The Journal of Sex Research

ISSN: 0022-4499 (Print) 1559-8519 (Online) Journal homepage: http://www.tandfonline.com/loi/hjsr20

A Methodological Review of Exploratory Factor Analysis in Sexuality Research: Used Practices, Best Practices, and Data Analysis Resources

John K. Sakaluk & Stephen D. Short

To cite this article: John K. Sakaluk & Stephen D. Short (2017) A Methodological Review of Exploratory Factor Analysis in Sexuality Research: Used Practices, Best Practices, and Data Analysis Resources, The Journal of Sex Research, 54:1, 1-9, DOI: 10.1080/00224499.2015.1137538

To link to this article: http://dx.doi.org/10.1080/00224499.2015.1137538

Published online: 17 Feb 2016.

JOURNAL OF SEX RESEARCH, *54*(1), 1–9, 2017

ISSN: 0022-4499 print/1559-8519 online
DOI: 10.1080/00224499.2015.1137538

Taylor & Francis
Taylor & Francis Group

A Methodological Review of Exploratory Factor Analysis in Sexuality Research: Used Practices, Best Practices, and Data Analysis Resources

John K. Sakaluk
Department of Psychology, University of Toronto Mississauga

Stephen D. Short
Department of Psychology, College of Charleston

Sexuality researchers frequently use exploratory factor analysis (EFA) to illuminate the distinguishable theoretical constructs assessed by a set of variables. EFA entails a substantive number of analytic decisions to be made with respect to sample size determination, and how factors are extracted, rotated, and retained. The available analytic options, however, are not all equally empirically rigorous. We discuss the commonly available options for conducting EFA and which options constitute best practices for EFA. We also present the results of a methodological review of the analytic options for EFA used by sexuality researchers in more than 200 EFAs, published in more than 160 articles and chapters from 1974 to 2014, in a sample of sexuality research journals. Our review reveals that best practices for EFA are actually those least frequently used by sexuality researchers. We introduce freely available analytic resources to help make it easier for sexuality researchers to adhere to best practices when conducting EFAs in their own research.

Exploratory factor analysis (EFA) is a statistical technique that helps researchers uncover the theoretical constructs underlying a set of variables, but it also requires researchers to make numerous analytical decisions. Namely, researchers must determine an appropriate sample size, and select among competing extraction, rotation, and retention methods. Understanding which analytic options constitute best practices for EFA is important, because failing to use them can negatively impact theory development and empirical outcomes. In the present article, we first present an overview of EFA, describing the options for sample size determination, factor extraction, rotation, and retention, and discussing which options constitute best practices and why. We then report on the results of a methodological review of the frequency with which the various options have been used by sexuality researchers. We conclude by describing a set of resources (located in the online supplemental materials) that should help sexuality researchers increase their use of best practices when conducting EFA.

Correspondence should be addressed to John K. Sakaluk, 3359 Mississauga Road, Deerfield Hall Rm. 4098, Department of Psychology, Mississauga, Ontario, Canada L5L 1C6. E-mail: jksakaluk@gmail.com

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HJSR.

A Brief Overview of Exploratory Factor Analysis

EFA is a statistical technique for analyzing patterns of correlations to uncover empirically distinct latent (i.e., unobserved) constructs. EFA is distinguishable from other types of latent variable analysis, such as confirmatory factor analysis (CFA). For EFA, researchers specify how many factors they want to extract, but otherwise leave it to the data to indicate which variables are important manifestations of—or "load onto"—which factors. For CFA, alternatively, researchers specify the exact pattern of which variables load onto which factors to see how well their a priori model accounts for the data. EFA can therefore be thought of as a means of *building* a theory of construct measurement (i.e., in cases where a theory of measurement does not already exist, or a theory of measurement has been refuted and theorizing must begin anew), whereas CFA can be thought of as a means of *testing* a theory of construct measurement.

Analytic Options and Best Practices for Exploratory Factor Analysis

EFA is often described as a flexible form of analysis, largely due to the numerous analytic options for how to determine appropriate sample sizes and how to extract, rotate, and retain factors. We now turn to reviewing these options and note best practices.

How Are Best Practices Evaluated?. Statistical methodologists determine best practices for EFA by conducting simulation studies (Bandalos, 2006). Simulations involve specifying a population EFA model with particular values (i.e., number of factors [e.g., three], factor loadings [e.g., ranging between 0.30 and 0.60], and correlations between factors [e.g., ranging between 0.20 and 0.37]). Statistical methodologists then simulate large numbers of random samples from this population to evaluate how accurately the prespecified population values are reproduced by the chosen analytic options. Sample sizes and extraction, rotation, and retention methods yielding factor solutions that closely align with population values constitute best practices, whereas those that do not are to be avoided. The best practices for EFA that we now discuss, in other words, are not merely opinion. Rather, they constitute what is known about EFA based on empirical study.

Sample Size Determination. Before performing EFA, researchers must first determine how large a sample they need. Numerous guidelines have been proposed to help researchers determine an appropriate minimum sample size when using EFA. Gorsuch (1983), for example, recommended a ratio of five participants for each variable included in the EFA, whereas others have suggested even larger ratios of participants:variables (e.g., Everitt, 1975). Another type of guideline stipulates absolute minimum sample sizes; recommendations of 100 to 250 are common (Cattell, 1978; Gorsuch, 1983). Also popular is Comrey and Lee's (1992) rating scale for evaluating EFA sample size quality (100 = *Poor*; 1,000 = *Excellent*).

MacCallum, Widaman, Zhang, and Hong (1999) evaluated the adequacy of these sample size heuristics and found that virtually all were ineffective; sample size needs for EFA are a product of a number of data-related conditions. Specifically, fewer participants are needed for EFA when communalities—the proportion of variance for a variable explained by the selected factor solution—for items are high, and only a few factors, defined by three or more strongly loading items, are extracted (MacCallum et al., 1999). Conversely, when communalities are low, and/or many factors with only a few low-loading items are extracted, sample size needs are much higher. Adequate sample sizes can span the range from as few as 50 participants to as many as 1,000, depending on these conditions. We therefore recommend that researchers plan for moderately optimal data conditions—requiring approximately 200 to 250 participants for an adequate sample size—and adjust the sample size accordingly depending on the actual observed nature of the data and factor solution (Fabrigar & Wegener, 2012; MacCallum et al., 1999).

Extraction. When extracting factors, there are a large number of parameters to estimate, including each variable's factor loading values, each variable's unique variance, and the correlation coefficients between factors. Though there are a number of different extraction options to choose from, they can be classified as conforming to either a common factor (CF) model (Thurstone, 1933), or a principal components analysis (PCA) model (Hotelling, 1933).

Common Factor Extraction. All CF extraction methods partition the variance of a given item into two sources: common and unique. Common variance is variance shared with other items included in the factor analysis (hence, common factor), whereas unique variance is variance that is attributable to idiosyncratic item characteristics as well as random and measurement error. Because common factors contain only shared item variance, they are error free, as all unique and error variance is left behind in unique factors.

Maximum likelihood (ML) is a popular CF extraction method, though the details of ML extraction are beyond the scope of this article (for reviews, see Fabrigar & Wegener, 2012; Jöreskog, 1977; and Lawley, 1940). ML extraction facilitates the estimation of standard errors for factor loadings and factor correlations, thereby enabling significance tests and confidence intervals to be calculated. Further, ML extraction enables the calculation of a χ^2 statistic that can be used to evaluate the fit of the extracted factor solution to the data.

Principal Components Analysis Extraction. The PCA extraction method, by comparison, does not distinguish between shared, unique, and error variance, or rather, may be thought as assuming that all the variance for a set of variables is explained by the selected set of components. One important consequence of this assumption is that principal components therefore contain both unique variance and, more troublingly, random and measurement error.

Extraction Best Practices. None of the other analytic options have generated as much spirited debate in the methodological literature as the choice between PCA and CF extraction methods (see Fabrigar & Wegener, 2012). We recommend that a CF method of factor extraction should always be used when researchers are interested in distilling a set of items into a smaller number of theoretically meaningful factors. This is because, as discussed, common factors contain only shared construct variance and are error free. Components, by comparison, are contaminated by unique and error variance alike. And when researchers eventually test their new factor structures via CFA (for a review, see Brown, 2006), it is essentially a CF method of extraction that is used, not a PCA (for a comparison of PCA, CF, and CFA models, see Figure 1). Further, PCA tends to overestimate the value of factor loadings, relative to CF approaches (Snook & Gorsuch, 1989) and underestimate the size of correlations among factors (Widaman, 1993). By using the PCA approach, researchers will therefore be

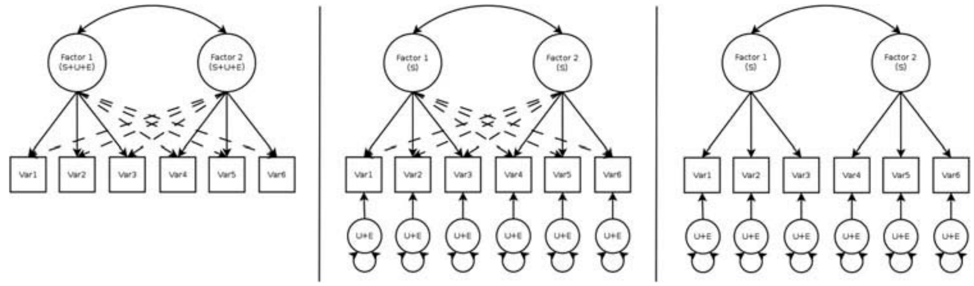

Figure 1. Differences between a principal components model, exploratory (common) factor analysis model, and confirmatory factor analysis model. S = shared variance; U = unique variance; E = error variance. A principal components model does not distinguish between shared, unique, and error variance, and a value is estimated for the loading of each item onto each component. An exploratory (common) factor analysis model partitions error and unique variance so that factors are comprised solely of shared variance, and a value is estimated for the loading of each item onto each factor. A confirmatory factor analysis model partitions error and unique variance so that factors are comprised solely of shared variance, and values are selectively estimated for the loading of particular items onto particular factors, on the basis of prior theory and research.

misled to keep variables in their factors that are not good indicators of the latent variables they are interested in studying and will miss out on associations between theoretically important variables. Though there are circumstances in which PCA and CF extraction methods will produce similar estimates, these conditions are rare in psychological research (Preacher & MacCallum, 2003; Widaman, 1993).

We further recommend that researchers use ML extraction whenever possible. When using ML extraction, researchers should first evaluate the assumption of multivariate normality by examining normality-related descriptive statistics (i.e., skewness and kurtosis) for each variable in the factor analysis and ensuring that the association between each pair of variables in the factor analysis is linearly related and bivariate normal (Kline, 2011). There are two advantages of ML extraction. First, ML extraction allows the χ^2 statistic to be computed, which is useful for making factor retention decisions. Second, ML extraction enables the calculation of standard errors for factor loadings and correlations (see Cudeck & O'Dell, 1994). This allows researchers to weigh the results of significance tests of factor loadings and correlations in determining which are important, as opposed to relying on arbitrary guidelines for interpreting these values (Tabachnick & Fidell, 2012). There are conditions, however, under which ML extraction is suboptimal. Namely, when assumptions for ML extraction are violated, its use is best avoided, though it is a process that is quite robust to even moderate violations of its assumptions (Fabrigar & Wegener, 2012; for suggested skewness and kurtosis limits, see Cohen, Cohen, West, & Aiken, 2002). When ML assumptions for normality are violated severely, however, other CF methods like principal axis factoring should be used.

Rotation. One of the most complicated issues in EFA is the problem of rotational indeterminacy. Simply stated, if more than one factor is extracted for an EFA solution, there is an infinite number of equally fitting ways those factors may be oriented in multidimensional space—a sexuality researcher using EFA must select one among the infinite possibilities by choosing from among a more limited number of empirical criteria to establish a desirable factor orientation.

Thurstone (1947) encouraged pursuing simple structure in factor solutions—solutions that would be the most interpretable and likely to replicate—as a way to choose between the endless orientations of factors. As explained by Fabrigar and Wegener (2012), simple structure essentially translates to a factor solution in which (a) each factor is represented by a subset of the analyzed items; (b) items representing differing factors should not overlap to a great extent; and (c) each item should be influenced by a limited number of factors. Researchers commonly misunderstand the pursuit of simple structure to require that items load substantially onto only *one* factor, but it is required only that items do not load substantially onto *all* factors.

Researchers can computationally rotate a factor solution in an attempt to achieve simple structure. Algorithms for rotation are numerous but can be classified as conforming to an orthogonal or oblique method of factor rotation, with each attempting to maximize different properties of the matrix of factor loadings in the pursuit of simple structure (see Browne, 2001).

Orthogonal versus Oblique Rotation. The crucial distinction between orthogonal and oblique methods of rotation is that orthogonal rotation assumes that extracted factors are uncorrelated, whereas oblique rotation does not invoke this assumption. Oblique rotation methods therefore enable researchers to estimate correlations between rotated factors, whereas orthogonal rotation methods do not. Among orthogonal rotation types, varimax rotation (Kaiser, 1958) is the default—and therefore most popular—orthogonal rotation method of many statistics programs. When interpreting an orthogonally rotated solution, researchers should examine the loading matrix. Oblique rotation methods, alternatively, include promax (Hendrickson & White, 1964), direct quartimin (Jennrich & Sampson, 1966), and geomin (Yates, 1987), among others. When utilizing an oblique rotation method, researchers should interpret the pattern matrix.

The distinction between orthogonal and oblique rotation has, unfortunately, become the source of a number of common misconceptions (Fabrigar & Wegener, 2012; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Preacher & MacCallum, 2003). Orthogonal rotation does not cause factors to become uncorrelated, nor does oblique rotation cause factors to become correlated. The two methods of rotation simply differ with regard to what they *assume* (in the case of orthogonal rotation) or *do not assume* (in the case of oblique rotation) of the correlations between factors. In fact, in the absence of any true correlations between factors, oblique rotation will essentially produce an orthogonally rotated solution (Floyd & Widaman, 1995).

Further, a correlated factor solution (i.e., obliquely rotated) is not, by definition, less fulfilling of simple structure than a factor solution that assumes no correlation (i.e., orthogonally rotated; Fabrigar & Wegener, 2012). When factors are truly correlated, oblique rotation will typically produce a greater level of simple structure than orthogonal rotation. Specifically, incorrectly assuming the factors are uncorrelated and using orthogonal rotation may artificially make items appear to be influenced by more than one factor (Osbourne, 2015). If researchers opt to retain items that load strongly onto only one factor, using orthogonal rotation methods when factors are correlated may therefore result in researchers mistakenly removing good indicators of the latent variables they are attempting to study. Indeed, in proposing simple structure, Thurstone (1947) lobbied for users of EFA to strongly consider using oblique methods of rotation, arguing the assumption of zero correlation between psychological factors to be highly unlikely.

Best Practices for Factor Rotation. Having previously clarified many of the common misconceptions about orthogonal versus oblique methods of factor rotation, it should come as no surprise that we strongly recommend sexuality researchers avoid the orthogonal rotation methods and instead use some form of oblique rotation when conducting EFA. Few psychological factors are truly and completely uncorrelated (an assumption of orthogonal rotation methods), and even if a set of factors is, in practice, uncorrelated, oblique methods of rotation will produce comparable factor loading estimates as orthogonal methods.

Retention. Once researchers have determined how they will extract and rotate their factors, they must decide how many factors to keep in their final solution. A number of criteria have been proposed to help guide researchers in selecting an appropriate number of factors to retain. These criteria can be grouped loosely into those that rely on interpreting the eigenvalues of factors and those that rely on the χ^2 (i.e., likelihood ratio) statistic.

Eigenvalue-Based Retention Criteria. Eigenvalues represent the amount of variance accounted for by a given factor, from a set of variables that have been factor analyzed. The number of calculated eigenvalues equals the number of variables in a factor analysis.

One of the most common eigenvalue-based methods of determining whether to retain a factor is the so-called Kaiser criterion, or eigenvalue-greater-than-one rule (Guttman, 1954; Kaiser, 1960). This guideline suggests—as its one name implies—that researchers should simply retain all factors with eigenvalues greater than one. The combined usage of PCA extraction, varimax rotation, and the eigenvalue-greater-than-one rule is occasionally referred to as "Little Jiffy" in the EFA literature (Kaiser, 1970; Preacher & MacCallum, 2003). Reviews of the use of EFA in other disciplines have consistently found Little Jiffy to be the most commonly selected combination of analytic options for EFA (Conway & Huffcutt, 2003; Fabrigar et al., 1999; Ford, MacCallum, & Tait, 1986; Henson & Roberts, 2006; Hetzel, 1996).

Another popular eigenvalue-based retention criterion, the scree test, involves interpreting a plot of eigenvalues (Cattell, 1966). To construct a scree plot, all eigenvalues are plotted in order of descending value. Researchers then attempt to discern where the scree—a term referencing rubble at the base of a hill—begins, or in other words, where within the plot the last major drop in eigenvalues occurs, after which the plotted eigenvalues are essentially uniform. Researchers are encouraged to retain the factors preceding the scree.

A final eigenvalue-based retention criterion is parallel analysis (Horn, 1965). As in the scree test, parallel analysis begins with the computation and plotting of eigenvalues. A random data set (or sets) is then simulated, containing the same number of items and of the same sample size. The same number of eigenvalues is then calculated and plotted for the random data set on the same scree plot (for an example macro for SAS and SPSS, see O'Connor, 2000). Researchers are encouraged to retain the number of factors that have eigenvalues from their real data that are larger than those from the randomly generated data set. The rationale is that factors should be retained only if they account for more meaningful variance than random statistical noise.

χ^2-Based Retention Criteria. ML extraction enables the calculation of the χ^2 statistic. The χ^2 statistic represents how well the extracted EFA model fits the observed data. Larger χ^2 values indicate worse fit of model to data, and increasing the number of factors retained will result in a lower χ^2 value (i.e., better fit). A significance test of the χ^2 statistic constitutes a χ^2 test of perfect model fit. The null hypothesis of this test is that the EFA model perfectly fits the data; a significant χ^2 test indicates that the model does not perfectly fit the data. Thus, researchers could extract factor solutions with an increasing numbers of factors retained, until they extract a solution with a sufficiently large enough number of factors such that the χ^2 test is nonsignificant.

Researchers can also use the χ^2 statistics and their associated degrees of freedom from two competing models to perform a nested model comparison. To conduct a nested model comparison, researchers subtract the smaller χ^2 statistic value (i.e., from the model with more factors) from the larger χ^2 statistic value (i.e., from the model with fewer factors). The same is done for the degrees of freedom from each model. The difference of χ^2 values is used as a new χ^2 statistic value, and the difference in degrees of freedom is used as that statistic's new degrees of freedom, thus enabling a significance test to be carried out. A significant comparison indicates that the model with the greater number of factors fits the data better than the model with fewer factors; a nonsignificant nested model comparison indicates that the model with the greater number of factors does not fit the data appreciably better than the model with fewer factors.

Finally, the χ^2 statistic facilitates the computation of other indices of model fit. For some indices, such as the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR), smaller values are indicative of better fit, whereas for others, like the comparative fit index (CFI) and the Tucker-Lewis index (TLI), larger values indicate better model fit (see Hu & Bentler, 1999). Researchers using EFA could therefore extract solutions varying in the number of factors retained and then select as their final model the solution with the fewest factors (i.e., the most parsimonious) that fits the data well.

Best Practices for Factor Retention. Though it is impossible to ever know, a priori, the "true" number of factors underlying a set of items, there are conceptual and practical consequences for extracting too few or too many factors to explain a set of items. Extracting too few factors (i.e., underfactoring) will result in sexuality researchers leaving theoretically meaningful item information on the proverbial table. Extracting too many factors (i.e., overfactoring), alternatively, will result in a less parsimonious representation of the data that will soak up valuable degrees of freedom and create problems of multicollinearity as predictors in subsequent analyses, or require additional analyses be conducted if factors are to be used as distinct outcomes. Although scholars may have slightly different recommendations, we are partial to guidelines offered by Fabrigar and Wegener (2012):

> Ideally, the number of major common factors is that number of common factors for a model in which: (1) the model does a good job accounting for the correlations among the measured variables, (2) a model with one fewer common factors would do substantially worse in accounting for the correlations, (3) a model with one more common factor would not do appreciably better in explaining the correlations, and (4) all common factors in the model are readily interpretable and can be related to constructs of theoretical utility to the domain of interest. (pp. 54–55)

Importantly, Fabrigar and Wegener's (2012) guidelines explicitly demarcate a major role for theory and interpretability to play in factor retention decisions. In other words, researchers should not use an entirely quantitative, atheoretical process when making factor retention decisions. Still, when using quantitative evidence to help inform factor retention decisions, sexuality researchers should ensure they are selecting criteria that are sufficiently rigorous.

Despite its popular use among sexuality researchers and other social scientists (e.g., Fabrigar et al., 1999), the eigenvalue-greater-than-one rule is a poor guide for factor retention (for reviews of simulations, see Fabrigar et al., 1999; Fabrigar & Wegener, 2012; Preacher & MacCallum, 2003). Specifically, it can lead to substantial underfactoring, or overfactoring, depending on certain features of the data, and conditions under which it will render accurate factor retention guidance are rare (Preacher & MacCallum, 2003). The Kaiser criterion is also an entirely arbitrary guideline; there is no clear conceptual rationale why a factor with an eigenvalue of 1.02, for example, should be considered conceptually superior to a factor with an eigenvalue of 0.97. As such, we encourage sexuality researchers to avoid its use in making factor retention decisions.

The scree test, similarly, is not an entirely objective test. In some EFA circumstances it may be clear to discern where meaningful factors end and the scree begin (Tucker, Koopman, & Linn, 1969). In other circumstances, however, a scree plot may not exhibit a decisive drop in factor eigenvalues, thereby resulting in a more ambiguous interpretation. As such, we strongly recommend that sexuality researchers supplement the scree test with parallel analysis.

Parallel analysis is a relatively accurate method of determining factor retention, though in some contexts it is prone to overfactoring (Fabrigar & Wegener, 2012). Sexuality researchers should therefore use parallel analysis as a way of establishing a probable number of maximum factors and then use other factor retention criteria (e.g., indices of model fit, interpretability, theory) to settle on a final factor solution. O'Connor (2000) has provided SPSS and SAS syntax for conducting parallel analysis, and parallel analysis can also be conducted in many other statistical programs (e.g., Mplus, R) with only a singleton line of code (see supplemental materials).

Although researchers could use the χ^2 test of perfect model fit itself as a criterion for model retention, we do not recommend it. Indeed, many statisticians have cautioned against putting much stock in the χ^2 test, because its null hypothesis (a perfectly fitting model) is conceptually untenable, and because it is extremely sensitive to sample size (Brown, 2006)—as samples get larger, it is more likely to reject models with trivial amounts of misspecification, and when sample sizes are superficially small, it is extremely unlikely to reject poorly specified models.

Unlike the χ^2 statistic, nested-model comparisons and other indices of model fit (e.g., RMSEA, CFI, TLI)—all of which rely on the initial computation of the χ^2 statistic—can serve as effective guides for factor retention. We therefore strongly recommend the use of nested-model comparisons and descriptive indices of model fit when making factor retention decisions for solutions extracted via maximum likelihood.

Summary of Best Practices for Efa

Our review has highlighted several of the analytic choices researchers face when conducting EFA, and we have provided recommended best practices for these choices based on empirical evidence. Specifically, we encourage researchers to anticipate a sample size of at least 200 to 250 participants, to use a common factor extraction method (preferably ML), to use an oblique method of rotation, and to use more rigorous empirical criteria, alongside theory and interpretability, when making factor retention decisions (i.e., parallel analysis, indices of model fit, and nested-model comparisons). However, an important methodological question remains: How frequently have sexuality researchers used these practices and others when conducting their EFAs?

Reviewing the Efa Practices Used by Sexuality Researchers

We conducted a methodological review of empirical articles using EFA published in sexuality research outlets (for similar methodological reviews of EFAs conducted in other disciplines, see Conway & Huffcutt, 2003; Fabrigar et al., 1999; Ford et al., 1986; Henson & Roberts, 2006; Hetzel, 1996). Specifically, we reviewed the frequency with which sexuality researchers used each option for extraction, rotation, and retention when conducting EFAs.

Method

Article Search Strategy. We reviewed articles using EFA from 1974 to 2014 in several prominent sexuality journals: *Archives of Sexual Behavior, International Journal of Sexual Health, Journal of Sex Research, Journal of Sex and Marital Therapy, Journal of Sexual Medicine*, and *Sex Roles*. Search terms for these journals included "factor anal*", "principle component", and "principal component" anywhere in the article. ("Principal" component is the correct spelling, but "principle" component was more commonly used.) This search yielded more than 600 articles. Given its popularity among researchers, we also reviewed scales published in the *Handbook of Sexuality-Related Measures* (Fisher, Davis, Yarber, & Davis, 2011), which includes more than 200 scales. Journal articles that did not conduct an original EFA and *Handbook* entries that did not make explicit mention of using EFA were subsequently excluded. This search strategy yielded a total of 139 journal articles and 24 *Handbook* entries that entailed a total of 216 EFAs (see online supplementary materials for the complete reference list for this review).

Coding of Articles. A team of four research assistants, who attended several training sessions teaching them about EFA, coded articles. The research assistants coded with respect to several characteristics, including (a) sample size, (b) factor extraction method, (c) rotation method, and (d) factor retention criteria. The second author coded articles that research assistants highlighted as particularly challenging to code (n = 26). The first author then checked a random 10% of entries for coding. This check revealed a high degree of coding agreement across extraction, rotation, and retention categories (average Cohen's κ = .88).

Results

Our findings are summarized in Figure 2. The median sample size used for EFAs conducted by sexuality researchers was 266.00 (min = 35, max = 11,543). PCA was clearly the most widely used factor extraction method among sexuality researchers, followed by non-ML-based CF methods, and finally ML extraction. One factor analysis was unclassifiable, as the authors described the impossible hybridization of PCA and CF methods. Orthogonal rotation methods were nearly twice as common as oblique methods, and three articles reported single-factor solutions that did not require rotation. Finally, in terms of factor retention criteria, more than half of sexuality researchers reported retaining factors with eigenvalues greater than one, followed by interpreting a scree plot. Other factor retention criteria were much less common.

Discussion

Though relatively straightforward, our methodological review revealed that most sexuality researchers relied on the combination of PCA extraction, orthogonal rotation, and the eigenvalue-greater-than-one guideline (i.e., Little Jiffy, Kaiser, 1970) when conducting EFAs. In other words, most sexuality researchers have not been using best practices for EFA. Another troubling trend worth noting is failure to report EFA practices; sexuality researchers did not discuss the extraction, rotation, or retention methods they used in 11.10% to 26.90% of EFAs that we reviewed. Though there is some room for disagreement regarding the best practices for EFA, we would argue that transparent reporting of analytical decisions—especially for an analysis as flexible as EFA—is essential for replicable sexual science.

Resources to Increase Adoption of Best Practices for Efa

Based on this review it is clear that, despite the popularity of EFA in sexuality research, many sexuality researchers are not using best practices for EFA. As other review papers —some of which are very highly cited—detailing many of the best practices we have discussed have been available for some time (e.g., Fabrigar et al., 1999), it is worth considering why these recommendations for EFA have not gained greater traction among sexuality researchers.

One possibility is that many of the previous reviews of best practices for EFA have been published in psychology methods journals (e.g., Costello & Osborne, 2005; Fabrigar

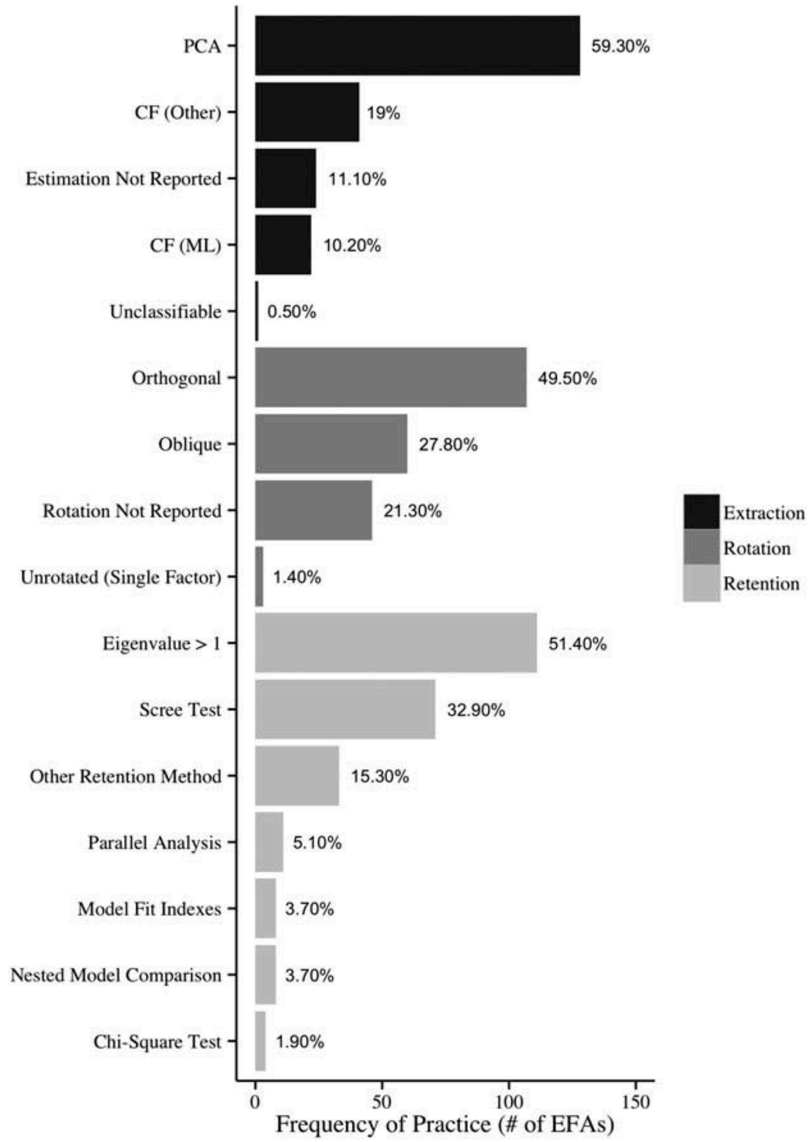

Figure 2. Frequencies and percentages of EFA practices in sexuality research articles. PCA = principal components analysis; CF = common factor extraction; ML = maximum likelihood estimation.

et al., 1999; Preacher & MacCallum, 2003). As psychology is but one discipline that contributes to the interdisciplinary composition of sexuality research, it is probable that many sexuality researchers do not rely on the psychology-based scholarly outlets in which best practices for EFA have been discussed. By publishing the present review in a sexuality research journal, we therefore hope this potential barrier to greater adoption of best practices for EFA can be reduced.

Another possibility, however, is that even if sexuality researchers are familiar with some best practices, they may not know how to enact them in their own research. To address this possibility, we have provided a group of EFA resources in our online supplementary materials. Included in this resource set are (a) syntax for conducting EFA with ML-based extraction, oblique rotation, and parallel analysis, in four different statistical analysis software packages (Mplus, R, SAS, and SPSS); (b) annotated output from each of the four statistical packages that is produced when our example syntax is run on our example data set; (c) a simple spreadsheet that, if ML-based extraction is used, calculates an index of model fit (RMSEA) and conducts nested model comparisons; and (d) an example methods and results section.

Conclusion

When best practices for EFA are not followed, sexuality researchers will be more likely to retain bad indicators and remove good indicators of the factors they are attempting to study, overlook correlations between factors, and make inferential errors during data analysis. To summarize our recommendations for authors of sexuality research articles using EFA, we would suggest that researchers (a) plan for at least a sample of 200 to 300 participants (MacCallum et al., 1999);

(b) use a common factor method of factor extraction, preferably ML (when data meet the required assumptions) to facilitate the computation of significance tests for confidence intervals for factor loadings and correlations, as well as indices of model fit; (c) use an oblique method of rotation; and (d) do not rely on arbitrary and subjective methods of determining factor retention, like the eigenvalue-greater-than-one rule and scree test. All EFAs should, at minimum, entail parallel analysis, and when possible (i.e., if ML extraction was used) the use of nested-model comparisons and/or the interpretation of descriptive indices of model fit. Further, sexuality researchers should be aware that two of the most popular statistical software packages, SAS and SPSS, default to the ill-advised Little Jiffy set of options (i.e., PCA extraction, orthogonal rotation, eigenvalue-greater-than-one retention).

We also encourage reviewers and editors to enforce these best practices when evaluating manuscripts using EFA. This is not to say that reviewers should universally recommend rejection of any article in which an EFA was conducted using PCA extraction, or orthogonal rotation, and so on. Rather, authors should be expected to provide a compelling statistical and theoretical rationale for why they are departing from best practices. Reviewers, however, will need to be vigilant in resisting previous precedent as adequate justification for deviation from best practices. In other words, as our review illustrates, many sexuality researchers have not been following best practices for EFA, but that is hardly justification to continue this practice.

The exploration of new theoretical constructs will almost certainly remain a cornerstone of sexuality research, and EFA will continue to provide sexuality researchers with a way of rigorously pursuing this goal. Even so, as our review has revealed, sexuality researchers have, thus far, typically not employed the best practices for conducting EFA in terms of factor extraction, rotation, and retention methods. With this article, we hope to aid the continuing tradition of scientific exploration in sexuality research by having provided more accessible explanations of extraction, rotation, and retention methods, discussing which constitute best practice, and creating analytic resources to share online, to make it easier for sexuality researchers to understand and employ best practices for EFA in their own research.

Funding

This work has been supported by a Social Science and Humanities Research Council (SSHRC) Postdoctoral Fellowship awarded to John Sakaluk.

Supplemental Material

Supplemental materials, including syntax for EFA, annotated output, the spreadsheet calculator, and example writeup, can be accessed here.

Acknowledgments

The authors would like to thank Amanda Gesselman, Robin Milhausen, Heath Schechinger, and Jessica Wood for their thoughtful feedback on earlier drafts of this article, as well as Alex Latham, Amanda Rixey, Blake Benson, and Kayla Mitchell for their valuable help as undergraduate research assistants for this project.

References

Bandalos, D. L. (2006). The role of simulation in structural equation modeling. In G. R. Hancock & R. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 385–426). Greenwich, CT: Information Age.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

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

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276. doi:10.1207/s15327906mbr0102_10

Cattell, R. B. (1978). *The scientific use of factor analysis*. New York, NY: Plenum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, *6*, 147–168. doi:10.1177/1094428103251541

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*, 1–8.

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*, 475–487. doi:10.1037/0033-2909.115.3.475

Everitt, B. S. (1975). Multivariate analysis: The need for data and other problems. *British Journal of Psychiatry*, *126*, 237–240. doi:10.1192/bjp.126.3.237

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. New York, NY: Oxford University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299. doi:10.1037/1082-989X.4.3.272

Fisher, T. D., Davis, C. M., Yarber, W. L., & Davis, S. L. (2011). *Handbook of sexuality-related measures* (3rd ed.). New York, NY: Routledge.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*, 286–299. doi:10.1037/1040-3590.7.3.286

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, *39*, 291–314. doi:10.1111/j.1744-6570.1986.tb00583.x

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*, 149–161. doi:10.1007/BF02289162

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17*, 65–70. doi:10.1111/j.2044-8317.1964.tb00244.x

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*, 393–416. doi:10.1177/0013164405282485

Hetzel, R. D. (1996). A primer on factor analysis with comments on patterns of practice and reporting. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 175–206). Greenwich, CT: JAI.

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

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*, 417–441. doi:10.1037/h0071325

Hu, L., & 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*, 1–55. doi:10.1080/10705519909540118

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, *31*, 313–323. doi:10.1007/BF02289465

Jöreskog, K. G. (1977). Factor analysis by least-squares and maximum-likelihood methods. In K. Enslein, A. Ralston, & H. S. Wilf (Eds.), *Statistical methods for digital computers* (pp. 125–153). New York, NY: Wiley.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200. doi:10.1007/BF02289233

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

Kaiser, H. F. (1970). A second generation Little Jiffy. *Psychometrika*, *35*, 401–415. doi:10.1007/BF02291817

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, Y: Guilford Press.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, *60A*, 64–82.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84–99. doi:10.1037/1082-989X.4.1.84

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers*, *32*, 396–402. doi:10.3758/BF03200807

Osbourne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research, and Evaluation*, *20*(2), 1–7.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*, 13–43. doi:10.1207/S15328031US0201_02

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, *106*, 148–154. doi:10.1037/0033-2909.106.1.148

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Upper Saddle River, NJ: Pearson.

Thurstone, L. L. (1933). *The theory of multiple factors*. Chicago, IL: University of Chicago Bookstore.

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

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*, 421–459. doi:10.1007/BF02290601

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*, 263–311. doi:10.1207/s15327906mbr2803_1

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