.56)	8.8 (0.50)
.50	1000	Scales	5.4 (0.6)	5.4 (0.6)	5.4 (0.6)	5.4 (0.6)	5.4 (0.6)	5.4 (0.6)	5.3 (0.6)
		% EV F1	11.1 (0.34)	11.5 (0.43)	12.1 (0.47)	12.8 (0.50)	13.5 (0.52)	14.1 (0.54)	14.8 (0.57)
		% EV F2	10.4 (0.32)	10.0 (0.36)	9.4 (0.36)	8.7 (0.34)	8.1 (0.32)	7.4 (0.29)	6.8 (0.27)
	500	Scales	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)	6.0 (0.1)
		% EV F1	11.4 (0.49)	11.7 (0.60)	12.3 (0.67)	12.9 (0.70)	13.6 (0.72)	14.2 (0.75)	14.9 (0.77)
		% EV F2	10.4 (0.44)	10.1 (0.49)	9.5 (0.50)	8.8 (0.48)	8.2 (0.44)	7.6 (0.41)	6.9 (0.37)

Notes: Scales = Average number of components generated by the procedure. % EV F1 = Percentage of variance explained by the first component. % EV F2 = Percentage of variance explained by the second component. ρ = Pearson correlation between the simulated factors.

APPENDIX Table 2. *Means (and standard deviations) for MSA estimations.*

Lambda	*Sample size*		*ρ = .0*	*ρ = .1*	*ρ = .2*	*ρ = .3*	*ρ = .4*	*ρ = .5*	*ρ = .6*
.90	1000	Scales	2.0 (0.0)	2.0 (0.0)	2.0 (0.1)	1.6 (0.5)	1.0 (0.1)	1.0 (0.0)	1.0 (0.0)
		H1 / Nº Items	.80 (0.0) / 11.0 (0.0)	.80 (0.0) / 11.0 (0.0)	.79 (0.0) / 11.2 (1.2)	.67 (0.1) / 15.6 (5.2)	.53 (0.0) / 21.9 (1.1)	.57 (0.0) / 22.0 (0.0)	.61 (0.0) / 22.0 (0.0)
		H2/ Nº Items	.80 (0.0) / 11.0 (0.0)	.80 (0.0) / 11.0 (0.0)	.80 (.011) / 10.8 (1.2)	.80 (0.0) / 6.4 (5.2)	.79 (0.0) / 0.1 (1.1)	- - / --	- - / --
	500	Scales	2.0 (0.0)	2.0 (0.0)	2.0 (0.2)	1.5 (0.5)	1.0 (0.2)	1.0 (0.0)	1.0 (0.0)
		H1 / Nº Items	.80 (0.0) / 11.0 (0.0)	.80 (0.0) / 11.0 (0.2)	.79 (0.1) / 11.5 (2.0)	.64 (0.1) / 16.8 (5.2)	.53 (0.0) / 21.7 (1.5)	.57 (0.0) / 13.9 (3.8)	.61 (0.0) / 22.0 (0.0)
		H2/ Nº Items	.80 (0.0) / 11.0 (0.0)	.80 (0.0) / 11.0 (0.2)	.80 (0.0) / 10.5 (2.0)	.80 (0.0) / 5.2 (5.3)	.82 (0.1) / 0.2 (1.5)	- - / --	- - / --
.70	1000	Scales	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	1.9 (0.2)	1.1 (0.3)
		H1 / Nº Items	.48 (0.0) / 11.0 (0.0)	.48 (0.0) / 11.0 (0.0)	.48 (0.0) / 11.0 (0.0)	.48 (0.0) / 11.0 (0.1)	.47 (0.0) / 11.1 (0.3)	.45 (0.0) / 12.8 (2.9)	.38 (0.0) / 20.6 (3.2)
		H2/ Nº Items	.47 (0.0) / 11.0 (0.0)	.47 (0.0) / 11.0 (0.0)	.47 (0.0) / 11.0 (0.0)	.47 (0.0) / 11.0 (0.1)	.47 (0.0) / 10.9 (0.3)	.46 (0.0) / 9.2 (3.0)	.43 (0.0) / 1.3 (3.3)
	500	Scales	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	2.0 (0.0)	2.0 (0.1)	1.8 (0.4)	1.3 (0.4)
		H1 / Nº Items	.48 (0.0) / 11.0 (0.0)	.48 (0.0) / 11.0 (0.0)	.48 (0.0) / 11.0 (0.1)	.48 (0.0) / 11.0 (0.3)	.47 (0.0) / 11.5 (1.2)	.44 (0.0) / 13.9 (3.8)	.39 (0.0) / 19.9 (3.5)
		H2/ Nº Items	.47 (0.0) / 11.0 (0.0)	.47 (0.0) / 11.0 (0.0)	.47 (0.0) / 11.0 (0.1)	.47 (0.0) / 10.9 (0.3)	.47 (0.0) / 10.5 (1.2)	.45 (0.0) / 8.0 (4.0)	.43 (0.1) / 1.9 (3.5)
.50	1000	Scales	4.9 (1.0)	5.0 (1.0)	4.9 (1.1)	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.1 (1.0)
		H1 / Nº Items	.38 (0.1) / 4.1 (1.1)	.38 (0.1) / 4.1 (1.1)	.39 (0.1) / 4.1 (1.2)	.39 (0.1) / 4.1 (1.2)	.39 (0.1) / 4.1 (1.1)	.39 (0.1) / 4.1 (1.2)	.40 (0.1) / 4.1 (1.1)
		H2/ Nº Items	.40 (0.1) / 3.8 (1.2)	.39 (0.1) / 3.7 (1.1)	.388 (0.1) / 3.8 (1.2)	.39 (0.1) / 3.7 (1.1)	.40 (0.1) / 3.7 (1.2)	.39 (0.1) / 3.7 (1.2)	.39 (0.1) / 3.7 (1.1)
	500	Scales	4.9 (1.1)	4.9 (1.1)	4.8 (1.1)	4.9 (1.1)	4.8 (1.2)	4.9 (1.1)	4.9 (1.1)
		H1 / Nº Items	.40 (0.1) / 4.5 (1.5)	.39 (0.1) / 4.5 (1.4)	.39 (0.1) / 4.6 (1.4)	.41 (0.1) / 4.3 (1.4)	.40 (0.1) / 4.5 (1.4)	.40 (0.1) / 4.5 (1.4)	.41 (0.1) / 4.5 (1.4)
		H2/ Nº Items	.41 (0.1) / 3.8 (1.3)	.40 (.1) / 3.9 (1.4)	.41 (0.1) / 3.8 (1.4)	.41 (0.1) / 4.9 (1.5)	.41 (0.1) / 3.9 (1.4)	.41 (0.1) / 3.8 (1.4)	.41 (0.1) / 3.8 (1.4)

Notes: Scales = Average number of components generated by the procedure. H1 = Average value of H for the first scale; H2 = Average value of H for the second scale; Nº items = Average numbers of items of the respective scale; ρ = Pearson correlation between the simulated factors.

APPENDIX Table 3. *Means (and standard deviations) for IFA estimations*

Lambda	*Sample size*		*ρ = .0*	*ρ = .1*	*ρ = .2*	*ρ = .3*	*ρ = .4*	*ρ = .5*	*ρ = .6*
.90	1000	λ F1	0.89 (0.01)	0.89 (0.01)	0.89 (0.01)	0.89 (0.01)	0.89 (0.01)	0.89 (0.01)	0.89 (0.01)
		λ F2	0.88 (0.01)	0.88 (0.01)	0.88 (0.01)	0.88 (0.01)	0.88 (0.01)	0.88 (0.01)	0.88 (0.01)
		SRMR	0.034 (0.00)	0.034 (0.00)	0.034 (0.00)	0.034 (0.00)	0.034 (0.00)	0.034 (0.00)	0.034 (0.00)
		RMSEA	0.043 (0.00)	0.043 (0.00)	0.043 (0.00)	0.043 (0.00)	0.043 (0.00)	0.043 (0.00)	0.042 (0.00)
	500	λ F1	0.88 (0.01)	0.88 (0.01)	0.88 (0.01)	0.88 (0.01)	0.88 (0.01)	0.88 (0.01)	0.87 (0.01)
		λ F2	0.87 (0.01)	0.87 (0.01)	0.87 (0.01)	0.87 (0.01)	0.87 (0.01)	0.87 (0.01)	0.86 (0.01)
		SRMR	0.044 (0.00)	0.044 (0.00)	0.044 (0.00)	0.044 (0.00)	0.044 (0.00)	0.043 (0.00)	0.043 (0.00)
		RMSEA	0.053 (0.01)	0.053 (0.01)	0.052 (0.01)	0.052 (0.01)	0.052 (0.01)	0.051 (0.01)	0.051 (0.01)
.70	1000	λ F1	0.71 (0.01)	0.71 (0.01)	0.71 (0.01)	0.71 (0.01)	0.70 (0.01)	0.70 (0.03)	0.70 (0.01)
		λ F2	0.69 (0.01)	0.69 (0.01)	0.69 (0.01)	0.69 (0.01)	0.69 (0.01)	0.69 (0.03)	0.69 (0.01)
		SRMR	0.041 (0.00)	0.042 (0.00)	0.041 (0.00)	0.041 (0.00)	0.041 (0.00)	0.041 (0.00)	0.042 (0.00)
		RMSEA	0.057 (0.01)	0.057 (0.01)	0.057 (0.01)	0.057 (0.01)	0.057 (0.01)	0.057 (0.01)	0.057 (0.01)
	500	λ F1	0.71 (0.01)	0.71 (0.01)	0.71 (0.01)	0.71 (0.01)	0.70 (0.01)	0.70 (0.01)	0.69 (0.02)
		λ F2	0.69 (0.02)	0.69 (0.02)	0.69 (0.02)	0.69 (0.02)	0.69 (0.02)	0.68 (0.02)	0.67 (0.02)
		SRMR	0.057 (0.00)	0.058 (0.00)	0.058 (0.00)	0.058 (0.00)	0.058 (0.00)	0.057 (0.00)	0.057 (0.00)
		RMSEA	0.078 (0.01)	0.078 (0.01)	0.079 (0.01)	0.078 (0.01)	0.078 (0.01)	0.08 (0.01)	0.078 (0.01)
.50	1000	λ F1	0.51 (0.01)	0.51 (0.01)	0.51 (0.01)	0.51 (0.01)	0.50 (0.01)	0.50 (0.01)	0.49 (0.02)
		λ F2	0.49 (0.01)	0.49 (0.01)	0.49 (0.01)	0.49 (0.01)	0.49 (0.02)	0.48 (0.02)	0.47 (0.02)
		SRMR	0.048 (0.00)	0.048 (0.00)	0.049 (0.00)	0.049 (0.00)	0.049 (0.00)	0.048 (0.00)	0.048 (0.00)
		RMSEA	0.069 (0.00)	0.069 (0.00)	0.069 (0.01)	0.069 (0.01)	0.069 (0.01)	0.069 (0.01)	0.069 (0.01)
	500	λ F1	0.51 (0.02)	0.51 (0.02)	0.51 (0.02)	0.51 (0.02)	0.50 (0.02)	0.50 (0.02)	0.49 (0.02)
		λ F2	0.49 (0.02)	0.49 (0.02)	0.49 (0.02)	0.48 (0.02)	0.48 (0.02)	0.47 (0.02)	0.46 (0.03)
		SRMR	0.069 (0.00)	0.069 (0.00)	0.068 (0.00)	0.068 (0.00)	0.068 (0.00)	0.068 (0.00)	0.068 (0.00)
		RMSEA	0.098 (0.01)	0.098 (0.01)	0.098 (0.01)	0.097 (0.01)	0.097 (0.01)	0.097 (0.01)	0.097 (0.01)

Notes: λ F1 = Average loads first factor; λ F2 = Average loads second factor; RMSEA = Root Mean Square Error of Approximation; SRMR = standardized root mean square residual; ρ = Pearson correlation between the simulated factors.

APPENDIX Table 4. *Table 3. Means (and standard deviations) of NOHARM estimations*

Lambda	*Sample size*		*P = .0*	*ρ = .1*	*ρ = .2*	*ρ = .3*	*ρ = .4*	*ρ = .5*	*ρ = .6*
.90	1000	λ F1	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.89 (0.01)
		λ F2	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.89 (0.01)
		SRMR	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)
		RMSEA	0.010 (0.01)	0.010 (0.01)	0.010 (0.01)	0.010 (0.01)	0.010 (0.01)	0.008 (0.01)	0.008 (0.01)
	500	λ F1	0.90 (0.01)	0.90 (0.01)	0.88 (0.02)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.89 (0.01)
		λ F2	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.90 (0.01)	0.89 (0.01)
		SRMR	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)
		RMSEA	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)
.70	1000	λ F1	0.70 (0.01)	0.70 (0.01)	0.70 (0.01)	0.70 (0.01)	0.70 (0.01)	0.70 (0.01)	0.69 (0.01)
		λ F2	0.70 (0.01)	0.70 (0.01)	0.70 (0.01)	0.70 (0.01)	0.70 (0.01)	0.70 (0.01)	0.69 (0.01)
		SRMR	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)	0.004 (0.00)
		RMSEA	0.006 (0.01)	0.006 (0.01)	0.006 (0.01)	0.006 (0.01)	0.006 (0.01)	0.006 (0.01)	0.006 (0.01)
	500	λ F1	0.70 (0.02)	0.70 (0.02)	0.68 (0.02)	0.70 (0.02)	0.70 (0.02)	0.70 (0.02)	0.69 (0.02)
		λ F2	0.70 (0.02)	0.70 (0.02)	0.70 (0.02)	0.70 (0.02)	0.70 (0.02)	0.70 (0.02)	0.69 (0.02)
		SRMR	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.006 (0.00)
		RMSEA	0.008 (0.01)	0.008 (0.01)	0.008 (0.01)	0.008 (0.01)	0.008 (0.01)	0.008 (0.01)	0.008 (0.01)
.50	1000	λ F1	0.50 (0.02)	0.50 (0.02)	0.49 (0.02)	0.50 (0.02)	0.50 (0.02)	0.50 (0.02)	0.49 (0.02)
		λ F2	0.50 (0.02)	0.50 (0.02)	0.50 (0.02)	0.50 (0.02)	0.50 (0.02)	0.50 (0.02)	0.49 (0.02)
		SRMR	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)	0.005 (0.00)
		RMSEA	0.022 (0.00)	0.022 (0.00)	0.022 (0.00)	0.022 (0.00)	0.022 (0.00)	0.022 (0.00)	0.022 (0.00)
	500	λ F1	0.50 (.002)	0.50 (0.02)	0.48 (0.03)	0.50 (0.02)	0.50 (0.02)	0.50 (0.02)	0.49 (0.03)
		λ F2	0.50 (0.02)	0.50 (0.02)	0.49 (0.03)	0.50 (0.02)	0.50 (0.02)	0.50 (0.02)	0.49 (0.03)
		SRMR	0.007 (0.00)	0.007 (0.00)	0.007 (0.00)	0.007 (0.00)	0.007 (0.00)	0.007 (0.00)	0.007 (0.00)
		RMSEA	0.031 (0.01)	0.031 (0.01)	0.031 (0.01)	0.030 (0.01)	0.030 (0.01)	0.030 (0.01)	0.030 (0.01)

Notes: λ F1 = Average loads first factor; λ F2 = Average loads second factor; RMSEA = Root Mean Square Error of Approximation; SRMR = standardized root mean square residual; ρ = Pearson correlation between the simulated factors

Endnotes

[i] In this study, the term construct means the theoretical concept (e.g. political efficacy), and the term instrument means the set of items designed to measure a construct or theoretical concept. We use the term latent dimensions to refer to the latent structures grouping different items, called factors in factor analysis, components in principal component analysis and scales in Mokken Scale Analysis. Even though results are usually discussed in terms of component scores in PCA, we will call them *factor scores* (the transformed variable values corresponding to particular data points) and *loadings* (the weight by which each standardized original variable must be multiplied to obtain the component score) in order to simplify the terminology.

[ii] The same fit criteria are used in both studies.