6-2022

Unrestricted Factor Analysis: A Powerful Alternative to Confirmatory Factor Analysis

Jan-Benedict E.M. Steenkamp

Alberto Maydeu-Olivares
University of South Carolina - Columbia, AMAYDEU@SC.EDU

Publication Info

Postprint version. Published in *Journal of the Academy of Marketing Science*, 2022.

https://doi.org/10.1007/s11747-022-00888-1.

UNRESTRICTED FACTOR ANALYSIS:
A POWERFUL ALTERNATIVE TO CONFIRMATORY FACTOR ANALYSIS

Jan-Benedict E.M. Steenkamp

Alberto Maydeu-Olivares

Journal of the Academy of Marketing Science (2022) 10.1007/s11747-022-00888-1

Jan-Benedict E. M. Steenkamp is the C. Knox Massey Distinguished Professor of Marketing and Marketing Area Chair, University of North Carolina at Chapel Hill, Campus Box 3490, McColl Building, Chapel Hill, NC 27599-3490, 1-919-962-9579 (e-mail: jbs@unc.edu). Alberto Maydeu-Olivares is Professor of Psychology, University of South Carolina, Barnwell College, 1512 Pendleton St. Columbia, SC 29208 (e-mail: amaydeu@sc.edu) & Professor of Psychology. University of Barcelona. We thank the editor and the reviewers for their constructive comments, AiMark for providing the Schwartz Value Survey data and Arvind Malhotra for providing the E-S-QUAL data. The participation of Alberto Maydeu-Olivares was supported, in part, by the Research Center for Child Well-Being (NIGMS P20GM130420), grant PID2020-119755GB-I00 funded by AEI 10.13039/501100011033, and AGAUR grant 1237SGR2017. The cross-national data for the Schwartz Value Survey as well as Mplus and R input files for both empirical applications can be found on: https://www.dropbox.com/sh/tg6573rsx27o21c/AAApQpxdvtZwWoKgosPVUVu-a?dl=0

ABSTRACT

The gold standard for modeling multiple indicator measurement data is confirmatory factor analysis (CFA), which has many statistical advantages over traditional exploratory factor analysis (EFA). In most CFA applications, items are assumed to be pure indicators of the construct they intend to measure. However, despite our best efforts, this is often not the case. Cross-loadings incorrectly set to zero can only be expressed through the correlations between the factors leading to biased factor correlations and to biased structural (regression) parameter estimates. This article introduces a third approach, which has emerged in the psychometric literature, viz., unrestricted factor analysis (UFA). UFA borrows strengths from both traditional EFA and CFA. In simulation studies, we show that ignoring cross-loadings even as low as .2 can substantially bias factor correlations when CFA is used and that even the commonly used guideline RMSEA ≤ .05 may be too lenient to guard against non-negligible bias in factor correlations in CFA. Next, we present two empirical applications using Schwartz's value theory, and electronic service quality. In the first case, UFA leads to much better model fit and more plausible regression estimates. In the second case, the difference is less dramatic but nevertheless, UFA provides richer results. We provide recommendations on when to use UFA vs. CFA.

Social scientists, including researchers in marketing and other business disciplines are keenly aware of the importance of using multiple items (metrics, observables, indicators) to measure constructs because each item is at best an imperfect indicator of the underlying construct (Bollen 1989). A key task for these scholars is to establish that their measures have adequate psychometric properties. This applies regardless of whether these items are obtained through surveys (Hulland et al. 2018), experiments (Bagozzi and Yi 1989), scraped from the internet (Unnava and Aravindakshan 2021), or based on "objective" data such as macroeconomic statistics (Roth 1995).

Currently, the gold standard for modeling (reflective) multi-item data is confirmatory factor analysis (CFA). CFA has many advantages over traditional exploratory factor analysis (EFA). By EFA we mean exploratory factor analysis as described in graduate textbooks like Hair et al. (2018) and Tabachnick and Fidell (2018), and as implemented in major statistical suites such as SPSS. CFA provides indices to assess the fit of the model and allows for correlated errors. CFA enables the study of measurement invariance across groups and group comparisons free of measurement error (Steenkamp and Baumgartner 1998), and the study of constructs over time (Steenkamp and Maydeu-Olivares 2015). Drivers of latent constructs can be incorporated in CFA models by means of multiple indicator multiple cause (MIMIC) models (Jöreskog and Goldberger 1975) and CFA can be seamlessly integrated into latent variable structural equation modeling (SEM), thus allowing for theory testing using structural (regression) parameter estimates that are free of measurement error (MacKenzie 2001; Steenkamp and Baumgartner 20002001).

Yet, these strengths of CFA come at a cost: CFA imposes a highly restrictive model on the data. Most CFA applications involve an independent-clusters model, in which each item loads on one factor only. In other words, cross-loadings are constrained to zero. Unfortunately, despite our best efforts, items are rarely pure indicators of the constructs they were intended to measure (Morin et al. 2016). As Marsh et al. (2013a, p. 258) put it:

> "For real data, unidimensionality and pure indicators are an ideal to strive toward (i.e., a convenient fiction), but are rarely if ever achieved. …almost all indicators are factorially

> complex if actually put to the test…. In studies based on multidimensional constructs or a variety of related factors, most indicators will cross-load on at least some of the other factors (if allowed to do so)."

For example, De Luca et al. (2021) propose a new theory on the impact of big data investments on service innovation and performance. They conceptualize and operationalize three types of big data marketing affordances. Yet, despite their careful, multi-study scale development efforts, aggregated across four studies, traditional EFA reveals that 78% of the cross-loadings exceed .2 and 14% exceed .3. To compound the problem, constructs that are well fit by a unidimensional model when analyzed by themselves may need cross-loadings when modeled together (Marsh et al. 2013a).

When an item provides information on constructs other than the one it was designed to measure and its cross-loadings are set to zero, the misspecified cross-loadings can only be expressed through the correlations between the factors, leading to biased factor correlations (Marsh et al. 2014). As we will show later, the direction of the bias depends on the sign of the omitted cross-loadings and the sign of the true correlation between the constructs. If these are in the same direction, the magnitude of the correlation is upwardly biased; if they are in the opposite direction, the magnitude of the correlation is downwardly biased. For example, De Luca et al. (2021; study 1c) report CFA correlations ranging from .580 to .702 between their three big data marketing affordances constructs, while all cross-loadings in their EFA are positive. This suggests that the CFA correlations reported between the constructs are upwardly inflated. Thomson et al.'s (2005) study on emotional attachment to brands allows us to get a sense of the bias because they report factor correlation estimates based on traditional EFA and CFA, albeit on different samples. EFA shows that 50% of the cross-loadings exceed .2, including 20% exceeding .3. The estimated average correlation between their three emotional attachment factors is .32. The CFA model, on the other hand, in which all cross-loadings are suppressed, gives an average inter-factor correlation estimate of .75.

Why should potentially biased factor correlations matter to business scholars? For one, upwardly biased correlations may call into question whether the constructs are conceptually truly distinct, and even lead the researcher to conclude that they are manifestations of a higher-order structure (e.g., a second-order factor model). The converse also applies—downwardly biased correlations may lead the researcher to believe that the constructs are conceptually (more) distinct than they really are and prevent further work to account for the associations among those constructs (e.g., failing to hypothesize a second-order factor model). Even more problematic, if the measurement model is embedded within a SEM model, biased correlations between the latent factors will lead to biased structural parameter estimates and biased inferences (Asparouhov and Muthén 2009; Asparouhov, Muthén and Morin 2015).[1]

Thus, business researchers are faced with a conundrum. Ignoring cross-loadings if they are substantial can lead to biased estimates and incorrect statistical inferences. However, traditional EFA is not an attractive option, given the limitations noted earlier. In this article, we discuss a powerful third approach, which has emerged in the psychometric literature, viz., contemporary EFA, which, unlike traditional EFA, has many of the strengths of CFA while retaining much of the flexibility of EFA. In fact, in contemporary EFA the measurement model is no longer "exploratory" in nature. Rather, as we will discuss, theory can be seamlessly incorporated into the model. Therefore, we follow Cudeck and O'Dell (1994) and will use the term unrestricted factor analysis (UFA) to refer to contemporary EFA. The term "unrestricted" comes about because when measuring *m* factors, in UFA the *maximum* number of parameters linking the *m* constructs to the observed variables (items) is estimated. In contrast, the standard independent-clusters CFA corresponds to the *minimum* number of parameters linking the *m* constructs to the observed variables. Thus, *restricted* factor analysis is probably a more accurate term than *confirmatory* factor analysis. However, given the prevalent use of the term CFA,

[1] Recall that in regression/SEM analysis, $\boldsymbol{\beta} = \mathbf{R}^{-1}\mathbf{r}$, where $\boldsymbol{\beta}$ is the vector of standardized regression coefficients, **R** is the correlation matrix of the independent variables and **r** is the vector of correlations between each independent variable and the dependent variable. Bias in **R** leads to a different estimate of its inverse ($\mathbf{R}^{-1}$), and hence to bias in the regression coefficients used in theory testing. Standard errors of the regression coefficients are also biased because they involve the square multiple correlation of each predictor *i* with the other predictors, R_i^2:
$SE_{\beta i} = \sqrt{(1 - R_Y^2)/(N - k - 1)} * \sqrt{1/(1 - R_i^2)}$ where R_Y^2 is the explained variance in the dependent variable, *N* is the sample size, and *k* is the number of predictors (Cohen et al. 2003, p. 86).

here we will use CFA to refer to this model.

The purpose of this article is to introduce UFA to substantive empirical business scholars. We draw upon the psychometrics, multivariate statistics, and specialized psychology literatures to provide an integrative discussion of the relevant issues that must be considered when choosing between CFA and UFA, how to use and evaluate UFA models, and how to integrate UFA into latent variable structural equation modeling. Table 1 summarizes the key differences between traditional EFA, CFA, and UFA. To be clear, our argument is not that unidimensionality is not valuable and a goal to strive for. Rather, we argue that despite our most conscientious efforts, this is often difficult to achieve in applications, and ignoring this problem by forcing the data into a CFA model in the presence of cross-loadings can be (highly) problematic. UFA has its own set of limitations (which we will discuss) and our argument is not that UFA is always preferable, but rather that it can – and in our experience often does – provide a powerful alternative to CFA.

--- Table 1 about here ---

The remainder of this article is organized as follows. First, we describe the key features of UFA. Next, we conduct two simulation studies investigating the magnitude of bias in estimated factor correlations in UFA vs. CFA in the presence of cross-loadings under a range of conditions, and evaluate the relation between goodness of fit and the extent of parameter bias in misspecified CFA models. Subsequently, we provide two empirical illustrations. We conclude our article with recommendations and suggestions for further research. All analyses reported in this article were performed with Mplus (Muthén and Muthén 2017).

UNRESTRICTED FACTOR ANALYSIS

UFA incorporates two important ideas. First, researchers often have a theory about how many factors underlie the set of items and which items should primarily reflect each factor. Indeed, in most applications, items are written to conform to a theory. Second, simple (measurement) structure does not require that cross-loadings are zero; rather the actual requirement is that their magnitude be small

(McDonald 1985; Morin et al. 2013).

In UFA, the researcher specifies the number of factors a priori and can indicate which items are substantive indicators of each factor, while allowing for cross-loadings on the other factors. Cross-loadings are aimed to be as close to zero as possible, rather than being constrained to zero. While the targets influence the final rotated solution, the actual estimated cross-loadings can be very different from zero if the zero-target loadings are inappropriate (Marsh et al. 2013a). This strategy reflects a compromise between the mechanical approach to rotation in traditional EFA and the a priori, independent-clusters CFA model (Browne 2001; McDonald 1999). When used in this fashion, for confirmatory rather than for exploratory purposes, UFA is an attractive alternative to CFA, as it suffices to have an approximate, rather than full, knowledge of the factor structure. We note that rotational indeterminacy also holds for a target rotated UFA solution. The fit of a UFA m-factor model with target rotation equals the fit of a traditional oblique EFA m-factor model. The difference between them is that a target-rotated UFA gives a very specific, theory-based direction to the rotation.

For the one-factor model ($m = 1$), UFA and CFA models are equivalent: they yield the same model fit and parameter estimates. If $m > 1$, UFA can be described as follows. First, an m-factor model is estimated using the minimum number of m^2 constraints among the factor loadings and inter-factor correlations that is required to achieve identification (Jöreskog 1969).[2] In a second step, this initial solution is rotated towards a target pattern matrix (Browne 2001), which reflects the researcher's a priori theory.[3] Rotated factors can be specified as being all correlated, or all uncorrelated. A p (items) × m (factors) target matrix **T**, with some specified elements and some unspecified elements is required. For example:

[2] These identification constraints are automatically implemented by Mplus. The researcher can review them with the command TECH1 on the OUTPUT line.

[3] For ease of exposition, we describe the estimation of UFA models as a two-step process as it is the approach used in traditional EFA. In fact, the rotated solution can be estimated directly (Browne 2001), and whether a one-step or two-step estimation approach is used, depends on computational/programming convenience.

$$\mathbf{T} = \begin{bmatrix} t_{11} & \approx 0 & \approx 0 \\ t_{21} & \approx 0 & \approx 0 \\ t_{31} & \approx 0 & \approx 0 \\ \approx 0 & t_{42} & \approx 0 \\ \approx 0 & t_{52} & \approx 0 \\ \approx 0 & t_{62} & \approx 0 \\ \approx 0 & \approx 0 & t_{73} \\ \approx 0 & \approx 0 & t_{83} \\ \approx 0 & \approx 0 & t_{93} \end{bmatrix}$$

where, based on a priori theory, t_{ij} indicates an unknown substantive loading and ≈ 0 indicates a cross-loading targeted to be close to zero. If the elements of matrix **T** are given by t_{ij}, Browne (1972, 2001) proposed as rotation criterion to minimize the squared differences between the estimated unrotated factor loadings λ_{ij} and prespecified (rotated) target values t_{ij}

(1) $$\min f\left(L\right) = \sum_{j=1}^{m} \sum_{i \in I_j} \left(\lambda_{ij} - t_{ij}\right)^2$$

where the I_j set includes the subscripts for specified target loadings in column j.

Target rotation resembles standard CFA as values for some factor loadings must be specified in advance. However, in CFA, specified factor loadings are usually forced to be zero. Misspecified constraints (e.g., factor loadings incorrectly set to zero) may only be detected by modification indices. In a target rotation, corresponding elements of the rotated factor pattern matrix are only made as close to the specified zeros as possible. There is no guarantee that the target structure is recovered. After all, the researcher's theory may be wrong, or the data may be ill-behaved.

Misspecified constraints on the factor loadings are not the only common source of misspecification in factor models. Another prevalent source is excess correlation between items that cannot be adequately fully modeled through the factor loadings and factor correlations. Correlated uniquenesses may be due to a variety of causes, such as item wording, item ambiguity, and location in the survey. Their presence is easily identified by examining modification indices and standardized expected parameter change estimates. Although researchers are advised to use correlated error terms sparingly, a factor model that does not account for substantial idiosyncratic associations between items risks producing a distorted factor solution (Cole, Ciesla and Steiger 2007), and can inflate estimates of reliability (Green and Hershberger 2000). Unlike in traditional EFA, excess correlation between items can be modeled in UFA by allowing the error terms of items in question to covary.

Choosing Between UFA and CFA

The choice between CFA and UFA should be based on informed assessment of model fit and parameter estimates rather than on mechanical application of a set of cutoffs for goodness of fit indices. However, that does not mean that the choice between UFA and CFA is just in the eyes of the beholder. We advocate that the researcher systematically compares the two solutions based on the following criteria.

Model fit. Three indices have emerged as the dominant measures to evaluate model fit: the Root Mean Square Error of Approximation (RMSEA), the Standardized Root Mean Squared Residual (SRMR), and the Comparative Fit Index (CFI) (Chen 2007). Models with RMSEA ≤ .05, SRMR ≤ .05, and CFI ≥ .95 provide close fit to the data, while RMSEA ≤ .08, SRMR ≤ .08, and CFI ≥ .90 indicate reasonable fit (Steenkamp and Maydeu-Olivares 2021). SRMR and CFI improve monotonically with the number of parameters estimated while RMSEA includes a penalty function for overfitting. These cutoffs should be treated with some caution (Niemand and Mai 2018; see also our first simulation study) because model fit depends on a number of factors such as the number of items being modeled (Moshagen 2012), the magnitude of factor loadings (Shi et al. 2018), and sample size (Shi et al. 2019). However, if model fit clearly falls short of these cutoffs, the burden is on the researcher to show that inferences based on their relatively ill-fitting model are nevertheless valid, since inferences drawn on poorly fitting models are often incorrect (Maydeu-Olivares 2017).

Difference in model fit. CFA is nested in UFA, and choosing between the CFA and UFA model constitutes a test on whether all cross-loadings set to zero in the CFA model are indeed equal to zero. The formal statistical test is whether the difference in chi-square between the two models is significant, with the degrees of freedom for this test being the difference in degrees of freedom between the UFA and CFA models (Pavlov et al. 2020). If the difference in chi-square is not significant, we can conclude that the cross-loadings do not significantly differ from zero, and thus, the CFA should be chosen. This is the only objective statistical test. On the other hand, if the difference in chi-square is significant, CFA's restrictive assumptions on the cross-loadings are not supported by the data. However, in large samples, it is still possible that the differences between the two model specifications are substantively small, since the chi-square difference test may have excessive power (Marsh et al. 2005). Therefore, *if and only if* the chi-square difference test is significant, we recommend that the substantive researcher consider whether the difference in model fit is practically meaningful. Several influential articles (Cheung and Rensvold 2002; Chen 2007) have established benchmarks for *relative* fit criteria, i.e., for change in absolute fit indices to determine whether the deterioration in fit between the nested versus the more general model is sufficiently small that the restricted model should not be rejected. These benchmarks are: ΔRMSEA $\leq .015$, ΔSRMR $\leq .030$, and ΔCFI $\leq .010$. Deterioration in model fit that clearly exceeds these cutoffs points toward UFA.

Examination of the factor structure. Each factor should be characterized by multiple factor loadings that are statistically significant. Again, statistical significance is the only objective criterion, but in large samples, even small factor loadings will be significant. Thus, an additional criterion that is often used is that the standardized factor loadings be substantial, where values ranging from .4 (Steenkamp and Maydeu Olivares 2021) to .7 (Hulland et al. 2018) have been proposed. It is obvious that higher factor loadings are preferable, *ceteris paribus*. However, in actual research, the desire to obtain high factor loadings may lead to reduced content validity: a number of very similarly worded items will enhance loadings but often results in poor content domain sampling. Further, it is our experience that higher factor loadings (in the .5 to .7 range, and beyond) are more realistic for conceptually relatively "narrow" constructs that are often found in marketing while lower factor loadings (in the .4 to .6 range) are often more common for conceptually "broader" constructs (e.g., personal values, personality traits, regulatory focus, or horizontal/vertical individualism/collectivism).

Sampling items across a broad domain almost always results in lower item intercorrelations. It is also possible that the broader domain is actually a second-order factor on which multiple, more narrowly defined first-order factors (aka facet scales; Church and Burke 1994) load. If the items achieve adequate content validity by sampling various first-order factors, their loading on the second-order factor is the product of the loading of the item on the first-order factor times the loading of the first-order factor on the second order factor. Thus, two large loadings of .7 still result in an observed loading below .5. Regardless, a key consideration is that the set of items result in a composite reliability (ρ_c, aka coefficient omega; McDonald 1999) that achieves a signal-to-noise ratio of at least 2:1 (which translates into a composite reliability of at least .67), or, still better, a composite reliability of .7 (Nunnally 1978). The composite reliability increases with the magnitude of standardized loadings and the number of items.

An additional consideration in UFA is the structure of the cross-loadings. The goal of UFA is to approximate (rather than force) simple structure, the golden standard in factor analysis. To aid researchers, we present guidelines proposed by Marsh et al. (2013a). They labeled a solution in which all cross-loadings vary between 0 and .10 as a close approximation to simple structure. Applied researchers will rarely encounter this in their data. Good approximation is characterized by a factor structure in which about one-third of the cross-loadings is close to 0, another third is close to .10, and the remaining third is close to .20. This might still be better than might be expected in many applied studies. Finally, a moderate approximation to simple structure is where a) one fourth of the cross-loadings is close to .10, .20, .30, and .40, respectively and b) where all cross-loadings are smaller than the corresponding loading on the factor it was designed to measure. Marsh et al.'s descriptions should be taken as labels for applied researchers to use in describing their results rather than as iron-clad rules. In this spirit, we recommend that researchers try to understand and explain using their theory why cross-loadings of substantial size (> |.3|; Brown 2015) do occur. We will illustrate this in the empirical applications.

Difference in factor correlations. A key consideration in choosing UFA over CFA is that cross-loadings can affect factor correlations, depending on the pattern of the cross-loadings. When some cross-loadings are negative while others are positive, the biasing effect of ignoring cross-loadings on factor correlations may cancel out. Thus, if UFA provides a better fit to the data, it is useful to assess the magnitude of the difference in factor correlations. What would constitute an appreciable difference in factor correlations is a matter of judgment. We will show in our second empirical application that differences in the range of .15 to .20 can already lead to different results in the structural estimates. On the other hand, using Cohen's (1988) cutoff of .1 for a small standardized effect size, differences in factor correlations less than |.10| are unlikely to be meaningful.

Difference in structural results. In many substantive studies, measurement analysis is not a goal in itself but an integral part of the researcher's interest in testing substantive theory by estimating structural relations between constructs in a nomological net. If that is the case, conduct the structural analysis using both CFA and UFA. Are the results basically the same or not? Which set of results is most consistent with theory and prior results, if any? Is the magnitude of the (standardized) regression estimates reasonable? Assess the statistical precision of both solutions by comparing the standard errors of the parameter estimates. Higher correlations among predictors yield larger standard errors. A benchmark is provided by McDonald (1985), who pointed out that for uncorrelated data, the standard error of a standardized parameter estimate is close to $1/\sqrt{N}$. If standard errors vastly exceed $1/\sqrt{N}$ and/or if individual standard errors fluctuate widely, these are indications of multicollinearity (Pedhazur 1982).

Structural Equation Modeling with UFA

Asparouhov and Muthén (2009) extended UFA to a general system of latent variable equations which they called Exploratory Structural Equation Modeling (ESEM), similar to how Jöreskog (1978) extended CFA into SEM. When the goal is to estimate the structural relations between multiple latent variables in the nomological net, the first step is to conduct CFA and UFA measurement analysis on all latent variables simultaneously, and compare results. If the CFA measurement model is supported, the researcher can estimate the structural relations with SEM. How to do this is well-known among substantive business scholars.

If UFA is the more appropriate model, the next thing the researcher needs to do, is to divide the constructs into sets of variables as specified in the nomological net, and estimate a set-UFA model (Marsh et al. 2020). For instance, in many applications, constructs can be divided into independent, mediating, and dependent constructs. A set-UFA model is a model in which the two or three (or more) sets of constructs are modeled simultaneously using UFA, but cross-loadings are only permissible for factors within the same set while they are constrained to be zero for constructs in different sets. Thus, set-UFA is a restricted case of full-UFA. This set-UFA analysis is useful because the current state of psychometric theory is that relations between factors that belong to the same UFA model can only be modeled through non-directed paths (correlations), not with regression coefficients. On the other hand, relations between factors belonging to different sets can be modeled using directed paths (regression coefficients), using ESEM (the direct counterpart of SEM).

If, however, the single, full-UFA model in which items can cross-load on every other construct, regardless whether they are independent, mediating, or dependent constructs, is preferred over set-UFA, the researcher needs to fit the ESEM model as a SEM model (EwSEM; Marsh et al. 2013b; Morin et al. 2013). ESEM models incorporate hidden constraints among the estimated parameters and the same number of constraints involved in the ESEM model needs to be incorporated in the SEM model to arrive at the EwSEM model. Fortunately, this is easy to do. To get the EwSEM model, recall that the UFA model has m^2 identification constraints among the model parameters, with m being the number of factors. Thus, after saving the model parameter estimates obtained in the single, full-UFA model, the researcher needs to fix m^2 parameters as follows: 1) constrain the m factor variances to 1; 2) select a reference indicator for each factor that has a large target loading and small cross-loadings and fix the small cross-loadings to their estimated values as obtained in the single, full-UFA solution. This introduces $m(m-1)$ additional model constraints. The other factor loadings will be freely estimated.

This results in a highly saturated CFA. In fact, it is the most saturated CFA model that is still identified, with the same number of degrees of freedom and--except for numerical error--identical model fit to the UFA model. But it being a CFA model, structural relations can be straightforwardly estimated with standard SEM. We will illustrate this in the second empirical application. See Figure 1 for the decision tree.

--- Figure 1 about here ---

Alternatively, some researchers prefer to estimate a standard regression/path analysis model using factor scores. However, factor scores contain error, and the OLS estimator is biased even asymptotically in the presence of measurement error in the predictors and/or dependent variables (Greene 2003) unless a correction is implemented (Croon 2002). Only if the factor scores are highly reliable will the uncorrected OLS results be close to the results obtained with SEM, ESEM, or EwSEM. We will illustrate this in our second empirical example. The sum score is the most widely used factor score in substantive research. It is obtained as an unweighted sum on the target items after items have been coded in the same direction. Using sum scores in regression analysis is an option if CFA is the retained model. Sum scores should never be used if UFA is the retained model since sum scores do not account for cross-loadings. For UFA, the researcher should use factor scores that account for cross-loadings such as empirical Bayes factor scores, aka regression factor scores (Skrondal and Rabe-Hesketh 2004).

SIMULATION STUDIES

Simulation studies are useful in documenting the extent to which parameters of interest are biased because we know their true values. For example, when comparing EFA and CFA factor correlations reported by Thomson et al. (2005) in the presence of cross-loadings, we cannot know for sure whether the average EFA correlation of .32 is closer to the true value than the CFA correlation of .75 because we do not know the true value. Here, we present the results of two simulation studies using a two-factor model to shed light on the magnitude of bias in estimating the correlation between constructs when cross-loadings are set to zero as in CFA.

Simulation Study 1: Bias in Population Factor Correlation

We manipulated the following factors: true factor correlation (-.5, -.3, -.1, .1, .3, .5), number of substantive (target) indicators per factor (4, 8, 12), magnitude of the substantive factor loadings (.4, .6), and magnitude of the cross-loadings (-.3, -.2, -.1, 0, .1, .2, .3). The percentage of items with cross-loadings was set at 25% for all models. Error variances were set so that the resulting covariance matrix has a unit diagonal. This gives 252 combinations. For each of these conditions we obtained the model-implied covariance matrix and fitted a two-factor CFA model without cross-loadings and a two-factor UFA model with target rotation, using maximum likelihood estimation. Doing so enables us to determine how well CFA without cross-loadings and UFA can recover inter-factor correlations at the population level.

Web Appendix A gives the ANOVA results on the absolute bias in the factor correlation. We find that the main drivers of the bias in the population estimates are the measurement model (CFA vs. UFA, explaining 42.3% of the variance), magnitude and direction of the cross-loadings (22.4%), and the interaction cross-loading x measurement model (18.0%). Table 2 summarizes these effects. Each cell in Table 2 corresponds to the mean bias over 12 conditions (number of substantive indicators per factor x magnitude of substantive indicators). Based on Cohen (1988), we consider bias $\leq |.1|$ as substantively negligible.

Panel 2A provides the results for CFA. We observe that ignoring cross-loadings of magnitude of $|.1|$ leads to negligible bias in the inter-factor correlations. However, ignoring larger cross-loadings, and especially cross-loadings of $|.3|$, leads to appreciable bias. Moreover, the direction of the bias depends on whether the sign of the factor correlation and cross-loadings are aligned. If they have the same sign, the estimated factor correlation is inflated. For example, a population factor correlation of .5 yields on average an estimated factor correlation of .80 in the presence of 25% cross-loadings of magnitude .3. The magnitude of the bias increases as the size of cross-loadings increases and the effect is symmetrical: a population factor correlation of -.5 yields on average an estimated factor correlation of -.80 in the presence of 25% cross-loadings of magnitude -.3. On the other hand, if the direction of the population correlation and the direction of the cross-loadings is different, this attenuates the estimate of the factor correlation. For example, a population factor correlation of .5 yields on average an estimated factor correlation of .35 in the presence of 25% cross-loadings of magnitude -.3. Panel 2B provides the results for UFA. The bias is well below |.1| in all conditions.

This simulation study shows that ignoring cross-loadings in one's measurement model can lead to bias in population factor correlations. The direction (inflation or deflation) and magnitude of the bias depends on the magnitude and sign of the cross-loadings and the population factor correlation.

--- Table 2 about here ---

As mentioned earlier, the discrepancy between the data generating process and the fitted model can be summarized using goodness of fit indices. Can the goodness-of-fit indices be used to predict the extent of parameter bias in misspecified CFA models? The sample RMSEA was constructed to be an (approximately) asymptotically unbiased estimator of the population RMSEA, but the sample SRMR and CFI are asymptotically biased estimators (Ximénez et al. 2022). Therefore, we will focus on RMSEA as the results apply to sample settings, too. The results are displayed in Figure 2. Across the different panels of Figure 2, the percentage of variance of log absolute bias explained by RMSEA ranges between 70% and 88%; although when there are only 4 indicators per factor and the magnitude of the substantive loading is only .4 (an admittedly weak model), the R^2 drops to 59% only. RMSEA for various conditions is reported in Table 3. This analysis shows that poorer values of RMSEA are indeed associated with higher levels of parameter bias, and that the commonly used guideline of RMSEA ≤ .05 to retain a CFA model may be too liberal if one strives for parameter bias in the range of .2 or less.

--- Figure 2 and Table 3 about here ---

Simulation Study 2: Bias in Factor Correlation for Varying Degrees of Approximation to Simple

Structure and Different Sample Sizes

The first simulation study ascertained that the effects of cross-loadings on estimated correlations are symmetric. Hence, in this second simulation study, we only consider positive factor correlations when simulating data. In real data, cross-loadings are not all 0 or, say, .2 or .3, but an assortment of essentially zero, small, and larger cross-loadings. Moreover, the performance of UFA and CFA in recovering the population factor correlation will depend on sample size as the statistical theory for these models is asymptotic. This second simulation involved a two-factor model with 108 empirical conditions: 3 types of approximation to simple structure (nearly pure, good, moderate) x 3 sample sizes (200, 600, 1,000) x 3 factor correlations (.1, .3, .5) x 2 magnitudes of substantive loadings (.4, .6) x 2 different number of substantive items per factor (4, 8). Inspired by the criteria set forth by Marsh et al. (2013a), nearly pure approximation to simple structure was operationalized as a condition where 50% of cross-loadings were set to zero, 25% to .05, and 25% to .10. Good approximation was operationalized as 25% of cross-loadings set to 0, .10, .15, and .20 each. Moderate approximation was operationalized as 25% of cross-loadings at 0, .10, .20, and .35, each. In all conditions, the variances of the errors of the items were set so that the covariance matrix had unit diagonals in the population. For each experimental condition, we generated 1,000 random samples drawn from a normal distribution with mean zero and the covariance structure implied by the above conditions leading to 108,000 data sets to which we fitted CFA and UFA with a target rotation, using maximum likelihood. Web Appendix B provides the average bias in the factor correlation for each of the 108 conditions for UFA and CFA.

We performed an ANOVA model on the absolute bias estimates using six factors and up to four-way interactions, obtaining an R^2 of .999. Three effects explain nearly 90% in the variance in model performance: measurement model (UFA vs. CFA, 60.0% of the variance), approximation to simple structure (16.0%) and the interaction approximation to simple structure x measurement model (12.5%). Figure 3 plots these results. It shows that on average, UFA performs much better than CFA –

across all conditions, the average bias in the estimated factor correlation for UFA is .02 versus .35 for CFA. Further, as may be expected, the poorer the approximation to simple structure, the greater the bias in CFA estimates. The difference on average between true and estimated inter-factor correlations in CFA is .14 for nearly pure approximation, which is close to being negligible, versus .38 for good approximation, and .53 for moderate approximation. In contrast, the bias in UFA estimates is only slightly affected by the degree of approximation to simple structure. As a result, the improvement obtained by using UFA instead of CFA ranges from a relatively modest .13 (nearly pure approximation) to a large .50 (moderate approximation).

---Figure 3 about here ---

Overall, sample size has little effect on the performance of the models – its main effect and all the interactions together account for only 1.6% of variance. However, we find that 51.3% of the (admittedly small) variance in UFA performance involves sample size. Under a specific set of conditions, UFA's performance is only a little better than that of CFA. These conditions are 1) the sample size is small (N=200), 2) approximation to simple structure is good or moderate, 3) the substantive factor loadings are .4, 4) the number of substantive loadings per factor is 4, and 5) the factor correlation is .5. Under these conditions, the two-factor structure is weakly defined and the sample size is too small to estimate the highly parameterized UFA model with sufficient precision. As a result, the absolute bias with which UFA estimates the factor correlation is .33 vs. .38 for CFA.

These two simulation studies show that ignoring cross-loadings in the range of |.2| or higher leads to non-negligible--and in some cases large--bias in factor correlations when CFA is used, while the bias in UFA estimates is, with rare exceptions, small.

EMPIRICAL APPLICATION TO SCHWARTZ VALUE SURVEY

Researchers have long been interested in the role of values in consumer behavior. Values are cognitive beliefs about desirable goals and modes of conduct to promote these goals, which vary in importance, and serve as standards to guide attitudes and behavior. The content and structure of human values has been most thoroughly elucidated by Schwartz (1992). Schwartz's theory has been used by marketing scholars in market segmentation (Branguele-Vlagsma et al. 2002), to understand why some consumers are more innovative (Steenkamp et al. 1999) and higher on optimal stimulation level (Steenkamp and Burgess 2002), to explore the relation between materialism and well-being (Burroughs and Rindfleisch 2002), to understand the relationship between brands and ideology (Shepherd et al. 2015), for positioning of multi-country brands (Batra et al. 2017), to develop global brand concepts (Torelli et al. 2012) and brand personality (Geuens et al. 2009), and as driver of consumer attitudes toward global and local consumer culture (Steenkamp and De Jong 2010), social desirability bias (Steenkamp et al. 2010a), salesperson performance (Swenson and Herche 1994), repeat purchase behavior (Paul et al. 2009), charitable behavior (Winterich and Zhang 2014), and consumer responses to corporate green and non-green actions (Xie et al. 2015). We will use Schwartz and Boehnke's (2004) adaptation of Schwartz's (1992) theory, which distinguishes between ten motivationally distinct types of values organized in a circular structure: concern for nature, social concern, benevolence, conformity/tradition, security, power, achievement, hedonism, stimulation, and self-direction (Figure 4).

--- Figure 4 about here ---

We will explore the effects of the Schwartz value types (i.e., common factors underlying the value ratings) on materialism, a construct that has received a significant amount of attention by marketing scholars (e.g., Ruvio et al. 2014; Steenkamp and Maydeu-Olivares 2021). Related research has documented the negative effect of materialism on satisfaction with life (Richins and Dawson 1992; Ryan and Dziurawiec 2001). Consistent with a stream of marketing research that has argued that general values affect consumer domain-specific values (e.g., Batra et al. 2001; Steenkamp et al. 1999), we will estimate a mediating framework in which the Schwartz value types impact materialism, and materialism is a predictor of satisfaction with life.

The purpose of this empirical application is not to break new theoretical ground. For example, consumer materialism overlaps with power (Schwartz et al. 2012), an issue we leave for future research. Rather, our purpose is to show what UFA can do, while CFA fails.

Method

We use survey data that were collected by the global marketing research agencies Kantar and GfK in national samples of 1,178 U.S. respondents (in English) and 547 respondents in Norway (in Norwegian). Data were collected on the Schwartz values, materialism, satisfaction with life, sociodemographics, and a number of other variables that are not of interest to this empirical illustration. The Schwartz values were measured with the Schwartz Value Survey (SVS: Schwartz and Sagiv 1995), consisting of 45 values (items), covering the ten value domains. Each value was listed by a descriptive name (e.g., "equality") with a short explanation in parentheses (e.g., "equal opportunity for all"). For materialism we used Richins and Dawson's (1992) six items that measure possession-defined success. Life satisfaction was measured with five items developed by Diener et al. (1985). See Web Appendix C for the items.

The U.S. sample will be our main focus. As the data were non-normally distributed, we use maximum likelihood estimation with standard errors robust to non-normality and a mean adjusted likelihood ratio test statistic (MLR in Mplus). We first examine the Schwartz value structure. Past research indicates the need to correct for method bias in responses to the SVS, such as the tendency of some respondents to score high on all values, regardless of their true value priorities (Schwartz 1992; Schwartz et al. 2012). Such mean differences between respondents in importance make little sense as some values are opposed to each other (see Figure 4). To control for this common method bias, we follow Schwartz et al. (2012) and include a latent method factor on which we fixed the loadings of all items to 1 (Maydeu-Olivares and Coffman 2006). Then, we estimate the structural relations between the value types, materialism, and life satisfaction, and perform a latent variable mediation analysis. Finally, we use the U.S. and Norwegian samples to illustrate cross-national invariance testing with UFA and estimation of a cross-national MIMIC model.

Factor Structure and Correlations

We illustrate the choice between UFA and CFA using the criteria introduced earlier. Regarding the first criterion, the CFA model does not achieve reasonable fit by conventional standards: $\chi^2(899) = 4256.9$, RMSEA = .056, SRMR = .093, CFI = .829. On the other hand, UFA achieves very good fit: $\chi^2(584) = 1195.9$, RMSEA = .030, SRMR = .019, CFI = .969. Turning to the second criterion, the difference in model chi-square between CFA and UFA is highly significant ($\Delta\chi^2\,(315) = 2796.8$, $p < .0001$), but this is only indicative given the large sample size.[4] However, the change in all three fit indices clearly exceeds the frequently used benchmarks of ΔRMSEA $\leq .015$, ΔSRMR $\leq .030$, and ΔCFI $\leq .010$.

The third criterion involves an examination of the factor structure. Table 4 shows that most value types are quite well-defined in the UFA solution. Forty-two out of 45 target loadings are significant, and for each value type, multiple target loadings are around or exceed .4. The composite reliability of all factors exceeds the signal-to-noise ratio of 2:1. The large majority of cross-loadings are below |.2| and only four cross-loadings exceed |.3|. Thus, the UFA solution constitutes roughly a 'good' approximation to simple structure. All four cross-loadings > |.3| are consistent with theory. The self-direction value "creativity" exhibits a cross-loading of .41 on stimulation, which is consistent with previous research (Steenkamp and Baumgartner 1992). That achievement is related to the importance attached to "choosing one's own goals" (cross-loading of .37) and that "capable" (cross-loading of .41) and "wisdom" (.32) are related to independent thought and action (self-direction) has face validity too. These cross-loadings reflect the fact that various values can tap into different value types. The CFA structure is reasonably well-defined too, and only security does not meet the signal-to-noise ratio of 2:1. Only three CFA target loadings are not significant.

--- Table 4 about here---

The fourth criterion concerns the difference in factor correlations. Table 5 provides the

[4] Because we use MLR, we use a modified chi-square difference test (Satorra and Bentler 2001). A convenient program to do the calculations is available on https://www.thestatisticalmind.com/calculators/SBChiSquareDifferenceTest.htm.

correlations free of measurement error among the 10 value types for UFA and CFA. The UFA factor correlations are modest in magnitude, the largest correlation being .52 between the adjacent values types of tradition/conformity and benevolence. On the other hand, 12 out of 45 CFA factor correlations exceed .5, and the CFA factor correlation matrix is not positive definite. The difference in four factor correlations between the two solutions exceeds .5, Cohen's (1988) cutoff for a large effect size. One such example involves stimulation and self-direction, which are correlated .97 in CFA but only .25 in UFA. As Table 4 shows, the self-direction values "curious" and "creativity" have cross-loadings of .29 and .41 respectively on stimulation while the stimulation value "a varied life" has a cross-loading of .23 on self-direction. There are no appreciable negative cross-loadings involving these two value types to mitigate the effect of these large positive cross-loadings and as a result, the CFA factor correlation is vastly inflated. Taken together, UFA provides stronger evidence that the value types are conceptually distinct than CFA. In fact, as we will see below, high correlations between the value types estimated using CFA virtually precludes their use in structural analyses (the fifth criterion).

We conclude by noting that there is a substantial amount of method variance in the SVS. This is consistent with Schwartz et al. (2012). The average method loading in UFA is .38 and .50 in CFA, indicating that 14% of the item variance in UFA is common method variance and 25% in CFA. This suggests that CFA pushes part of the effect of suppressed cross-loadings into the method factor. Failure to model common method variance leads to much larger factor correlations, especially in CFA. In UFA, the average correlation between the factors increases from .15 (with method factor) to .24 (without method factor). In CFA, the average correlation is .28 and .64, respectively. We note that the difference in chi-square between the models with vs. without a method factor is highly significant for both UFA and CFA. Web Appendix D provides the factor correlations for the models without the method factor.

--- Table 5 about here ---

Structural Model Analysis

As per our previous discussion, before conducting a structural analysis, we need to choose the most appropriate measurement model for the 10 value types, materialism, and life satisfaction (Figure 1). Not surprisingly, the 12-factor CFA model fits the data poorly: χ^2 (1417) = 5307.7, RMSEA = .048, SRMR = .078, CFI = .847. In contrast, the 12-factor UFA model achieved close fit: χ^2 (933) = 1799.5, RMSEA = .028, SRMR = .017, CFI = .966. The difference in model fit is significant and substantial. The UFA model revealed low cross-loadings between the SVS items, the materialism items, and the satisfaction with life items, the exception being a large cross-loading of .40 of the power item "wealth" (presented in the survey with a short explanation: "material possessions, money" see Web Appendix C) on materialism, which makes sense, given the two are nearly tautological. Set-UFA with three blocks of variables—independent variables (ten value types), mediator (materialism), and dependent variable (life satisfaction) provided also a close fit to the data: χ^2 (1102) = 2064.9, RMSEA = .027, SRMR = .027, CFI = .962. Although the difference in chi-square is significant ($\Delta\chi^2$ (169) = 281.4, $p <$ .001), the other fit indices change little or actually improve (RMSEA). Hence, we continue with ESEM (Figure 1), but we add a cross-loading of "wealth" on materialism.

The ESEM model achieves close fit: χ^2(1102) = 2085.7, RMSEA = .028, SRMR = .028, CFI = .961. Table 6 reports the structural coefficients. The standard errors are reasonable. With a sample size of 1,178, we can expect a standard error for standardized parameters of around .03. The ESEM standard errors are in this ballpark and do not fluctuate widely. Our structural findings are broadly consistent with the correlational results reported by Richins (2004). Consumers that place more emphasis on power, achievement and hedonism, and less emphasis on tradition/conformity are more materialistic, with power having the greatest effect. It is noteworthy that if we would have ignored the cross-loading of wealth on materialism, the effect of power would be inflated by 63%. This is yet another illustration for the importance of allowing for cross-loadings to arrive at precise structural effects. Furthermore, we find that materialism has a strong negative effect on satisfaction with life, and that it completely mediates the effects of achievement, hedonism, and tradition/conformity. For power, we observe competitive mediation: it has a significant positive direct effect on life satisfaction, and a significant negative effect through materialism. Security has a positive effect on life satisfaction, which is not mediated by materialism. The other value types have no effect on life satisfaction.

--- Table 6 about here ---

For comparison, we also estimated the traditional CFA/SEM model (Table 6). Although the model converged, the latent variable covariance matrix was not positive definite. This is due to extremely high correlations between the value types, which result in the standard errors for the structural coefficients that are large and fluctuate widely, indicating lack of statistical precision. None of the structural effects is significant. Thus, the SEM findings are neither robust nor plausible. Hence, it is not surprising that previous (pre-UFA) research often examined relations between the Schwartz value types and other constructs using bivariate correlations rather than multivariate regression (e.g., Richins 2004; Steenkamp and Burgess 2002). As Table 6 shows, the bivariate correlations between the value types and materialism look much more reasonable. To conclude, also our last criterion to choose between UFA and CFA, viz., assessment of the structural results, points toward UFA.

Cross-national analysis

With the surge of cross-cultural research, the issue of cross-national comparability of data is of great importance to business scholars. Multigroup CFA analysis is a "go-to" method to investigate cross-national measurement invariance (Steenkamp and Baumgartner 1998). According to these authors, the most basic substantive question is whether the constructs have the same basic factorial structure across countries. This requires configural invariance. The second question is whether the items designed to measure a factor have the same meaning across countries, which requires metric invariance. It is achieved when the unstandardized factor loadings are invariant across countries. Metric invariance is required for comparing unstandardized regression coefficients across groups. A third substantive question is whether the countries are different in their mean standing on the factors. To investigate this issue, scalar invariance is required, which is achieved when item intercepts are invariant across groups. Metric invariance is nested in configural invariance, and scalar invariance is nested in metric invariance. Traditional EFA lacks rigorous procedures for invariance testing. Since establishing measurement invariance has become a *sine qua non* in international business research, this more or less had forced substantive researchers to rely on CFA, even if their model does not fit well. However, all three types of invariance can be tested with UFA. We illustrate this by performing a multigroup analysis on the U.S. and Norwegian samples.

We conduct the series of invariance tests detailed by Steenkamp and Baumgartner (1998). The results are reported in Table 7, Panel A. Even the highly restrictive scalar invariance UFA model in which 450 factor loadings and 45 item intercepts are invariant, achieves close fit. The difference in chi-square between successive models is always significant, which is not surprising given the large sample size. More instructively, the deterioration in SRMR is substantially below commonly accepted benchmarks (less than .03 for metric invariance, less than .01 for scalar invariance; Chen 2007). The decline in CFI is close to the recommended threshold of .01 (Steenkamp and Maydeu-Olivares 2021), while RMSEA, which takes parsimony into account, does not change. Thus, we conclude that the SVS-UFA structure exhibits cross-national invariance. Table 7, Panel B reports latent means and variances for the U.S. We find that Americans on average attach less importance to power, stimulation, self-direction, concern for nature, and benevolence and more importance to achievement and tradition/conformity than Norwegians. Further, compared to Norwegians, Americans are more heterogeneous in the importance attached to power and concern for nature, and less heterogeneous in the importance attached to security. Finally, there is more method bias among American respondents and they are less heterogeneous in the amount of method bias than Norwegians.

We note that the cross-national invariance analysis with CFA runs into multiple problems. The configural invariance model does not converge. Turning to the scalar invariance model, the matrix of the factor correlations is not positive definite in either country, and model fit is quite poor: χ^2 (1867) =

7,882.0, RMSEA = .061, SRMR = .087, CFI = .852. Finally, we only observe significant differences in latent means for achievement and tradition/conformity. Regardless, results based on such a bad-fitting model cannot be trusted (Maydeu-Olivares 2017).

--- Table 7 about here ---

The scalar invariance model can be used in multigroup ESEM (or alternatively, the factor scores can be saved and used in multigroup path models). For illustrative purposes, we estimate a Multiple Indicators Multiple Causes (MIMIC) model specifying the effects of two key sociodemographics, gender (1=female, 0=male) and age (in years), on the ten value types. The MIMIC model achieves good fit: χ^2 (1688) = 3629.5, RMSEA = .037, SRMR = .033, CFI = .934. Table 8 reports the unstandardized regression coefficients, whose magnitude can compared across countries. Comparing the significance between the two countries can be misleading because of the differences in sample size. Hence, Table 8 identifies two instances where the Norwegian parameter estimates are reasonably close to being statistically significant at the 5% level. Before turning to these results, we test whether the effects of gender and age are the same across the two countries by constraining the respective parameter estimates to be cross-nationally invariant. The test on cross-national equality of the effects of gender and age is rejected ($\Delta\chi^2$ (22) = 366.3, $p < .001$)

The most striking result is the difference in the role of gender in value importance. Gender has a significant effect on the importance of 7 value types in the U.S. versus 2 in Norway. We speculate that this is due to the cross-national differences in the role of gender. While Norway ranks 6th in gender equality in the world, the U.S. ranks 46th, according to the United Nations (http://hdr.undp.org/en/content/gender-inequality-index-gii). The effect of age on value importance ratings is much more aligned across countries, with older people in both countries attaching more importance to Schwartz's value domains of self-transcendence (concern for nature, social concern, benevolence) and conservation (tradition/conformity, security).

--- Table 8 about here ---

EMPIRICAL APPLICATION TO ELECTRONIC SERVICE QUALITY

Context

The first empirical application represents a rather challenging empirical context because of the complexity of the Schwartz value theory. We now turn to a less complex model, involving four predictors and two outcome variables. Parasuraman et al. (2005) developed an influential 22-item scale[5] for electronic service quality called E-S-QUAL comprising of four dimensions (efficiency, system availability, fulfillment, and privacy), and related them to two outcomes (perceived value, and loyalty intentions). We reanalyze data collected by these authors among 593 Amazon users in the U.S. for whom there were no missing data. Table 9 provides the items. For brevity, we focus on the structural model.

Results

Following Figure 1, we first estimate the overall six-factor CFA and UFA models. CFA fits the data reasonably well: χ^2 (419) = 1400.5, RMSEA = .063, SRMR = .051, CFI = .912. In fact, based on this model fit, most researchers would feel comfortable using this model. However, UFA achieves a much better fit: χ^2 (294) = 753.7, RMSEA = .051, SRMR = .019, CFI = .959. The difference in model chi-square is highly significant $\Delta\chi^2$ (125) = 616.4, $p < .0001$. The change in RMSEA, SRMR, and CFI is also large. Table 9 presents the factor loadings for both models. Both factor solutions are well-defined with significant and large target loadings. For both models, the composite reliability of all factors exceeds .80. UFA achieves a good approximation to simple structure. There are several substantial cross-loadings, the highest being the cross-loading of EFF5 (“It loads its pages fast”) and EFF7 (“This site enables me to get on to it quickly”) on system availability. This makes sense since system availability refers to “the correct technical functioning of the site” (Parasuraman et al. 2005, p. 220). Turning to the factor correlations (Table 10), we see differences in the range of .2 to .3 between UFA and CFA for system availability with the other E-S-QUAL dimensions, perceived value and loyalty

[5] By June 2022, the article had garnered nearly 6,000 Google Scholar citations.

intentions, as well as a difference of .16 in the factor correlation between perceived value and efficiency between the two factor solutions. The other differences in correlations are minor. In sum, detailed examination of difference in model fit and the parameter estimates suggests that UFA is the more appropriate factor solution, although the difference with CFA is not nearly as dramatic as in the first empirical application.

--- Tables 9 and 10 about here ---

The next question is whether the single UFA model can be split into two models, one for the four independent constructs (efficiency, system availability, fulfillment, privacy) and one for the two dependent constructs (value, loyalty) (Figure 1). This set-UFA model achieves good fit: χ^2 (358) = 966.3, RMSEA = .053, SRMR = .027, CFI = .946. The difference in model chi-square with the single UFA model is highly significant $\Delta\chi^2$ (64) =215.8, p < .0001, but the change in the other fit indices is not large (only CFI falls a little short). Here we have an example where detailed inspection of the results is particularly useful. Although the set-UFA model quite performs well, Table 9 reveals cross-loadings exceeding .3 for VAL2 (“The overall convenience of using this site”) and VAL3 (“The extent to which the site gives you a feeling of being in control”) on the E-S-QUAL dimension of efficiency. This makes sense as these items have an efficiency connotation, which is defined as “The ease and speed of accessing and using the site” (Parasuraman et al. 2005, p. 220). In the set-UFA model, these cross-loadings are suppressed which leads to an inflated correlation between efficiency and value of .79 (vs. .64 in the single UFA model). We continue with the single UFA model because it accommodates these substantial cross-loadings, thus leading to a more accurate estimate of the effect of efficiency on value.

We next estimate the structural model using EwSEM (Figure 1). Consistent with our earlier description, since the UFA model contains *m*=6 factors, we have to impose 6^2 =36 restrictions on the parameters. We constrain all factor variances to one and select a reference indicator for each factor (EFF4, SYS3, FUL1, PRI2, VAL4, LOY3) and fix their small cross-loadings to their estimated values as obtained in the overall UFA solution. This introduces $m(m-1) = 30$ additional model constraints. The other factor loadings are freely estimated. With this EwSEM model specification, we can now estimate the structural effects of the four E-S-QUAL factors on perceived value and loyalty intentions, while accounting for cross-loadings within and between independent and dependent constructs. As expected (Marsh et al. 2013b) the fit of this model is very close to that of the single UFA model: χ^2 (294) = 754.0 vs. χ^2 (294) = 753.7.

Table 11 provides the standardized regression coefficients for EwSEM. For comparison, we also report the standard SEM estimates. We find some interesting differences. In SEM, two of the four dimensions have no significant effect on either outcome of E-S-QUAL. This calls into question the usefulness of system availability and privacy as constructs of interest and suggests that firms should focus on efficiency and fulfillment. The absence of an effect of privacy on perceived value is also inconsistent with Steenkamp and Geyskens (2006). EwSEM yields a more nuanced picture. Three of the four dimensions matter, and EwSEM replicates the previous finding that privacy affects perceived value of websites. The overwhelming dominance of efficiency on value as found in SEM is reduced by about 50%, while fulfillment plays a larger role.

Finally, we also present the path analysis results for the regression factor scores based on the overall UFA model. The results are close to EwSEM. This is because the reliability of all factor scores was extremely high, the lowest being .84 for system availability.

--- Table 11 about here---

DISCUSSION

Researchers using multiple indicators to measure unobserved constructs need to document the psychometric quality of their measures. This applies regardless of the data collection method. In fact, many so-called objective measures also contain measurement error (e.g., Aruoba et al. 2016; Bollen and Schwing 1987). Since the 1980s, the gold standard in the psychometric analysis of (reflective) multi-item data has been CFA. While CFA constitutes a powerful model, it also imposes very restrictive assumptions on the measurement model, viz., cross-loadings are constrained to zero. Our simulation studies show that if non-zero cross-loadings are present and constrained to zero, this can substantially bias factor correlations. When the source of model misfit is due to omitted loadings, UFA provides an alternative applied researchers should consider to avoid making incorrect substantive conclusions. UFA has been extended to ESEM and we have illustrated in this article that many of the research questions that were addressed using CFA measurement models can now be addressed using the less restricted UFA measurement model.

We have provided two detailed applications to illustrate the differences that can be expected when using UFA vs. CFA as measurement model. Table 12 summarizes findings for several other data sets that we analyzed for the purposes of this article. These analyses provide additional empirical evidence that: 1) UFA often provides a substantially better fit than CFA; 2) the difference in factor correlations between CFA and UFA is often appreciable, and that as a result: 3) CFA and UFA often lead to different results in structural analyses. The difference in factor correlations and structural estimates is especially large when the cross-loadings are substantial and the sign of (most) cross-loadings is aligned with the sign of the factor correlations, an observation that is consistent with our simulation studies. However, there are also instances where the difference in factor correlations is negligible and CFA and UFA lead to similar conclusions. In these instances, cross-loadings are small, and/or positive and negative cross-loadings are approximately equal in size and occurrence.

--- Table 12 about here ---

Recommendations

We offer six recommendations to substantive business scholars. We start with two general recommendations regarding measurement analysis and next introduce four recommendations about UFA versus CFA (Table 13).

--- Table 13 about here ---

First, always conduct an in-depth measurement analysis and always provide information on model fit. This applies regardless whether established scales are used or new scales are developed. Although this recommendation may seem obvious, rigorous measurement analysis is often found wanting. Without information on model fit, it is simply impossible to be confident about the validity of the inferences drawn from the measures used. It could well be that an alternative model provides a substantial better fit, and that under the alternative model different conclusions are drawn (Maydeu-Olivares 2017). Even more problematic, some studies apparently conduct no measurement analysis at all in that they only provide coefficient α to inform the quality of their measures. We observed this practice in a substantial number of recent articles that appeared in leading marketing journals. Coefficient α is a normed index of a sum score's precision in measuring a unidimensional construct. Coefficient α increases with the number of items but provides no information about whether a set of items measures a single construct (Sijtsma 2009). Rather, it assumes that a single construct is measured. It is even insufficient to just report coefficient α if you use established scales because there is no guarantee that previous findings on measurement unidimensionality are replicated in your sample (Marsh et al. 2013a). This matters as your results depend on the psychometric characteristics of the measures in your sample (Morin et al. 2016).

Second, use maximum likelihood with standard errors and goodness of fit tests robust to non-normality (MLR) as the default estimation method. Rarely have we analyzed data that did not deviate from normality. If MLR is used, a scale correction factor is provided in the computer output, which is indicative of the degree of non-normality of the data (Satorra and Bentler 1994). The more it deviates from one, the greater the deviation from multivariate normality. By using MLR rather than ML under normality, the researcher can assess whether it is necessary to use MLR simply by inspecting the scale correction factor.

Third, analyze the data with both CFA and UFA, and use informed judgment to identify the preferred model. Criteria you can use for this include model fit, difference in model fit between UFA and CFA, examination of the factor structure, comparison of the factor correlations between UFA and

CFA, and (if applicable) comparison of the structural parameter estimates and their standard errors. We realize that this requires a little extra effort. But apart from consideration that it is a scholarly obligation to produce findings that are as accurate as possible, the time investment in these additional analyses pales against the time spent on developing one's theory and on data collection.

Fourth, test your theory and hypotheses within a latent variables structural equation framework. Measurement error in predictors biases OLS regression coefficients, and this bias is not mitigated with large samples (Greene 2003). In the not-too-distant past, SEM was restricted to estimating linear effects, using interval-scaled dependent variables. However, advances in psychometric theory now allow researchers to estimate quadratic effects and interactions for latent constructs, use bootstrapping for indirect effects, and analyze dependent variables that are censored, categorical, or count data within a CFA/SEM framework. As currently developed in psychometrics, these models cannot be estimated with ESEM but can be estimated without difficulty with EwSEM. So, if one's model is ESEM, simply convert it to EwSEM—a highly saturated SEM model--as described in this article. Web Appendix E provides two illustrations, one involving quadratic effects and interactions between latent variables, and another involving a negative binomial regression of count data on three latent consumer constructs where we also include interactions between the latent constructs. These models are a far cry from standard SEM.

Fifth, if the researcher nevertheless wants to use factor scores as observed variables in regression/path analysis, the choice of the type of factor scores depends on the measurement model. Sum scores might be used if measurement analysis supports the use of CFA because each item loads only on one factor. However, sum scores should never be used when the measurement analysis supports UFA as they do not account for cross-loadings. In the presence of non-negligible cross-loadings, researchers should use more complex factor scores such as regression factor scores. But we caution that factor scores may lead to different conclusions than obtained using latent variables structural equation modeling unless a bias correction is performed (Croon 2002). A key factor here is the reliability of the factor scores. If the reliability is low, the results can be very different. In our experience, the standard reliability cutoff of .7 might be too lenient.

Sixth, the above mentioned advances in the psychometric analysis have all been incorporated in Mplus. At the time of this writing, they are not available in other popular packages such as LISREL or AMOS. Another option is R. To the best of our knowledge, R is currently restricted to a single block UFA. Neither set-UFA nor ESEM models can be estimated. Until these models are incorporated in R, we recommend R users to estimate the measurement model with a single, full-UFA and next estimate the structural relations using EwSEM or, less preferable, standard regression/path analysis with regression factor scores (Figure 1).

Limitations and Future Research

UFA is theory-based and more rigorous than traditional EFA while being more flexible than CFA. It allows substantive marketing and business scholars to apply the advanced statistical methods available in CFA and SEM. However, UFA is neither a panacea for all measurement challenges nor should it be used or seen as such. There is no guarantee that a UFA model achieves at least moderate approximation to simple structure. If UFA using a target rotation leads to many substantial cross-loadings, it calls into question the quality of the measures per se. Such a solution should not be used for theory testing. Of course, forcing these data into a CFA model is not the solution either. Rather, it requires the researcher to go back to the drawing board and reconceptualize their theory.

An important area for future research is how to deal with method variance in factor models. The latent factor approach was effective in capturing method variance in the SVS application but in this case, high vs. low scores on all values can be credibly interpreted as method variance (Schwartz et al. 2012). We did not add such a method factor to E-S-QUAL because here high vs. low scores on the overall factor have a straightforward and valid behavioral interpretation. A related area is to what extent cross-loadings are due to substantive overlap between items or to method effects. Therefore, it is useful to evaluate major cross-loadings in detail and assess whether they can be explained by theory or whether they appear to be the result of method variance between selected items. In the first case, they are to be modeled as cross-loadings; in the latter case, they are to be modeled using correlated errors. Failing to account for method effects among selected items via correlated errors in UFA and CFA models is another area for future research.

Our simulation studies did not include scenarios of different proportions of positive versus negative cross-loadings. Such conditions dampen the effect of suppression of cross-loadings in CFA on factor correlations (see Table 12). Although our second simulation study showed little impact of sample size except when the factor structure is weak, additional simulation studies are needed, as the UFA model involves more parameters than the corresponding CFA model. We further study the relation between RMSEA and the extent of parameter bias (absolute difference between true and predicted factor correlation) in misspecified CFA models. Future research using simulation studies should provide recommendations under a wider range of scenarios than those considered here and other goodness-of-fit indices.

Some features such as higher-order factor models, interactions between latent variables, quadratic effects of latent variables, mediation analysis with bootstrapping, and structural analysis with non-interval scaled dependent variables are only available in UFA/ESEM if the researcher reformulates the model as EwSEM model (Table 1). We show that this is easy to do and model fit is the same as the corresponding UFA/ESEM model. Future research is needed to develop statistical methods to allow such models to be directly estimated as ESEM models.

Our presentation has focused on frequentist methods. Alternatively, a Bayesian framework may be adopted. Some research has started to compare the performance of Bayesian Structural Equation Modeling (BSEM: Muthén and Asparouhov 2012) with SEM and ESEM (e.g., Guo et al. 2019) but much more research is needed before we know under which conditions a frequentist versus a Bayesian approach is preferable.

To conclude, UFA is <u>not</u> the solution for badly designed or ill-behaved measurement instruments. However, as we show in this article, it offers a powerful alternative to substantive business scholars if CFA is simply too restrictive, leading to model anomalies which carry over into factor correlations and structural parameter estimates. We hope that this article will encourages other marketing researchers to include UFA in their arsenal of measurement analysis tools to test their substantive theories.

REFERENCES

Aruoba, S. Borağan, Francis X. Diebold, Jeremy Nalewaik, Frank Schorfheide, Dongho Song (2016), "Improving GDP Measurement: A Measurement-error Perspective," *Journal of Econometrics*, 191 (2), 384-397.

Asparouhov, Tihomir and Bengt Muthén (2009), "Exploratory Structural Equation Modeling," *Structural Equation Modeling*, 16 (3), 397-438.

Asparouhov, Tihomir, Bengt Muthén, and Alexandre J.S. Morin (2015), "Bayesian Structural Equation Modeling with Cross-Loadings and Residual Covariances," *Journal of Management*, 41 (September), 1561-1577.

Bagozzi, Richard P. and Youjae Yi (1989), "On the Use of Structural Equation Models in Experimental Designs," *Journal of Marketing Research*, 26 (3), 271-284.

Batra, Rajeev, Pamela M. Homer, and Lynn R. Kahle (2001), "Values, Susceptibility to Normative Influence, and Attribute Importance Weights: A Nomological Analysis," *Journal of Consumer Psychology*, 11 (2), 115-128.

Batra, Rajeev, Y. Charles Zhang, Nilüfer Z. Aydinoğlu, and Fred M. Feinberg (2017), "Positioning Multicountry Brands: The Impact of Variation in Cultural Values and Competitive Set," *Journal of Marketing Research*, 44 (6), 914-931.

Baumgartner, Hans and Jan-Benedict E.M. Steenkamp (1996), "Exploratory Consumer Buying Behavior: Conceptualization and Measurement," *International Journal of Research in Marketing*, 13 (2), 121-137.

Bearden, William O., Richard G. Netemeyer, and Jesse E. Teel (1989), "Measurement of Consumer Susceptibility to Interpersonal Influence," *Journal of Consumer Research*, 15 (March), 473-481.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Bollen, Kenneth A. and Richard C. Schwing (1987), "Air Pollution-Mortality Models: A Demonstration of the Effects of Random Measurement Error," *Quality and Quantity*, 21, 37-48.

Brangule-Vlagsma, Kristine, Rik G.M. Pieters, and Michel Wedel (2002), "The Dynamics of Value Segments: Modeling Framework and Empirical Illustration," *International Journal of Research in Marketing*, 19 (3), 267-285.

Brown, Timothy A. (2015), *Confirmatory Factor Analysis for Applied Research*, New York: Guilford Press, 2nd ed.

Browne, Michael W. (1972), "Oblique Rotation to a Partially Specified Target," *British Journal of Mathematical and Statistical Psychology*, 25 (2), 207–12.

Browne, Michael W. (2001), "An Overview of Analytic Rotation in Exploratory Factor Analysis," *Multivariate Behavioral Research*, 36 (1), 111–50.

Burroughs, James E. and Aric Rindfleisch (2002), "Materialism and Well-Being: A Conflicting Values Perspective," *Journal of Consumer Research*, 29 (December), 348-369.

Chen, Fang Fang (2007), "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance," *Structural Equation Modeling*, 14 (3), 464-504.

Cheung, Gordon W. and Roger B. Rensvold (2002), "Evaluating Goodness-of Fit Indexes for Testing Measurement Invariance," *Structural Equation Modeling*, 9 (2), 233-255.

Church, Timothy A. and Peter J. Burke (1994), "Exploratory and Confirmatory Tests of the Big Five and

Tellegen's Three-and Four-Dimensional Models," *Journal of Personality and Social Psychology*, 66 (1), 93-114.
Cohen, Jacob (1988), *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale, NJ: Lawrence Erlbaum, 2nd ed. α
Cohen, Jacob, Patricia Cohen, Stephen G. West, and Leona S. Aiken (2003), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Mahwah, NJ: Erlbaum, 3rd ed.
Cole, David A., Jeffrey A. Ciesla, and James H. Steiger (2007), "The Insidious Effects of Failing to Include Design-Driven Correlated Residuals in Latent-Variable Covariance Structure Analysis," *Psychological Methods*, 12 (4), 381–398.
Croon, Marcel A. (2002), "Using Predicted Latent Scores in General Latent Structure Models," in *Latent Variable and Latent Structure Modeling*, G. A. Marcoulides and I. Moustaki, eds., Mahwah, NJ: Lawrence Erlbaum, 195–223.
Cudeck, Robert and Lisa L. O'Dell (1994), "Applications of Standard Error Estimates in Unrestricted Factor Analysis: Significance Tests for Factor Loadings and Correlations," *Psychological Bulletin*, 115 (3), 475–87.
De Jong, Martijn G., Jean-Paul Fox, and Jan-Benedict E.M. Steenkamp (2015), "Quantifying Under- and Overreporting In Surveys Through A Dual-Questioning-Technique Design," *Journal Of Marketing Research*, 52 (6), 737–53.
De Luca, Luigi M. Dennis Herhausen, Gabriele Troilo, and Andrea Rossi (2021), "How and When Do Big Data Investments Pay Off? The Role of Marketing Affordances and Service Innovation," *Journal of the Academy of Marketing Science*, 49 (4), 790-810.
Diener, Ed, Robert A. Emmons, Randy Larsen, and Sharon Griffin (1985), "The Satisfaction With Life Scale," *Journal of Personality Assessment*, 49 (February), 71-75.
Fischer, Marc, Franziska Völckner, and Henrik Sattler (2010), "How Important Are Brands: A Cross-Category, Cross-Country Study," *Journal of Marketing Research*, 47 (October), 823-839.
Geuens, Maggie, Bert Weijters, and Kristof De Wulf (2009), "A New Measure of Brand Personality," *International Journal of Research in Marketing*, 26 (1), 97-107.
Green, Samuel B. and Scott L. Hershberger (2000), "Correlated Errors in True Score Models and Their Effect on Coefficient Alpha," *Structural Equation Modeling*, 7 (2), 251–70.
Greene, William H. (2003) *Econometric Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 5th ed.
Guo, Jiesi, Herbert W. Marsh, Philip D. Parker, Theresa Dicke, Oliver Lüdtke, and Thierno M. O. Diallo (2019), "A Systematic Evaluation and Comparison Between Exploratory Structural Equation Modeling and Bayesian Structural Equation Modeling," *Structural Equation Modeling*, 26 (4), 529-556.
Hair, Joseph F., Barry J. Babin, Rolph E. Anderson, and William C. Black (2018), *Multivariate Statistics*, Cengage.
Haws, Kelly L., Utpal M. Dholakia, and William O. Bearden (2010), "An Assessment of Chronic Regulatory Focus Measures," *Journal of Marketing Research*, 47 (October), 967-982.
Hulland, John, Hans Baumgartner, and Keith M. Smith (2018), "Marketing Survey Research Best Practices: Evidence and Recommendations from a Review of JAMS Articles," *Journal of the Academy of Marketing Science*, 46, 92-108.
Jöreskog, Karl G. (1969), "A General Approach to Confirmatory Maximum Likelihood Factor Analysis," *Psychometrika*, 34 (2), 183-202.
Jöreskog, Karl G. (1978), "Structural Analysis of Covariance and Correlation Matrices," *Psychometrika*, 43 (4), 443-477.
Jöreskog, Karl G. and Arthur S. Goldberger (1975), "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association*, 70 (351), 631.
MacKenzie, Scott B. (2001), "Opportunities for Improving Consumer Research through Latent Variable Structural Equation Modeling," *Journal of Consumer Research*, 28 (1), 159-166.
Marsh, Herbert W., Jiesie Guo, Theresa Dicke, Philip D. Parker, and Rhonda G. Craven (2020), "Confirmatory Factor Analysis (CFA), Exploratory Structural Equation Modeling (ESEM), and Set ESEM: Optimal Balance Between Goodness of Fit and Parsimony," *Multivariate Behavioral Research*,

55 (1), 102-119.
Marsh, Herbert W., Kit-Tai Hau, and David Grayson (2005), "Goodness of Fit in Structural Equation Models," in: *Contemporary Psychometrics*, Alberto Maydeu-Olivares and John J. McArdle (eds.), Mahwah, NJ: Erlbaum, 275-340.
Marsh, Herbert W., Oliver Lüdtke, Benjamin Nagengast, Alexandre J. S. Morin, and Matthias von Davier (2013a), "Why Item Parcels Are (Almost) Never Appropriate: Two Wrongs Do Not Make a Right—Camouflaging Misspecification With Item Parcels in CFA," *Psychological Methods*, 18 (3), 257-284.
Marsh, Herbert, Benjamin Nagengast, and Alexandre J. S. Morin (2013b), "Measurement Invariance of Big-Five Factors Over the Life Span: ESEM Tests of Gender, Age, Plasticity, Maturity, and La Dolce Vita Effects," *Developmental Psychology*, 49 (6), 1194-1218.
Marsh, Herbert W., Alexandre J.S. Morin, Philip D. Parker, and Gurvinder Kaur (2014), "Exploratory Structural Equation Modeling: An Integration of the Best Features of Exploratory and Confirmatory Factor Analysis," *Annual Review of Clinical Psychology*, 85-110.
Maydeu-Olivares, Alberto (2017), "Assessing the Size of Model Misfit in Structural Equation Models," *Psychometrika*, 82 (3), 533–58.
Maydeu-Olivares, Alberto and Donna L. Coffman (2006), "Random Intercept Item Factor Analysis," *Psychological Methods*, 11 (4), 344–62.
McDonald Roderick P. (1985), *Factor Analysis and Related Methods*, Hillsdale, NJ: Erlbaum.
McDonald, Roderick P. (1999), *Test Theory: A Unified Approach*, Mahwah, NJ: Erlbaum.
Morin, Alexandre J S., A. Katrin Arens, and Herbert W. Marsh (2016), "A Bifactor Exploratory Structural Equation Modeling Framework for the Identification of Distinct Sources of Construct-Relevant Psychometric Multidimensionality," *Structural Equation Modeling,* 23 (1), 116–139.
Morin Alexandre J.S., Herbert W. Marsh, and Benjamin Nagengast (2013), "Exploratory Structural Equation Modeling: An Introduction," in: *Structural Equation Modeling: A Second Course*, ed. Gregory R. Hancock and Ralph O. Mueller (eds.), Greenwich, CT: IAP, 2nd ed., 395– 436.
Moshagen, Morten (2012), "The Model Size Effect in SEM: Inflated Goodness-of-Fit Statistics Are Due to the Size of the Covariance Matrix," *Structural Equation Modeling*, 19 (1), 86–98.
Muthén, Bengt and Tihomir Asparouhov (2012), "Bayesian Structural Equation Modeling: A More Flexible Representation of Substantive Theory," *Psychological Methods*, 17(3), 313-335.
Muthén, Linda K. and Bengt Muthén (2017), "Mplus 8 [Computer Program]," Los Angeles, CA: Muthén & Muthén.
Niemand, Thomas and Robert Mai (2018), "Flexible Cutoff Values for Fit Indices in the Evaluation of Structural Equation Models," *Journal of the Academy of Marketing Science*, 46 (6), 1148-1172.
Nunnally, Jum C. (1978), *Psychometric Theory*, New York, New York: McGraw-Hill, 2nd edition.
Parasuraman, A, Valarie A. Zeithaml, and Arvind Malhotra (2005), "E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality," *Journal Of Service Research*, 7 (3), 213–33.
Paul, Michael, Thorsten Hennig-Thurau, Dwayne D. Gremler, Kevin P. Gwinner, and Caroline Wiertz (2009), "Toward a Theory of Repeat Purchase Drivers for Consumer Services," *Journal of the Academy of Marketing Science*, 37 (2), 215-237
Pavlov, Goran, Dexin Shi, and Alberto Maydeu-Olivares (2020), "Chi-square Difference Tests for Comparing Nested Models: An Evaluation with Non-normal Data," *Structural Equation Modeling*, 27 (6), 908-917.
Pedhazur, Elazar J. (1982), *Multiple Regression in Behavioral Research*, New York: Holt, Rinehart and Winston, 2nd edition.
Richins, Marsha L. (2004), "The Material Values Scale: Measurement Properties and Development of a Short Form," *Journal of Consumer Research*, 31 (1), 209–19.
Richins, Marsha L. and Scott Dawson (1992), "A Consumer Values Orientation for Materialism and Its Measurement: Scale Development and Validation," *Journal of Consumer Research*, 19 (3), 303-316.
Roth, Martin S. (1995), "The Effects of Culture and Socioeconomics on the Performance of Global Brand Image Strategies," *Journal of Marketing Research*, 32 (May), 163-175.

Ruvio, Ayalla, Eli Somer, and Aric Rindfleisch (2014), "When Bad Gets Worse: The Amplifying Effect of Materialism on Traumatic Stress and Maladaptive Consumption," *Journal of the Academy of Marketing Science*, 42 (1), 90-101.

Ryan, Lisa and Suzanne Dziurawiec (2001), "Materialism and Its Relationship to Life Satisfaction," *Social Indicators Research*, 55 (2), 185-197.

Satorra, Albert and Peter M. Bentler (1994),"Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis," In A. von Eye and C.C. Clogg (Eds.), *Latent Variables Analysis: Applications for Developmental Research*, Thousand Oaks: Sage, 399–419.

Satorra, Albert and Peter M. Bentler (2001), "A Scaled Difference Chi-Square Test Statistic for Moment Structure Analysis," *Psychometrika*, 66 (4), 507-514.

Schwartz, Shalom H. (1992), "Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries," in *Advances in Experimental Social Psychology*, Vol. 25, ed. Mark Zanna, Orlando, FL: Academic Press, 1-65.

Schwartz, Shalom H. and Klaus Boehnke (2004), "Evaluating the Structure of Human Values with Confirmatory Factor analysis," *Journal of Research in Personality*, 38, 230-255.

Schwartz, Shalom H., Jan Cieciuch et al. (2012), "Refining the Theory of Basic Individual Values," *Journal of Personality and Social Psychology*, 103 (4), 663-688.

Schwartz, Shalom H. and Lilach Sagiv (1995), "Identifying Culture-Specifics in the Content and Structure of Values," *Journal of Cross-Cultural Psychology*, 26 (January), 92-116.

Shepherd, Steven, Tanya L. Chartrand, and Gavan J. Fitzsimons (2015), "When Brands Reflect Our Ideal World: The Values and Brand Preferences of Consumers Who Support versus Reject Society's Dominant Ideology," *Journal of Consumer Research*, 42 (1), 76-92.

Shi, Dexin, Alberto Maydeu-Olivares, and Christine DiStefano (2018), "The Relationship Between the Standardized Root Mean Square Residual and Model Misspecification in Factor Analysis Models," *Multivariate Behavioral Research*, 53 (5), 676–94.

Shi, Dexin, Taehun Lee, and Alberto Maydeu-Olivares (2019), "Understanding the Model Size Effect on SEM Fit Indices," *Educational and Psychological Measurement*, 79 (2), 310–34.

Sijtsma, Klaas (2009), "On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha," *Psychometrika*, 74 (March), 107-120.

Skrondal, Anders and Sophia Rabe-Hesketh (2004), *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, Boca Raton, FL: CRC Press.

Steenkamp, Jan-Benedict E.M. and Hans Baumgartner (1992), "The Role of Optimum Stimulation Level in Exploratory Consumer Behavior," *Journal of Consumer Research*, 19 (3), 434-448.

Steenkamp, Jan-Benedict E.M. and Hans Baumgartner (1998), "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research*, 25 (1), 78-90.

Steenkamp, Jan-Benedict E.M. and Hans Baumgartner (2000), "On the Use of Structural Equation Models in Marketing Modeling," *International Journal of Research in Marketing*, 17 (June-September), 195-202.

Steenkamp, Jan-Benedict E.M. and Steven M. Burgess (2002), "Optimum Stimulation Level and Exploratory Consumer Behavior in an Emerging Consumer Market," *International Journal of Research in Marketing*, 19 (2), 131-150.

Steenkamp, Jan-Benedict E.M. and Martijn G. de Jong (2010), "A Global Investigation into the Constellation of Consumer Attitudes Toward Global and Local Products," *Journal of Marketing*, 74 (November), 18-40.

Steenkamp, Jan-Benedict E.M., Martijn G. de Jong, and Hans Baumgartner (2010a) "Socially Desirable Response Tendencies in Survey Research," *Journal of Marketing Research*, 47 (April), 199-214.

Steenkamp, Jan-Benedict E.M. and Inge Geyskens (2006), "How Country Characteristics Affect the Perceived Value of Web Sites," *Journal of Marketing*, 70 (July), 136-150.

Steenkamp, Jan-Benedict E.M. and Alberto Maydeu-Olivares (2015), "Stability and Change in Consumer Traits: Evidence from a Twelve-Year Longitudinal Study, 2002-2013," *Journal of Marketing Research*, 52 (June), 287-308.

Steenkamp, Jan-Benedict E.M. and Alberto Maydeu-Olivares (2021), "An Updated Paradigm for Evaluating Measurement Invariance Incorporating Common Method Variance and Its Assessment," *Journal of the Academy of Marketing Science*, 49 (1), 5-29.

Steenkamp, Jan-Benedict E.M., Frenkel Ter Hofstede and Michel Wedel (1999), "A Cross-National Investigation into the Individual and National-Cultural Antecedents of Consumer Innovativeness," *Journal of Marketing*, 63 (April), 55-69.

Steenkamp, Jan-Benedict E.M., Harald van Heerde, and Inge Geyskens (2010b), "What Makes Consumers Willing to Pay a Price Premium for National Brands over Private Labels?," *Journal of Marketing Research*, 47 (December), 1011-1024.

Steenkamp, Jan-Benedict E.M. and Michel Wedel (1991), "Segmenting Retail Markets on Store Image Using a Consumer-Based Methodology," *Journal of Retailing*, 67 (3), 300-20.

Swenson, Michael J. and Joel Herche (1994), "Social Values and Salesperson Performance: An Empirical Examination," *Journal of the Academy of Marketing Science*, 22 (3), 283-289.

Tabachnick, Barbara and Linda Fidell (2018), *Using Multivariate Statistics*, Boston, MA: Pearson, 5th ed.

Thomson, Matthew, Deborah J. MacInnis, and C. Whan Park (2005), "The Ties That Bind: Measuring the Strength of Consumers' Emotional Attachments to Brands," *Journal of Consumer Psychology*, 15 (1), 77-91.

Torelli, Carlos J., Ayşegül Özsomer, Sergio W. Carvalho, Hean Tat Keh, and Natalia Maehle (2012), "Brand Concepts as Representations of Human Values: Do Cultural Congruity and Compatibility Between Values Matter?," *Journal of Marketing*, 76 (4), 92-108.

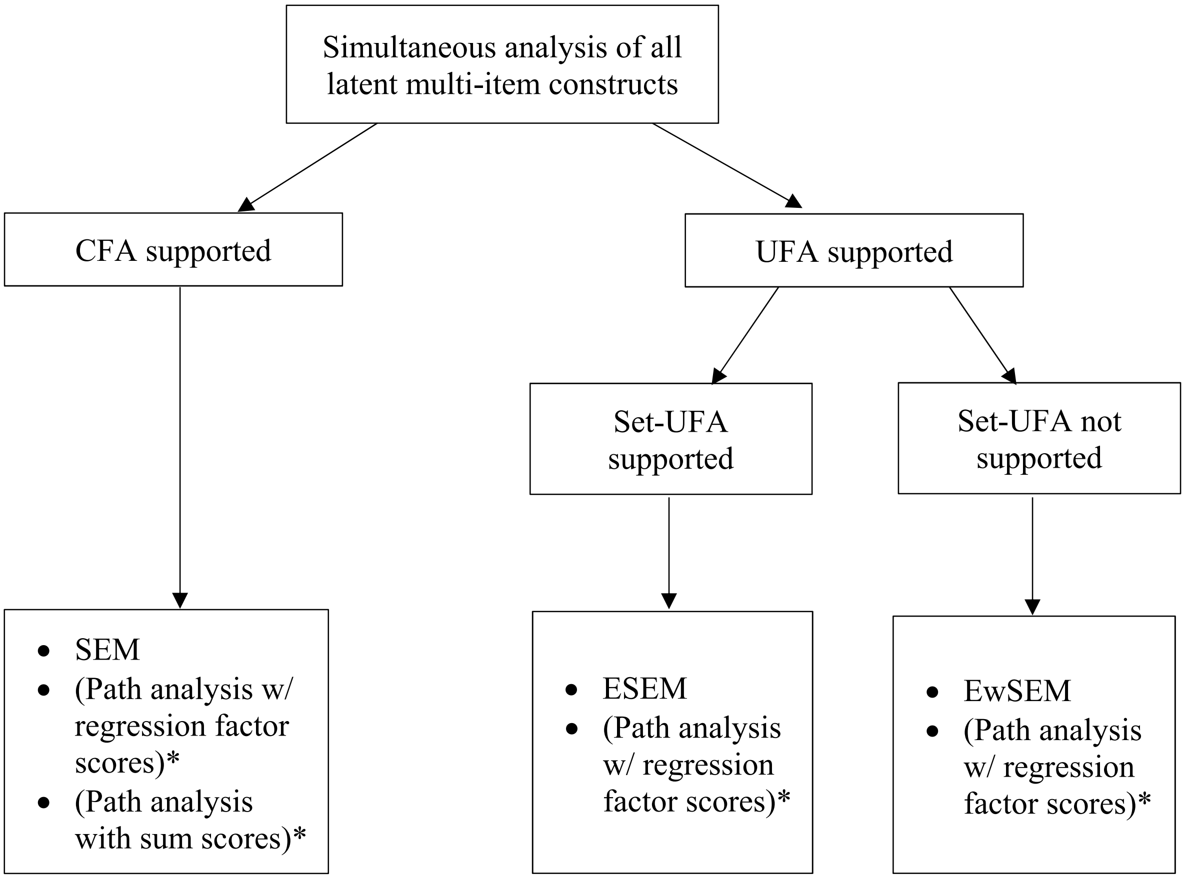
Unnava, Vasu and Ashwin Aravindakshan (2021), "How Does Consumer Engagement Evolve When Brands Post Across Multiple Social Media?" *Journal of the Academy of Marketing Science*, 49 (5), 864-881.

Winterich, Karen P. and Yinlong Zhang (2014), "Accepting Inequality Deters Responsibility: How Power Distance Decreases Charitable Behavior," *Journal of Consumer Research*, 41 (2), 274-293.

Xie, Chunyan, Richard P. Bagozzi, and Kjell Grønhaug (2015), "The Role of Moral Emotions and Individual Differences in Consumer Responses to Corporate Green and Non-green Actions," *Journal of the Academy of Marketing Science*, 43 (3), 333-356.

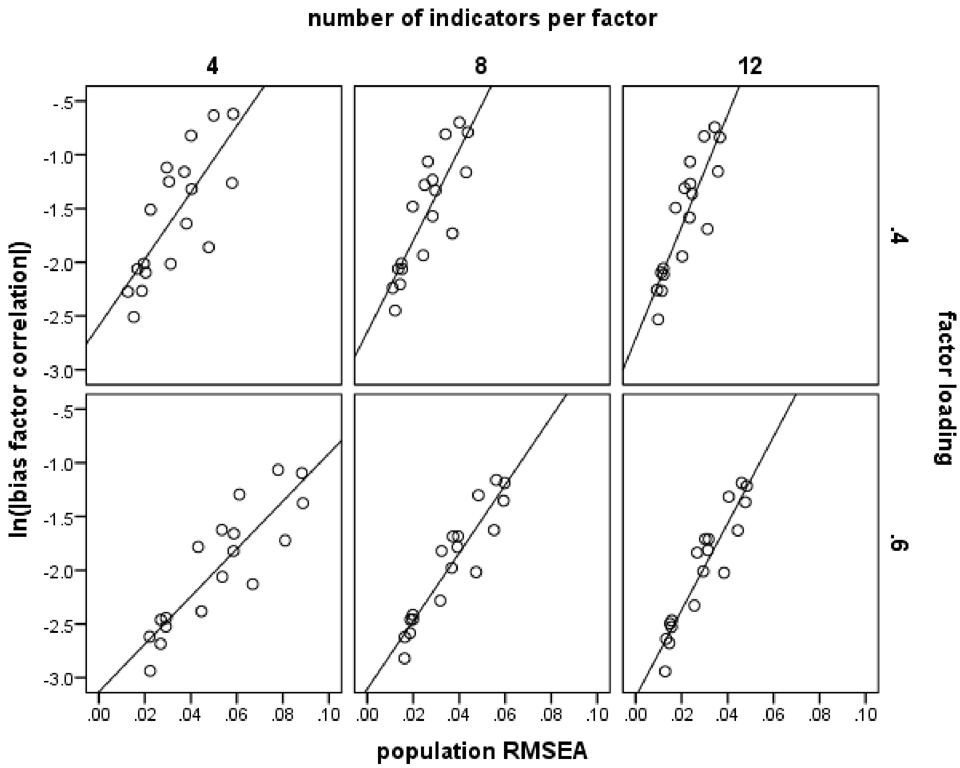
Ximénez, Carmen, Alberto Maydeu-Olivares, Dexin Shi, and Javier Revuelta (2022), "Assessing Cutoff Values of SEM Fit Indices: Advantages of the Unbiased SRMR Index and Its Cutoff Criterion Based on Communality," *Structural Equation Modeling*, 29 (3), 368-380.

Figure 1: Structural analysis with UFA and CFA



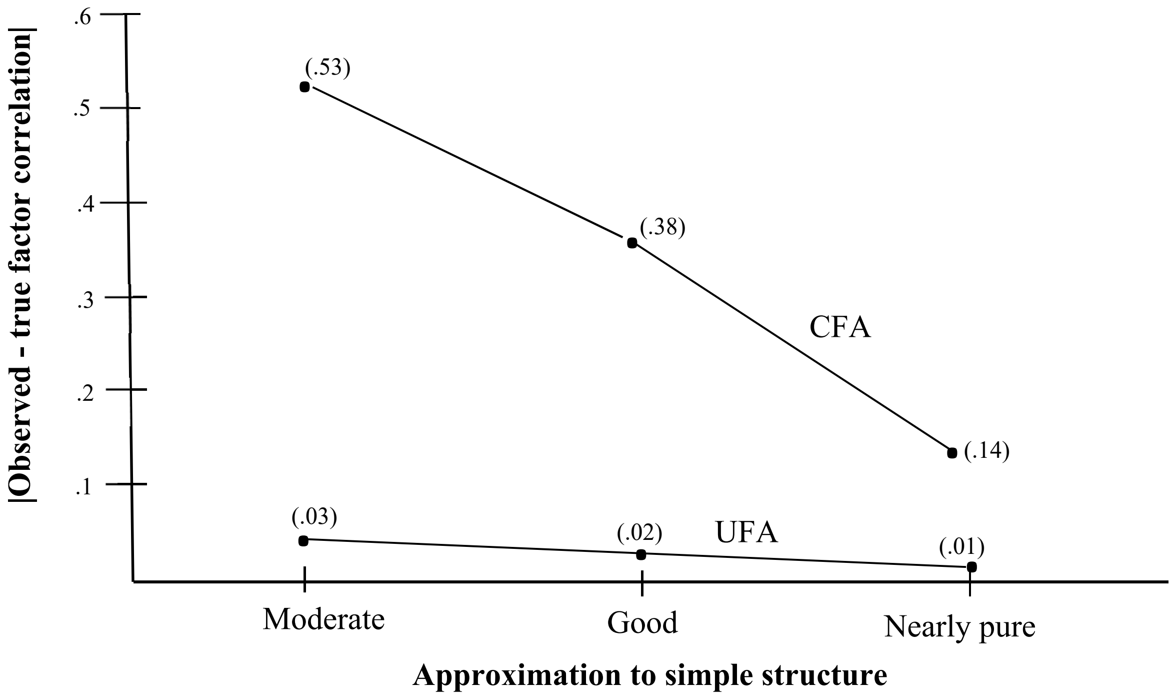
* Only to be used if scores are highly reliable.

Figure 2: Relationship between population RMSEA and parameter bias in CFA models when the data generating model involves cross-loadings for 25% of the items



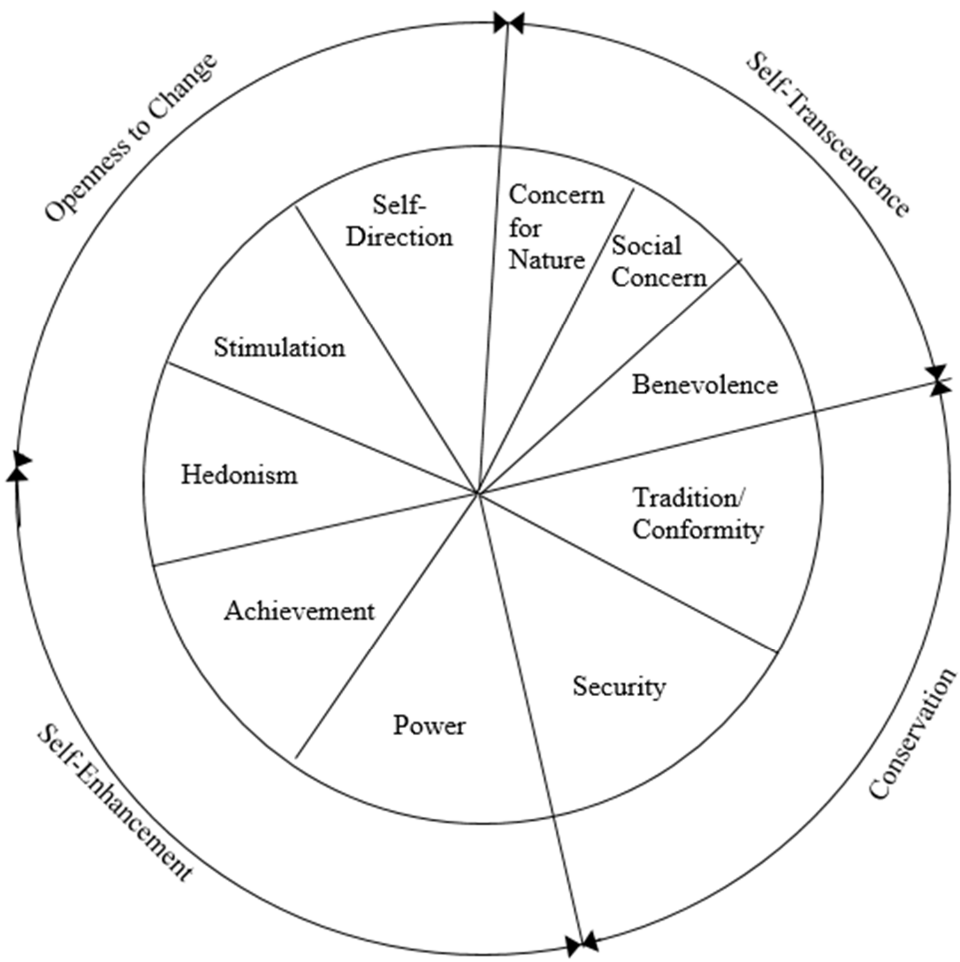
Note: Results across 216 conditions for a two factor model crossing factor correlation (-.5,-.3,-.1,.1, .3, .5) x magnitude of substantive loadings (.4, .6) x number of substantive items per factor (4, 8, 12) x minor loading (-.3, -.2, -.1, .1,.2,.3). Bias factor correlation = estimated – true factor correlation. Results are plotted using the two major drivers of the relationship between parameter bias and RMSEA.

Figure 3: Absolute bias in factor correlations for different degrees of approximation to simple structure



Note: Results of simulation study for a two-factor model with the following 108 empirical conditions: approximation to simple structure (nearly pure, good, moderate) x sample size (200, 600, 1,000) x true factor correlation (.1, .3, .5) x magnitude of substantive loadings (.4, .6) x number of substantive items per factor (4, 8). CFA = confirmatory factor analysis, UFA = unrestricted factor analysis

Figure 4: Schwartz's value types and structure



Value type	Definition	Value type	Definition
Power	Social status and prestige, control or dominance over people and resources.	Concern for nature	Understanding, appreciation, and protection of nature.
Achievement	Personal success through demonstrating competence according to social standards.	Social concern	Concern for and action to promote the welfare of people outside one's in-group.
Hedonism	Pleasure and sensuous gratification for oneself.	Benevolence	Preservation and enhancement of the welfare of close others with whom one is in frequent personal contact.
Stimulation	Excitement, novelty and challenge in life.	Tradition/conformity	Subordination of self in favor of socially imposed expectations.
Self-direction	Independent thought and action - choosing, creating, exploring.	Security	Safety, harmony, and stability of society, of relationships, and of self.

Table 1: Comparison of traditional EFA, CFA, and UFA

Features	Traditional EFA	CFA	UFA
Parameter estimates obtained by rotation	Yes	No	Yes
Independent-clusters model (constraints on factor loadings)	No	Yes	No
Test of exact fit under normality	Yes	Yes	Yes
Standard errors robust to non-normality	No	Yes	Yes
Test of close fit/goodness of fit indices	No	Yes	Yes
Use of a priori theory about factor structure	No	Yes	Yes
Correlated errors	No	Yes	Yes
Higher-order factor models	No	Yes	With EwSEM
Integration into structural equation model	No	Yes	Yes
Combination of CFA and UFA in structural equation model	No	No	Yes
Mediation analysis (Sobel test)	No	Yes	Yes
Mediation analysis (bootstrapping)	No	Yes	With EwSEM
Structural analysis with non-interval scaled dependent variables	No	Yes	With EwSEM
Interactions between latent variables	No	Yes	With EwSEM
MIMIC analysis	No	Yes	Yes
Multigroup invariance testing	No	Yes	Yes
Longitudinal (panel) models	No	Yes	Yes
Latent curve models	No	Yes	Yes

Note: EFA = exploratory factor analysis, CFA = confirmatory factor analysis, UFA = unrestricted factor analysis, EwSEM = ESEM within SEM, i.e., an ‘exploratory’ structural equation model (SEM) as a ‘standard’ SEM model.

Table 2: Bias in population factor correlation

Panel A: CFA							
	true ρ						cross-loadings
	-.5	-.3	-.1	.1	.3	.5	
Bias (estimated minus true ρ)	-.30	-.38	-.41	-.36	-.25	-.15	-.3
	-.19	-.23	-.24	-.21	-.17	-.12	-.2
	-.09	-.10	-.11	.10	-.09	-.07	-.1
	.00	.00	.00	.00	.00	.00	.0
	.07	.09	-.10	.11	.10	.09	.1
	.12	.17	.21	.24	.23	.19	.2
	.15	.25	.36	.41	.38	.30	.3
Panel B: UFA							
	true ρ						cross-loadings
	-.5	-.3	-.1	.1	.3	.5	
Bias (estimated minus true ρ)	-.01	-.01	-.02	-.03	-.03	-.03	-.3
	-.01	-.01	-.02	-.02	-.02	-.02	-.2
	-.01	-.01	-.01	-.02	-.02	-.02	-.1
	.00	.00	.00	.00	.00	.00	.0
	.02	.02	.02	.01	.01	.01	.1
	.02	.02	.02	.02	.01	.01	.2
	.02	.03	.03	.02	.01	.01	.3

Note: Results of simulation study for a two-factor model with the following 252 empirical conditions: true factor correlation (-.5, -.3, -.1, .1, .3, .5) x number of substantive (target) indicators per factor (4, 8, 12) x magnitude of the substantive factor loadings (.4, .6) x magnitude of the cross-loadings (-.3, -.2, -.1, 0, .1, .2, .3).

Table 3: RMSEA cutoff values corresponding to selected levels of factor correlation of bias |estimated – true correlation| as a function of factor loadings and number of indicators per factor

Factor loading	Number of indicators per factor		
	4	**8**	**12**
	|estimated - true factor correlation| = .1		
.4	.009	.008	.008
.6	.036	.026	.021
	|estimated - true factor correlation| = .2		
.4	.031	.025	.021
.6	.068	.048	.038
	|estimated - true factor correlation| = .3		
.4	.044	.034	.027
.6	.086	.061	.048

Note: These are population values obtained for a two-factor model across 252 conditions obtained by crossing: true factor correlation (-.5, -.3, -.1, .1, .3, .5) x number of substantive (target) indicators per factor (4, 8, 12) x magnitude of the substantive factor loadings (.4, .6) x magnitude of the cross-loadings (-.3, -.2, -.1, 0, .1, .2, .3).

Table 4: Modeling Schwartz Value Survey: Standardized factor loadings

Value types/Values	UFA											CFA	
	Power	Achiev	Hedon	Stim	Self-direct	Concern for nature	Social concern	Benev	Tradit/Conform	Secur	Method	SVS	Method
Power													
Social power	**.43**	.06	.03	-.03	-.09	.04	-.17	-.13	-.03	-.16	.33	**.52**	.38
Authority	**.42**	.18	-.02	.10	.02	.09	-.20	-.12	.23	-.01	.31	**.58**	.39
Wealth	**.51**	.12	.17	.02	.02	-.01	.02	-.11	-.21	.12	.31	**.55**	.39
Preserving my public image	**.49**	.11	.13	.07	-.10	.02	.01	-.10	.20	.05	.28	**.62**	.36
Achievement													
Successful	.23	**.54**	.05	.09	.07	-.10	.09	.17	-.11	.16	.39	**.45**	.53
Capable	-.06	**.34**	.08	-.10	*.41*	.03	.03	.14	.06	.02	.46	.14	.67
Ambitious	-.01	**.32**	.08	.13	.15	-.12	.03	.07	.25	.06	.41	**.29**	.57
Influential	.26	**.42**	-.11	.12	-.02	.14	-.01	-.08	.06	-.05	.33	**.60**	.42
Hedonism													
Pleasure	.17	-.11	**.54**	.23	.02	-.02	.03	-.09	-.04	-.03	.34	**.70**	.43
Enjoying life	.07	.10	**.54**	.10	.02	.15	-.02	.24	-.08	.06	.36	**.37**	.51
Stimulation													
An exciting life	.11	.02	.23	**.54**	.01	-.03	.08	-.05	-.02	.08	.35	**.54**	.45
Daring	-.05	.21	.14	**.53**	-.00	.08	-.09	-.10	.13	-.20	.30	**.59**	.40
A varied life	.03	-.03	-.01	**.41**	.23	.25	.06	.06	-.01	-.01	.37	**.50**	.50
Self-direction													
Creativity	.04	.06	-.17	*.41*	**.23**	.12	.10	.02	-.02	.03	.36	**.41**	.49
Freedom	-.13	-.02	.09	.03	**.21**	-.00	.17	-.10	-.06	.07	.48	.01	.59
Independent	.03	.15	.11	-.05	**.41**	.05	.08	.02	.05	.01	.43	.13	.60
Curious	-.04	.18	.02	.29	**.16**	.18	-.04	.23	-.14	-.12	.38	**.43**	.51
Choosing own goals	-.12	*.37*	.09	.02	**.39**	.07	-.01	.02	.06	.01	.43	**.23**	.62
Concern for nature													
A world of beauty	.11	-.10	-.01	.13	.07	**.53**	.12	.13	.03	.05	.34	**.64**	.46
Unity with nature	.06	-.14	.08	.03	.08	**.71**	.09	-.01	.07	.01	.31	**.72**	.42
Protecting the environment	-.05	.13	.04	-.04	-.08	**.70**	.19	-.05	.01	-.03	.31	**.70**	.42
Social concern													
Broad minded	-.16	.15	.04	.07	.14	.16	**.43**	.00	-.03	-.12	.37	**.40**	.52
Wisdom	.04	-.06	-.16	.13	*.32*	.11	.08	.21	.07	.08	.45	**.21**	.62
Social justice	.01	-.07	-.08	-.00	.00	.24	**.30**	.09	.12	.06	.40	**.49**	.52
Equality	-.08	-.03	.00	-.02	-.01	.02	**.52**	.07	-.01	-.02	.36	**.39**	.48
A world at peace	-.06	.06	.05	-.09	-.12	.20	**.46**	-.04	.03	.23	.37	**.46**	.49
Benevolence													
Helpful	-.09	.10	.04	.05	-.13	.11	.12	**.38**	.24	-.11	.40	**.55**	.54
Honest	-.18	.02	.09	-.08	.06	.04	.03	**.39**	.12	.04	.51	**.36**	.67
Forgiving	-.06	.08	-.11	.06	-.19	.04	.11	**.50**	.06	-.04	.40	**.50**	.51

Loyal	-.10	.02	.12	.00	.11	-.03	.05	**.24**	.26	.04	.51	**.28**	.68
Responsible	-.01	.08	.07	-.11	.17	.02	-.03	**.47**	.11	.10	.51	**.35**	.70
Tradition/Conformity													
Humble	-.04	.11	-.02	.11	-.02	-.06	.19	.08	**.41**	-.25	.36	**.32**	.48
Accepting my portion in life	.12	-.01	.16	-.08	.01	.09	.01	.10	**.43**	-.15	.29	**.36**	.39
Devout	.12	.04	-.24	-.07	-.24	.04	-.11	.27	**.38**	.09	.25	**.64**	.32
Respect for tradition	.03	.08	-.04	.07	.01	.15	-.03	-.13	**.51**	.21	.34	**.41**	.45
Moderate	.22	-.01	-.01	-.12	.17	-.03	.17	-.01	**.35**	-.26	.35	**.20**	.46
Politeness	-.07	-.11	.06	.16	.01	-.07	.13	.13	**.31**	.22	.46	**.29**	.61
Obedient	.06	.05	.08	-.03	-.09	-.01	-.04	.26	**.50**	.01	.37	**.57**	.50
Self-discipline	.06	-.06	-.18	.07	.22	.03	-.01	.10	**.40**	.17	.41	**.37**	.56
Honoring of parents and elders	-.23	.18	.01	-.00	-.07	.08	-.06	-.04	**.50**	.21	.43	**.46**	.57
Security													
Family security	-,18	.04	.00	-.03	.04	.05	.02	.03	.09	**.40**	.54	**.20**	.64
National security	.09	.09	.02	-.11	-.02	.02	.05	-.07	.14	**.42**	.38	**.35**	.49
Social order	.24	.01	-.08	.09	.00	-.08	.27	-.08	.04	.10	.30	**.22**	.39
Clean	.16	.24	.08	-.06	-.10	-.00	.13	.21	.21	**.19**	.34	**.39**	.47
Reciprocation of favors	.26	-.09	.04	.06	.15	-.01	.07	.02	.16	.05	.33	**.19**	.43

Note: Statistically significant target factor loadings ($p < .05$) are in bold; cross-loadings $\geq .30$ are in italics.

Table 5: Modeling Schwartz value types: Factor correlations

Value type	Power	Achievement	Hedonism	Stimula-tion	Self-direction	Concern for nature	Social concern	Benevol ence	Tradition/ Conformity	Security
Power		.06	.07	.24	-.01	-.01	-.04	-.15	.07	-.06
Achievement	.69		.33	.20	.16	.22	.08	.28	.36	.01
Hedonism	.56	.28		.11	.14	.04	.16	-.01	.05	-.00
Stimulation	.48	.57	.58		.25	.39	.21	.09	.04	-.04
Self-direction	.16	.61	.20	.97		.21	.21	.24	.12	.08
Concern for nature	.09	.24	.13	.53	.62		.43	.27	.31	.08
Social concern	-.08	.07	-.03	.25	.25	.78		.28	.25	.13
Benevolence	-.14	.25	-.25	-.05	.10	.30	.45		.52	.15
Tradition/Conformity	.21	.36	-.23	-.02	-.11	.27	.28	.77		.25
Security	.54	.43	-.04	-.02	-.34	.21	.37	.38	.80	

Note: CFA factor correlations are below the diagonal, UFA factor correlations are above the diagonal. We omitted 1's on the diagonal for readability.

Table 6: Modeling Schwartz value types: Standardized structural effects on materialism and satisfaction with life

Estimation model/ Predictors	Materialism			Satisfaction with life			Indirect effect through materialism		Type of mediation
	β	S.E.	r	β	S.E.	r	β	S.E.	
ESEM									
Power	.41*	.09	.47*	.16*	.06	.03	-.10*	.03	Competitive
Achievement	.17*	.05	.13*	.13	.07	.12*	-.04*	.02	Indirect only
Hedonism	.21*	.05	.26*	.01	.06	-.02	-.05*	.02	Indirect only
Stimulation	.02	.06	.11*	.02	.06	.03	<-.01	.01	No effect
Self-direction	-.09	.05	-.11*	-.01	.06	.05	.02	.01	No effect
Concern for nature	.02	.06	-.10*	-.05	.06	.01	-.01	.02	No effect
Social concern	-.11	.07	-.18*	-.05	.07	.01	.03	.02	No effect
Benevolence	-.08	.09	-.29*	.12	.06	.16*	.02	.02	No effect
Tradition/conformity	-.28*	.10	-.22*	-.06	.08	.14*	.07*	.03	Indirect only
Security	.09	.07	-.03	.15*	.06	.14*	-.02	.02	Direct only
Materialism				-.25*	.05	-.16*			
R^2	.34			.09					
SEM									
Power	.55	.48	.53*	-.32	.50	.01	-.02	.18	No effect
Achievement	.71	1.44	.35*	.45	.67	.10*	-.02	.20	No effect
Hedonism	.52	.75	.47*	.27	.36	-.10	-.01	.16	No effect
Stimulation	-.93	1.42	.27*	-.45	.74	.02	.02	.28	No effect
Self-direction	.18	.58	.08	.32	.34	-.06	-.01	.06	No effect
Concern for nature	-.45	1.52	-.03	-.55	.63	-.04	.01	.11	No effect
Social concern	.95	2.74	-.14	.66	1.10	-.04	-.03	.25	No effect
Benevolence	-1.09	3.13	-.23*	-.98	1.28	.08	.03	.28	No effect
Tradition/conformity	1.21	4.48	-.15*	1.31	1.96	.11*	-.03	.29	No effect
Security	-1.19	3.52	.16	-.57	1.47	.12	.03	.31	No effect
Materialism				-.03	.33	-.16*			
R^2	.30			Could not be reliably calculated					

* $p < .05$.

Note: β = standardized regression coefficient; S.E. = standard error; r = bivariate correlation; ESEM = Exploratory Structural Equation Model; SEM = Structural Equation Model. The standard error of the indirect effect is computed using robust standard errors (i.e., Sobel test accounting for non-normality).

Table 7: Cross-national analysis (U.S. vs. Norway) of Schwartz value types: Measurement model

Panel A: Model fit

Model	χ^2	df	RMSEA	SRMR	CFI
Configural invariance	2,304.9	1168	.034	.021	.960
Metric invariance	3,038.2	1518	.034	.032	.946
Scalar invariance	3,124.6	1552	.034	.031	.945

Panel B: Latent means and variances U.S.[1)]

Schwartz value type	Latent means	Latent variances
Power	-1.06*	1.74*
Achievement	.82*	.93
Hedonism	-.49	1.65
Stimulation	-.68*	.89
Self-direction	-.99†	.912
Concern for nature	-.70*	1.56*
Social concern	.23	.83
Benevolence	-1.43*	1.10
Tradition/conformity	.84*	.83
Security	-.44	.52*
Method bias	.83*	.32*

[1)] Versus Norway for which latent means are fixed at zero and latent variances at 1. * $p < .05$; † $p = .051$
Note: RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Squared Residual CFI = Comparative Fit Index.

Table 8: Cross-national analysis (U.S. vs. Norway) of Schwartz value types: MIMIC model

Criterion variables	**Predictors**			
	Gender (Female)		Age	
	U.S.	Norway	U.S.	Norway
Power	-.36*	-.10	<.01	.01
Achievement	.38*	.34[1)]	<.01	-.01
Hedonism	-.21	.14	<-.01	-.02*
Stimulation	-.31*	-.15	<-.01	-.01
Self-direction	-.10	-.19	<.01	.01*
Concern for nature	.25*	.22	.01*	.02*
Social concern	.73*	1.06*	.02*	.02[2)]
Benevolence	.18*	-.02	.04*	.02*
Tradition/Conformity	.43*	-.08	.01*	.02*
Security	.16	.21	.02*	.02*
Method factor	-.04	-.02	-.01	-.01

* $p < .05$. Note: Reported are unstandardized regression coefficients. [1)] $p = .079$; [2)] $p = .107$

Table 9: Modeling E-S-QUAL: Standardized factor loadings

Factor	Item wording	UFA						CFA
		Eff	SA	Ful	Priv	Val	Loy	
Efficiency (Eff)								
EFF1	This site makes it easy to find what I need.	**.80**	-.07	-.01	.07	.10	.02	**.86**
EFF2	It makes it easy to get anywhere on the site.	**.68**	.05	.02	.06	.12	.04	**.86**
EFF3	It enables me to complete a transaction quickly.	**.43**	.14	.16	.03	.06	.13	**.80**
EFF4	Information at this site is well organized.	**.90**	-.05	.02	.02	.05	-.05	**.86**
EFF5	It loads its pages fast.	**.26**	*.40*	.03	.07	.16	.07	**.76**
EFF6	This site is simple to use.	**.66**	.15	-.03	.08	-.01	.04	**.81**
EFF7	This site enables me to get on to it quickly.	**.39**	*.37*	.12	.05	.04	.06	**.81**
EFF8	This site is well organized.	**.80**	-.13	.15	.05	-.02	.01	**.82**
System availability (SA)								
SYS1	This site is always available for business.	.15	**.49**	.17	.06	.00	.03	**.76**
SYS2	This site launches and runs right away.	.24	**.50**	.01	.03	-.04	.17	**.74**
SYS3	This site does not crash.	.01	**.63**	.08	.14	.11	.01	**.80**
SYS4	Pages at this site do not freeze after I enter my order information.	.17	**.54**	.11	.06	.08	.04	**.82**
Fulfillment (Ful)								
FUL1	It delivers orders when promised.	.00	-.10	**.98**	-.04	-.01	.00	**.89**
FUL2	This site makes items available for delivery within a suitable time frame.	.06	.02	**.85**	-.05	.11	-.10	**.88**
FUL3	It quickly delivers what I order.	-.07	-.10	**.91**	.04	-.06	.04	**.82**
FUL4	It sends out the items ordered.	.07	.14	**.67**	-.06	.04	.01	**.79**
FUL5	It has in stock the items the company claims to have.	.11	.04	**.58**	-.00	.09	.07	**.79**
FUL6	It is truthful about its offerings.	.05	*.32*	**.37**	.08	.03	.12	**.72**
FUL7	It makes accurate promises about delivery of products.	-.07	-.06	**.90**	.11	-.05	.03	**.87**
Privacy (Priv)								
PRI1	It protects information about my Web-shopping behavior.	.10	-.06	-.04	**.80**	.01	-.01	**.82**
PRI2	It does not share my personal information with other sites.	-.08	-.02	-.02	**.97**	-.03	.04	**.87**
PRI3	This site protects information about my credit card.	.08	.08	.13	**.58**	.03	-.05	**.77**
Perceived value (Val)								
VAL1	The prices of the products and services available at this site (how economical the site is).	-.19	.01	.03	.04	**.91**	-.06	**.96**
VAL2	The overall convenience of using this site.	*.34*	-.01	-.04	-.04	**.51**	.15	**.88**
VAL3	The extent to which the site gives you a feeling of being in control.	*.33*	-.08	-.01	.05	**.49**	.11	**.85**
VAL4	The overall value you get from this site for your money and effort.	-.09	-.03	.07	-.02	**.96**	.01	**.84**
Loyalty intentions (Loy)								
How likely are you to …								
LOY1	Say positive things about this site to other people?	.02	-.06	.05	.00	-.04	**.94**	**.94**
LOY2	Recommend this site to someone who seeks your advice?	-.00	.02	.07	-.06	-.05	**.97**	**.95**
LOY3	Encourage friends and others to do business with this site?	.01	-.03	.00	-.01	-.03	**.91**	**.88**
LOY4	Consider this site to be your first choice for future transactions?	-.07	.01	-.04	.06	.16	**.69**	**.78**
LOY5	Do more business with this site in the coming months?	-.06	.03	-.04	.08	.13	**.64**	**.73**

Note: Statistically significant target loadings ($p < .05$) are in bold; cross-loadings ≥ .30 are in italics. Efficiency, system availability, fulfillment, and privacy items were scored on 5-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). Perceived value items were scored on a scale of 1 (*poor*) to10 (*excellent*). Loyalty intention items were scored on a 5-point scale (1 = *very unlikely*, 5 = *very likely*).

Table 10: Modeling E-S-QUAL: Factor correlations

E-S-QUAL dimension	Efficiency	System availability	Fulfillment	Privacy	Value	Loyalty
Efficiency		.56	.63	.58	.64	.63
System availability	.85		.48	.42	.40	.42
Fulfillment	.74	.72		.60	.58	.65
Privacy	.67	.64	.63		.53	.51
Value	.80	.67	.66	.59		.72
Loyalty	.69	.63	.69	.52	.77	

Note: CFA factor correlations are below the diagonal, UFA factor correlations are above the diagonal. We omitted 1's on the diagonal for readability.

Table 11: Modeling E-S-QUAL: Standardized structural effects on perceived value and loyalty intentions

E-S-QUAL dimensions/ fit	EwSEM				SEM				Factor scores			
	Value		Loyalty		Value		Loyalty		Value		Loyalty	
	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.
Efficiency	.41*	.08	.35*	.07	.75**	.09	.40**	.09	.45*	.05	.34*	.05
System availability	-.06	.07	-.05	.07	-.11	.09	.01	.08	-.04	.05	-.01	.05
Fulfillment	.25*	.08	.39*	.07	.16**	.06	.38**	.057	.24*	.05	.41*	.05
Privacy	.17*	.05	.08	.06	.05	.04	<.01	.053	.16*	.04	.08	.04
R^2	.47*		.511*		.67*		.55*		.53*		.54*	
χ^2 (df)	754.0 (294)				1400.5 (419)							
RMSEA	.051				.063							
SRMR	.019				.051							
CFI	.959				.912							

* p < .05.

Note: β = standardized regression coefficient; S.E. = standard error; EwSEM = ESEM within SEM; SEM = Structural Equation Model. RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Squared Residual CFI = Comparative Fit Index.

Table 12: Analyses of some other constructs used in marketing

Constructs	Example of use in marketing	Sample[1)]	# fact-ors	CFA			UFA			Max. Δr	Additional analyses
				RMSEA	SRMR	CFI	RMSEA	SRMR	CFI		
Exploratory consumer behavior tendencies (EBBT)	Baumgartner and Steenkamp (1996)	Nether-lands (N=3248)	2	.052	.042	.922	.054	.038	.926	.01	No cross-loadings > |.2|. Positive and negative cross-loadings were about equal in magnitude and occurrence canceling each other out. ESEM and SEM gave the same sign. effects of the EBBT factors on market mavenism.
Horizontal/ vertical individualism / collectivism	De Jong et al. (2015)	Nether-lands (N=1408)	4	.060	.049	.872	.062	.030	.913	.08	Only two cross-loading > |.10|; they were VC items loading on HC and had the opposite sign, canceling each other out. ESEM and SEM gave the same sign. effects of the H/V I/C dimensions on risk taking.
Material values: Success, acquisition, happiness	Richins (2004)	Great Britain (N=1534)	3	.097	.055	.920	.000	.004	1.000	.36	All cross-loadings > |.2| were aligned with direction of factor correlation, inflating the correlation between the MVS factors in CFA. In ESEM, the MVS factor “success” had a negative effect and the MVS factors “acquisition” and “happiness” had a positive effect on price consciousness. In SEM, the effect of acquisition was not significant and structural effects had standard errors 4-5x larger, indicating lack of estimation precision.
Consumer susceptibility to interpersonal influence	Bearden et al. (1989)	Nether-lands (N=3248)	2	.079	.047	.894	.075	.033	.924	.19	All cross-loadings > |.2| were aligned with direction of factor correlation, inflating the correlation between the two CSII factors in CFA. SEM and ESEM both gave sign. negative (positive) effect of SII (SNI) on market mavenism but SEM gave much higher standard errors.
Big Five	Steenkamp et al. (2010a)	Germany (N=1550)	5	.082	.082	.788	.070	.033	.904	.39	The direction of 88% of the cross-loadings > |.2| was aligned with direction of factor correlations, inflating correlations in CFA. ESEM, but not SEM, gave sign. negative effect of openness to new experience on brand loyalty.
Brand relevance in category (BRiC)	Fischer et al. (2010)	India (N=1503)	3	.047	.034	.975	.023	.010	.996	.18	All cross-loadings > |.2| were aligned with direction of factor correlation, inflating the correlation between the BRiC factors in CFA.
Positive and negative affect	Burroughs and Rindfleisch	Switzer-land (N=393)	2	.064	.061	.851	.067	.053	.852	.17	The direction of 80% of the cross-loadings > |.2| was aligned with direction of the factor correlation, inflating the correlation between PA

	(2002)										and NA in CFA. ESEM showed expected effects of PA (+) and NA (-) on life satisfaction; in SEM, effect of NA was not significant.
Regulatory focus	Haws et al. (2010)	Spain (N=1550)	2	.076	.070	.849	.069	.043	.903	.00	No cross-loadings > |.2|; positive and negative cross-loadings were about equal in magnitude and occurrence canceling each other out.
Store image attributes: price, quality, assortment, service, atmosphere	Steenkamp and Wedel (1991)	Nether-lands N=871)	5	.067	.067	.847	.054	.028	.931	.28	90% of the cross-loadings > |.2| were positive, inflating the positive correlations between the store image attributes in CFA. ESEM, but not SEM, showed a sign. effect of service on overall store image.
Marketing mix: Brand advertising, price promotion, functional positioning, emotional positioning	Steenkamp et al. (2010b)	China (N=2994)	4	.042	.016	.989	.031	.004	.998	.10	The direction of all cross-loadings > |.2| was aligned with direction of factor correlations, inflating correlations in CFA ESEM shows sign. positive effects of functional and emotional positioning, and brand advertising on brand loyalty; SEM only for functional positioning.

[1)] All samples are national samples of consumers collected by professional market research agencies. The second column provides a marketing reference and is not necessarily imply that the scale was developed in that article.

Table 13: Summary of recommendations

Topic	Recommendations
Measurement analysis	• Always to be performed, regardless of whether the scale(s) has (have) been shown to be unidimensional in extant research. • Always provide information on χ^2 (df), RMSEA, SRMR, and CFI. • Assess model fit against common benchmarks: Models with RMSEA ≤ .05, SRMR ≤ .05, and CFI ≥ .95 generally indicate close fit to the data; RMSEA ≤ .08, SRMR ≤ .08, and CFI ≥ .90 generally indicate reasonable fit.
Estimation	• Use maximum likelihood with standard errors and goodness of fit tests robust to non-normality (MLR) as the default estimation method. • For MLR, use a modified chi-square difference test to test differences in model fit.
Measurement model	• Use EFA if you have no firm a priori theory about the number of factors and target loadings. • Analyze data with UFA *and* CFA if you have a priori theory. • Choose between UFA and CFA using model fit, difference in model fit between UFA and CFA, examination of the factor structure, comparison of the factor correlations between UFA and CFA, and (if applicable) comparison of the structural parameter estimates and their standard errors.
Theory testing	• Test your theory and hypotheses within a latent variables structural equation framework. • Use SEM if CFA is supported; use ESEM when set-UFA is supported; use EwSEM when full-UFA is supported.
Factor scores	• Use of factor scores rather than latent variables is statistically inferior except when the constructs achieve very high reliability. • Sum/mean scores may only be used if CFA is supported. • Use factor scores that account for cross-loadings such regression factor scores if UFA is supported.
Implementation	• Currently, only Mplus and R allow for UFA (and many other advances in psychometrics). We recommend that the substantive researcher acquaint themselves with either of these programs.