Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 29 Number 15, November 2024 ISSN 1531-7714

The Circle of Methods for Evaluating Latent Variable Measurement Models: EFA, CFA, and ESEM

Kelvin T. Afolabi, *University of Virginia*
Timothy R. Konold, *University of Virginia*

Exploratory structural equation (ESEM) has received increased attention in the methodological literature as a promising tool for evaluating latent variable measurement models. It overcomes many of the limitations attached to exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), while capitalizing on the benefits of each. Given that the recent introduction of ESEM in statistical software has made its use more accessible to applied researchers, we describe the differences between these three approaches to evaluating measurement models and provide an illustration of their differences through an applied example. Syntax for running these models is also provided that can be modified for application by others.

Keywords: factor analysis, CFA, ESEM, structural equation modeling, school climate

Introduction

Unlike variables that are directly observable and measurable with physical instruments (e.g., time, height), many substantive questions in the social sciences involve the use of latent variables that are not directly observable (e.g., depression, self-esteem). In the case of latent variables, quantification of their presence is inferred through the collection of measurable characteristics (e.g., items on a survey or questions on an assessment) that are presumed to be indicators of them. Otherwise stated, the directly unobservable latent construct is presumed to influence those characteristics which can be observed, and the directly observable indicators are presumed to be influenced by the underlying latent variables that are not directly observable. For example, the construct of school climate is not an objectively measurable variable but might be measured through the use of survey items administered to students that ask about different experiences thought to be influenced by healthy school climates (Konold, et al., 2018).

Researchers seeking to measure latent variables begin by collecting observed measures (e.g., items) theoretically believed to be indicators of one or more latent variables, and these observed indicators are then evaluated within a latent variable measurement model. Historically, exploratory factor analysis (EFA; Spearman, 1904) has been the procedure of choice for understanding and quantifying the number and nature of factors alleged to exist within a set of indicators. While EFA remains a useful tool, it has largely given way to the use and popularity of confirmatory factor analysis (CFA; Joreskog, 1969) which is often described as constituting a more theory driven approach to model evaluation. However, in more recent years, CFA has been found to be too restrictive in evaluations of many latent variable measurement models, and attention has turned to exploratory structural equation models (ESEM; Asparouhov & Muthen 2009; Marsh et al. 2009). ESEM can be characterized as a hybrid of EFA and CFA that draws from the advantages of both. In comparison to CFA, ESEM is less likely to result in biased structural relationships among factors and is less likely to result in the rejection of trivially mis-specified models (e.g., Marsh et al., 2014). This has been demonstrated across a number of controlled simulation studies (Asparouhov & Muthén, 2009; Konold & Sanders, 2024; Steenkamp & Maydeu-Olivares, 2023); as well as in numerous substantive applications related to investigations of mental health (van Zyl & Klooster, 2022), student evaluations of university teaching (Marsh et al., 2009), bullying and victimization (2011); the Big Five personality traits (Marsh et al., 2010), and self-concept (Marsh et al., 2020).

Given ESEMs relatively recent re-introduction into the literature, that coincided with its recent availability to users in statistical computing software (Asparouhov & Muthén, 2009), the current paper describes ESEM along with considerations involved in conducting an ESEM for purposes of evaluating a latent variable measurement model. We supplement this with an applied example that illustrates its usefulness and provide example code that can be adapted for application. Because ESEM can be considered a hybrid of EFA and CFA, we briefly review these procedures and in doing so point to reasons for the evolution that has been taking place.

Latent Variable Measurement Models

As noted above, the quantification of some variables does not require a measurement model because they are directly observable and obtainable through a single measurement (e.g., the number of students in a school, the number of doctor visits in a given year). By contrast, latent variables are not directly observable and require the use of things we can observe (e.g., responses to questions) in order to draw inferences about their presence and magnitude. For example, the constructs of adult-student relationships and inclusive instruction shown in Table 1 might be measured by student (or teacher) response to the items that are listed under each in order to operationalize them. Various latent variable measurement models are graphically illustrated in Figure 1 that could be used to evaluate the psychometric characteristics of these measurement models. Common to each is a set of observed indicators (in boxes) and some number of latent variables (or factors, depicted in ovals) that might be correlated with one another (double headed curve arrows). The factors are presumed to influence the observed indicators (single headed arrows) and explain the common variance among the indicators that are linked to them. The unexplained portions of the observed variables are depicted by smaller single headed arrows below the boxes. In estimating these characteristics of the model through one of the methods described below, focus is often on the strength of the indicator-factor relationships (factor loadings), the proportion of observed variable variance that is unexplained by its factor(s), and how strongly the factors are associated (i.e., structural relationships).

Exploratory Factor Analysis (EFA)

EFA can be considered a dimension reduction procedure that seeks to identify latent dimensions (or factors) within a set of indicators based on their shared variance (Bartholomew et al., 2011). A key feature of EFA is that solutions result in as many factors as there are indicators entering into the analysis and all indicators are allowed to be associated with (i.e., load on) all factors, as illustrated in Panel A of Figure 1. Researchers attempting to evaluate the quality of their measurement models within an EFA framework simultaneously attempt to identify the number of factors that are present in the data along with which indicators are the best measures of them.

Spearman's (1904) seminal work in EFA spawned a proliferation of methodological advances for evaluating the number and nature of underlying factors believed to explain the covariance structure of a set of indicators. As a result, applications of EFA typically involve a number of considerations that could include different rotations (see below), use of different estimators, evaluations of different heuristics in settling on the number of factors to retain (e.g., parallel analysis, Horn, 1965; eigenvalues greater than one, Kaiser, 1960), and consideration of the resulting solutions with respect to theories of the measurement structure; all of which are further evaluated within the context of obtaining simple structure (Thurstone, 1947). Where, simple structure reflects that each factor has several target indicators with high loadings, that each indicator is strongly influenced by only one factor, and that non-target indicators have low (though not necessarily zero) loadings on non-target factors. See, for example, Panel B of Figure 1. Here, indicators are typically ignored when factor loadings are at or below some threshold value (e.g., $\leq .30$; Tabachnick & Fidell, 2007) or above some threshold to be considered salient (e.g., $> .40$; Steenkamp & Maydeu Olivares, 2021).

Factor extraction for continuous data is typically performed using maximum likelihood (ML) estimation that rests on the assumption of multivariate normality, and provides goodness-of-fit statistics for evaluating model quality (Brown, 2015; Schmitt, 2011).

EFA allows for cross-loadings, where variables can load onto multiple factors (Flora, 2017), and does not require pre-specification of common factors. The number of factor loadings estimated in an EFA framework can reduce clarity of the solution. As a result, rotations are used to help simplify the factor structure and increase interpretability (see for example, Flora, 2017; Osborne, 2015). While EFA is useful for exploring dimensionality, it has been criticized as being based on the subjective application of extraction and rotation criteria (Gorsuch, 1983; Grice, 2001). Different rotations will give rise to an infinite number of solutions with respect to the model estimated parameters (e.g., factor loadings and factor correlations), all of which will result in the same model implied covariance matrix when applied to the same data. Because factor scores are not uniquely defined, the resulting indeterminacy limits the use of EFA in calculating and interpreting factor scores, evaluating measurement invariance across groups, and modeling method effects (Gorsuch, 1983; Grice, 2001; Morin et al, 2016; van Zyl & Ten Klooster, 2022).

Confirmatory Factor Analysis (CFA)

The development of confirmatory factor analysis (CFA; Joreskog, 1969) and structural equation modeling (SEM; Joreskog; 1978) paved the way for addressing many of the limitations inherent to EFA. In contrast to EFA, applications of CFA require users to specify in advance the number of factors they believe to exist within a set of variables along with which observed variables are believed to be associated with each factor. The independent cluster model (ICM) confirmatory factor model has been considered the gold standard for evaluating latent variable measurement models (Marsh et al., 2014). It assumes that each observed indicator should only be related to a single latent factor, and that these indicators are unrelated to other factors in the model (i.e., simple structure). This is illustrated in Panel B of Figure 1. Here, the researcher posits that the covariances among the observed indicators are best explained by two (correlated) latent factors. Each factor has three target indicator variables that are believed to be uniquely associated with their respective target factors, and associations (i.e., factor loadings) between non-target indicators and non-target factors are fixed to zero.

In contrast to EFA models, CFA models result in a single solution when conditioned on a particular estimator (e.g., maximum likelihood) and overcome the aforementioned limitations associated with EFA. Notably they allow for tests of more restricted models that better align with theory, provide for more stable factor scores, allow for tests of measurement invariance, provide formal tests for different factor solutions, and provide a better mechanism to evaluate method effects (Brown, 2015). Moreover, constraining some parameters to be zero (or imposing equality constraints) results in a restricted model implied variance-covariance matrix that can be used to obtain a formal test of model fit when compared to the observed sample variance-covariance matrix. For example, the likelihood ratio (LR; also known as the χ^2 test) is based on the ML fit function (Fml) that reflects the similarity of the model implied covariance matrix to that of the observed sample covariance matrix, weighted by N–1 (Bollen, 1989): $LR = (N-1)Fml$. However, given the tendency of this test statistics to reject what are often believed to be trivially mis-specified models, researchers are more likely to rely on more approximate measures of fit (McDonald & Ho, 2002) such as the Comparative Fit Index (CFI: Bentler, 1990) or the Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980) when evaluating model quality.

Despite the many advantages of CFA, it has recently been shown that these models may in fact be too overly restrictive, and that the idea of unidimensional indicators may be more of a "convenient fiction" than reality (Marsh, et al., 2013, p. 258). Notably, recent research has demonstrated that constraining even small cross-loadings to zero in a CFA framework results in biased structural relationships (e.g., factor correlations used for evaluating convergent and discriminant validity), and the rejection of trivially mis-specified models (Asparouhov & Muthén, 2009; Konold & Sanders, 2024; Marsh et al., 2014; Steenkamp & Maydeu-Olivares, 2023)

Exploratory Structural Equation Modeling (ESEM)

Exploratory structural equation modeling (ESEM)[1] provides an integration of the best features of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA)/structural equation modeling (SEM; Marsh et al., 2009, 2010; Steenkamp & Maydeu-Olivares, 2023), and its recent introduction into software (e.g., *Mplus*; Asparouhov & Muthén, 2009) has made it more accessible to users. All three methods evaluate the same measurement model but differ with respect to assumptions that are made about how the indicators are related to the factors. Whereas EFA estimates all indicator-factor relationships and CFA only estimates these relationships for each factor's target indicators, ESEM provides for both. ESEM allows researchers to overcome the limitations of traditional CFAs, which are often too restrictive (i.e., reject models that may be trivially mis-specified), and result in inflated factor correlations that undermine discriminant validity (Asparouhov & Muthén, 2009; Marsh et al., 2010; Steenkamp & Maydeu-Olivares, 2023). By contrast, ESEM estimates factors with both target loadings (TLs) and cross-loadings (CLs)[2] while also incorporating key features of CFA/SEM such as tests of model fit, tests of predictive relations, measurement invariance (MI), and complex structural models (Marsh et al., 2009, 2013).

Panel C of Figure 1 illustrates a hypothetical two-factor ESEM model. Where, each factor is defined by three target indicators (solid arrows) and non-target indicators are permitted to cross-load on their non-target factors (dashed arrows). In estimating this model, the magnitude of cross-loadings is typically constrained to be below some pre-specified threshold (e.g., less than 0.5; Marsh et al., 2009). ESEM's accommodation of CLs in multidimensional factor models has been shown to reduce bias in factor correlations, that are often used to gauge convergent/discriminant validity, when compared to results that would be obtained by constraining these relationships to zero through a CFA investigation. The direction of this bias (positive or negative) depends on both the sign of the omitted CL in relation to the sign and magnitude of the true correlation among the constructs (De Luca et al., 2021; Konold & Sanders, 2024; Steenkamp & Maydeu-Olivares, 2023). For example, when CLs are positive and the true correlation is positive, constraining the CLs to zero in a CFA framework results in an upwardly biased factor association. Conversely, when CLs are negative and the true correlation is positive, the CFA correlation has been found to be downwardly biased.

Factor Rotations

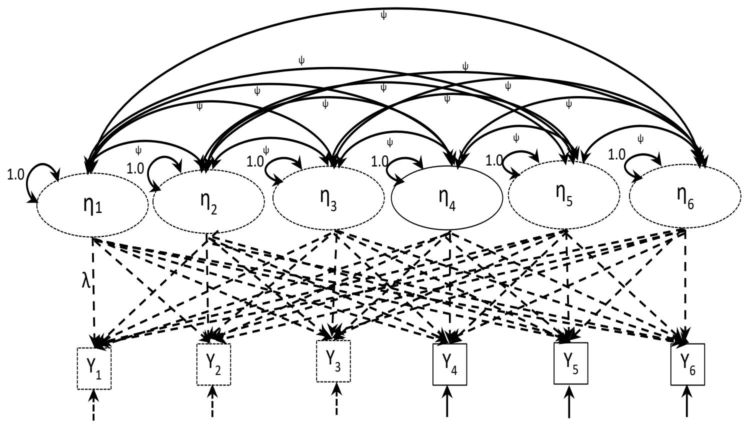
Similar to a kaleidoscope creating ever-changing patterns, rotations present researchers with an array of possible structures underlying their data. Rotations are motivated by a quest for simple structure, shifting alignments to clarify associations. Similar to EFA, the interpretability of ESEM solutions can be facilitated through the use of different factor rotations. Rotations refer to the movements of factor axes in multidimensional space to improve their nearness to the observed indicators. Rotations can be specified to maintain right angles (i.e., 90°) of the axes (i.e., orthogonal or uncorrelated factors) or not (i.e., non-orthogonal, oblique, or correlated factors). Geometrically, factor loadings can be conceptualized as the distance of an indicator to a factor line. Consequently, a goal of rotation is often to place these factor lines as close to the center of a cluster of indicators as possible (under constraints of the approach) such that the factor loading for a cluster of indicators on a given factor become larger and the factor loadings for that same cluster of indicators on a different factor become smaller. Different rotations for a given number of factors will result in explaining the same amount of total shared variance among a set of indicators, but typically results in differences in the distribution of this variance to the factors in the model. See Osborne (2015) for a good gentle introduction to rotation. Although there are many types of rotations, Geomin and Target rotations are the most frequently used approaches in ESEM frameworks (Asparouhov & Muthén, 2009; Marsh et al., 2014; Steenkamp & Maydeu-Olivares, 2023). However, an advantage of

---

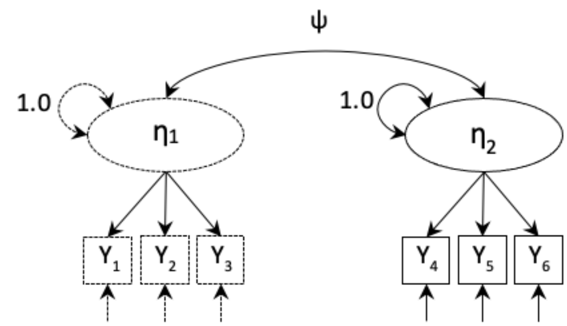
[1] ESEM is sometimes referred to as unrestricted factor analysis (UFA; Steenkamp & Maydeu-Olivares, 2023) when applied to latent variable measurement models (vs. full structural equation models).
[2] It also remains possible to constrain some CLs to zero.

Figure 1.

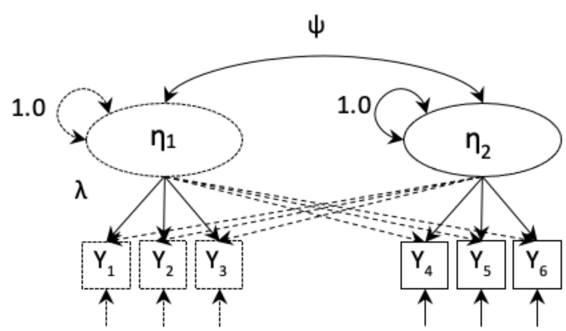
Panel A: EFA Model



Panel B: ICM-CFA with Target Indicators (no CLs)



Panel C: ESEM Model with Target and Non-Target Indicators (CLs)



target rotation is that target indicators (e.g., items) can be specified to load onto their corresponding factors and non-target indicators can be constrained to be as close to zero as possible (Asparouhov & Muthén, 2009). Salient characteristics of these forms of rotations are described below, along with an orthogonal rotation that might be useful in some unique instances (e.g., bi-factor models).

Geomin is an oblique rotation that allows factors to correlate and is the default in Mplus (Asparouhov & Muthén, 2009). It has been found to work well in a variety of circumstances (Marsh et al., 2014) and provides a good balance between achieving simple structure and allowing for correlations among factors (Browne, 2001; Costello & Osborne, 2005; Fabrigar & Wegener, 2011). It is generally considered a good choice when factors are expected to be distinct, but associated (Marsh et al., 2014; Asparouhov & Muthén, 2009). It is considered a bit more exploratory than target rotation (Marsh et al., 2014) and may be more useful for simpler models with fewer factors or in the development of new instruments where factor structure is less clear (Howard et al., 2017).

Geomin permits regulation of cross-loading values through utilization of an epsilon (ε) parameter to restrict the threshold for loadings targeted for rotation (Browne, 2001). Factor loadings with an absolute value less than epsilon are estimated without restriction, while loadings greater than epsilon are adjusted downward through rotation. Larger epsilon values allow for larger cross-loading values (Asparouhov & Muthén, 2009). In ESEM applications, epsilon values are typically within the 0.01 to 0.50 range, with 0.50 being suitable for significant but not excessive cross-loadings in most cases (Marsh et al., 2014). Adopting smaller epsilon values restricts more loadings to zero, enhancing simplicity and interpretability. Conversely, the use of larger epsilon values may enhance model fit (Marsh et al., 2014) by decreasing model restrictions.

Target rotation (Asparouhov & Muthén, 2009) was developed specifically for use in ESEM frameworks. As the name implies, target indicators are pre-specified to load on their respective factors in order to guide rotation toward a simplified solution (Marsh et al., 2014). Target rotation has been described as being "superior to traditional rotations in its ability to identify the proper solution" in ESEM models (Asparouhov & Muthén, 2009; p. 431) because it rotates directly toward a hypothesized pattern. As such, it requires a stronger a priori theory about the expected factor structure than is required for goemin rotations, making it more suitable for more confirmatory evaluations (Osborne, 2015). Although sensitivity to the specified target indicators can be a limitation (Morin & Asparouhov, 2018), appropriate attention to these specifications from prior research or theory allows target rotation to maximize the likelihood of recovering the hypothesized factor structure (Fabrigar & Wegener, 2011). Target rotation is also recommended for more complex models to improve model conformability while retaining flexibility (Marsh et al., 2014). It also allows users to explicitly set certain cross-loadings to zero if desired in order to test subtle distinctions between factors, without restrictive independent cluster model constraints (Morin et al., 2013).

Orthogonal rotation produces factors that are uncorrelated with each other (Osborne, 2015), and are said to uphold Thurstone's (1947) original principle of simple structure which sought to simplify factors while retaining independence (Fabrigar & Wegener, 2011). Within ESEM frameworks, the most common orthogonal approach is varimax rotation. The goal of varimax rotation is to maximize the variance of squared loadings on each factor, resulting in some very high and very low loadings that enhance interpretability (Costello & Osborne, 2005). However, Browne (2001) cautions that varimax solutions often involve splitting factors that contain heterogeneous clusters of variables. Moreover, Osborne (2015) notes that forcing orthogonality when factors are correlated can distort the underlying factor structure. As a result, Fabrigar and Wegener (2011) suggest varimax may be best viewed as a starting point before moving on to other rotation methods if the varimax-rotated solution is not satisfactory. Within ESEM frameworks, orthogonal rotations are typically needed in the analysis of bifactor models (Morin et al., 2016), but have also been found to be helpful in the analysis of multidimensional psychiatric scales (Prokofieva et al., 2023) and in psychometric evaluations of positive psychology measures (van Zyl & ten Klooster, 2022).

A Pedagogical Example

We demonstrate the usefulness of ESEM, as an alternative to CFA, for evaluating the psychometric characteristics of multi-factor measurement models in instances in which 1) small cross-loadings can give rise to biased structural coefficients (e.g., factor correlations) when they are constrained to zero in a traditional CFA, and 2) that such constraints can result in poorer fitting models that are only trivially mis-specified. In doing so, we

These ideas are illustrated through applications of CFA and ESEM to a substantive question pertaining to the psychometric characteristics of items related to school climate. Healthy and supportive school climates are important priorities at national and state levels (Darling-Hammond & DePaoli, 2020). We draw from items located on the Authoritative School Climate Survey (ASCS; Cornell et al., 2016) that characterizes positive school climates as those that hold high disciplinary expectations for their students and have supportive adult-student relationships. We examine 8 items that might be considered indicators of (and influenced by) two dimensions of school climate: Adult-Student Relationships and Inclusive Instruction[3], as shown in Table 1. In other words, we examine the measurement characteristics of 8 items designed to measure two distinct (but related) factors. This hypothesis allows us to specify a CFA model with four target indicators for each factor, and cross-loadings set to zero; and to also specify salient items for each factor in an ESEM target rotation.

Data Source

Data were obtained from the Virginia Center for School and Campus Safety's 2023 survey of climate and working conditions. The cohort of respondents consisted of 49,350 classroom instructors from 430 schools in the state, which equated to a response rate of 65.6% of all teachers. The composition of teacher's gender identity included males (10.6%), females (83.2%), and non-binary (0.3%); with 5.8% opting to not disclose their gender identity. Participants also identified as American Indian or Alaskan Native, (0.7%), Asian (2.6%), Black or African American (11.8%), Native Hawaiian or Pacific Islander (0.2%), White (82.1%), and other racial categories (2.5%).

Methods

Because school climate is a school level construct and informants within schools provide valuable indicators of these factors (Konold & Sanders, 2020; Stapleton et al., 2016), individual reports from teachers within schools were aggregated to the school level for analysis. Consequently, our examination of measurement structure of the 8 items was conducted on a sample size of 430 (schools).

To illustrate differences among some of the methods described above, data were analyzed through use of CFA and four different ESEM specifications. These included an orthogonal rotation (forcing the factor correlations to zero), two Geomin rotations (with epsilon values set at 0.0001 and 0.5 to accommodate differences in cross loading magnitudes), and a target rotation. Evaluation of model quality typically includes consideration of several quantitative indicators of fit. For this purpose, we consider the LRT (i.e., χ^2 statistic) but caution that it is well-known to reject reasonably specified models (Cheung & Rensvold, 2002; Schumacker & Lomax, 2010). We place greater emphasis on the comparative fit index (CFI) and Tucker-Lewis index (TLI) where values $\geq$ .95 are often associated with good fit (Hu & Bentler, 1999); the RMSEA where values < .05 and < .08 are often associated with good and mediocre fit, respectively, (MacCallum et al., 1996); and the AIC and BIC for model comparisons. Although strict adherence to these thresholds should be avoided because they were derived from simulations on a finite set of conditions that may not apply to the specific conditions present in a particular model, they are often a useful starting point in model evaluations. All models were estimated with Robust Maximum Likelihood Estimation (MLR) in *Mplus* 8.10 on a macOS operating system. Mplus code for our different model specifications is provided in the Appendix. Although Mplus currently provides the most robust platform in terms of functionality and efficiency for estimating ESEM models, the open-source packages of R (`lavann` and `psych`) and JASP also provide some functionality for ESEM.

Results and Discussion

Model fit estimates for the various models are shown on the bottom of Table 2. As described above, all ESEM rotations resulted in the exact same estimates of fit regardless of rotation type. Consequently, they are of little help in differentiating between the different ESEM solutions when the same number of factors are specified. Neither the CFA, $\chi^2(19) = 90.33$, $p < .05$; or the ESEM models, $\chi^2(13) = 44.75$, $p < .05$, were found to provide a good approximation to the observed unstructured variance-covariance matrix as gauged by the LRT. Other measures of model fit for the CFA were somewhat mixed in that the CFI of .95 was at a

[3] These domain names are constructions of the authors of this paper for illustrative purposes. They should not be taken to reflect the intended uses of these items as constructed by the authors of the Virginia Center for School and Campus Safety survey.

threshold value considered good, the TLI was only slightly below this at .93, and the RMSEA confidence interval (CI) contained .08 (90% CI: .075, .113). Consequently, a case could be made that the CFA solution provides a reasonable approximation to the observed data, when non-zero CLs are constrained to zero. Further evidence of a reasonable fit for this model could be obtained from inspection of the factor loadings that reveal strong associations between each indicator and their target factors. In such circumstances, it is likely that many researchers would conclude with this model and not consider alternative specifications that allow for the estimation of CLs[4].

At the same time, relaxing the assumption that CLs be fixed to zero and allowing them to be estimated to varying degrees (through the use of different rotations) within an ESEM model resulted in improved, and more uniform fit across the CFI (= .98), TLI (= .96), and RMSEA (= .08) values. Moreover, the lower AIC (ΔAIC = 80.89) and BIC (ΔBIC = 56.50) values associated with the ESEM models (vs. the CFA model) favored the more saturated ESEM models. As can be seen in Table 2, model fit values are the same across different rotations and are of little help in adjudicating among the different solutions. However, evaluation of the factor loadings with the perspective of their behavior under controlled and known conditions (i.e., prior simulation work) can help elucidate what we are seeing.

Orthogonal rotations fix factor correlations to zero with the goal of obtaining distinct variable separation across factors (Reise et al., 2010) and they can be particularly useful in bifactor models. In our school climate analysis, however, we do not consider a bifactor model and different dimensions of schools of school climate are often expected to be associated with one another (Wang & Degol, 2016). In the current illustration, two things are evident that should alert researchers to the fact that factor correlations are not likely to be zero in the data. First, setting the factor correlation to zero (when it is not) forces the misspecification to other locations in the model. Here, when the two sets of items cannot express their association through their factor correlations, the result is CL values that disguise a more simple structure that may be present in the data, where non-target items produce more moderate factor loadings on their non-target factors. Second, a comparison of the orthogonal solution with the factor correlation produced from all the other non-orthogonal rotations shown in Table 2 (that range from .61 to .81) suggests that the factor correlation is a non-zero value.

Geomin and target rotations both allow for CLs to be estimated (unlike most applications of CFA), factor correlations to be non-zero (unlike orthogonal rotations), and both attempts to clarify the factor structure through reducing variable complexity by yielding larger TLs and smaller CL values. As can be seen in Table 2, the Geomin solutions resulted in larger loadings for the target indicators and smaller loadings for the non-target indicators, when compared to the orthogonal rotation. A key benefit of Geomin rotation is more direct control over variable complexity through different specifications of the epsilon (ε) parameter. Smaller values of ε constrain non-target indicator loadings to be smaller and higher values enables them to be more freely estimated across factors[5]. This is illustrated in Table 2 between ε values of 0.0001 and 0.5. Restricting the non-target indicators to have smaller loadings on their non-target factors (ε = 0.0001) results in a larger factor correlation (r = .77) than when these are more freely estimated (ε = 0.5; r = .61), as more of the indicator relationships with one another are absorbed into the factor loadings. The use of different ε values can be helpful in examining potentially different representations of the data and applying targeted constraints (Quilty et al., 2014). However, caution is needed to avoid overfitting solutions (through the use of larger ε values) and unnecessarily attenuating other model parameters (Hopwood & Donnellan, 2010; Marsh et al., 2014; Sass & Schmitt, 2010) like factor correlations, as can be seen in the current illustration.

---

[4] The number of items in this example limits the ability to test a more saturated three factor model, and specification of a more restrictive single factor model would be expected to result in a worse fit than the two-factor model.

[5] We also note that larger values of ε may be needed to aid in convergence in more complex measurement models with a larger number of factors. See Asparouhov & Muthen (2009).

Table 1. Item Stems, Correlations, and Means

Items	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Mean
Adult-Student Relationships									
Adults at this school…									
Q1: want students to do well	1								5.49
Q2: listen to what students have to say	0.84	1							5.13
Q3: recognize and value each individual's cultural background	0.79	0.87	1						5.14
Q4: are treated with respect by students	0.57	0.69	0.58	1					4.13
Inclusive Instruction									
Teachers at this school…									
Q5: want students to think about different ways to solve problems	0.64	0.69	0.67	0.54	1				5.14
Q6: encourage students to value and search for a diversity of opinions, perspectives, and abilities	0.60	0.67	0.73	0.51	0.81	1			4.93
Q7: often connect what students are learning to life outside the Classroom	0.65	0.74	0.72	0.57	0.75	0.79	1		4.95
Q8: The content taught at this school reflects multiple cultural backgrounds, ethnicities, and identities	0.56	0.65	0.75	0.53	0.71	0.82	0.74	1	4.77

Table 2. Standardized Factor Loadings and Estimates of Model Fit across Measurement Models and Rotations

			ESEM							
	CFA		Orthogonal		Geomin (ε = 0.0001)		Geomin (ε = 0.5)		Target	
	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2
Q1	**0.87**	0.00	**0.78**	0.36	**0.85**	0.01	**0.78**	0.12	**0.89**	-0.04
Q2	**0.96**	0.00	**0.90**	0.39	**1.01**	-0.03	**0.92**	0.10	**1.06**	-0.10
Q3	**0.91**	0.00	**0.73**	0.53	**0.67**	0.28	**0.66**	0.34	**0.69**	0.25
Q4	**0.69**	0.00	**0.61**	0.35	**0.62**	0.10	**0.59**	0.17	**0.65**	0.06
Q5	0.00	**0.86**	0.44	**0.73**	0.12	**0.76**	0.23	**0.69**	0.10	**0.77**
Q6	0.00	**0.92**	0.36	**0.89**	-0.11	**1.04**	0.07	**0.91**	-0.15	**1.07**
Q7	0.00	**0.87**	0.51	**0.69**	0.25	**0.66**	0.33	**0.62**	0.24	**0.66**
Q8	0.00	**0.86**	0.39	**0.78**	0.01	**0.86**	0.15	**0.77**	-0.02	**0.88**
$r_{F1,F2}$	0.82		0.00		0.77		0.61		0.81	
Model Fit										
χ^2	90.33*		44.75*		44.75*		44.75*		44.75*	
df	19.00		13.00		13.00		13.00		13.00	
CFI	0.95		0.98		0.98		0.98		0.98	
TLI	0.93		0.96		0.96		0.96		0.96	
RMSEA	0.09		0.08		0.08		0.08		0.08	
SRMR	0.03		0.02		0.02		0.02		0.02	
AIC	-2949.23		-3030.12		-3030.12		-3030.12		-3030.12	
BIC	-2847.64		-2904.14		-2904.14		-2904.14		-2904.14	

CFA = Confirmatory Factor Analysis, ESEM = Exploratory Structural Equation Modeling, ε = epsilon.

*$p < 0.05$. Note. For a given rotation, EFA would produce the same results as those of ESEM.

Target rotation, as the name implies, rotates directly toward a hypothesized pattern. As such, it requires a stronger a priori theory about the expected factor structure than is required for goemin rotation, making it somewhat more suitable for more confirmatory evaluations (Osborne, 2015). Here, target indicators are pre-specified to load on their respective factors in order to guide rotation toward a simplified solution (Marsh et al., 2014) and has been described as being "superior to traditional rotations in its ability to identify the proper solution" in ESEM models (Asparouhov & Muthén, 2009; p. 431). Target rotation simplifies solutions by constraining smaller cross-loadings while retaining model complexity (Asparouhov & Muthén, 2009). In the current illustration, there is good separation in factor loadings for each item with respect to their target and non-target factors (i.e., TLs are much higher than their CL values), see right side of Table 2.

Given the discussion above on how the suppression of non-zero CLs (in a CFA model) can produce biased factor correlations, readers may be surprised to see that the factor correlations from the CFA and target rotation models are nearly identical (.82 and .81, respectively). However, this is the case in the current illustration because the CLs are a mix of negative and positive values, that have been shown to cancel out bias in factor correlations (Konold & Sanders, 2024). More generally, the direction of this bias (positive or negative) depends on both the sign of the omitted CL in relation to the sign and magnitude of the true correlation among the constructs (De Luca et al., 2021; Konold & Sanders, 2024; Steenkamp & Maydeu-Olivares, 2023). For example, when CLs are positive, and the true correlation is positive, constraining CLs to zero in a CFA framework results in an upwardly biased factor association. Conversely, when CLs are negative and the true correlation is positive, the CFA correlation has been found to be downwardly biased.

Summary

Across all model results in Table 2, we would tend toward advocating for the ESEM target rotation. Support for this could come from the ESEM models providing somewhat better fit to the data than the CFA model. Among the ESEM models, the orthogonal rotation is furthest from simple structure and a factor correlation value of 0 is implausible on the basis of both theoretical grounds that school climate variables tend to be related, and non-zero factor associations across all other rotations. The choice between Geomin and target rotation is a bit more nuanced. Because we had a hypothesis about the factor structure entering into the analyses (i.e., which factors would influence which items), target rotation provides more of a mechanism to incorporate this prior theory into the analyses, without the potentially over-restrictive assumption that CL values be zero. This is because the target indicators are specified in advance, and the solution rotates toward that specification. Moreover, others have found it to work well in recovering known structures (Asparouhov & Muthén, 2009; Marsh et al., 2014). In other applications where target indicators for factors are less salient, Geomin rotations could offer a viable alternative.

The aim of the current paper was to elucidate distinctions that exist among methods for evaluating latent variable measurement models. The effects of different model constraints through CFA and ESEM rotations illustrated the differences in results one might expect and highlighted the importance of examining multiple solutions. Historically, standardized factor loadings < |.30| were often overlooked (Brown, 2015; Cudeck & O'Dell, 1994; Tabachnick & Fidell, 2007) and considered ignorable. However, ESEM research has highlighted the importance of cross-loadings and their impact on structural relationships (Asparouhov & Muthen, 2009; Konold & Sanders, 2024; Marsh et al., 2009; Marsh et al., 2013; Marsh et al., 2014; Steenkamp & Maydeu-Olivares, 2023). Namely, that factor correlations tend to be biased when non-zero cross-loadings are forced to be zero in a simple structure CFA specification, even for CLs as small as |.10| to |.30| (Asparouhov et al., 2015; Hsu et al., 2014; Marsh et al., 2013, 2014; Steenkamp & Maydeu-Olivares, 2023).

When there is reasonable theory for the measurement model, we recommend that researchers evaluate both a CFA model and a target ESEM model. If the CFA model provides good fit to the data, better fit than the ESEM specification, and little difference in factor correlations, preference for this model would be indicated on the basis of parsimony. On the other hand, when these conditions are not met, results from ESEM are likely to provide for less biased factor associations.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural equation modeling: a multidisciplinary journal*, *16*(3), 397-438. https://doi.org/10.1080/10705510903008204

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management*, *41*(6), 1561-1577. https://doi.org/10.1177/0149206315591075

Bartholomew, D., Knotts, M., and Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach*. (3rd ed.). John Wiley & Sons.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238. https://doi.org/10.1037/0033-2909.107.2.238

Bollen, K. A. (1989). *Structural equations with latent variables* (Vol. 210). John Wiley & Son

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111-150. https://doi.org/10.1207/S15327906MBR3601_05

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Cornell, D., Shukla, K., & Konold, T. R. (2016). Authoritative school climate and student academic engagement, grades, and aspirations in middle and high schools. *AERA Open*, *2*(2), 1 -18

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7), 1-9. https://doi.org/10.7275/jyj1-4868

Cudeck, R., & O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: significance tests for factor loadings and correlations. *Psychological Bulletin*, *115*(3), 475. https://doi.org/10.1037/0033-2909.115.3.475

Darling-Hammond, L., & DePaoli, J. (2020). Why school climate matters and what can be done to improve it. *Standard: The Journal of the National Association of State Boards of Education, 20*, 6-11.

de Luca, L. M., Herhausen, D., Troilo, G., & Rossi, A. (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. https://doi.org/10.1007/s11747-020-00739-x

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.

Flora, D. B. (2017). *Statistical methods for the social and behavioural sciences: A model-based approach*. Sage

Gorsuch, R. (1983). *Factor analysis*. L. Erlbaum Associates.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430-450. https://doi.org/10.1037/1082-989X.6.4.430

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, *14*(3), 332-346. https://doi.org/10.1177/1088868310361240

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185. https://doi.org/10.1007/BF02289447

Howard, J. L., Gagné, M., & Bureau, J. S. (2017). Testing a continuum structure of self-determined motivation: A meta-analysis. *Psychological Bulletin, 143*(12), 1346–1377. https://doi.org/10.1037/bul0000125

Hsu, H. Y., Skidmore, S. T., Li, Y., & Thompson, B. (2014). Forced zero cross-loading misspecifications in measurement component of structural equation models. *Methodology*, *10*(4), 138–152. https://doi.org/10.1027/1614-2241/a000084

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183-202. https://doi.org/10.1007/BF02289343

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*(4), 443-477. https://doi.org/10.1007/BF02293808

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141-151. https://doi.org/10.1177/001316446002000116

Konold, T. R., Cornell, D. Jia, Y. & Malone, M. (2018). School climate, student engagement, and academic achievement: A latent variable, multilevel multi-informant examination. *AERA Open,* *4*, 1-17. https://doi.org/10.1177/2332858418815661

Konold, T. R., & Sanders, E. A. (2024). On the behavior of fit indices for adjudicating between exploratory structural equation and confirmatory factor analysis models. *Measurement: Interdisciplinary Research and Perspectives*, 1-20. https://doi.org/10.1080/15366367.2023.2264605

Konold, T. R., & Sanders, E. A. (2023). The SEM reliability paradox in a Bayesian framework. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-21. https://doi.org/10.1080/10705511.2023.2220915

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods,* *1*(2), 130–149.https://doi.org/10.1037/1082-989X.1.2.130

Marsh, H.W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big-five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*, 471–491. https://doi.org/10.1037/a0019227

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—Camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, *18*(3), 257-284. https://doi.org/10.1037/a0032773

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10, 85-110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 439-476. https://doi.org/10.1080/10705510903008220

Marsh, H.W., Guo, J., Dicke, T., Parker, P., & Craven, R. (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*, 102-119. https://doi.org/10.1080/00273171.2019.1602503

Marsh, H.W., Nagengast, B., Morin, A. J. S., Parada, R. H., Craven, R. G., & Hamilton, L. R. (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. *Journal of Educational Psychology, 103*(3), 701–732. https://doi.org/10.1037/a0024122

McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64-84. https://doi.org/10.1037/1082-989X.7.1.64

Morin A. J. S., Marsh H., Nagengast B. (2013). Exploratory structural equation modeling: An introduction. In Hancock G. R., Mueller R. O. (Eds.), *Structural equation modeling: A second course* (2nd ed.): 395-436. Greenwich, CT: IAP.

Morin, A. J. S., & Asparouhov, T. (2018). Estimation of a hierarchical Exploratory Structural Equation Model (ESEM) using ESEM-within-CFA. *Montreal, QC: Substantive Methodological Synergy Research Laboratory.*

Morin, A.J.S., Arens, A.K., & Marsh, H.W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 116–139. http://doi.org/10.1080/10705511.2014.961800

Osborne, Jason W. (2015). What is Rotating in Exploratory Factor Analysis? *Practical Assessment, Research & Evaluation*, *20*(2), 1-8. https://doi.org/10.7275/hb2g-m060

Prokofieva, M., Zarate, D., Parker, A., Palikara, O., & Stavropoulos, V. (2023). Exploratory structural equation modeling: a streamlined step by step approach using the R Project software. *BMC* *psychiatry*, *23*(1), 546. https://doi.org/10.1186/s12888-023-05028-9

Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg self-esteem scale method effects. *Structural Equation Modeling*, *13*(1), 99-117. https://doi.org/10.1207/s15328007sem1301_5

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544-559. https://doi.org/10.1080/00223891.2010.496477

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, *45*(1), 73-103. https://doi.org/10.1080/00273170903504810

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304-321. https://doi.org/10.1177/0734282911406653

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). Routledge/Taylor & Francis Group.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Steenkamp, J. B. E., & Maydeu-Olivares, A. (2021). An updated paradigm for evaluating measurement invariance incorporating common method variance and its assessment. *Journal of the Academy of Marketing Science*, *49*, 5-29. https://doi.org/10.1007/s11747-020-00745-z

Steenkamp, J. B. E., & Maydeu-Olivares, A. (2023). Unrestricted factor analysis: A powerful alternative to confirmatory factor analysis. *Journal of the Academy of Marketing Science*, *51*(1), 86-113. https://doi.org/10.1007/s11747-022-00888-1

Steiger, J. H. & Lind, J. C. (1980). *Statistically based tests for the number of factors.* Papers presented at the annual Spring meeting of the Psychometric society. Iowa City, IA.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics.* Pearson Education Inc

Thurstone, L. L. (1947). *Multiple-factor analysis; a development and expansion of The Vectors of Mind.* University of Chicago Press.

van Zyl, L. E., & Ten Klooster, P. M. (2022). Exploratory structural equation modeling: practical guidelines and tutorial with a convenient online tool for mplus. *Frontiers in Psychiatry*, *12*, 795672. https://doi.org/10.3389/fpsyt.2021.795672

Wang, M. T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315-352. https://doi.org/10.1007/s10648-015-9319-1

Citation:
Afolabi, K. T., & Konold, T. (2024). The circle of methods for evaluating latent variable measurement models: EFA, CFA, and ESEM. *Practical Assessment, Research, & Evaluation*, 29(15). Available online: https://doi.org/10.7275/pare.2061

Corresponding Author:
Kelvin Tosin Afolabi
University of Virginia
Email: wrx8mv@virginia.edu

Appendix 1.

Mplus Code for the Substantive Illustration

```
TITLE: Two-factor CFA with four indicators each.
!The dataset used for the analyses.
DATA: FILE IS "VDOE Aggregates.dat";
!Variables in the dataset.
VARIABLE:
     NAMES= id Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;
!Variables used for the analyses.
     USEVARIABLES = Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;
!Each factor is defined by its target indicators.
MODEL: f1 BY Q1 Q2 Q3 Q4;
                f2 BY Q5 Q6 Q7 Q8;
!Maximum likelihood was used as the estimator.
ANALYSIS:
     ESTIMATOR = MLR;
!Additional output of potential interest.
OUTPUT: STANDARDIZED RESIDUAL CINTERVAL MODINDICES (3.0) TECH2 TECH4
SAMPSTAT;

TITLE: Two-factor ESEM (Orthogonal Rotation).
DATA: FILE IS "VDOE Aggregates.dat";
VARIABLE:
     NAMES = id Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;
     USEVARIABLES = Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;
MODEL: f1-f2 BY Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 (*1);
!Orthogonal rotation.
ANALYSIS:
     ESTIMATOR = MLR;
     ROTATION = CF-Varimax(Orthogonal);
OUTPUT: STANDARDIZED RESIDUAL CINTERVAL MODINDICES (3.0) TECH2 TECH4
SAMPSTAT;
```

```
TITLE: Two-factor ESEM (Geomin Rotation, Epsilon = 0).
DATA: FILE IS "VDOE Aggregates.dat";
VARIABLE:
     NAMES = id Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;
     USEVARIABLES = Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;
!The model is specified using Geomin (oblique) rotation.
MODEL: f1-f2 BY  Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8  (*1);
!Geomin rotation with epsilon set to default value of 0.
ANALYSIS:
     ESTIMATOR = MLR;
     ROTATION = GEOMIN (OBLIQUE);
OUTPUT: STANDARDIZED RESIDUAL CINTERVAL MODINDICES (3.0) TECH2 TECH4
SAMPSTAT;

TITLE: Two-factor ESEM (Geomin Rotation, Epsilon = 0.5).
DATA: FILE IS "VDOE Aggregates.dat";
VARIABLE:
     NAMES = id Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;.
     USEVARIABLES = Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;
MODEL: f1-f2 BY  Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 (*1);
!Geomin rotation with epsilon set to 0.5.
ANALYSIS:
     ESTIMATOR = MLR;
     ROTATION = GEOMIN (OBLIQUE, .5);
OUTPUT: STANDARDIZED RESIDUAL CINTERVAL MODINDICES (3.0) TECH2 TECH4
SAMPSTAT;

TITLE: Two-factor ESEM (Target Rotation) with target indicators specified
for each factor.
DATA: FILE IS "VDOE Aggregates.dat";
VARIABLE:
     NAMES=    id Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;
     USEVARIABLES=  Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8;
!The model is specified with target indicators, and CL values to be small.
MODEL: f1 BY Q1 Q2 Q3 Q4  Q5~0 Q6~0 Q7~0 Q8~0  (*1);
                f2 BY Q1~0 Q2~0 Q3~0 Q4~0  Q5 Q6 Q7 Q8 (*1);
ANALYSIS:
     ESTIMATOR = MLR;
     ROTATION = TARGET(OBLIQUE);
OUTPUT: STANDARDIZED RESIDUAL CINTERVAL MODINDICES (3.0) TECH2 TECH4
SAMPSTAT;
```