Journal of Business Research 164 (2023) 113969

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

A systematic literature review of exploratory factor analyses in management☆

Matt C. Howard *

The University of South Alabama, Mitchell College of Business, USA

ARTICLE INFO

Keywords:
Exploratory factor analysis
Factor analysis
One-factor test
Principal components analysis
Confirmatory factor analysis
Factor rotation

ABSTRACT

We review the application and reporting of exploratory factor analysis (EFA) in management. First, we integrate recommendations from relevant reviews and simulation studies to provide modern guidelines regarding EFA for psychometric investigation, and we highlight severe concerns associated with EFA for Harman's one-factor test. Second, we conduct a systematic literature review including 1,790 articles and 3,396 EFAs from 23 premier journals since the year 2000. Our review reveals that many widespread guidelines for EFA are infrequently applied. We also show that researchers regularly fail to report essential aspects of their EFAs, leaving it unknown whether they are correctly conducting their analyses. Third, our discussion reinforces modern recommendations that are infrequently applied, exposes researchers to emerging developments, and provides new factor loading interpretations based on our systematic review results. To conclude, we provide step-by-step visual guides and a checklist to aid future researchers on their application and reporting of EFA.

Exploratory factor analysis (EFA) is a powerful statistical technique that enables researchers to use their judgement and interpretation to identify a set of latent factors that meaningfully and parsimoniously represent a set of indicators (Goretzko et al., 2021; Hair et al., 2019; Howard, 2016; Watkins, 2018). The technique estimates the number of latent factors underlying the indicators as well as the association of each indicator to each latent factor, which is known as factor loadings. Researchers can interpret the conceptual meaning of the emergent factors by qualitatively assessing the content of indicators that load strongly onto the factors. Researchers can likewise assign indicators to latent factors based on their factor loadings and identify indicators with problematic properties, such as failing to substantially load onto any factor or substantially loading onto multiple factors (i.e., cross-loadings). Given the important information obtained from EFA, it is among the most common statistical techniques applied in the field of management (Conway & Huffcutt, 2003; Fabrigar et al., 1999; Ford et al., 1986).

Conducting EFA is not a straightforward process, and researchers must make several analytical decisions based on the attributes of their research questions and data (Luo et al., 2019; Peterson, 2000; Steiner & Grieder, 2020; Watkins, 2021). Many of these decisions are not made on purely objective criteria, and new statistical advances regularly produce updated guidelines for determining the most appropriate EFA for a given context (Howard & Henderson, 2023; Ledesma et al., 2021; Osborne, 2015; Reio & Shuck, 2015; Sakaluk & Short, 2017; Watkins, 2018). For these reasons, it is imperative to routinely assess the use of EFA to ensure that it is being applied in accordance with modern recommendations. Despite its popularity, however, considerable time has elapsed since the last reviews focused on EFA in management (Conway & Huffcutt, 2003; Fabrigar et al., 1999; Ford et al., 1986). The typical management researcher may be unaware of modern recommendations, and they may commonly apply analytical decisions that have been denounced for decades. These practices may routinely produce inappropriate and misleading theoretical insights that obfuscate the true nature of phenomenon, such as identifying an incorrect number of latent factors represented by a measure. In the current article, we rectify these concerns by reviewing the use and reporting of EFA in management, which can update researchers' toolkits and curb the proliferation of improper analytical practices – ultimately resulting in a field of study with more accurate empirical research and theoretical insights.

Our first goal is to summarize modern guidelines for EFA by integrating relevant reviews and simulation studies. We emphasize the

☆ I would like to thank Jennifer Henderson, Justin Scott, Emory Serviss, and Elise Van Zandt for their assistance in the coding of studies. I would also like to thank Mickey Smith, Kayla Follmer, I-Heng (Ray) Wu, and Matthew Crayne for their valuable feedback on prior versions of the present work.

* Address: 5811 USA Drive S., Rm. 337, Mitchell College of Business, University of South Alabama, Mobile, AL, 36688, USA.
E-mail address: MHoward@SouthAlabama.edu.

https://doi.org/10.1016/j.jbusres.2023.113969
Received 6 June 2022; Received in revised form 30 March 2023; Accepted 17 April 2023
Available online 5 May 2023
0148-2963/

importance of proper EFA procedures by discussing a research topic for which questionable EFA practices hampered theoretical development, work engagement (Mills et al., 2012), and specifying that confirmatory factor analysis (CFA) is not a remedy for poor EFA practices (Brown, 2015). We then provide guidelines regarding EFA for psychometric investigation, which we separate into five steps: (1) data quality checks, (2) factor extraction method, (3) factor retention approach, (4) factor rotation method, and (5) interpreting factor loadings. We also discuss Harman's one-factor test. We stress concerns associated with this application of EFA, and we highlight that the analysis often provides support for problematic research designs (Fuller et al., 2016; Malhotra et al., 2006; Podsakoff et al., 2012).

Our second goal is to provide a comprehensive overview of modern applications of EFA by conducting a systematic literature review. Our review includes 1,790 articles and 3,396 EFAs from 23 premier journals since 2000. We show that widespread guidelines for EFA are infrequently applied, indicating that a significant gap exists between articles on EFA and articles applying EFA. We also show that most researchers fail to report essential aspects of their EFAs, leaving it unknown whether they are correctly conducting their analyses. Such concerns are troubling considering efforts to encourage open science practices and enhance replicability (OSC, 2015, 2017), as most reported EFAs cannot be reproduced. We lastly find that *no author* has found common-method bias to be a concern when applying Harman's one-factor test.

Our third goal is to reinforce modern recommendations that are infrequently applied, provide new recommendations for steps without clear guidance, and expose researchers to recent developments (Cosemans et al., 2022; Iacobucci et al., 2022; Montoya & Edwards, 2021; Zhang et al., 2019). These recent developments have yet to be widely adopted by management researchers (e.g., comparison data method; Auerswald & Moshagen, 2019), and we discuss how these developments can improve the sophistication and accuracy of analyses. To conclude, we provide a checklist to guide future researchers on the use of EFA, and we call for authors to no longer apply Harman's one-factor test without further investigation on the approach.

The current article provides several benefits for research and practice, and we presently highlight four. First, we provide evidence that misinterpretations of constructs may be common due to substandard EFA practices. Future researchers should improve current practices using our recommendations, but they should also reinvestigate measures that were originally tested with inappropriate EFA methods. Our coding database in the supplemental materials identifies measures created via particularly problematic approaches, enabling focused investigations that may produce new interpretations of constructs. Second, we review recent developments in EFA, and we offer a new development by reinterpreting factor loading cutoffs. Based on our systematic literature review results, we provide new factor loading guidelines that move beyond dichotomous assessments, and they instead enable researchers to determine more conservative or liberal cutoffs based on the nature of their studied construct(s). Third, we encourage researchers to abandon Harman's one-factor test and reconsider assessments of common-method variance, emphasizing the need to both investigate and use more appropriate assessments. Fourth, EFA is regularly used in applied contexts, such as the creation of selection and assessment procedures (Ellingwood et al., 2020; Schmit & Ryan, 1993). The current article can provide guidance for these endeavors, resulting in better organizational functioning and immediate practical impacts.

1. Recommended standards for conducting exploratory factor analysis

1.1. Exploratory factor analysis for psychometric investigation

Incorrectly performed EFAs can hamper theoretical development, which is particularly evident in research on the Utrecht Work Engagement Scale (UWES) (Kulikowski, 2017; Saks & Gruman, 2014). The three-dimension UWES was developed via a two-study process using CFA (Schaufeli et al., 2002), but several subsequent authors have reanalyzed the UWES with EFA perhaps due to concerns that the use of CFA was premature (discussed below) (e.g., Sonnentag, 2003). These authors regularly failed to produce a three-factor solution and/or found poor factor loadings, and the factor structures from these EFAs showed little agreement. Mills et al. (2012) argued that, "reasons for why exploratory factor analytic solutions such as those mentioned above cannot be replicated include inappropriate factor extractions criteria, incorrect rotation method, and the use of principal components analysis instead of exploratory factor analysis" (p. 521). In their EFA using recommended approaches, Mills et al. (2012) identified a fourth factor outside the scope of work engagement, indicating construct contamination. Recent authors use the full UWES less often, devoting increased attention to a shortened version without items that load onto this fourth dimension (Fong & Ho, 2015; Kulikowski, 2017; Willmer et al., 2019).

Incorrect uses of EFA produced considerable confusion in the work engagement literature, causing reviews of work engagement to dedicate significant space to psychometric investigations of the UWES (Cole et al., 2012; Saks & Gruman, 2014). Kulikowski (2017) even reported a focused review solely on these studies. Such distractions delayed theoretical insights into work engagement, and some authors question whether studies using the UWES need to be reconducted with alternative measures (Cole et al., 2012; Saks & Gruman, 2014). These difficulties in the work engagement literature stress the importance of reviewing recommended standards of EFA, such that similar concerns can be avoided in other research domains.

Further, EFA and CFA should be differentiated, as they are often complementary applied. EFA identifies latent factors and their indicator relations without an a priori model, whereas CFA tests whether an a priori model represents underlying latent factors and their indicator relations. EFA is believed to be more apt at detecting unanticipated latent factors and factor loadings, whereas CFA is believed to be more apt at testing the validity of specific models (Brown, 2015). The difference in model testing causes EFA and CFA to be applied in different circumstances.

EFA is usually applied in earlier phases of scale development to identify latent factors and detect poor indicator loadings, whereas CFA is applied in later phases with a separate sample to ensure that EFA results were not idiosyncratic to the original sample (Hinkin, 1995, 1998). Either analysis is used to support that applied measures are appropriate representations of constructs when investigating independent research questions (i.e., not scale development), and the choice of analysis is typically determined by the extent of prior psychometric investigation on the measure. EFA is commonly used when a measure is created for the purposes of the study (whether from scratch or adapted), even if the researcher has expectations for the factor structure. It is also common to conduct EFAs on existing measures with unclear factor structures, such as the UWES example above. While rare, some researchers use EFA to jointly assess all applied measures, even if all have been supported in prior research (Flouri et al., 2015; Guo et al., 2017; Truninger et al., 2020). Unexpected factors and indicator loadings may occur, but CFA is more commonly applied to evaluate established measures. Thus, the decision to apply EFA or CFA is "shades of grey" rather than "black and white" (see Hurley et al., 1997)[1], but it is preferred to apply EFA when factor structures have yet to be tested or psychometric properties are unclear.

Researchers may believe that inappropriate inferences from

[1] Researchers may be in the later phases of construct validation and perform a CFA, only to discover poor fit and/or multiple models with indistinguishable fit. In such cases, the construct validation process may not be linear. It is acceptable to perform an EFA, utilizing the EFA results in a similar manner to modification indices. The researcher should report their entire process (Greene et al., 2022), recognizing that the CFA is no longer purely confirmatory.

incorrectly performed EFAs are addressed by CFA. This notion is untrue. Some authors conduct EFA and CFA on the same sample, but CFA provide limited information beyond EFA in such instances. Factor structures derived from EFA can produce exceptional model fit via CFA when using the same sample due to overfitting, such that misleading estimates are produced by spurious effects that do not reflect the population (Fokkema & Greiff, 2017). Improper EFAs can also produce incorrect models that adequately reproduce covariances, and CFAs testing such models on the same sample can produce acceptable results (Marsh et al., 2004). While it does provide more assurances, CFA with separate samples does not guarantee correct factor solutions. If two samples are taken from the same source, biases arising from that source can produce misleading EFA and CFA estimates (Hurley et al., 1997). Even when tested on alternative samples, incorrect CFA models derived from improperly performed EFAs can still produce adequate results (Marsh et al., 2004).

It should also be noted that researchers most often utilize EFA results to determine which indicators should represent their construct(s), and researchers fully report their EFA methods and results when doing so. Averages of these indicators are then used in subsequent analyses. Some researchers apply EFA to solely calculate factor scores to use in subsequent analyses. In doing so, they often do not report their EFA methods and results, as they are only using the analysis as a data reduction technique. For this reason, the current review discusses EFAs for psychometric investigations rather than data reduction due to the differences in expected reporting.

1.2. Five-step process of performing exploratory factor analysis for psychometric investigations

Recommendations for analytical and interpretation decisions when conducting EFA for psychometric investigations can be summarized as a five-step process consisting of (1) data quality checks and sample size, (2) factor extraction method, (3) factor retention approach, (4) factor rotation method, and (5) interpreting factor loadings (Conway & Huffcutt, 2003; Costello & Osborne, 2005; Howard, 2016; Watkins, 2018). We review these five steps below.

1.2.1. Data quality checks and sample size

Before conducting their EFA, researchers should consider their data quality checks and sample size. Typical data quality should be applied checks when performing EFA (e.g., missing data), but two data quality checks are specific to EFA: Bartlett's test of sphericity (Bartlett, 1951) and Kaiser-Meyer-Olkin (KMO) test for sampling adequacy (Kaiser, 1970). Indicators with insufficient common variance can produce factor structures that appear accurate but explain little regarding the indicators (Fokkema & Greiff, 2017), indicating that seemingly appropriate factors have little theoretical importance. Bartlett's test assesses the similarity of indicators' correlations to an identity matrix, which is a correlation matrix entirely of null values on the off-diagonals. A correlation matrix that is not significantly different from an identity matrix suggests that the indicators do not share sufficient covariance to conduct an EFA, and a non-significant Bartlett's test result signifies that EFA should not be performed (Bartlett, 1951; Dziuban & Shirkey, 1974). Alternatively, KMO's test assesses whether sufficient common variance is shared among indicators by comparing the correlations of each indicator to the partial correlations of each indicator when controlling for all other indicators (Beavers et al., 2013; Kaiser, 1970). Values closer to 0 indicate less common variance is shared between indicators, whereas values closer to 1 indicate greater common variance is shared. Values below 0.50 indicate that EFA should not be performed, values between 0.50 and 0.60 draw concerns, and values above 0.60 support the use of EFA (Kaiser, 1974).

Furthermore, small sample sizes in EFA can cause the misidentification of factors and factor loadings (Hirschfeld et al., 2014). Popular sample size guidelines can be categorized as absolute (e.g., >300 participants) and participant-to-indicator ratio (e.g., >10 participants per indicator) (Anthoine et al., 2014), but studies have failed to support absolute and participant-to-indicator ratio guidelines. Rouquette and Falissard (2011) conducted a series of simulations that showed "short scales do not allow smaller sample size" and sample sizes "should be increased when the number of factors within the scale is large" (p. 235). These authors suggest that sample sizes should be based on the number of indicators and extracted factors, which is reflective of common variance. MacCallum et al. (1999) supported via simulations that,

> "Our theoretical framework and results show clearly that common rules of thumb regarding sample size in factor analysis are not valid or useful. The minimum level of *N*, or the minimum *N:p* ratio, needed to assure good recovery of population factors is not constant across studies but rather dependent on some aspects of the variables and design in a given study. Most importantly, level of communality plays a critical role" (p. 96).

From these empirical results and expressions of concern, it should be recognized that absolute and participant-to-indicator ratio guidelines often recommend smaller sample sizes than those supported by simulation studies, and sample size requirements depend on common variance.

Common variance can be observed via communalities, factor loadings, and other aspects (Kahn, 2006). Unfortunately, it is difficult to anticipate these aspects before conducting EFA. Authors have suggested sample size guidelines based on expected indicator-to-factor ratios (Hirschfeld et al., 2014; MacCallum et al., 1999; Rouquette & Falissard, 2011), but this suggestion has yet to be widely adopted. Perhaps the most straightforward and empirically supported indicator-to-factor ratio guidelines for sample size are those of Rouquette and Falissard (2011), who conducted a simulation study to produce sample size tables. In the current article, we recommend the sample size guidelines of Rouquette and Falissard (2011), as they are more closely based on aspects of EFA that determine sample size requirements.

1.2.2. Factor extraction method

Three factor extraction methods are currently dominant in research. Principal axis factoring (PAF) identifies factors to best account for common variance and represent the shared underlying structure of indicators (Jain & Shandliya, 2013; Matsunaga, 2010). PAF does not assume multivariate normality, and it can provide accurate results under a wider range of circumstances than many other factor extraction methods (De Winter & Dodou, 2012; Fabrigar et al., 1999). PAF is also less likely than other factor extraction methods to produce Heywood cases[2] or fail to converge, both of which prevent the interpretation of factor solutions (Finch & West, 1997). Due in part to its relative resilience, prior empirical studies and reviews of other research domains have repeatedly endorsed its application in favor of other techniques (Beavers et al., 2013; Costello & Osborne, 2005; Howard, 2016; Watkins, 2018).

Maximum likelihood (ML) also identifies a set of factors that explains common variance (Conway & Huffcutt, 2003; Costello & Osborne, 2005), and it is often assessed by its relative benefits and detriments compared to PAF. A benefit of ML is the provision of fit indices that are not natively provided by PAF, both indices of absolute (e.g., SRMR and CFI) and relative fit (e.g., AIC and BIC). These fit indices can be used to identify the ideal number of factors to retain, but a recent study draws this benefit into question by showing that comparing fit indices may produce inaccurate EFA solutions (Montoya & Edwards, 2021). On the other hand, ML has stronger assumptions regarding normality, which can produce inaccurate results when deviations from normality occur

[2] Heywood cases are impermissible values in which the residual variance of an indicator is zero or negative because the amount of its explained variance (e.g., commonality) is 1.00 or greater. Readers should refer to Cooperman and Waller (2022) and Lorenzo-Seva and Ferrando (2021) for information on how to address Heywood cases.

(De Winter & Dodou, 2012; Fabrigar et al., 1999).

Principal components analysis (PCA) differs from PAF and ML. PCA identifies a set of components that account for the total variances of indicators, which contains both common and unique variance (including error) (Hair et al., 2019; Park et al., 2002). Most researchers consider factor analytic methods to identify latent factors that account for common variance in indicators (e.g., PAF and ML), and PCA is often not considered a factor analytic method because the resultant components do not solely reflect common variance (Jain & Shandliya, 2013; Velicer & Jackson, 1990). For this reason, PAF and ML are factor-based approaches (e.g., factor analysis), whereas PCA is a components-based approach (e.g., components analysis). Most frequently, these approaches produce similar results. As Velicer and Jackson (1990) note, "when the same number of components or factors are extracted, the results from different types of component or factor analysis procedures typically yield highly similar results. Discrepancies are rarely, if ever, of any practical importance in subsequent interpretations" (p. 9). Discrepancies are most likely to occur when too many or few factors are extracted (Matsunaga, 2010; Trendafilov et al., 2013; Velicer & Jackson, 1990), causing the following step of the EFA process to pivotal.

Together, researchers must weigh the relative benefits and detriments of their chosen factor extraction method. If the assessment of common variance is desired, then researchers should then determine if fit indices would benefit the interpretation of results. If not, then researchers should apply PAF due to its fewer assumptions. If so, then researchers should apply ML. If the assessment of total variance is desired, then researchers should apply PCA.

1.2.3. Factor retention approach

Factor retention is the process of determining the number of factors to retain. Most factor retention approaches are based on eigenvalues, which represent the amount of variance that a factor explains in the indicators (Ruscio & Roche, 2012). It is necessary to retain factors that explain a meaningful amount of variance while excluding those that do not. Factor retention approaches often give differing recommendations for the number of factors to retain, and it is therefore necessary to apply multiple factor retention approaches to identify correct solutions.

The Kaiser criterion specifies that researchers should retain components with eigenvalues greater than one, based on the notion that "a component having an eigenvalue greater than one accounts for more variance than would a single item, thus suggesting merit for combining those items into a factor" (Beavers et al., 2013, p. 7). The Kaiser criterion only applies to the analysis of total variance (e.g., PCA), and its usage for common variance (e.g., PAF or ML) defies the logic of the criterion; indicators have total variance of one but common variance of less than one, and factors with eigenvalues less than one can still explain more common variance than an indicator. While the Kaiser criterion is the most common factor retention approach in reviews within and outside of management (Goretzko et al., 2021; Howard, 2016; Watkins, 2018), its popularity is unfortunately due to its ease of use rather than accuracy. Simulations have shown that the Kaiser criterion is among the least accurate factor retention approaches, and it routinely overestimates the number of factors that should be retained (Auerswald & Moshagen, 2019; Yeomans & Golder, 1982). The Kaiser criterion should not be used in modern research.

Cattell's visual scree plot analysis can provide accurate factor retention decisions, especially compared to the Kaiser criterion (Costello & Osborne, 2005; Howard, 2016; Watkins, 2018). To conduct a visual scree plot analysis, researchers plot their eigenvalues on a chart, identify the final substantial drop-off in the magnitude of eigenvalues, and retain the number of factors just before the final drop-off. Some authors may be skeptical of visual scree plot analysis because it is subjective, and factor retention decisions are less clear with visual scree plot analysis than the Kaiser criterion, parallel analysis, or comparison data (described below). For this reason, it is imperative to perform scree plot analyses in conjunction with other approaches (as recommended for all retention techniques) and report eigenvalues for both retained and excluded factors, such that readers can ensure that the drop-off was correctly identified.

Parallel analysis provides objective cutoffs, and simulation studies have shown that it is among the most accurate factor retention approaches (Goretzko et al., 2020; Lim & Jahng, 2019). When performing a parallel analysis, a number of datasets (e.g., 10,000) consisting entirely of random numbers are generated with an equal number of participants and indicators as the researcher's original dataset, and the eigenvalues for each factor are recorded from each randomly generated dataset (Buja & Eyuboglu, 1992; Hayton et al., 2004). Because these datasets are random numbers, the produced factors (and their eigenvalues) represent nothing more than entirely random variance (i.e., noise). The researcher then compares eigenvalues from their original dataset to the parallel analysis 95th percentile eigenvalues, and eigenvalues that are larger than the parallel analysis 95th percentile eigenvalues are considered to represent more than random variance alone. Lim and Jahng (2019) supported that Horn's original parallel analysis approach more often produced accurate factor retention decisions than proceeding parallel analysis variations, but they also supported – perhaps more importantly – that researchers should assess the validity of one factor more and one factor less than the proposed solution via parallel analysis to ensure that correct factor solutions are identified and account for the margin of error.

We also highlight the comparison data method due to its accuracy and accessibility. The comparison data method sequentially generates datasets with an increasing number of underlying latent factors, such as 500 datasets with one factor, 500 datasets with two factors, etc. It identifies the number of latent factors that produced datasets with eigenvalues most closely approximating the observed eigenvalues, providing support for retaining that number of latent factors (Goretzko et al., 2021; Ruscio & Roche, 2012). This method was supported to be a particularly accurate factor retention approach (Auerswald & Moshagen, 2019), and Courtney (2013) provides step-by-step instructions for conducting the comparison data method in SPSS. Thus, the analysis provides accurate results and is accessible to management researchers.

Extant findings result in the suggestion to together perform scree plot analysis, parallel analysis, and comparison data method. Alternative approaches can be applied with these, but we do not presently discuss alternatives due to specific barriers (e.g., only initial support, difficulty of application). They are featured in our discussion section as emerging advancements.

1.2.4. Factor rotation method

When performing EFA, factor loadings are produced to accurately describe the relation between indicators and latent factors; however, an infinite number of equally fitting solutions exists for any solution with two or more factors, and they are also correct interpretations of the relation between indicators and latent factors (Osborne, 2015). The goal of factor rotations is to thereby identify an equally fitting solution that produces more interpretable factor loadings.

Factor rotations are categorized into two families, orthogonal and oblique, and deciding the family of rotation to apply typically poses greater implications than deciding the exact rotation within a family (Browne, 2001; Park et al., 2002). Orthogonal rotations do not allow factors to be correlated, whereas oblique rotations allow factors to be correlated. Orthogonal rotations do not eliminate relations between the latent factors, but they instead fail to model it. This inability causes orthogonal rotations to produce inaccurate estimates when latent factors are indeed correlated. On the other hand, oblique rotations do not force factors to be correlated, and oblique rotations can produce accurate results both when factors are correlated and uncorrelated. These features cause oblique rotations to be recommended in favor of orthogonal rotations.

Authors rarely provide justification for orthogonal rotations, but they sometimes claim that orthogonal rotations are justified because emergent factors produced small intercorrelations (often < |30|) (Loo, 1979; Tabachnick et al., 2007). The present authors and others (Osborne, 2015; Osborne et al., 2014) are unaware of statistical evidence supporting the superiority of orthogonal rotations under these circumstances. Nevertheless, the preferred orthogonal rotation is varimax (Ford et al, 1986; Goretzko et al., 2021; Sakaluk & Short, 2017). Varimax increases the variance of factor loadings, such that large factor loadings become larger and small factor loadings are "minimized" (Osborne et al., 2014, p. 33). Large factor loadings are differentiated from small factor loadings, enabling researchers to identify indicators more easily for retention.

Researchers justify oblique rotations due to the possibility of correlated factors (Osborne, 2015; Schmitt & Sass, 2011), and oblimin is the most popular oblique rotation (Ford et al, 1986; Gaskin & Happell, 2014; Sakaluk & Short, 2017). Like varimax, oblimin differentiates large factor loadings from small factor loadings to enable researchers to identify substantial loadings more easily. It achieves this by "minimiz[ing] the correlations of loadings across factors" (Spicer, 2005, p. 188), which produces a simpler structure to the EFA results. Another popular oblique rotation, promax, begins by conducting a varimax rotation and then a second rotation via a target matrix to enable correlated factors (Finch, 2006), which causes the large factor loadings to become easier to differentiate from small factor loadings. The ability of oblique rotations to estimate correlated factors causes them to be recommended in favor of orthogonal rotations, but there is no widespread recommendation regarding which orthogonal rotation to apply.

1.2.5. Interpreting factor loadings

Factor loadings represent the variance shared between an indicator and the respective latent factor (Yong & Pearce, 2013). Researchers should retain indicators that strongly relate to their respective latent factors and remove indicators that weakly relate to their respective latent factors, indicating that cutoffs are needed that are high enough to pare irrelevant indicators but low enough to ensure that the entire criterion space of a construct is gauged. Arguments for strong and weak cutoffs are subjective, however. Most researchers do not consider statistical significance tests to be useful for identifying strong loadings because (a) not all factor extraction methods lend themselves to statistical significance tests and (b) large sample sizes cause very small factor loadings to be statistically significant (Fabrigar et al., 1999). Instead, the size of factor loadings is used to differentiate strong from weak, which has caused researchers to propose a very wide range of cutoffs ranging from 0.30 to 0.70 (Beavers et al., 2013; Hair et al., 2019; Williams, Hartman et al., 2010; Williams,Onsman et al., 2010). Among the most common recommendations in management is 0.30, 0.40, and 0.50 (Conway & Huffcutt, 2003; Fabrigar et al., 1999; Ford et al., 1986; Hinkin, 1998).

Researchers should also remove indicators that cross-load onto multiple factors, which refers to possessing strong factor loadings with more than one factor (Hair et al., 2019; Hinkin, 1998). Identifying strong factor cross-loadings is also subjective. Many authors apply similar cutoffs to identify cross-loadings as they do to determine strong primary loadings, and typical cross-loading cutoffs are 0.20 to 0.40 (Beavers et al., 2013; Hair et al., 2019; Williams, Hartman et al., 2010; Williams,Onsman et al., 2010).

A growing number of authors recognize that cross-loadings should be considered based on differences between primary and alternative loadings (Hair et al., 2019; Sakaluk & Short, 2017). For instance, an indicator with a primary loading of 0.41 and alternative loading of 0.39 may not be flagged as problematic when using primary and cross-loading cutoffs of 0.40, but many researchers would consider this indicator to be problematic because 0.02 is not a meaningful difference in magnitude. Researchers have begun to identify criteria regarding the difference between primary and alternative loadings. For instance, Hair et al. (2019) proposed that researchers should calculate a ratio of variance explained by the primary factor and secondary factor ($\frac{\lambda^2_{primary}}{\lambda^2_{secondary}}$); ratios greater than 2.0 are ignorable, ratios between 1.5 and 2.0 are potentially problematic, and ratios below 1.5 are problematic. While no guideline regarding factor loading differences is widespread, they should be considered to improve EFA interpretations.

These factor loading guidelines, however, are based on subjective recommendations rather than empirical observations (Ford et al., 1986; Hinkin, 1998), which is perhaps why a wide range of cutoffs have been recommended in prior research. For this reason, the current article assesses the smallest retained primary loading, largest retained cross-loading, and smallest retained loading difference for each EFA. We provide percentiles for the strength of loadings to offer empirically based criteria to judge factor loadings, similar to efforts identifying correlational effect size benchmarks (Bosco et al., 2015; Gignac & Szodorai, 2016). We offer percentile ranges of ≤10th (very small), 10th–33rd (small), 33rd–67th (medium), 67th–90th (large), and ≥90th (very large), such that researchers can both determine a priori cutoffs based on specific percentiles and more accurately describe their loadings. From these results, we suggest that researchers should default to using cutoffs associated with moderately sized primary loadings, cross loadings, and their difference (33rd–67th percentile), but they may apply alternative cutoffs with sufficient theoretical justification. Researchers should then describe the size of their loadings (e.g., small, medium, and large) by comparing them to our percentile ranges. This process enables researchers to move away from dichotomous assessments and interpret loadings in a continuous manner, which can produce more accurate and empirically based interpretations.

1.3. Exploratory factor analysis for one-factor tests

Harman's one-factor test is the use of EFA to determine whether study results represent common-method bias rather than substantive effects (Fuller et al., 2016; Malhotra et al., 2006). To perform a one-factor test, an EFA is conducted with all indicators, and the most common criteria for discerning common-method bias is the emergence of a one-factor solution and/or the first factor explaining 50% or more of the indicators' variance (Podsakoff et al., 2003, 2012). If findings fail to satisfy these criteria, then common method variance is not considered a concern.

Despite widespread use of Harman's one-factor test, it has many critics. These authors claim that it is supportive of almost any research design, including those that produce significant common method variance (e.g., cross-sectional, single-source) (Malhotra et al., 2006; Podsakoff et al., 2012). Fuller et al. (2016) even showed that the technique routinely produces both false positives and false negatives, causing the authors to argue that it "cannot consistently produce an accurate conclusion about biasing levels of CMV [common method variance]" (p. 3197). These discoveries indicate that Harman's one-factor test should not be applied, and authors using the analysis may be claiming support for research designs with significant common method bias.

1.4. Additional considerations

Our systematic literature review focuses on EFA in management since the year 2000. As this range of years is wider than prior reviews of EFA in management (Conway & Huffcutt, 2003; Fabrigar et al., 1999; Ford et al., 1986), we determine whether EFA practices have been changing over the past 20 years. EFA practices may be problematic when assessed over the entire period, but researchers may be trending away from these practices. If true, then concerns would not be as severe as initially assumed, as the field may be heading in the right direction.

Also, it is possible that researchers routinely fail to report essential aspects of their EFAs because word and page lengths are limited, and only reporting portions of EFAs may be a common approach to reducing manuscript length (although other analyses are expected to be reported in full). A possible solution may be to succinctly report EFAs in the primary text and fully report them in supplemental materials that typically do not count towards word or page limits. Thus, we assess the frequency that EFAs are reported in supplemental materials.

2. Method

2.1. Search and coding procedures

The current article now reports a systematic literature review to assess the extent that researchers adhere to recommended practices for EFA. By doing so, our article determines whether modern uses of EFA produce accurate and replicable results or whether inferences in the field may be obfuscated by improper analytical procedures and reporting. To identify our sources, we used the Academic Journal Guide of the Charted Association of Business Schools (2018). Journals ranked 4* or 4 are premier outlets, and we included all 23 journals ranked 4* or 4 from six categories relevant to management (Table 1). The number of articles recorded from each journal can be found in Supplemental Material A. In August 2021, we used Google Scholar to create a list of all articles published in these outlets since 2000 that include the term, “Exploratory Factor Analysis”, “EFA”, “Principal Components Analysis”, “PCA”, “Harman’s Test”, or “One-Factor Test”. Cataloguing these sources resulted in an initial list of 2,244 articles. Coders then developed coding procedures, independently coded sets of 20 articles until inter-rater agreement standards were met (e.g., Cohen’s $\kappa = 0.80$), and one coder coded the remaining articles. These procedures were based on guidelines for systematic reviews (Aguinis et al., 2020; DeSimone et al., 2021; Fisch & Block, 2018; Xiao & Watson, 2019).

We coded whether each source applied PCA and/or EFA for psychometric investigation or Harman’s one-factor test. We did not include the use of PCA or EFA for data reduction (e.g., parceling for structural equation modeling [SEM]), as authors are not expected to fully report PCAs or EFAs for such analyses. This reduced our list of 2,244 articles to 1,790 articles. Then, we coded the attributes of each reported EFA. Our final dataset included 1,790 articles with 3,396 EFAs. The attributes coded for each EFA are detailed in Supplemental Material B.

Table 1
Journals represented in systematic literature review.

Entrepreneurship and Small Business Management
Entrepreneurship, Theory and Practice
Journal of Business Venturing
Strategic Entrepreneurship Journal
General Management, Ethics, Gender, and Social Responsibility
Academy of Management Journal
Administrative Science Quarterly
British Journal of Management
Business Ethics Quarterly
Journal of Management
Journal of Management Studies
International Business and Area Studies
Journal of International Business Studies
Journal of World Business
Organizational Studies
Human Relations
Leadership Quarterly
Organization Science
Organizational Studies
Psychology (Organizational)
Journal of Applied Psychology
Journal of Occupational and Organizational Psychology
Journal of Occupational Health Psychology
Journal of Organizational Behavior
Journal of Vocational Behavior
Organizational Behavior and Human Decision Processes
Personnel Psychology
Strategy
Strategic Management Journal

Note. Categories taken from the Academic Journal Guide of the Charted Association of Business Schools (2018). Included journals were ranked 4* or 4 from these six categories.

2.2. Analyses

Datasets are provided in Supplemental Material A. We performed several analyses to assess the differences by year (Supplemental Material C). For dichotomous EFA attributes, we used binomial logistic regression. For categorical EFA attributes, we used binomial logistic regression to determine the relation of year with the reporting of each category, but we also performed multinomial logistic regression to determine the relation of year with the reporting of all categories. For continuous EFA attributes, we used linear regression. Year was the sole predictor in each analysis. Supplemental Material C also includes tables that provide all regression results, charts demonstrating yearly changes in EFA practices, and frequencies of EFA practices separated by the years 2000–2004 ($n = 503$), 2005–2009 (n = 939), 2010–2014 (n = 876), and 2015–2021 ($n = 1,078$). The primary text focuses on the regression results, as these provide a statistical significance test regarding whether researchers are trending away from certain practices. If the regressions associated with a certain practice is statistically significant, then researchers are either significantly more or less likely to apply the practice over time.

The average number of EFAs reported per included article was 1.90 ($SD = 2.44$). We first identified outlier articles regarding the number of reported EFAs by calculating Z-scores. Three articles were outliers with 62 (*Z-score* = 24.61), 33 (*Z-score* = 12.74), and 32 (*Z* = *score* = 12.33) EFAs, whereas the following articles reported much fewer EFAs with 20 (*Z-score* = 7.41), 18 (*Z-score* = 6.60), 16 (*Z-score* = 5.78), and 15 (*Z-score* = 5.37) reported EFAs. To ensure that these outliers did not sway our results, we removed these three from our reporting in the primary text, and Supplemental Material D includes analyses with these articles.

Our primary text details all reported EFAs from our sources. Doing so weights each EFA equally; however, it causes the articles to be weighted differently, as articles that reported more EFAs are included more than articles that reported fewer EFAs. We provide alternative results in Supplemental Material E, wherein we average the EFA results within each article before reporting our results. By doing so, each article is represented once in our dataset, providing an equal weighting to articles and unequal weighting to the EFAs themselves.

We coded the EFAs exactly as they were reported, but some articles reported multiple EFAs and only detailed the attributes for one. For instance, an article may have noted that they conducted an EFA in Study 1, but they may have specified that they used PAF in Study 2. In the primary text, we would consider the factor extraction method for the first EFA to be unknown. In Supplemental Material F, we report a reanalysis wherein we assume that all EFAs in an article were conducted identically unless otherwise stated. In the example above, these supplemental analyses consider both EFAs to have used PAF. Our inferences were consistent between the primary text and all alternative analyses, strongly supporting the robustness of our findings.

Some readers may be interested in EFA practices for only the domains of Organizational Behavior and Applied Psychology (OB/AP) or Strategy. Supplemental Material G provides the results of all analyses when separately conducted for OB/AP and Strategy. Both domains produced similar results, supporting our decision to study them together in the primary analyses. Similarly, some researchers may be interested in narrower subsets of management research. Supplemental Material G also provides our results separated by each category of the Academic Journal Guide of the Charted Association of Business Schools (2018), and readers should refer to Table 1 to identify the journals included within each category. These results were relatively consistent across each domain, again supporting our decision to study them together.

3. Results depicting current practices

Table 2, Table 3, and Table 4 include statistics regarding EFAs for psychometric investigations. A CFA was performed in conjunction with 1,276 EFAs (46%), wherein 671 of these were performed with the same sample (24%) and 605 with a different sample (22%). Recent articles were more likely to perform CFA with EFA ($\beta = 0.08$, $S.E. = 0.01$, $Z = 11.04$, $p < .01$), but the choice of sample did not have a significant relation with year ($\beta =$ -0.02, $S.E. = 0.01$, $Z =$ -1.78, $p = .07$). Before conducting any EFA, Bartlett's test and KMO test were only reported 103 (4%) and 140 (5%) times, respectively. Bartlett's test was supportive in all applications, whereas KMO test supported the use of EFA in all but one. Year did not have a statistically significant relation with the reporting of Bartlett's test or KMO test (both $p > .05$).

The median sample size was 217. Most EFAs included samples sizes larger than 200 (1,502; 55%) but not 300 (988; 37%). The median participant-to-indicator ratio was 19.90. Almost all EFAs had participant-to-indicator ratios larger than 5 (2,338; 90%), whereas most had ratios larger than 10 (1,899; 73%). The publication year did not have a significant relation with sample size ($p = .59$). It did have a significant relation with the participant-to-indicator ratio ($\beta = 6.67$, $S.E. = 2.16$, $t = 3.09$, $p < .01$), likely because researchers are including fewer indicators in their EFAs over time ($\beta = -0.37$, $S.E. = 0.09$, $t = -4.00$, $p < .01$). Table 4 includes median sample sizes separated by the number of indicators and emergent factors. Most combinations failed to meet the minimum sample size suggestions of Rouquette and Falissard (2011). The number of indicators and factors were also not significant predictors of sample size (both $p > .05$).

Half of reported EFAs did not include a specified factor extraction method (1,361; 50%). The most common reported approach was PCA (855; 31%), followed by PAF (353; 13%), ML (166; 6%), and other (15; 1%). Year had a significant relation with the factor extraction method. Earlier studies were more likely to report the use of PCA ($\beta = -0.05$, $S.E. = 0.01$, $Z = -6.34$, $p < .01$). Recent studies were less likely to report their factor extraction method ($\beta = 0.03$, $S.E. = 0.01$, $Z = 5.00$, $p < .01$) and more likely to use PAF ($\beta = 0.02$, $S.E. = 0.01$, $Z = 2.05$, $p < .01$).

To make factor retention decisions, most authors did not state their approach (1,847; 66%). The most popular reported method was the Kaiser criterion (687; 25%), followed by visual scree plot analysis (335; 12%). Parallel analysis (100; 4%) and other approaches (52; 2%) were rarely applied. The most popular combination of approaches was the Kaiser criterion and visual scree plot analysis (179; 7%). More recent studies were significantly more likely to have applied parallel analysis than older studies ($\beta = 0.07$, $S.E. = 0.02$, $Z = 3.56$, $p < .01$), but they were also significantly less likely to report their factor retention approach ($\beta = 0.04$, $S.E. = 0.01$, $Z = 5.41$, $p < .01$). Perhaps for this reason, they were also less likely to report the application of the other three types of factor retention decisions: Kaiser criterion ($\beta = -0.05$, $S.E. = 0.01$, $Z = -5.86$, $p < .01$), visual scree plot analysis ($\beta = -0.06$, $S.E. = 0.01$, $Z = -5.49$, $p < .01$), and other ($\beta = -0.12$, $S.E. = 0.03$, $Z = -4.40$, $p < .01$). More recent studies were also less likely to apply the combination of Kaiser criterion and visual scree plot analysis together ($\beta = -0.08$, $S.E. = 0.01$, $Z = -5.65$, $p < .01$).

Only 21% of EFAs reported Eigenvalues for all retained factors (584), 19% reported some Eigenvalues (527), and 60% reported none (1,639). Year had a significant relation with the reporting of Eigenvalues. Recent students were less likely to report Eigenvalues for all ($\beta = -0.05$, $S.E. = 0.01$, $Z = -5.69$, $p < .01$) or some ($\beta = -0.02$, $S.E. = 0.01$, $Z = -0.212$, $p = .02$) retained factors.

To assess rotation methods, we only analyzed EFAs that did not specify the retention of a single factor ($n = 1,826$). More than one-third of EFAs did not report their rotation method (767; 42%), about one-third applied an orthogonal rotation (539; 30%), and about one-fourth applied an oblique rotation (499; 27%). The most popular orthogonal rotation method was varimax (509; 28%), whereas oblimin (187; 10%) and promax (130; 7%) were the most popular oblique rotation methods. The failure to report a rotation method was more common in recent articles ($\beta = 0.04$, $S.E. = 0.01$, $Z = 4.22$, $p < .01$). The use of orthogonal rotations was more common in older EFAs ($\beta = -0.05$, $S.E. = 0.01$, $Z = -5.34$, $p < .01$), as was the use of varimax ($\beta = -0.05$, $S.E. = 0.01$, $Z = -5.21$, $p < .01$). The reporting of promax was more common in recent EFAs ($\beta = 0.08$, $S.E. = 0.02$, $Z = 5.09$, $p < .01$), but the use of oblique rotations and oblimin was not (both $p > .05$).

For primary loadings, the most popular cutoffs were 0.30 (30; 15%), 0.40 (66; 33%), 0.50 (44, 22%), 0.60 (13; 7%) and 0.70 (16, 8%). The smallest primary loading was more often larger than 0.40 (1,177; 87%) and 0.50 (874; 66%) but less often larger than 0.60 (565; 43%), 0.70 (290; 22%), or 0.80 (101; 8%). For cross-loadings, the most popular a priori cutoffs were 0.20 (7; 8%), 0.30 (31; 37%), 0.35 (7; 8%), and 0.40 (19; 23%). The largest cross-loading was most often larger than 0.20 (539; 87%) and 0.30 (359; 58%), whereas it was less often larger than

Table 2
EFA attributes in management.

	Frequency (Relative Frequency)	
	CFA with EFA	
Same Sample	671 (24%)	
Different Sample	605 (22%)	
	Pre-Analysis Data Checks	
KMO Test	140 (5%)	
Bartlett's Test	103 (4%)	
	Factor Extraction Method	
PCA	855 (31%)	
PAF	353 (13%)	
ML	166 (6%)	
Other	15 (1%)	
Unknown	1,361 (50%)	
	Factor Retention Determination[1]	
	Inclusive	Exclusive
Kaiser Criterion (KC)	687 (25%)	498 (18%)
Scree Plot (SP)	335 (12%)	131 (5%)
Parallel Analysis (PA)	100 (4%)	49 (2%)
Other (O)	52 (2%)	9 (0%)
KC & SP	179 (7%)	133 (5%)
KC & PA	28 (1%)	4 (0%)
KC & O	34 (1%)	6 (0%)
SP & PA	45 (2%)	18 (1%)
SP & O	35 (1%)	4 (0%)
PA & O	11 (0%)	2 (0%)
KC & SP & PA	24 (1%)	18 (1%)
KC & SP & O	28 (1%)	22 (1%)
KC & PA & O	6 (0%)	0 (0%)
SP & PA & O	9 (0%)	3 (0%)
KC & SP & PA & O	6 (0%)	6 (0%)
Unknown	1,847 (67%)	1,847 (67%)
	Factor Rotation Method[2]	
Orthogonal	**539 (30%)**	
Varimax	509 (28%)	
Other	12 (1%)	
Oblique	**499 (27%)**	
Oblimin	187 (10%)	
Promax	130 (7%)	
Other	7 (0%)	
None	**23 (1%)**	
Unknown	**767 (42%)**	

Note. The first number represents the number of studies that applied the method designated by the row, and the second number in parentheses represents the percent of studies that applied the method relative to the total number EFAs. Total n = 2,750.

[1] The first number and percentage reports the categories as inclusive, whereas the second number and percentage reports the categories as exclusive. For the first number and percentage, an EFA that reported the use of both the Kaiser criterion and visual scree plot analysis would be included in the count for the Kaiser criterion, visual scree plot analysis, as well as both the Kaiser criterion and visual scree plot analysis. For the second number and percentage, an EFA that reported the use of both the Kaiser criterion and visual scree plot analysis would solely be included in the count for both the Kaiser criterion and visual scree plot analysis.

[2] Only included studies that did not specify one retained factor (n = 1,826).

Table 3
Additional EFA attributes in management.

Percentile	Sample Size	# of Indicators	Ratio	Primary Loading Cutoff	Smallest Primary Loading	Cross-Loading Cutoff	Largest Cross-Loading	Factor Loading Difference Cutoff	Largest Factor Loading Difference	% of Variance All Factors
n[1]	2,707	2,625	2,592	200	1,327	83	620	24	459	1,203
10th	91	4	5.31	0.30	0.40	0.20	0.19	0.10	0.02	48
20th	121	5	7.92	0.35	0.43	0.30	0.24	0.13	0.09	54
25th	138	6	9.54	0.40	0.46	0.30	0.26	0.15	0.12	56
30th	151	7	11.30	0.40	0.49	0.30	0.29	0.15	0.14	58
33rd	163	8	12.46	0.40	0.50	0.30	0.29	0.15	0.15	60
40th	187	9	15.09	0.40	0.51	0.30	0.30	0.15	0.18	62
50th	217	11	19.90	0.40	0.56	0.30	0.34	0.15	0.21	65
60th	280	14	27.35	0.50	0.61	0.33	0.36	0.20	0.27	68
67th	321	16	34.22	0.50	0.64	0.35	0.38	0.20	0.30	70
70th	351	18	38.69	0.50	0.66	0.40	0.39	0.20	0.32	71
75th	401	20	46.84	0.50	0.69	0.40	0.40	0.20	0.36	73
80th	503	24	57.48	0.50	0.71	0.40	0.42	0.20	0.40	75
90th	1,105	36	117.65	0.65	0.77	0.40	0.48	0.20	0.50	80

[1] This row represents the number of EFAs in our dataset that were able to be coded for this category. Because not all articles reported each piece of information represented by the column headings, not all EFAs were able to be included in the calculation of the percentile values provided in the following rows. For instance, only 200 EFAs clearly stated an a priori primary loading cutoff, resulting in the percentiles for a priori primary loading cutoff to be calculated from these 200 EFAs.

Table 4
Reported EFA sample sizes separated by number of indicators and retained factors.

	# of Indicators									
# of Factors	5	10	15	20	25	30	35	40	45	50+
1	247 (80)	203 (42)	185 (12)	296 (9)	–	–	–	–	–	–
2	208 (18)	157 (35)	351 (15)	287 (6)	**400 (6)**	–	–	–	–	–
3	–	260 (23)	180 (15)	224 (13)	188 (6)	–	–	–	–	–
4	–	–	141 (8)	268 (13)	248 (11)	339 (6)	244 (9)	–	–	185 (7)
5+	–	–	260 (5)	212 (12)	211 (10)	334 (16)	404 (10)	288 (8)	–	360 (22)

Note. First number in each cell represents the median reported sample size for EFAs with the number of indicators indicated by the column and the number of factors represented by the row. The number in parentheses in each cell represents the number of EFAs that were included in the calculation of the median. Indicator and factor combination must have been represented by four or more studies to be included, and studies must have likewise stated the use of EFA to be included (e.g., no studies with PCA or unknown factor extraction methods). The number of indicators included in each EFA was rounded to the nearest fifth. Cells in bold exceed the sample size minimum recommended by Rouquette and Falissard (2011), which only apply to cells with the number of indicators between 10 and 45 as well as the number of factors between 2 and 4 (i.e., outer cells excluded).

0.40 (147; 24%) and 0.50 (50; 8%). For the difference between primary and secondary loadings, the most popular cutoffs were 0.10 (5; 21%), 0.15 (8; 33%), and 0.20 (9; 38%). The difference was greater than 0.10 in 77% of cases (355) and 0.20 in 53% of cases (245). This difference was greater than 0.30 (151; 33%), 0.40 (90; 20%), and 0.50 (44; 10%) in less than half. Year only had a significant relation with primary loading cutoff ($\beta = 0.01$, $S.E. = 0.00$, $t = 3.77$, $p < .01$).

Most EFAs explained more than 50% (1,062; 88%) and 60% (789; 66%) of variance, and less than half explained more than 70% (391; 33%). Year had a significant relation with variance explained ($\beta = 0.39$, $S.E. = 0.07$, $t = 5.45$, $p < .01$), with newer studies explaining more. Any communalities were reported for 50 EFAs, and supplemental material was provided for 31 EFAs.

Table 5 includes statistics regarding EFA for Harman's one-factor test. The most common cutoffs for Harman's one-factor test were a one-factor solution (52; 10%), more than 50% of variance explained by the first factor (113; 22%), as well as a one-factor solution and/or more than 50% of variance explained by the first factor (113; 22%). Forty-five percent of studies did not report their cutoffs (232). More recent studies were more likely to use the 50% cutoff ($\beta = 0.11$, $S.E. = 0.02$, $t = 4.55$, $p < .01$), whereas older studies were more likely to apply both cutoffs ($\beta = -0.09$, $S.E. = 0.02$, $t = -3.83$, $p < .01$). No study met their applied cutoff, and thus no author found common method variance to have caused their results. Year did not have a significant relation with variance explained by the first factor ($p = .32$).

Table 5
Harman's one-factor test attributes.

Percentile	Sample Size	# of Indicators	Ratio	% of Variance All Factors	% of Variance First Factor
n[1]	517	496	495	171	338
10th	100	13	3.14	56	15
20th	120	17	4.38	62	17
25th	133	18	4.84	64	18
30th	144	20	5.48	65	19
33rd	156	20	5.83	66	20
40th	177	23	6.67	67	21
50th	202	26	7.83	68	23
60th	231	29	9.66	70	25
67th	280	32	11.41	72	26
70th	302	33	12.72	72	27
75th	369	36	14.47	73	28
80th	448	39	18.73	75	29
90th	946	49	43.69	80	34

[1] This row represents the number of Harman's One-Factor Tests in our dataset that were able to be coded for this category. Because not all articles reported each piece of information represented by the column headings, not all Harman's One-Factor Tests were able to be included in the calculation of the percentile values provided in the following rows. For instance, only 338 EFAs clearly stated the amount of variance explained by the first factor, resulting in the percentiles for the variance explained by the first factor to be calculated from these 338 Harman's One-Factor Tests.

4. Discussion

Perhaps our most consistent finding was that authors regularly do not report essential aspects of their EFAs, which has become more frequent over the past 20 years. This finding is concerning. For the majority of EFAs performed in management, it is unclear whether authors are conducting their analyses correctly, and the field may be largely based on measures with poor or misleading psychometric properties. Likewise, the complete reporting of analyses is necessary for authors to replicate prior results. Without being able to do so, a cornerstone of ensuring research validity in the social sciences is lost (OSC, 2012, 2015, 2017). Perhaps our strongest recommendation is for authors, reviewers, and editors to ensure that EFAs are reported in entirety. We recognize that page lengths are limited, but authors are also not utilizing supplemental materials or online repositories. Thus, we urge future researchers to include the full reporting of their EFA, even if in supplemental materials or online repositories.

Further, by failing to report all aspects of EFA, promising trends in management may also not be promising at all. Authors are less frequently applying some problematic EFA approaches in the past 20 years; however, it could be interpreted that these authors are still applying these problematic approaches at the same frequency and instead failing to report their analytical approaches. For instance, decreased reporting of the Kaiser criterion may not be due to authors applying the method less often, but it may instead be due to authors neglecting to report their factor retention method. Even apparent improvements in the use of EFA may not be actualized, reinforcing our call for the more complete reporting of EFA.

Some readers may assume that misestimations from improperly performed EFAs are addressed by a subsequent CFA. We found that this is not the case. CFA was conducted on the same sample in 24% of cases, and CFA was conducted on a different sample in 22% of cases. Less than half of articles performed a subsequent CFA, and it was regularly performed on the same sample which poses concerns (Fokkema & Greiff, 2017). While CFA is not a panacea, researchers are often not performing robust follow-up analyses. Misestimations arising from improper EFAs may persist to bias subsequent analyses. Researchers should perform their EFAs correctly and conduct follow-up CFAs with alternative samples when resources permit.

We now discuss our findings regarding the five steps of EFAs to assess psychometric properties. First, Bartlett's test and KMO test are rarely applied. Researchers may neglect to perform these analyses, perhaps because they assume that common variance is not an issue if they obtain interpretable EFA results. This is not the case. EFA results may be interpretable when Bartlett's test and KMO test indicate that the analysis should not be performed, but the resultant factor structure may explain little variance in the indicators and provide few theoretical insights (Tobias & Carlson, 1969; Wilson & Martin, 1983). It would be more useful to identify and remove indicators that cause poor Bartlett's test and KMO test results. If the removal of indicators does not improve results, then the researcher should reassess whether their sampling source produced poor quality data. If the data can be assured to be sufficient, then the researcher may be analyzing indicators that are unsuitable for EFA, such as largely independent indicators.

Further, researchers rarely met sample size guidelines based on common variance (e.g., indicators per factor), and both the number of indicators and factors did not have a significant association with sample size. These results are troubling. Many EFAs have likely produced inaccurate estimates due to small sample sizes, and researchers are not basing their sample sizes on known properties of EFA that necessitate larger sample sizes. Moving forward, researchers should adhere to sample size recommendations based on common variance, and we recommend those of Rouquette and Falissard (2011). While it is difficult to estimate indicators of common variance beforehand (e.g., communalities), it is reasonable to know indicator-to-factor ratios a priori, which is a benefit of the sample size tables provided by Rouquette and Falissard (2011).

Second, PCA is more popular than PAF or ML. This suggests that researchers in management prefer to identify components that model total variance rather than factors that model common variance (Park et al., 2002; Trendafilov et al., 2013); however, in our review, we found almost no articles that justified their use of any extraction method, and researchers may more frequently apply PCA because it is the default option in popular statistics programs (e.g., SPSS). We urge future researchers to justify their applied extraction approach, whether factor- or components-based, which echoes prior calls that have seemingly been ignored (Preacher & MacCallum, 2003). By providing justifications, the field of management can ensure that EFAs reflect researchers' intended approach to modeling variance rather than practical convenience.

Third, the Kaiser criterion was the most popular factor retention approach, even in conjunction with the study of common variance that defies the logic of the criterion. Researchers are seemingly unaware of the Kaiser criterion's limitations as well as the benefits of other factor retention approaches. Because the Kaiser criterion routinely produces inaccurate results, editors and reviewers should no longer accept the use of the Kaiser criterion. Also, researchers regularly engaged in the problematic practice of applying only one factor retention approach, and future researchers should apply a combination of approaches. We recommend the application of visual scree plot analysis, parallel analysis, and the comparison data method due to their accuracy and ease to conduct, along with other emergent factor retention methods that are discussed below.

Fourth, our sources were almost evenly split between orthogonal rotations and oblique rotations. This is surprising given widespread concerns regarding orthogonal rotations (Browne, 2001; Costello & Osborne, 2005; Norris & Lecavalier, 2010; Park et al., 2002). All EFAs produce factors that correlate to some extent, and orthogonal rotations are unable to account for this covariance. Many authors are performing rotations that provide inaccurate depictions of their factor loadings. Reviewers and editors should no longer accept the use of orthogonal rotations, and researchers should apply oblique rotations (e.g., oblimin or promax).

Fifth, researchers use a wide range of primary loading cutoffs, with common cutoffs spanning from 0.30 to 0.70. Because factor loading interpretations are presently subjective, this wide range may be due to uncertainties regarding the appropriateness of any specific cutoff. Using the percentiles listed in Table 3, our results can progress interpretations of factor loadings by enabling researchers to both interpret their results on a gradient and provide empirical justification for cutoffs. Researchers can consider primary factor loadings less than 0.40 to be very small ($\leq 10^{th}$ percentile), between 0.40 and 0.50 to be small (10th–33rd percentile), between 0.50 and 0.64 to be moderate (33^{rd}–67th percentile), between 0.64 and 0.77 to be large (67th–90th percentile), and above 0.77 to be very large ($\geq 90^{th}$ percentile). These labels, though, are relative and based on retained indicators. A loading of 0.41 may be small relative to retained indicators in prior research, but it may be sufficiently large to retain the indicator in certain circumstances.

Moving forward, researchers should utilize these percentiles to determine their primary loading cutoff. We recommend that researchers should default to utilizing a cutoff of ≥ 0.50, as this would ensure that retained indicators have a primary factor loading that is at least moderate in magnitude relative to prior research. It should be acknowledged, however, that the nature of studied measures may call for other cutoffs. Authors have long recognized that measures for broad, multidimensional, and/or elusive constructs are more likely to produce smaller primary loadings and larger cross loadings than measures for narrow, unidimensional, and/or well-defined constructs (Hair et al., 2019; Marsh, 2007; Marsh et al., 2014). For instance, a measure for all Big Five personality factors would be expected to produce smaller primary loadings and larger cross-loadings than a measure for a specific facet of conscientiousness (Marsh et al., 2010). The cutoff for small primary factor loadings (≥ 0.40) may be suitable for measures of constructs that are broad, multidimensional, and/or elusive, whereas the cutoff for large factor loadings (≥0.64) may be suitable for measures of constructs that are narrow, unidimensional, and/or well-defined; however, researchers should only deviate from the cutoff for moderate primary factor loadings (≥0.50) with sufficient theoretical justification.

Our results likewise progress the interpretation of cross-loadings. A wide range of cross-loading cutoffs are applied with modest agreement, and few researchers provide a cross-loading difference cutoff perhaps because they are unaware of any recommendations. Moving forward, researchers should reference Table 3 to interpret their cross-loadings and cross-loading differences. Researchers could identify cross-loadings less than 0.19 to be very small (≤10th percentile), between 0.19 and 0.29 to be small (10th–33rd percentile), between 0.29 and 0.38 to be moderate (33rd–67th percentile), between 0.38 and 0.48 to be large (67th–90th percentile), and above 0.48 to be very large (≥90th percentile). Regarding cross-loading differences, a researcher could identify differences less than 0.02 to be very small (≤10th percentile), between 0.02 and 0.15 to be small (10th–33rd percentile), between 0.15 and 0.30 to be moderate (33rd–67th percentile), between 0.30 and 0.50 to be large (67th–90th percentile), and above 0.50 to be very large (≥90th percentile).

Again, researchers should utilize these percentiles to determine their cutoffs. Researchers should default to the cutoffs of ≤0.38 for cross-loadings and ≥0.15 for cross-loading differences, as these benchmarks ensure that retained indicators' cross-loadings are no more than moderate and cross-loading differences are at least moderate in magnitude relative to prior research. Large cross-loading cutoffs (≤0.48) and small difference cutoffs (≥0.02) may be suitable for broad, multidimensional, and/or elusive constructs, whereas small cross-loading cutoffs (≤0.29) and large difference cutoffs (≥0.30) may be suitable for narrow, unidimensional, and/or well-defined constructs. Proper theoretical rationale must be provided for the use of alternative cutoffs.

EFA practices and findings were remarkably similar between OB/AP and Strategy, and they were consistent across the narrower categories of the journal list. Similar improvements should be made to EFA practices in all areas of management. Also, Ford et al. (1986), Fabrigar et al. (1999), as well as Conway and Huffcutt (2003) found that the most common factor retention approach was the Kaiser criterion and the most popular factor rotation was orthogonal. We found the same results despite suggestions to transition away from these problematic EFA approaches. To aid in applying appropriate EFA methods, we provide a checklist in Table 6. This checklist is intended to be a "starting point". Deviations from this checklist can be appropriate, such as justifying the use of smaller sample sizes based on the communalities; however, the intent of the checklist is to provide easy-to-follow guidelines. We also provide Supplemental Material H, which provides step-by-step guides for applications of EFA in SPSS and R. Researchers should review these guides to ensure that their analyses are conducted correctly, and they should not assume that default settings reflect recommended practices. Readers should seek guides for EFA with other programs, such as Stata, from other sources to ensure the application of presently recommended practices, as it cannot be assumed that the treatment of EFA in these programs is consistent with SPSS or R.

Finally, we found that researchers are still regularly applying Harman's one-factor test, despite its known concerns (Fuller et al., 2016; Malhotra et al., 2006; Podsakoff et al., 2012). Most authors considered common method variance to be a concern only if a one-factor solution emerged and/or the first factor explained more than 50% of the indicators' variance. No article failed their applied cutoff, suggesting that cutoffs for one-factor tests are much too high to discover problematic common method variance. Researchers may be claiming support for designs that include substantial common method bias, and they may also be producing inaccurate insights based on their biased empirical findings. We call for a moratorium on one-factor tests until research has identified conditions in which it can provide accurate results.

Table 6
Checklist for conducting exploratory factor analysis.

Step	Recommendation	Check
Step 0 – Review Literature	Review literature for newly recognized EFA concerns and recommended approaches. Choose approaches that provide the most accurate results.	☐
	Identify whether the application calls for specific rather than general EFA approaches, as best practices may differ from current recommendations. Below are three occasional situations, but others exist. • Nested data calls for multilevel EFA. • Very small sample sizes (e.g., <50) call for unweighted least squares or regularized EFA. • Factor structures that can be partially assumed call for target rotation.	☐
Step 1 – Data Quality Checks	Obtain a sample size following guidelines based on indicators of common variance (e.g., ratio of indicators to intended factors), which are typically above 300 participants and a 10-to-1 participant to indicator ratio. We recommend the sample size tables of Rouquette and Falissard (2011).	☐
	Perform typical assessments of data quality with appropriate corrections. Below are two examples, but researchers should use other data quality corrections. • Removing participants that failed attention checks. • Use regression-based imputation techniques to address missing data.	☐
	Perform Bartlett's test (p <.05) and KMO test (>0.50).	☐
	Assess and correct aspects of data quality (e.g., nonnormality) that may influence intended EFA techniques (e.g., maximum likelihood extraction).	☐
Step 2 – Factor Extraction	If assessment of common variance is desired, use principal axis factoring.	☐
	If assessment of total variance is desired, use principal components analysis.	☐
	If fit indices are desired and multivariate normality can be supported, use maximum likelihood factor extraction.	☐
Step 3 – Factor Retention	Apply visual scree plot analysis, parallel analysis, and comparison data method.	☐
	Consider additional factor retention techniques, such as the Hull method and machine learning estimations.	☐
	Do not use the Kaiser criterion. Apply more than one retention method.	☐
Step 4 – Factor Rotation	Apply an oblique rotation, such as direct oblimin or promax.	☐
	Do not use orthogonal rotations unless special circumstances can strongly support their application based on sound justifications.	☐
Step 5 – Factor Loadings	Utilize cutoffs provided in the primary text associated with moderate primary loadings (≥0.50), cross-loadings (≤0.38), and cross-loading differences (≥0.15). Remove items that fail any of the cutoffs. Alternative cutoffs associated with small or large percentile cutoffs can be applied, but these should only be used with sufficient theoretical rationale based on the underlying construct's breadth, dimensionality, and/or elusiveness.	☐
	Interpret factor loadings continuously rather than dichotomously, such that the strength of factor loadings is considered on a spectrum. Utilize Table 3 to identify ranges for very small, small, medium,	☐

(continued on next page)

Table 6 (*continued*)

Step	Recommendation	Check
	large, and very large factor loadings, cross-loadings, and their differences for retained items.	
Post-Analyses	Report all methodological approaches, justifying rationale, and statistical results. Use supplemental materials and online repositories if needed.	☐

4.1. Emerging advancements in exploratory factor analysis

With increasing frequency, researchers are recognizing that data quality concerns and associated remedies are not consistent across all EFAs. Certain data quality concerns, such as missing data, have larger biasing influences when sample sizes are small, and McNeish (2017) identified that predictive means matching, a regression-based imputation technique, provides the most accurate EFA estimates with small sample sizes. Likewise, Goretzko et al. (2020) showed that missing data imputation techniques differently impact the performance of parallel analysis. Predictive means matching performed well, but random forest imputation, a machine-learning technique, performed the best of the tested approaches. Overall, modern recommendations for data quality remedies (e.g., regression-based & machine learning imputations; Newman, 2014) perform well across most EFA applications (Goretzko et al., 2020; McNeish, 2017), but researchers should monitor advancements relevant to their applied EFA approaches (Table 7).

Three factor extraction approaches may also prove useful in future research. While PAF and ML are preferred and more supported, researchers should monitor developments to Bayesian EFA (Conti et al., 2014) and regularized EFA (Jung et al., 2020) to determine whether they are more suitable for their applications. For instance, Jung and Lee (2011) showed that regularized EFA provides more accurate estimates than ML when sample sizes are very small (e.g., <50). Further, many contexts call for multilevel EFA, but the factor extraction approach is rarely applied in management. Multilevel EFA creates separate between- and within-group factor structures (Kim et al., 2016), which can determine whether factor structures replicate at the appropriate level of analysis. Researchers should consider theoretical implications of these factor structures and decide whether multilevel EFA is appropriate for their research questions.

Factor retention may be the most active research topic of EFA. We believe that our recommended approaches have the most ideal combination of present empirical support and practical ease, but certain emergent method may join these recommended methods in the future. Goretzko and Bühner (2020) proposed and supported the use of machine learning for factor retention, and promising initial results have been provided for exploratory graph analysis (Golino & Epskamp, 2017; Golino et al., 2020). In addition to our recommended approaches, Auerswald and Moshagen (2019) performed a series of simulations that also supported the empirical Kaiser criterion, Hull method, and sequential model tests. Researchers should monitor developments in statistical programs that may cause these approaches to be more accessible, such that they can be applied with scree plot analysis, parallel analysis, and comparison data method.

Varimax, oblimin, and promax have dominated research for decades, but other rotations have been recently developed. Scharf and Nestler (2019) supported that regularized EFA can provide comparable estimates to rotated EFA; however, research is needed before regularized EFA can be recommended beyond rotated EFA. Zhang et al. (2019) further developed target rotation, wherein the technique utilizes the researcher's a priori specifications for a presumed factor structure in obtaining estimates. Such an approach is beneficial for situations in which a portion of the factor structure can be assumed, such as analyzing multitrait-multimethod data.

Fewer developments have been made regarding our understanding of factor loadings. Some evidence is apparent for researchers placing a

Table 7
Emerging Advancements in Exploratory Factor Analysis and Relevant Key Citations.

Area of Advancement	Emerging Advancement	Relevant Key Citations
Data Quality Checks	Identifying biasing effects of specific data quality concerns under certain conditions	Caron (2019); Cooperman and Waller (2022); McNeish (2017)
	Testing data quality remedies under specific conditions and/or EFA approaches	Goretzko (2022); Goretzko et al., 2020; McNeish (2017); Nassiri et al. (2018); Xiao et al. (2019)
Factor Extraction Method	Regularized EFA	Jung (2013); Jung and Lee (2011); Jung et al. (2020); Jung and Takane (2008); Scharf and Nestler (2019)
	Bayesian EFA	Chen (2021); Conti et al. (2014)
	Multilevel EFA	D'haenens et al. (2010); Ji et al. (2021); Kim et al. (2016)
Factor Retention Approach	Parallel Analysis Variations	Courtney (2013); Greene et al. (2022); Iacobucci et al (2022); Levy et al. (2021); Lim & Jahng (2019); Steiner & Grieder (2020)
	Empirical Kaiser Criterion	Auerswald and Moshagen (2019); Braeken and Van Assen (2017); Li et al (2020); Steiner and Grieder (2020)
	Hull Method	Auerswald and Moshagen (2019); Lorenzo-Seva and Ferrando (2021); Steiner and Grieder (2020)
	Model Fit Tests	Auerswald and Moshagen (2019); Finch (2020); Li et al (2020); Montoya and Edwards (2021); Steiner and Grieder (2020)
	Machine Learning	Goretzko and Bühner (2020)
	Exploratory Graph Analysis	Cosemans et al. (2022); Golino and Epskamp (2017); Golino et al., 2020
Factor Rotation Method	Regularized EFA	Scharf and Nestler (2019)
	Target Rotation	Zhang et al. (2019)
Interpreting Factor Loadings	Reassessing Factor Loading Strength	Current Article
Other	Exploratory Structural Equation Modeling	Asparouhov and Muthén (2009); Mai et al. (2018); Marsh et al. (2004); Morin et al. (2020); Perry et al. (2015)

Note. Citations are marked in the reference list of the primary text with an asterisk.

greater focus on cross-loadings, as a priori cutoffs for cross-loadings have increased over time. This focus on cross-loadings may be spurred by broader research on discriminant validity in the social sciences (Henseler et al., 2015; Rönkkö & Cho, 2020). While these approaches assess the discriminant validity of constructs, they do not identify indicators that may represent multiple constructs. Researchers are reliant on EFA to identify cross-loading indicators, emphasizing the importance of the analysis.

Almost all recent investigations are unsupportive of Harman's one-factor test (Fuller et al., 2016), including the present study. We call on researchers to investigate whether any EFA approach is suitable to assess common method bias. One-factor tests may work well with different cutoffs or specific EFA techniques. Researchers should use Table 5 to assess varying cutoffs. One-factor test results may even need to be assessed on a gradient, such that researchers can compare the variance explained by their first factor to our percentiles. Authors also differ regarding their use of rotation with one-factor tests, and one-factor tests should only be applied after simulation studies support specific specifications. Until then, authors should instead apply the marker technique to assess common method bias, which presently has firmer support (Richardson et al., 2009; Williams, Hartman et al., 2010; Williams, Onsman et al., 2010). Researchers should also preventatively address common method bias via research designs (e.g., longitudinal and multi-source), as preventive approaches are superior to post hoc assessments (Podsakoff et al., 2003, 2012).

We also highlight the integration of EFA with SEM due to its growing popularity and significant impact. Exploratory SEM utilizes EFA with rotations for the measurement portion of SEM, enabling the simultaneous modeling of cross-loadings and structural paths (Asparouhov & Muthén, 2009). This less restrictive approach can perform the same analyses as SEM, and it is known to perform more accurate assessments and produce better model fit in the study of multidimensional constructs (among other benefits) (Mai et al., 2018; Marsh et al., 2014; Morin et al., 2020; Perry et al., 2015). Our recommendations for conducting EFA are also important for exploratory SEM. Further, this integration of EFA and SEM is performed with covariance-based structural equation modeling (Mai et al., 2018; Marsh et al., 2014; Morin et al., 2020; Perry et al., 2015). We are not aware of research presently integrating EFA with partial least-squares structural equation modeling, but we expect our recommendations to hold across this application (Hair et al., 2020).

While open science practices are an established rather than emergent advancement (OSC, 2015, 2017), we feel the need to stress them again. It is likely that many authors fail to sufficiently report their EFAs because page lengths are limited in journals, but engaging in open science practices can resolve the tension between page limits and reporting requirements. Researchers should remain cognizant of their access to supplemental materials and online repositories, as these resources can ensure that EFAs are sufficiently reported and replicable.

Lastly, our review assumed that authors sought simple structure because this is most often the goal of researchers in management and associated fields of study, but some authors may not have sought simple structure. While we believe that such instances were rare, it should nevertheless be considered, and the current results should be interpreted with this caveat in mind. Future researchers should investigate when simple structure may not be desired as well as the EFA approaches that may be most ideal for such instances.

CRediT authorship contribution statement

Matt C. Howard: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data is provided as supplemental materials.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbusres.2023.113969.

References

(* Indicates reference included in Table 7)

Aguinis, H., Ramani, R., & Alabduljader, N. (2020). Best-practice recommendations for producers, evaluators, and users of methodological literature reviews. *Organizational Research Methods, 26*(1), 46–76.

Anthoine, E., Moret, L., Regnault, A., Sébille, V., & Hardouin, J. B. (2014). Sample size used to validate a scale: A review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes, 12*(1), 1–10.

*Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397–438.

*Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods, 24*(4), 468.

Bartlett, M. S. (1951). The effect of standardization on a Chi-square approximation in factor analysis. *Biometrika, 38*(3/4), 337–344.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., & Huck, S. W. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation, 18*(1), 1–13.

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology, 100*(2), 431.

*Braeken, J., & Van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods, 22*(3), 450.

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

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate behavioral research, 36*(1), 111–150.

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27*(4), 509–540.

*Caron, P. (2019). Minimum average partial correlation and parallel analysis: The influence of oblique structures. *Communications in Statistics-Simulation and Computation, 48*(7), 2110–2117.

*Chen, J. (2021). A Bayesian regularized approach to exploratory factor analysis in one step. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(4), 518–528.

Cole, M. S., Walter, F., Bedeian, A. G., & O'Boyle, E. H. (2012). Job burnout and employee engagement: A meta-analytic examination of construct proliferation. *Journal of Management, 38*(5), 1550–1581.

*Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., & Piatek, R. (2014). Bayesian exploratory factor analysis. *Journal of Econometrics, 183*(1), 31–57.

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods, 6*(2), 147–168.

*Cooperman, A. W., & Waller, N. G. (2022). Heywood you go away! Examining causes, effects, and treatments for Heywood cases in exploratory factor analysis. *Psychological Methods, 27*(2), 156–176.

*Cosemans, T., Rosseel, Y., & Gelper, S. (2022). Exploratory graph analysis for factor retention: Simulation results for continuous and binary data. *Educational and Psychological Measurement, 82*(5), 880–910.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(1), 1–9.

*Courtney, M. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-menu v2 0 to make more judicious estimations. *Practical Assessment, Research, and Evaluation, 18*(1), 1–14.

De Winter, J. C., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics, 39*(4), 695–710.

DeSimone, J. A., Brannick, M. T., O'Boyle, E. H., & Ryu, J. W. (2021). Recommendations for reviewing meta-analyses in organizational research. *Organizational Research Methods, 24*(4), 694–717.

*D'haenens, E., Van Damme, J., & Onghena, P. (2010). Multilevel exploratory factor analysis: illustrating its surplus value in educational effectiveness research. *School Effectiveness and School Improvement, 21*(2), 209–235.

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin, 81*(6), 358–361.

Ellingwood, H., Williams, K. M., Sitarenios, G., & Solomon, J. (2020). Psychometric properties of a contextualized, actuarially informed assessment for law enforcement personnel selection: The M-PULSE inventory. *Journal of Police and Criminal Psychology, 35*, 191–206.

Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299.

Finch, J. F., & West, S. G. (1997). The investigation of personality structure: Statistical models. *Journal of Research in Personality, 31*(4), 439–485.

Finch, W. (2006). Comparison of the performance of varimax and promax: Factor structure recovery for dichotomous items. *Journal of Educational Measurement, 43*(1), 39–52.

*Finch, W. H. (2020). Using fit statistic differences to determine the optimal number of factors to retain in an exploratory factor analysis. *Educational and Psychological Measurement, 80*(2), 217–241. * Indicates reference included in Table 7.

Fisch, C., & Block, J. (2018). Six tips for your (systematic) literature review in business and management research. *Management Review Quarterly, 68*(2), 103–106.

Flouri, E., Tsivrikos, D., Akhtar, R., & Midouhas, E. (2015). Neighbourhood, school and family determinants of children's aspirations in primary school. *Journal of Vocational Behavior, 87*, 71–79.

Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble. *European Journal of Psychological Assessment, 33*(6), 399–402.
Fong, T. C., & Ho, R. T. (2015). Dimensionality of the 9-item Utrecht Work Engagement Scale revisited: A Bayesian structural equation modeling approach. *Journal of Occupational Health*, 15–0057.
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*(2), 291–314.
Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common methods variance detection in business research. *Journal of Business Research, 69*(8), 3192–3198.
Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies, 51*(3), 511–521.
Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78.
*Golino, H., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS one, 12*(6), e0174035. * Indicates reference included in Table 7.
*Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., … Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods, 25*(3), 292–320.
*Goretzko, D. (2022). Factor Retention in Exploratory Factor Analysis With Missing Data. *Educational and Psychological Measurement, 82*(3), 444–464.
*Goretzko, D., & Bühner, M. (2020). One model to rule them all? Using machine learning algorithms to determine the number of factors in exploratory factor analysis. *Psychological Methods, 25*(6), 776–786.
*Goretzko, D., Heumann, C., & Bühner, M. (2020). Investigating parallel analysis in the context of missing data: A simulation study comparing six missing data methods. *Educational and Psychological Measurement, 80*(4), 756–774.
Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology, 40*(7), 3510–3521.
Greene, A. L., Watts, A. L., Forbes, M. K., Kotov, R., Krueger, R. F., & Eaton, N. R. (2022). Misbegotten methodologies and forgotten lessons from Tom Swift's electric factor analysis machine: A demonstration with competing structural models of psychopathology. *Psychological Methods*. Advance online publication.
Guo, L., Baruch, Y., & Russo, M. (2017). Work resources and civic engagement: Benefits to employee physical and mental health. *Journal of Vocational Behavior, 100*, 56–66.
Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2019). *Multivariate data analysis*. UK: Cengage.
Hair, J. F., Jr, Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research, 109*, 101–110.
Hayton, J., Allen, D., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191–205.
Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135.
Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of management, 21*(5), 967–988.
Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–121.
Hirschfeld, G., von Brachel, R., & Thielsch, M. (2014). Selecting items for Big Five questionnaires: At what sample size do factor loadings stabilize? *Journal of Research in Personality, 53*, 54–63.
Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction, 32*(1), 51–62.
Howard, M. C., & Henderson, J. (2023). A review of exploratory factor analysis in tourism and hospitality research: Identifying current practices and avenues for improvement. *Journal of Business Research, 154*, Article 113328.
Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior, 18* (6), 667–683.
Iacobucci, D., Ruvio, A., Román, S., Moon, S., & Herr, P. M. (2022). How many factors in factor analysis? New insights about parallel analysis with confidence intervals. *Journal of Business Research, 139*, 1026–1043.
Jain, P. M., & Shandliya, V. K. (2013). A survey paper on comparative study between principal component analysis (PCA) and exploratory factor analysis (EFA). *International Journal of Management, IT and Engineering, 3*(6), 415–424.
*Ji, Y., Luo, W., Lai, M. H., Yoon, M., Chen, L. S., & Kwok, O. M. (2021). Model-selection-based approaches to identifying the optimal number of factors in multilevel exploratory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(5), 763–777.
*Jung, S. (2013). Exploratory factor analysis with small sample sizes: A comparison of three approaches. *Behavioural Processes, 97*, 90–95.
*Jung, S., & Lee, S. (2011). Exploratory factor analysis for small samples. *Behavior Research Methods, 43*(3), 701–709.
*Jung, S., Seo, D. G., & Park, J. (2020). Regularized exploratory bifactor analysis with small sample sizes. *Frontiers in Psychology, 11*, 507.
*Jung, S., & Takane, Y. (2008). Regularized common factor analysis. *New trends in psychometrics*, 141–149.
Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist, 34*(5), 684–718.
Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika, 35*(4), 401–415.
Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36.
*Kim, E. S., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. *Multivariate Behavioral Research, 51*(6), 881–898.
Kulikowski, K. (2017). Do we all agree on how to measure work engagement? Factorial validity of Utrecht Work Engagement Scale as a standard measurement tool–A literature review. *International Journal of Occupational Medicine and Environmental Health, 30*(2), 161–175.
Ledesma, R. D., Ferrando, P. J., Trógolo, M. A., Poó, F. M., Tosi, J. D., & Castro, C. (2021). Exploratory factor analysis in transportation research: Current practices and recommendations. *Transportation Research Part F: Traffic Psychology and Behaviour, 78*, 340–352.
*Levy, R., Xia, Y., & Green, S. B. (2021). Incorporating uncertainty into parallel analysis for choosing the number of factors via Bayesian methods. *Educational and Psychological Measurement, 81*(3), 466–490.
Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods, 24*(4), 452–467.
Loo, R. (1979). The orthogonal rotation of factors in clinical research: A critical note. *Journal of Clinical Psychology, 35*(4), 762–765.
Lorenzo-Seva, U., & Ferrando, P. J. (2021). Not positive definite correlation matrices in exploratory item factor analysis: Causes, consequences and a proposed solution. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(1), 138–147.
Luo, L., Arizmendi, C., & Gates, K. M. (2019). Exploratory factor analysis (EFA) programs in R. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(5), 819–826.
MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99.
*Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(5), 737–749.
Malhotra, N. K., Kim, S. S., & Patil, A. (2006). Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management Science, 52*(12), 1865–1883.
Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of Sport Psychology* (3rd ed., pp. 774–798). New York: Wiley.
Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11* (3), 320–341.
Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*(3), 471–491.
*Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85–110. * Indicates reference included in Table 7.
Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research, 3*(1), 97–110.
*McNeish, D. (2017). Exploratory factor analysis with small samples and missing data. *Journal of Personality Assessment, 99*(6), 637–652.
Mills, M. J., Culbertson, S. S., & Fullagar, C. J. (2012). Conceptualizing and measuring engagement: An analysis of the Utrecht Work Engagement Scale. *Journal of Happiness Studies, 13*(3), 519–545.
*Montoya, A. K., & Edwards, M. C. (2021). The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. *Educational and Psychological Measurement, 81*(3), 413–440.
*Morin, A. J., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM), and bifactor-ESEM. *Handbook of Sport Psychology, 51*, 1044–1073.
Nassiri, V., Lovik, A., Molenberghs, G., & Verbeke, G. (2018). On using multiple imputation for exploratory factor analysis of incomplete data. *Behavior research methods, 50*(2), 501–517.
Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods, 17*(4), 372–411.
Norris, M., & Lecavalier, L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychological research. *Journal of Autism and Developmental Disorders, 40*(1), 8–20.
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251).
Open Science Collaboration. (2017). Maximizing the reproducibility of your research. Psychological science under scrutiny: Recent challenges and proposed solutions. In S. Lillenfeld, & I. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 1–21). Wiley.
Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research, and Evaluation, 20*(1), 1–7.
Osborne, J. W., Costello, A. B., & Kellow, J. T. (2014). *Best practices in exploratory factor analysis*. Louisville, KY: CreateSpace Independent Publishing Platform.
Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research, 28*(4), 562–577.
*Perry, J. L., Nicholls, A. R., Clough, P. J., & Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural

equation modeling. *Measurement in Physical Education and Exercise Science, 19*(1), 12–21.
Peterson, R. A. (2000). A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing Letters, 11*, 261–275.
Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.
Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*(1), 13–43.
Reio, T. G., Jr, & Shuck, B. (2015). Exploratory factor analysis: Implications for theory, research, and practice. *Advances in Developing Human Resources, 17*(1), 12–25.
Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods, 12*(4), 762–800.
Rönkkö, M., & Cho, E. (2020). An updated guideline for assessing discriminant validity. *Organizational Research Methods, 1094428120968614*.
Rouquette, A., & Falissard, B. (2011). Sample size requirements for the internal validation of psychiatric scales. *International Journal of Methods in Psychiatric Research, 20*(4), 235–249.
Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment, 24*(2), 282.
Sakaluk, J. K., & Short, S. D. (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.
Saks, A. M., & Gruman, J. A. (2014). What do we really know about employee engagement? *Human Resource Development Quarterly, 25*(2), 155–182.
*Scharf, F., & Nestler, S. (2019). Should regularization replace simple structure rotation in exploratory factor analysis? *Structural Equation Modeling: A Multidisciplinary Journal, 26*(4), 576–590.
Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies, 3*(1), 71–92.
Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement, 71*(1), 95–113.
Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of applied psychology, 78*(6), 966.
Sonnentag, S. (2003). Recovery, work engagement, and proactive behavior: A new look at the interface between nonwork and work. *Journal of Applied Psychology, 88*(3), 518–528.
Spicer, J. (2005). *Making sense of multivariate data analysis: An intuitive approach*. Sage.
*Steiner, M. D., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software, 5*(53), 2521. * Indicates reference included in Table 7.
Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics*. Boston, MA: Pearson.
Tobias, S., & Carlson, J. E. (1969). Brief report: Bartlett's test of sphericity and chance findings in factor analysis. *Multivariate Behavioral Research, 4*(3), 375–377.
Trendafilov, N. T., Unkel, S., & Krzanowski, W. (2013). Exploratory factor and principal component analyses: Some new aspects. *Statistics and Computing, 23*(2), 209–220.
Truninger, M., Ruderman, M. N., Clerkin, C., Fernandez, K. C., & Cancro, D. (2020). Sounds like a leader: An ascription–actuality approach to examining leader emergence and effectiveness. *The Leadership Quarterly, 101420*.
Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25*(1), 1–28.
Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219–246.
Watkins, M. W. (2021). *A step-by-step guide to exploratory factor analysis with Stata*. Routledge.
Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*(3), 1–13.
Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method variance and marker variables: A review and comprehensive CFA marker technique. *Organizational Research Methods, 13*(3), 477–514.
Willmer, M., Westerberg Jacobson, J., & Lindberg, M. (2019). Exploratory and confirmatory factor analysis of the 9-Item Utrecht Work Engagement Scale in a multi-occupational female sample: A cross-sectional study. *Frontiers in Psychology, 10*, 2771.
Wilson, G. A., & Martin, S. A. (1983). An empirical comparison of two methods for testing the significance of a correlation matrix. *Educational and Psychological Measurement, 43*(1), 11–14.
*Xiao, C., Bruner, D. W., Dai, T., Guo, Y., & Hanlon, A. (2019). A comparison of missing-data imputation techniques in exploratory factor analysis. *Journal of Nursing Measurement, 27*(2), 313–334.
Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research, 39*(1), 93–112.
Yeomans, K. A., & Golder, P. A. (1982). The Guttman-Kaiser criterion as a predictor of the number of common factors. *The Statistician*, 221–229.
Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79–94.
*Zhang, G., Hattori, M., Trichtinger, L. A., & Wang, X. (2019). Target rotation with both factor loadings and factor correlations. *Psychological Methods, 24*(3), 390–402.