11-1-2011

A Comparison of Factor Rotation Methods for Dichotomous Data

W. Holmes Finch
Ball State University, whfinch@bsu.edu

A Comparison of Factor Rotation Methods for Dichotomous Data

W. Holmes Finch
Ball State University,
Muncie, IN

Exploratory factor analysis (EFA) is frequently used in the social sciences and is a common component in many validity studies. A core aspect of EFA is the determination of which observed indicator variables are associated with which latent factors through the use of factor loadings. Loadings are initially extracted using an algorithm, such as maximum likelihood or weighted least squares, and then transformed - or rotated - to make them more interpretable. There are a number of rotational techniques available to the researcher making use of EFA. Prior work has discussed the advantages of a number of these criteria from a theoretical perspective, but few previous studies compare their performance across a broad range of conditions. This simulation study compared eight factor rotation criteria in terms of how well they were able to group dichotomous indicator variables correctly on the same factor, order the indicators by the magnitude of the factor loadings (identifying those indicators that were most strongly associated with the factors) and estimate the inter-factor correlations. Results reveal a mixed pattern of performance among the various rotations with the orthogonal Equamax consistently near the top in terms of correctly grouping and ordering indicator variables, and the orthogonal Facparsim performing well with more observed indicators. Advice regarding possible rotations to use for researchers conducting EFA with dichotomous indicators is provided.

Key words: Factor rotation, dichotomous data, exploratory factor analysis, EFA.

Introduction

Exploratory Factor Analysis (EFA) of items on an instrument is a tool employed by psychometricians in the investigation of validity evidence for cognitive and affective measures (Zumbo, 2007; McDonald, 1999). In conjunction with subject matter expertise regarding the purpose of the instrument and its assumed structure, EFA can be used to identify the latent constructs underlying the observed items (McLeod, Swygert & Thissen, 2001). When items are found to group in conceptually meaningful ways based on content, instrument developers can conclude that the traits the scale is intended to measure are actually being represented. Conversely, when individual items are found to load on multiple factors - or to group in ways that do not conform to their content or intent - developers may target them for revision or removal from the instrument (Sass & Schmitt, 2010). Given its role in validity assessment, psychometricians must have a full understanding regarding the performance of EFA in the context of item level data under a variety of conditions. The objective of this simulation study was to investigate one important aspect of the EFA analysis process: factor rotation. A variety of factor rotation methods were compared with respect to how well they recovered the underlying latent structure for a set of dichotomous indicators like those that might comprise a psychological or educational scale. (Readers interested in learning more about the basic factor analysis model are encouraged to read one of several excellent references including: Gorsuch, 1983; Thompson, 2004; McLeod, Swygert & Thissen, 2001.)

W. Holmes Finch is a Professor of Psychology in the Department of Educational Psychology, and Educational Psychology Director of Research in the Office of Charter School. Email him at: whfinch@bsu.edu.

Factor Analysis of Dichotomous Data

The original EFA model was based on the presumption that observed indicators were continuous variables, calling into question its applicability for dichotomous data such as that from item responses (Gorsuch, 1983). Early analyses applying the standard linear EFA model to dichotomous item response data consistently identified a factor reflecting item difficulty, having nothing to do with substantive dimensions related to item content (Hattie, 1985; Guilford, 1941; Spearman, 1927). Furthermore, the use of linear factor analysis with dichotomous items was found to produce distorted factor loading estimates for very difficult and very easy items (Hattie, 1985).

In response to these problems, McDonald introduced nonlinear factor analysis based on the normal ogive (McDonald, 1967; 1962). In the case of dichotomous variables such as item responses, this factor model takes the form

$$P\{U_j = 1 \mid \theta\} = N(\beta_{j0} + \beta_{j1}\theta + \beta_{21}\theta + ... + \beta_{jm}\theta)$$
(1)

where U_j is the response to item with a 1 indicating correct, β_{j0} is the intercept for item j and β_{j1} is the factor loading for item j with latent trait m. Parameter estimation in this Normal Ogive Harmonic Analysis Robust Method (NOHARM) is conducted using unweighted least squares (ULS), allowing for analysis of large sets of items exhibiting high dimensionality (McDonald, 1981; 1967). This model was implemented in the NOHARM software package (Fraser & McDonald, 1988) and features both Varimax and Promax rotations.

Bock and Aitkin (1981) developed an alternative model for the factor analysis of dichotomous item response data that takes the form:

$$P\left(x_{ij} = 1 \mid \theta_j\right) = \frac{1}{\sqrt{2\pi}} \int_{-z_i(\theta_j)}^{\infty} e^{-\frac{t^2}{2}} dt$$
(2)

where $z_i\left(\theta_j\right) = a_j\left(\theta_{ti} - b_j\right)$, a_j is the slope for item j, b_j is the threshold for item j, and θ_{tj} is the latent trait for subject i on item j. In this conceptualization of the model, a_j corresponds to item discrimination and b_j corresponds to item difficulty, in the context of item response theory. This full information factor model underlies the TESTFACT software (Bock, et al., 2003) and is estimated using marginal maximum likelihood (MML), in contrast to the ULS used with NOHARM. Researchers comparing these approaches have found that ULS tends to provide more accurate parameter estimation for a smaller number of items, although MML is generally more accurate for more items (Gosz & Walker, 2002). As with NOHARM, TESTFACT allows for either VARIMAX or PROMAX rotations.

Christofferson (1975) also introduced a factor model for item response data based on the normal ogive model, as was McDonald's approach. The Christofferson model is expressed as

$$P\left(u_i = 1\right) = \int_{z_i}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$
(3)

where z_i is the threshold for item i. This model was expanded upon by Muthén (1978) and has been shown to be equivalent to McDonald's model (McDonald, 1997).

Another approach to factor analysis for dichotomous data, such as item responses, is based on robust weighted least squares (RWLS). Weighted least squares (WLS) estimation has been shown to perform poorly for categorical variables in the context of factor analysis with small to moderate sample sizes (Flora & Curran, 2004). Muthén, du Toit and Spisic (1997) and Muthén (1993) extended the WLS approach in the form of RWLS, which does not require the inversion of the weight matrix used in the standard WLS approach, leading to very stable parameter estimation for samples as small as 100 with dichotomous indicator variables (Flora & Curran, 2004). The RWLS approach can also be used in the context of EFA with the MPLus software package (Muthén & Muthén, 2007) as was done herein.

Factor Rotation

The estimation of factor loadings in EFA typically occurs in two stages, the first of

which - factor extraction - involves the initial estimation of loadings based on the covariance matrix for the indicator variables. The second step in an EFA - factor rotation - involves the transformation of the initial factor loadings in order to make them more interpretable in terms of (ideally) clearly associating an indicator variable with a single factor (Gorsuch, 1983). Although a large number of rotation algorithms have been described in the literature, these criteria all have the common goal of reducing a complexity function, $f(\Lambda)$, so that the loadings approximate a simple structure and are thus more interpretable in practice.

The notion of simple structure has been discussed extensively in the factor analysis literature, and though there is a common sense as to its meaning, there is no agreement regarding exact details. Thurstone (1947) first described simple structure as occurring when each row in the factor loading matrix has at least one zero, where rows represent indicator variables and columns represent factors. He also included 4 other rules that were initially intended to yield over-determination and stability of the factor loading matrix, but which were subsequently used by other researchers to define simple structure for methods of rotation (Browne, 2001). Subsequent to Thurstone's work, others varying definitions of simple structure have been provided. For example, Jennrich (2007) defined perfect simple structure as occurring when each indicator has only one nonzero factor loading and compared it to Thurstone simple structure in which there are a fair number of zeros in the factor loading matrix, but not as many as in perfect simple structure. Conversely, Browne (2001) defined the complexity of a factor pattern as the number of nonzero elements in the rows of the loading matrix. These many varying definitions of simple structure have led to the development of a number of rotational criteria with the overarching goal of obtaining the most interpretable solution possible for a set of data (Asparouhov & Muthén, 2009).

Factor rotations are broadly classified as either: (1) orthogonal, in which the factors are constrained to be uncorrelated, or (2) oblique, in which this constraint is relaxed. Within each of these classes, several options are available.

Browne (2001) provides an excellent discussion of a number of these rotational criteria along with a history of their development and concluded that, when the factor pattern in the population conformed to what is termed above as pure simple structure, most of the rotation methods reviewed produced acceptable solutions. However, when there was greater complexity in the factor pattern, the rotations did not all perform equally well and - in some cases - the majority of them produced unacceptable results (Browne, 2001). For this reason, he argued for the need of educated human judgment in the selection of the best factor rotation solution for a given problem. In a similar vein, Yates (1987) stated that some rotations are designed to find a perfect simple structure solution in all cases, even when this may not be appropriate for the data at hand.

Several excellent discussions of these rotation criteria are available in the literature, including two recently published manuscripts which provide detailed descriptions for interested readers (Sass & Schmitt, 2010; Asparouhov & Muthén, 2009). The rotations included in this study are summarized in Table 1. Many of these methods are readily available in common statistical software packages such as MPlus (Muthén & Muthén, 2007), which is featured in this study, as well as SAS and SPSS. Perhaps the most popular method in applied practice is the orthogonal Varimax rotation (Kaiser, 1958), which is a member of a larger group of criteria known collectively as the Orthomax family of rotations. The goal in Varimax rotation is to create simple structure by maximizing differences among loadings within factors across variables. Other notable Orthomax rotations include Quartimax, Equamax, Parsimax and Factor Parsimony. Promax is a two-stage oblique Procrustean rotation in which loadings are first obtained from the orthogonal Varimax rotation and then transformed based upon a target matrix of loadings raised to a particular power (typically the 4th power), after which a transformation matrix is obtained using least squares (Hendrickson & White, 1964). Other Procrustean rotations include Promaj (Trendafilov, 1994) and Promin (Lorenzo-Seva, 1999). Another group of factor rotations is the Crawford-Ferguson (CF) family (Crawford &

A COMPARISON OF FACTOR ROTATION METHODS FOR DICHOTOMOUS DATA

Table 1: Summary of Studied Rotation Methods*

Rotation Criteria	Definition	Comments
Varimax	$$f(\Lambda) = \left[p \sum_{i=1}^{p} \left(\lambda_{ij}^2 \right)^2 - \left(\sum_{i=1}^{p} \left(\lambda_{ij}^2 \right)^2 \right)^2 \right] / p^2$$	Spreads variance across factors
Quartimin	$$f(\Lambda) = \sum_{i=1}^{p} \sum_{j=1}^{m} \sum_{l \neq j}^{m} \lambda_{ij}^2 \lambda_{il}^2$$	Designed to minimize complexity of loadings across indicator variables.
Quartimax	$$f(\Lambda) = \sum_{i=1}^{p} \sum_{j=1}^{m} \lambda_{ij}^4 + \sum_{i=1}^{p} \sum_{j=1}^{m} \sum_{l \neq j}^{m} \lambda_{ij}^2 \lambda_{il}^2$$	Spreads variance across indicators
Equamax	$$\left(1 - \frac{m}{2p} \right) \sum_{i=1}^{p} \sum_{j=1}^{m} \sum_{l \neq j}^{m} \lambda_{ij}^2 \lambda_{il}^2 + \frac{m}{2p} \sum_{i=1}^{m} \sum_{i=1}^{p} \sum_{l \neq j}^{p} \lambda_{ij}^2 \lambda_{lj}^2$$	Combines Quartimax and Varimax criteria
Parsimax	$$f(\Lambda) = \left(1 - \frac{m-1}{p+m-2} \right) \sum_{i=1}^{p} \sum_{j=1}^{m} \sum_{l \neq j}^{m} \lambda_{ij}^2 \lambda_{il}^2 + \frac{m-1}{p+m-2} \left(\sum_{i=1}^{m} \sum_{i=1}^{p} \sum_{l \neq j}^{p} \lambda_{ij}^2 \lambda_{il}^2 \right)$$	Equal weight is given to factor and indicator complexity.
Geomin	$$f(\Lambda) = \sum_{i=1}^{p} \left[\prod_{j=1}^{m} \left(\lambda_{ij}^2 + \in \right) \right]^{\frac{1}{m}}$$	Accommodates factor complexity while still providing interpretable solution.
Promax	Raise loadings from Varimax to some power (e.g., 4) and rotate the resulting matrix allowing for correlated factors.	Based on Varimax rotation, but allows for correlated factors.
Facparsim	$$f(\Lambda) = \sum_{i=1}^{m} \sum_{i=1}^{p} \sum_{l \neq j}^{p} \lambda_{ij}^2 \lambda_{il}^2$$	Minimizes loading complexity across factors.

*p=Number of indicators, m=Number of factors, λ=Extracted factor loading

Ferguson, 1970). This criterion accounts for complexity across both the indicator variables and the factors. Members of the CF family differ in terms of a parameter, k, ranging between 0 and 1, where larger values of k place greater weight on minimizing factor complexity, whereas lower values emphasize the minimization of indicator variable complexity (Crawford & Ferguson, 1970). Other rotations that have been discussed widely in the literature are oblique Quartimin (Carroll, 1957), which seeks to minimize complexity only within the indicator variables, and Geomin (Yates, 1987) which also was designed to minimize variable complexity, but which allows for more such complexity than does Quartimin. There are a number of other rotation criteria extant in the literature. However, given that the current study is focused on comparing methods that are available to practitioners in commonly available software, they will not be discussed here. The interested reader is invited to read Mulaik (2010) and Browne (2001) for excellent descriptions of these alternative methods of rotation.

Prior Research on Factor Rotations

As noted, a large number of rotational criteria are available to a researcher interested in using EFA. Some of these, such as Varimax and Promax, are well known and frequently used, while others may be less well known but offer statistical advantages over the more commonly used approaches (Asparouhov & Muthén, 2009). Despite the abundance of available rotational methods, a great deal of empirical research has not been conducted regarding which might be best in a given research context (Sass & Schmitt, 2010). In addition, virtually none of the prior work examining the performance of these various rotation methods has been conducted with dichotomous indicator variables (the focus of this study). Therefore, earlier work using continuous indicators provides the only extant evidence regarding the comparative behavior of factor rotation methods, all of which can be applied to both EFA with continuous or dichotomous indicators. Thus, although they did not utilize dichotomous indicators, earlier studies provide researchers with some insights into what might be expected with regard to the performance of these rotation methods in

general. Nevertheless, it is not clear to what extent earlier research with continuous indicators may be applicable. Therefore, this article builds upon this earlier research in an attempt to extend these results based on continuous variables to the case of dichotomous indicators.

One recent Monte Carlo study (Sass & Schmitt, 2010) compared the ability of four rotational methods in terms of their abilities to reproduce the population factor loadings used to generate the data. This study involved 30 standard normally distributed observed indicators with 2 factors, and 4 different types of factor structure including perfect simple, approximate simple, complex and general (a single common factor) structures; note that the variables used in this study were continuous and not categorical. Sass and Schmidt focused on the performance of these rotation methods for normally distributed indicator variables; however, their study is relevant to this research with dichotomous indicators in that it is one of the few to systematically compare multiple rotational criteria. Furthermore, several of the rotations considered by Sass and Schmidt are also included in this study. Therefore, although their results with continuous, normally distributed variables may not be directly applicable to situations involving dichotomous indicators, their study does provide some potential insights into the performance of the rotational criteria that may in turn inform this research.

Sass and Schmidt generated a sample of 300, with correlations between the factors (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7) and used four oblique single stage rotational criteria, including Quartimin, CF-Equamax, CF-Facparsim and Geomin. They found that in the perfect simple structure condition all of the methods performed equally well, echoing Browne (2001). In the more complex cases, however, CF-Equamax and CF-Facparsim demonstrated somewhat less bias in factor loading estimates than did the other rotations. These authors concluded that researchers must be careful not to think of a particular rotational solution as inherently right or wrong, given that model fit does not change based on rotation. Echoing Browne (2001), Sass and Schmidt argued that the selection of the best

rotation must be made by the researcher using informed judgment, and cannot be done deterministically based solely on statistical results.

A similar finding was reported by Asparouhov & Muthén (2009), who stated that based on their own simulated comparisons of the Geomin and Quartimin rotation criteria with loading bias as the primary outcome variable, the researcher in the end is responsible for determining what constitutes a simple and interpretable solution. Consistent with Sass and Schmitt (2010), they found that for simple factor patterns the rotation criteria performed similarly, but for more complex patterns the results across rotational methods (and even for the same method using different settings) might differ substantially. As noted, although the previous simulation research comparing factor rotation performance was focused on continuous indicator variables, it remains relevant for this study in that it provides the only published evidence regarding the behavior of these rotation criteria, all of which can be used with dichotomous indicators.

The goal of this simulation study was to extend upon this earlier work by comparing the performance of several methods of factor rotation with dichotomous, rather than normally distributed continuous, indicator variables, and by including several more rotation criteria, including the very popular Varimax and Promax methods as well as others that have been shown to be effective previously. Furthermore, the current study extends upon these earlier efforts by including a broader range of conditions with respect to number of indicator variables, sample sizes and number of factors. Finally, the focus of this study in terms of outcomes is different than that of the previously mentioned research.

Methodology

A Monte Carlo simulation study was conducted to compare the performance of several methods of factor rotation in four areas: (1) proportion of correctly grouped indicator variables, (2) proportion of incorrectly grouped indicator variables, (3) proportion of indicator variables correctly ordered based on their population factor loading values, and (4) for oblique rotations, bias in the estimates of inter-factor correlations. Outcome 1 was the proportion of all item pairs that should have been grouped together that actually were, and outcome 2 was the proportion of all item pairs that should have been kept separate that actually were. Outcome 3 was the proportion of cases in which the item with the larger factor loading in the population also had the larger loading in the sample. Outcome 4 was the degree of accuracy of the inter-factor correlation estimate, which was calculated as $r_o - r_p$, where r_o = sample estimate of inter-factor correlation between two factors and r_p = population inter-factor correlation used in data simulation. In addition, the standardized bias of the correlation estimates was also calculated as the bias defined previously divided by the standard deviation of the correlation estimates.

These outcomes were selected because they reflect issues that applied researchers might be interested in; that is, how accurately are the factors defined by appropriately grouped variables, how well ordered are the indicators in terms of the magnitude of their relationships to the factors and how well estimated are the correlations among the factors. Although all of these outcomes may be important in specific contexts, one could argue that the ability to accurately identify the factor structure by correctly grouping the items together may be the most crucial. Given that validity assessment is typically based on the extent to which the empirically identified factors reflect what would be expected for the constructs in question based on substantive content of the items, the accuracy of an EFA solution from a sample to reproduce the population factor structure would seem to be paramount. However, in certain circumstances each of these outcomes would be important to researchers using EFA.

For each combination of the simulation conditions, 1,000 replications were generated using MPlus, version 5.1 (Muthén & Muthén, 2007) and all study conditions were completely crossed with one another. Dichotomous indicators were generated in MPlus using threshold values of 0.25 and were held constant across the observed variables. The relationship between the threshold (τ) value and the probability (P_i) of a respondent endorsing a

dichotomous item is $P_i = \frac{1}{1+e^{-\tau}}$. The threshold value of 0.25 corresponds to a probability of endorsing an item of 0.56 and was selected because it has been used in other simulation research involving factor analysis of dichotomous data (French & Finch, 2006; Meade & Lautenschlager, 2004).

For each replication, exploratory factor analysis with Robust Weighted Least Squares (WLSMV) extraction was conducted using the MPlus software because it has been supported for use with categorical data in prior research (e.g., Muthén & Muthén, 2007; Flora & Curran, 2004). In conducting EFA with dichotomous data, MPlus first calculates the tetrachoric correlation matrix among the variables and then uses it to estimate the factor analysis parameters (factor loadings, inter-factor correlations). The commands to run the analysis requested the extraction of the correct number of factors (2 or 4) for a given replication but because the analysis was EFA, individual indicators were not linked to specific factors as they would have been in a confirmatory factor analysis. For example, when the data generated were from a 2 factor condition, the MPlus commands to run the EFA on the sample requested the extraction of 2 factors, but the individual indicators were not linked to a given factor.

Data were generated for either 2 or 4 factors in the population, and for each factor there were either 6 or 12 observed indicator variables, leading to the following combinations: 2 factors with 6 indicators each, 2 factors with 12 indicators each, 4 factors with 6 indicators each and 4 factors with 12 indicators each. Four inter-factor correlation conditions were simulated: 0.1, 0.3, 0.5 and 0.7. All pairs of factors were correlated at the same level for a given combination of study conditions. For example, in the 4 factor, 6 indicator condition with $r = 0.3$, each pair of the 4 factors were generated with a correlation of 0.3. Four sample size conditions were simulated, 100, 200, 500 and 1,000. Prior research studying the minimum sample size necessary for EFA to provide reliable results with continuous indicators has found that when communalities are relatively high (e.g., 0.5), and most of the factors have a large number of indicators population factor are recovered well with samples as small as 100 subjects (MacCallum, et al., 1999).

Conversely, MacCallum, et al. (1999), found that for low communalities and many factors, each of which has a small number indicators, samples of 500 or more are necessary. Preacher and MacCallum (2002) found that for sample sizes as low as 30, factor structure recovery was good (low root mean square error) provided that communalities were high (e.g., 0.8), the number of factors retained was 4 or fewer and the total number of indicators was 25 or more.

Subsequently, other researchers investigating the impact of sample size on factor analysis have reported similar findings with regard to the need for larger samples with relatively poorly conditioned solutions (fewer indicators with low factor loadings, low communalities and many factors), and the positive performance with smaller samples (fewer than 50) when factors are well conditioned (de Winter, Dodou & Wieringa, 2009; Gagné & Hancock, 2006; Mundfrom, Shaw & Ke, 2005). Of particular interest given the inclusion of non-simple structure conditions in the current research are the results of de Winter, et al., who found that in the presence of non-simple structure, EFA performs worse with relatively smaller samples in terms of factor structure recovery, particularly when factors are correlated at 0.5 or greater. Given these earlier studies, sample sizes selected for the current research range from what might be considered somewhat small (100) to very large (1,000).

Finally, the data were generated with 4 levels of factor structure complexity, reflecting different degrees to which individual indicators cross-loaded with a secondary factor. Table 2 provides an example of these patterns for each level of structural complexity in the 2 factor 6 indicator condition. For example, in complexity condition 1 each indicator has non-zero loadings for only one factor, whereas in the other 3 conditions, each indicator has an additional non-zero loading on one other factor with complexity conditions differing based upon the magnitude of these non-zero loadings. In the 4 factor conditions, each indicator variable had only 2 non-zero loadings, one for its primary factor and

Table 2: Example Factor Loading Patterns Used In the Simulations

Complexity Condition 1			Complexity Condition 2		
Indicator	Factor 1	Factor 2	Indicator	Factor 1	Factor 2
Y1	0.8	0	Y1	0.8	0.1
Y2	0.8	0	Y2	0.8	0.1
Y3	0.6	0	Y3	0.6	0.1
Y4	0.6	0	Y4	0.6	0.1
Y5	0.4	0	Y5	0.4	0.1
Y6	0.4	0	Y6	0.4	0.1
Y7	0	0.8	Y7	0.1	0.8
Y8	0	0.8	Y8	0.1	0.8
Y9	0	0.6	Y9	0.1	0.6
Y10	0	0.6	Y10	0.1	0.6
Y11	0	0.4	Y11	0.1	0.4
Y12	0	0.4	Y12	0.1	0.4
Complexity Condition 3			Complexity Condition 4		
Indicator	Factor 1	Factor 2	Indicator	Factor 1	Factor 2
Y1	0.8	0.2	Y1	0.8	0.3
Y2	0.8	0.2	Y2	0.8	0.3
Y3	0.6	0.2	Y3	0.6	0.3
Y4	0.6	0.2	Y4	0.6	0.3
Y5	0.4	0.2	Y5	0.4	0.3
Y6	0.4	0.2	Y6	0.4	0.3
Y7	0.2	0.8	Y7	0.3	0.8
Y8	0.2	0.8	Y8	0.3	0.8
Y9	0.2	0.6	Y9	0.3	0.6
Y10	0.2	0.6	Y10	0.3	0.6
Y11	0.2	0.4	Y11	0.3	0.4
Y12	0.2	0.4	Y12	0.3	0.4

the other for a single secondary factor. For example, in complexity condition 2 with 4 factors and 12 indicators for each, indicator 1 had a loading of 0.8 for factor 1, a loading of 0.1 for factor 2 and loadings of 0 for factors 3 and 4. On the other hand, indicator 48 had a loading of 0.4 for factor 4, a loading of 0.1 for factor 3 and 0 loadings for factors 1 and 2. The decision to allow indicators in the 4 factor conditions to cross load with only one other factor was made to avoid confounding the number of cross loadings with the number of factors, making it impossible to directly compare results in the 2 and 4 factors cases. Similar factor loading patterns were used with the other factor and indicator combinations included in this study. Although a very large number of different such factor patterns could have been simulated using the number of factors and indicators included in this study, these patterns were selected because it was felt that they represented a range of non-simple structure conditions, were few enough so as to keep the study manageable and allowed for investigation of the impact of progressively greater factor complexity.

The methods of factor rotation included the study were Quartimin (oblique), Varimax (orthogonal), Quartimax (orthogonal), Equamax (orthogonal), Parsimax (oblique), Geomin (oblique), Promax (oblique) and Facparsim (oblique). The selection of these particular rotations was made based upon a combination of prior research results, popularity in use and availability in statistical software. Again, though prior research comparing performance of rotational criteria used continuous indicators, these are the only available studies examining this issue; therefore, it was determined that these earlier studies did provide some insights into which rotations should be used. Sass and Schmitt (2010) used only oblique rotations, including Quartimin, oblique CF-Equamax, CF-Facparsim and Geomin, and found that Geomin and Quartimin performed slightly better in a pure simple structure condition (Complexity condition 1 in the current study), whereas oblique CF-Equamax and CF-Facparsim were somewhat better in the more complex cases. Asparouhov and Muthén (2009) compared Quartimin with Geomin using two values of the constant ε, 0.01 and 0.0001 and reported that

Geomin with $\varepsilon = 0.001$ consistently produced the least bias in factor loading estimates. Based on these results, the current study included Geomin with $\varepsilon = 0.001$, Quartimin, and Facparsim. In addition, three orthogonal rotations (i.e., Varimax, Quartimax and Equamax) were included because heretofore their performance has not been investigated in such a study and they are very commonly used in practice. Similarly, Promax was included in the study because of its popularity and ubiquity in statistical software, and the fact that it was not included in the earlier work. For each included rotation criterion, except for Geomin as noted above, the default settings in MPlus were used in conducting the analyses in order to mimic what researchers are likely to do in practice.

In addition to the Monte Carlo simulation, this study also included the use of EFA with item responses from a sample of 1,000 examinees who took the Law School Admissions Test (LSAT). These data, which have been discussed previously in the literature, have been shown to contain 4 separate factors corresponding to the 4 reading passages contained in the exam (Stout, et al., 1996). For these data, EFA using the RWLS method of extraction was followed with each of the rotations included in the simulation study. Note that analysis was conducted on the raw binary data.

Results

Because an initial examination of the simulation outcomes revealed that the results for factors 1, 2, 3 and 4 were similar in terms of the grouping of indicators and the ordering of indicators by factor loading magnitude, results are presented for the first factor only. Similarly, estimates of the inter-factor correlation between factors 1 and 2 were similar to those for the other factor pairs (where applicable), thus, only the results for this correlation will be presented.

Factor Grouping

A repeated measures Analysis of Variance (ANOVA) was used to identify which of the manipulated conditions and their interactions were significantly associated with the proportion of item pairs correctly grouped together, which served as the dependent

variable. These conditions included type of rotation, number of observed indicators per factor, number of factors, factor complexity, sample size and inter-factor correlation. Assumptions of equality of variance and normality of errors were assessed using Levene's test and QQ plots, respectively, and were found to have been met. The results of the ANOVA indicated that the highest order significant (α = 0.05) interaction was type of rotation by number of factors, number of indicators and factor complexity (η^2 = 0.112). In addition, the interaction among type of rotation, inter-factor correlation and factor complexity was also significantly related to the proportion of indicators correctly grouped (η^2 = 0.482), as was the main effect of sample size (η^2 = 0.801). All other significant main effects and interactions were subsumed in one of these three terms and will therefore not be discussed.

Table 3 shows the proportion of observed indicator variables correctly and incorrectly grouped by the number of factors, number of indicators per factor, factor complexity and method of rotation. An examination of these results reveals that across methods of rotation the proportion of variables correctly grouped declined as the factor structure became more complex, but the proportion incorrectly grouped together increased. (Note that the numbers for complexity conditions presented in subsequent tables correspond to the numbers in Table 2). This decrease in indicator grouping accuracy with increased structural complexity was less marked for the Quartimin (QMIN) rotation across the number of factors and number of indicators, and the Facparsim (FAC) when there were 12 indicators per factor, regardless of the number of factors. Indeed, when there were 12 indicators per factor the decline in grouping accuracy for QMIN was very small, 0.04 for 2 factors and 0.02 for 4 factors. By contrast, QMIN also demonstrated a much higher rate of incorrectly grouping indicator variables together for more complex factor patterns, across numbers of factors and indicators. The other rotations generally demonstrated comparable levels of grouping accuracy across the conditions contained in Table 3. The only exceptions to this general

result were for Varimax (VAR) and Parsimax (PAR) with 4 factors, both of which had somewhat larger declines in the proportion of correctly grouped indicators than the other approaches in the presence of 4 factors, and for Equamax (EQU), which consistently demonstrated among the lowest rates of incorrectly grouping indicators together, and comparable rates of correctly grouping indicators with one another.

Table 4 presents the proportions of correctly and incorrectly grouped indicators by method of rotation, inter-factor correlation and factor complexity. As evident in Table 3, with increasing model complexity QMIN displayed a smaller decline in the proportion of correctly grouped indicators and a greater increase in the proportion of incorrectly indicators, than did the other rotation methods. Of particular interest is that two of the orthogonal rotations, VAR and EQU, did not show any greater diminution in the proportion of correctly grouped indicators than the oblique rotations as the inter-factor correlations increased, nor did they have greater increases in the proportion of incorrectly grouped items. By contrast, the orthogonal method QUA exhibited among the highest rates of incorrectly grouped indicators for the more complex factor patterns when the inter-factor correlation was 0.5 or 0.7. EQU and PAR consistently demonstrated among the lowest rates of incorrect indicator grouping, while being comparable to the other rotational methods (except QMIN) in terms of correctly grouped indicator variables.

The impact of the factor pattern on correct indicator grouping was essentially the same regardless of the inter-factor correlation, with decreases in the proportion of correctly grouped item pairs and increases in the proportion of correctly grouped item pairs. For all methods of rotation, the proportion of correctly grouped indicator variables increased concomitantly with increases in sample size, whereas the proportion of incorrectly grouped indicators declined (see Table 5).

Factor Loading Magnitudes

As with the proportion of correctly grouped items, repeated measures ANOVA was used to determine which of the study conditions

Table 3: Proportion of Variables Correctly | Incorrectly Grouped into Factors by Number of Factors (F), Number of Indicators per Factor (I) and Population Factor Complexity (C)

F	I	C	EQU*	GEO	PAR	PRO	QUA	QMIN	VAR	FAC
2	6	1	.94|.10	.94|.11	.94|.10	.91|.10	.94|.11	.93|.14	.94|.10	.88|.10
		2	.91|.16	.90|.18	.91|.16	.87|.17	.90|.18	.91|.31	.91|.16	.79|.17
		3	.86|.27	.85|.31	.86|.28	.82|.29	.85|.33	.88|.57	.85|.27	.69|.26
		4	.78|.45	.77|.51	.77|.45	.74|.48	.78|.56	.87|.83	.75|.46	.66|.49
2	12	1	.97|.02	.97|.03	.97|.02	.95|.02	.97|.03	.96|.12	.97|.02	.99|.03
		2	.95|.04	.95|.05	.95|.04	.93|.05	.95|.05	.96|.32	.95|.04	.98|.06
		3	.89|.10	.88|.11	.89|.09	.86|.10	.89|.22	.92|.66	.88|.10	.98|.13
		4	.80|.24	.80|.28	.80|.21	.77|.23	.83|.48	.92|.96	.78|.22	.95|.29
4	6	1	.92|.13	.91|.14	.91|.13	.90|.14	.91|.14	.91|.21	.90|.13	.82|.15
		2	.90|.17	.89|.17	.89|.17	.87|.16	.89|.18	.90|.30	.88|.16	.73|.19
		3	.86|.25	.86|.26	.85|.25	.83|.24	.86|.27	.90|.43	.83|.25	.63|.26
		4	.82|.38	.82|.41	.79|.38	.82|.41	.85|.42	.90|.59	.73|.42	.51|.43
4	12	1	.96|.05	.95|.05	.95|.05	.95|.15	.95|.05	.95|.07	.95|.14	.99|.06
		2	.94|.06	.94|.07	.94|.06	.94|.19	.94|.06	.95|.13	.94|.18	.96|.08
		3	.89|.13	.92|.18	.88|.12	.92|.32	.90|.16	.94|.28	.93|.31	.95|.18
		4	.82|.22	.88|.31	.79|.20	.88|.45	.85|.28	.93|.41	.83|.45	.93|.31

*EQU = Equamax, GEO = Geomin, PAR = Parsimax, PRO = Promax, QUA = Quartimax, QMIN = Quartimin, VAR = Varimax, FAC = Facparsim.

Table 4: Proportion of Variables Correctly | Incorrectly Grouped into Factors by Inter-Factor Correlations (r) and Population Factor Complexity (C)

r	C	EQU*	GEO	PAR	PRO	QUA	QMIN	VAR	FAC
0.1	1	.97|.04	.97|.04	.97|.04	.95|.08	.97|.04	.97|.04	.97|.07	.98|.03
	2	.96|.05	.95|.05	.95|.05	.94|.08	.95|.05	.95|.05	.95|.07	.95|.05
	3	.93|.10	.92|.11	.92|.10	.91|.13	.92|.10	.92|.11	.92|.12	.92|.11
	4	.85|.21	.84|.24	.84|.21	.83|.30	.84|.23	.91|.60	.85|.30	.80|.26
0.3	1	.96|.05	.96|.05	.96|.05	.95|.07	.96|.05	.96|.05	.96|.07	.96|.05
	2	.94|.07	.94|.08	.94|.08	.92|.10	.94|.08	.94|.08	.94|.09	.92|.08
	3	.91|.13	.90|.15	.90|.14	.88|.18	.89|.15	.93|.40	.89|.17	.86|.16
	4	.83|.26	.84|.31	.84|.26	.82|.36	.82|.30	.91|.68	.86|.36	.84|.37
0.5	1	.95|.08	.94|.08	.94|.08	.93|.10	.94|.08	.94|.08	.94|.09	.95|.08
	2	.94|.10	.93|11	.93|.10	.91|.13	.93|.11	.92|.22	.93|.12	.91|.10
	3	.88|.19	.87|.23	.87|.20	.85|.27	.87|.22	.93|.66	.87|.27	.86|.24
	4	.80|.34	.85|.41	.78|.33	.83|.42	.85|.55	.94|.73	.84|.41	.80|.48
0.7	1	.91|.14	.90|.16	.90|.15	.88|.17	.90|.16	.94|.38	.89|.16	.87|.16
	2	.87|.21	.87|.24	.87|.21	.84|.27	.87|23	.94|.75	.86|.26	.85|.24
	3	.82|.32	.85|.38	.81|.31	.82|.38	.85|.52	.92|.80	.83|.37	.79|.51
	4	.76|.49	.75|.55	.74|.43	.74|.49	.77|.69	.88|.82	.74|.48	.73|.50

*EQU = Equamax, GEO = Geomin, PAR = Parsimax, PRO = Promax, QUA = Quartimax, QMIN = Quartimin, VAR = Varimax, FAC = Facparsim.

Table 5: Proportion of Variables Correctly | Incorrectly Grouped into Factors by Sample Size

N	EQU*	GEO	PAR	PRO	QUA	QMIN	VAR	FAC
100	.83|.31	.85|.33	.82|.30	.82|34	.84|.33	.88|.45	.84|.33	.84|.32
200	.87|.21	.88|.24	.86|.21	.86|.25	.88|.24	.92|.40	.88|.24	.86|.23
500	.92|.11	.92|.14	.91|.11	.90|.17	.93|.17	.95|.37	.92|.16	.89|.13
1000	.94|.07	.94|.09	.94|.07	.93|.13	.94|.14	.97|.37	.94|.13	.94|.12

*EQU = Equamax, GEO = Geomin, PAR = Parsimax, PRO = Promax, QUA = Quartimax, QMIN = Quartimin, VAR = Varimax, FAC = Facparsim.

and their interactions were significantly related to the proportion of correctly ordered factor indicators based on their loading magnitudes in the sample. The highest order significant interaction was the rotation by inter-factor correlation by factor pattern ($\eta^2 = 0.201$). In addition, the 2-way interactions of rotation by number of indicators per factor ($\eta^2 = 0.236$) and rotation by number of factors ($\eta^2 = 0.275$) were also statistically significant, as was the main effect of sample size ($\eta^2 = 0.858$).

For all of the rotations, results demonstrate (see Table 6) that the proportion of correctly ordered factor indicators by loading magnitude declines with increases in the inter-factor correlation and with increased factor complexity (reflected through higher numbers for the factor complexity condition). In addition, the deleterious impact of greater factor complexity was more pronounced for larger values of the inter-factor correlation. For example, in the simple structure condition (C = 1) with correlations of 0.1 and 0.3, the rotations performed similarly with respect to correct ordering of the factor indicators by loading magnitude, whereas for r = 0.5 FAC displayed a higher proportion of correctly ordered factor loadings, and for r = 0.7, FAC, QMIN and VAR all had somewhat higher proportions of correctly ordered loadings. On the other hand, for the

greatest factor complexity (C = 4) VAR consistently had the highest proportion of correctly ordered loadings, with a variety of other rotations performing comparably for a given inter-factor correlation. For example, QMIN performed similarly to VAR in the most complex case for inter-factor correlations of 0.1, 0.3 and 0.7, and FAC had similar values to VAR for proportion of correctly ordered loadings in the most complex case when r = 0.3.

Results in Table 7 show that all of the rotations were more accurate in terms of correctly ordering indicators by the magnitude of factor loadings for 12 indicators, for 2 factors and for larger sample sizes. FAC was the rotation method whose performance was most strongly influenced by the number of indicators. For 6 indicators per factor, it performed the worst in terms of correctly ordering loadings, whereas for 12 indicators it performed the best. QMIN and VAR consistently produced among the most accurate ordering of loadings by magnitude across all of the conditions contained in Table 7. The performances of the other rotation methods were generally similar to one another, and somewhat worse than that of QMIN and VAR.

Inter-Factor Correlation Bias

A repeated measures ANOVA identified the 3-way interaction of rotation method by

Table 6: Proportion of Factor Loadings Correctly Ordered by Magnitude by Inter-Factor Correlations (r) and Population Factor Complexity (C)

r	C	EQU*	GEO	PAR	PRO	QUA	QMIN	VAR	FAC
0.1	1	.94	.94	.94	.93	.94	.94	.94	.96
	2	.93	.92	.93	.91	.93	.93	.93	.94
	3	.90	.89	.89	.88	.91	.93	.90	.92
	4	.83	.81	.81	.81	.81	.84	.84	.81
0.3	1	.93	.92	.93	.91	.92	.93	.93	.94
	2	.91	.90	.90	.89	.90	.91	.91	.91
	3	.87	.85	.85	.84	.85	.87	.87	.84
	4	.78	.76	.75	.77	.76	.81	.81	.82
0.5	1	.90	.89	.89	.88	.89	.90	.90	.95
	2	.89	.87	.88	.86	.87	.89	.90	.92
	3	.81	.79	.79	.79	.78	.84	.83	.81
	4	.73	.70	.68	.68	.73	.70	.77	.70
0.7	1	.83	.81	.81	.81	.80	.84	.85	.84
	2	.79	.75	.75	.77	.75	.82	.81	.80
	3	.72	.69	.67	.71	.70	.76	.76	.69
	4	.65	.63	.57	.65	.64	.70	.70	.64

*EQU = Equamax, GEO = Geomin, PAR = Parsimax, PRO = Promax, QUA = Quartimax, QMIN = Quartimin, VAR = Varimax, FAC = Facparsim.

Table 7: Proportion of Factor Loadings Correctly Ordered by Magnitude by Number of Indicators per Factor (I), Number of Factors (F), and Sample Size

I	EQU*	GEO	PAR	PRO	QUA	QMIN	VAR	FAC
6	.75	.72	.72	.74	.72	.76	.77	.66
12	.93	.92	.91	.90	.92	.94	.93	.97

F	EQU	GEO	PAR	PRO	QUA	QMIN	VAR	FAC
2	.89	.86	.87	.86	.88	.91	.90	.86
4	.78	.77	.76	.79	.76	.79	.80	.78

N	EQU	GEO	PAR	PRO	QUA	QMIN	VAR	FAC
100	.69	.67	.66	.67	.67	.71	.70	.68
200	.80	.78	.77	.78	.78	.82	.82	.79
500	.91	.89	.89	.90	.89	.94	.93	.90
1000	.95	.94	.94	.94	.94	.96	.96	.92

*EQU = Equamax, GEO = Geomin, PAR = Parsimax, PRO = Promax, QUA = Quartimax, QMIN = Quartimin, VAR = Varimax, FAC = Facparsim.

inter-factor correlation by factor complexity ($\eta^2 = 0.049$) as the highest order significant term. In addition, the main effects of number of factors ($\eta^2 = 0.313$), number of indicators per factor ($\eta^2 = 0.041$), and sample size ($\eta^2 = 0.021$) were also statistically significant. Table 8 contains the mean raw bias and the standardized bias values across replications by the inter-factor correlation and the degree of model complexity. For $r = 0.1$, the sample correlation estimates displayed a positive bias across rotations, except for the simple structure condition (C = 1). In addition, as the degree of complexity increased, so did both raw and standardized bias, except for PRO. When $r = 0.3$, the negative bias in the simple structure condition was greater than for $r = 0.1$, and the positive bias for more complex models was lower, across rotation methods. For $r = 0.5$ and 0.7, bias was uniformly negative across levels of factor complexity, with greater negative bias associated with the largest population correlation. In addition, for $r = 0.5$ all rotation methods, except PAR, displayed greater negative bias for simple structure data (C = 1) or for the most complex structure (C = 4). In contrast, when $r = 0.7$, bias was generally higher for simple structure than for the next level of factor complexity (C = 2), after which bias increased concomitantly with increased model complexity. None of the rotation criteria consistently produced the least raw or standardized biased estimates.

Table 9 shows that inter-factor correlation bias was more pronounced (and negative) when more indicators were present. In addition, the degree of bias for most of the rotation methods was slightly greater (and negative) for 4 factors as compared to 2, where the bias was positive. Finally, bias in the inter-factor correlation estimates declined with increased sample size, and across all conditions PAR produced somewhat more negatively biased estimates than the other criteria. Otherwise, differences in estimation accuracy across the conditions were relatively minor.

Analysis of LSAT Data

In order to demonstrate the relative performance of the rotation criteria on an actual, well studied data set, EFA was run on the LSAT data described in Stout, et al. (1996). Given that these authors, and others, reported the presence of 4 stable dimensions, 4 factors were extracted in this analysis, and each rotation was applied. Table 10 contains the factor loadings only for the primary factor for each item in order to save space. There were no cross-loaded items for any of the rotation criteria, defined as having multiple factors for which the loading values were great than 0.32 (Tabachnick & Fidell, 2007). A perusal of these results demonstrates that across items and factors, the loading values for the 8 different rotations were very similar to one another. There is no discernible pattern of difference in loadings by rotation, suggesting that a researcher using any of these criteria would reach the same substantive conclusions regarding both how items grouped together, and the strength of relationships between items and factors.

Table 11 includes the correlation estimates for the 4 factor solution of the LSAT data for each of the oblique rotations studied here, and their standard errors with the exception of PROMAX, for which standard errors are not calculated in MPlus. These results demonstrate a greater degree of variation across rotation criteria than was evident for the factor loadings. For example, PROMAX had much larger inter-factor correlation estimates than the other methods for factor 1 with 3, 1 with 4 and 3 with 4. By contrast, PARSIMAX had much lower correlation estimates than the other methods for factors 1 with 3, 1 with 4, 2 with 4 and 3 with 4. GEOMIN, QUARTIMIN and FACPARSIM had very similar inter-factor correlation estimates to one another for this sample.

Conclusion

This study extends previous research comparing rotations in EFA, which focused on continuous factor indicator variables by comparing the performance of 8 factor rotation criteria with dichotomous indicator variables using the WLSMV initial extraction method in MPlus across a variety of conditions. Among the rotations included were some that had previously been found to be promising in terms of accuracy of factor loading estimates such as Geomin and Facparsim, and others that had not been studied before but which are very commonly used in practice, including Varimax and Promax. The outcomes of interest included

Table 8: Inter-Correlation Bias (Standardized Bias) by Inter-Factor Correlations (r) and Population Factor Complexity (C)

r	C	GEO*	PAR	PRO	QMIN	FAC
0.1	1	-0.04 (-0.43)	-0.05 (-0.75)	-0.02 (-0.23)	-0.04 (-0.44)	-0.03 (-0.33)
	2	0.08 (0.45)	0.05 (0.34)	0.12 (0.69)	0.08 (0.49)	0.09 (0.50)
	3	0.18 (0.65)	0.14 (0.55)	0.22 (0.84)	0.19 (0.73	0.19 (0.73)
	4	0.24 (0.73)	0.21 (0.73)	0.17 (0.52)	0.25 (0.80)	0.26 (0.79)
0.3	1	-0.12 (-0.84)	-0.16 (-0.93)	-0.10 (-0.71)	-0.11 (-0.79)	-0.11 (-0.78)
	2	-0.01 (-0.04)	-0.07 (-0.38)	0.03 (0.13)	-0.01 (-0.02)	0.01 (0.01)
	3	0.07 (0.22)	0.01 (0.05)	0.08 (0.30)	0.08 (0.27)	0.09 (0.29)
	4	0.09 (0.25)	0.07 (0.23)	-0.02 (-0.07)	0.12 (0.35)	0.11 (0.32)
0.5	1	-0.21 (-0.95)	-0.27 (-1.53)	-0.18 (-0.94)	-0.20 (-0.92)	-0.21 (-0.92)
	2	-0.09 (-0.34)	-0.17 (-0.77)	-0.08 (-0.31)	-0.09 (-0.32)	-0.09 (-0.33)
	3	-0.08 (-0.22)	-0.12 (-0.46)	-0.14 (-0.43)	-0.06 (-0.18)	-0.08 (-0.19)
	4	-0.13 (-0.35)	-0.09 (-0.29)	-0.20 (-0.58)	-0.19 (-0.60)	-0.20 (-0.59)
0.7	1	-0.31 (-1.00)	-0.37 (-1.65)	-0.31 (-1.20)	-0.30 (-1.07)	-0.32 (-1.06)
	2	-0.26 (-0.78)	-0.31 (-1.15)	-0.32 (-1.00)	-0.25 (-0.76)	-0.27 (-0.79)
	3	-0.30 (-0.80)	-0.28 (-0.80)	-0.35 (-1.06)	-0.36 (-1.06)	-0.36 (-1.08)
	4	-0.38 (-0.99)	-0.31 (-0.81)	-0.33 (-1.05)	-0.33 (-1.54)	-0.36 (-1.44)

*GEO = Geomin, PAR = Parsimax, PRO = Promax, QMIN = Quartimin, FAC = Facparsim.

Table 9: Inter-Correlation Bias by Magnitude by Number of Indicators Per Factor (I), Number of Factors (F), and Sample Size

I	GEO*	PAR	PRO	QMIN	FAC
6	0.03	-0.03	0.02	0.04	0.06
12	-0.13	-0.15	-0.15	-0.14	-0.13

F	GEO	PAR	PRO	QMIN	FAC
2	0.17	0.10	0.16	0.11	0.12
4	-0.16	-0.17	-0.17	-0.14	-0.15

N	GEO	PAR	PRO	QMIN	FAC
100	-0.11	-0.12	-0.08	-0.10	-0.11
200	-0.10	-0.12	-0.08	-0.09	-0.10
500	-0.06	-0.09	-0.08	-0.06	-0.08
1000	-0.03	-0.08	-0.08	-0.04	-0.06

*GEO = Geomin, PAR = Parsimax, PRO = Promax, QMIN = Quartimin, FAC = Facparsim.

the proportion of accurately grouped indicator variables, the proportion of indicators correctly ordered by the magnitude of their loading values and, for the oblique methods, the accuracy of inter-factor correlation estimates. It is hoped that this study builds upon earlier work by focusing on dichotomous indicators (i.e., items), by including outcomes that would be of interest to practitioners interested in using these methods to identify potential latent variables in existing measures and by expanding the range of conditions under which the rotations are examined, including the rotations themselves.

Implications for Practice

One implication of this study for researchers using EFA with categorical indicator variables is that when they know, or suspect, that the correlations among the factors will be upwards of 0.5, they should expect to have problems not only with appropriately grouping variables together, but also with accurately ordering variables in terms of the importance of

their relationships with the factors. These problems are likely to be particularly acute if the factor pattern structure is very complex. It does seem however, that having a larger sample may ameliorate these problems to some extent, so that when it is likely the factors will be highly correlated and/or the factor pattern may be complex in nature, researchers should ideally try to obtain samples of 500 or more. These results are similar to those reported in de Winter, Dodou and Wieringa (2009) for continuous data.

A second implication is that - for the oblique methods of rotation studied - there may be problems with accurately estimating inter-factor correlations across conditions like those simulated here. When these correlations were greater than 0.3, all of the criteria produced underestimates of r, whereas for lower correlations r was overestimated for more complex factor patterns and underestimated for the less complex patterns. These correlation estimation bias results are similar to those

Table 10: Rotated Factor Loading Matrices for LSAT Data

Item	EQU*	GEO	PAR	PRO	QUA	QMIN	VAR	FAC
Factor 1								
1	0.35	0.33	0.33	0.33	0.34	0.33	0.35	0.32
2	0.40	0.40	0.40	0.41	0.40	0.40	0.40	0.39
3	0.43	.045	0.45	0.47	0.43	0.45	0.43	0.45
4	0.36	0.39	0.38	0.40	0.36	0.38	0.36	0.38
5	0.40	0.38	0.38	0.39	0.39	0.38	0.39	0.38
6	0.51	0.53	0.52	0.55	0.51	0.53	0.51	0.53
7	0.33	0.30	0.31	0.30	0.31	0.30	0.33	0.30
Factor 2								
8	0.52	0.54	0.54	0.56	0.51	0.54	0.52	0.53
9	0.38	0.40	0.40	0.41	0.37	0.39	0.38	0.39
10	0.52	0.55	0.55	0.57	0.51	0.55	0.53	0.54
11	0.28	0.27	0.28	0.28	0.27	0.27	0.28	0.27
12	0.37	0.40	0.39	0.42	0.37	0.40	0.37	0.39
13	0.38	0.37	0.37	0.39	0.38	0.38	0.37	0.38
Factor 3								
14	0.54	0.55	0.54	0.58	0.54	0.56	0.54	0.55
15	0.53	0.54	0.53	0.56	0.53	0.54	0.53	0.54
16	0.44	0.46	0.45	0.48	0.44	0.46	0.44	0.46
17	0.16	0.15	0.15	0.15	0.16	0.15	0.16	0.15
18	0.48	0.48	0.45	0.49	0.49	0.49	0.47	0.49
19	0.51	0.50	0.47	0.51	0.52	0.51	0.50	0.51
Factor 4								
20	0.42	0.41	0.38	0.41	0.43	0.41	0.42	0.41
21	0.56	0.56	0.53	0.57	0.57	0.57	0.55	0.56
22	0.59	0.60	0.56	0.61	0.60	0.60	0.58	0.60
23	0.47	0.48	0.45	0.49	0.48	0.48	0.47	0.48
24	0.50	0.52	0.49	0.53	0.50	0.52	0.50	0.52

*EQU = Equamax, GEO = Geomin, PAR = Parsimax, PRO = Promax, QUA = Quartimax, QMIN = Quartimin, VAR = Varimax, FAC = Facparsim.

Table 11: Inter-Factor Correlation (Standard Error) Estimates for LSAT Data by Oblique Rotations

Factor Pair	GEO*	PAR	PRO	QMIN	FAC
1 with 2	0.35 (0.05)	0.30 (0.04)	0.32 (NA)	0.34 (0.06)	0.34 (0.06)
1 with 3	0.28 (0.05)	0.20 (0.04)	0.42 (NA)	0.28 (0.05)	0.29 (0.05)
1 with 4	0.26 (0.05)	0.18 (0.04)	0.35 (NA)	0.26 (0.05)	0.26 (0.06)
2 with 3	0.32 (0.05)	0.35 (0.04)	0.36 (NA)	0.33 (0.05)	0.31 (0.05)
2 with 4	0.42 (0.05)	0.23 (0.04)	0.38 (NA)	0.42 (0.05)	0.42 (0.05)
3 with 4	0.30 (0.04)	0.20 (0.03)	0.50 (NA)	0.32 (0.04)	0.33 (0.05)

*EQU = Equamax, GEO = Geomin, PAR = Parsimax, PRO = Promax, QMIN = Quartimin,
FAC = Facparsim.

reported by Sass and Schmitt (2010) for the case of continuous indicators.

A third implication for practitioners is that including more indicator variables (assuming that they are of good quality) will yield better solutions both in terms of correctly grouping the indicators and accurately ordering them in terms of their relationships to the factors. This result seems reasonable given that including more indicators for each factor provides a greater amount of information for the EFA extraction algorithm as well as for the rotations. The number of indicators was particularly important for the FAC technique, particularly in the case of a more complex factor pattern structures with more factors. Based on these results, researchers may consider using FAC when they have at least 12 indicators per factor, as it demonstrated better performance in terms of grouping the variables as well as ordering them, particularly in the 4 factor case. On the other hand, FAC would not appear to be optimal with fewer indicators per factor.

A final implication of these results is that, in terms of both indicator grouping and ordering of importance in terms of factor relationships, researchers may generally find orthogonal and oblique rotations will produce

similar results. Indeed, one of the consistently best performers in this study was the orthogonal rotation EQU. This result is not completely surprising, as EQU was designed to spread loading variation more equally across factors than several of the other rotations studied here (Saunders, 1962) by combining the VAR and QUA criteria. Thus, although VAR seeks to maximize the variation of loadings for factors, and QUA seeks to simplify loadings for the observed variables, EQU combines these two goals. This is not to suggest that researchers should only use EQU as the rotation of choice for all problems. When factors are thought to be correlated, the choice of an orthogonal rotation may not be appropriate, regardless of how well it performs. However, when the inter-factor correlation is low and the primary goal of a study is to identify which indicators are associated with which factors, EQU would be a reasonable choice.

When a researcher is interested in estimating inter-factor correlations, or they believe that these correlations may be fairly large (greater than 0.5), several of the oblique rotations studied here would appear to be appropriate. In particular, PAR and FAC (for situations with a larger number of indicator

variables) demonstrated consistently strong performance in terms of correctly grouping and ordering indicator variables. On the other hand, QMIN may not be reliable for researchers interested in finding the correct groupings of factor indicators, as it (or the equivalent methods of oblique Quartimax and Oblimin) appears to reduce dimensionality in the sample too much by grouping most of the variables into a single factor. As a consequence, researchers using QMIN may come to the conclusion that, based on the sample there are a smaller number of factors present than is actually true for the population.

Limitations

As with any research effort, limitations to this study that must be considered when interpreting the results. First, for all of the rotations the MPlus system defaults were used. This was a decision made for two reasons: (1) It was desired to mimic what might be most commonly done in practices, and (2) In many cases there are a very large number of alternative settings that could have been used for some of the rotations. Therefore, in order to keep the study to a manageable size and the interpretation of the results fairly straightforward, it was felt that only a limited number of options could be used. Nonetheless, in practice researchers can choose from a broader range of settings when using many of these rotational criteria.

A second limitation relates to the conditions simulated, including the factor patterns used and the number of indicators. In both cases, the selections made for this study were designed to mimic what would be seen in practice. However, clearly many other factor patterns and numbers of indicators could have been included, which may well have provided different results. Future studies should focus on both of these issues in order to expand upon what was learned here.

Finally, these results were based on dichotomous indicator variables, which may not translate directly to ordinal data, such as that commonly found in many psychological scales. It should be noted that because rotations focus on loadings rather than the raw data, it is not clear how important this issue might be.

Nonetheless, future research should verify to what extent the nature of the categorical data has an impact on the performance of rotational criteria.

Summary

In the final analysis, the admonition offered by Browne (2001) for researchers to use their expert judgment in conjunction with statistical results is definitely supported by these results. It is clearly not possible to state that any single rotational criterion will fit all EFA problems adequately, although in practice researchers often appear to use favorites regardless of the context. However, these results do suggest that certain features of the data will support the use of one or more such methods studied here. Clearly the ubiquitous VAR and PRO rotations must be used with caution when at all, as often they do not produce optimal results in terms of accurately reflecting the underlying factor structure. With the increased availability of other rotations in software packages such as MPlus, researchers are no longer limited to a small number of available options, and can thus experiment with a broader array of tools than could be done previously.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397-438.

Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*(*4*), 443-459.

Bock, R. D., Gibbons, R., Schilling, S. G., Muraki, E., Wilson, D. T. & Wood, R. (2003). *TESTFACT 4*. Lincolnwood, IL: Scientific Software International.

Browne, M. W. (2001). An overview of analytic rotations in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(*1*), 111-150.

Carroll, J. B. (1957). Biquartimin criterion for rotation to oblique simple structure in factor analysis. *Science*, *126*, 1114-1115.

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

Crawford, C. B., & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika*, *35*, 321-332.

de Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*, 147-181.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466-491.

Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research*, *23*, 267-269.

French, B. F. & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling*, *13*(*3*), 378-402.

Gagné, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmation factor models. *Multivariate Behavioral Research*, *41*, 65-83.

Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gosz, J. K., & Walker, C. M. (2002). *An empirical comparison of multidimensional item response data using TESTFACT and NOHARM*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Guilford, J. P. (1941). The difficulty of a test and its factor composition. *Psychometrika*, *6*, 67-78.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*(*2*), 139-164.

Jennrich, R. I. (2007). Rotation methods, algorithms, and standard errors. In R. Cudek & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions*, 315-335. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(*3*), 187-200.

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research*, *34*, 347-365.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(*1*), 84-99.

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika*, *27*, 297-415.

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

McDonald, R. P. (1981). The dimensinality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*, 100-117.

McDonald, R. P. (1997). Normal-ogive multidimensional model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*, 257-269. New York: Springer.

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

McLeod, L. D., Swygert, K. A., & Thissen, D. (2001). Factor analysis for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring*, 189-216. Mahwah, NJ: Lawrence Erlbaum Associates.

Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling*, *11*, 60-72.

Mulaik, S. A. (2010). *Foundations of factor analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, *5*, 159-168.

Muthén, B. O. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, *43*, 531-560.

Muthén, B. O. (1993). Goodness of fit with categorical and other non-normal variables. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models*, 205-243. Newbury Park, CA: Sage.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript.

Muthén, L. K., & Muthén, B. O. (2007). *MPlus, version 5: User's guide*. Los Angeles: Author.

Preacher, K. J. & MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior Genetics*, *32*(*2*), 153-161.

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, *45*, 73-103.

Saunders, D. R. (1962). *Trans-varimax: Some properties of the Ratiomax and Equamax criteria for blind orthogonal rotation*. Paper presented at the Meeting of the American Psychological Association, St. Louis, MO, September.

Spearman, C. (1929). *The abilities of man.* New York: Macmillan.

Stout, W., Habing, B., Douglas, J., Kim, H. R., Roussos, L., & Zhang, J. Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement*, *19*(*4*), 331-354.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston: Pearson.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago press.

Trendafilov, N. T. (1994). A simple method for procrustean rotation in factor analysis using majorization theory. *Multivariate Behavioral Research*, *29*, 385-408.

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis.* Albany: State University of New York Press.

Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao (Series Ed.) & S. Sinharay (Volume Ed.), *Handbook of statistics: Vol 25 Psychometrics*, 45-80. Amsterdam: Elsevier.