Teaching English as a Second Language Quarterly
(TESLQ)
(Formerly Journal of Teaching Language Skills)
43(1), Winter 2024, pp. 151-178
https://doi.org/10.22099/tesl.2024.49564.3264

Online ISSN: 2717-1604
Print ISSN: 2008-8191
Research Paper

Considerations and Praxis of Exploratory Factor Analysis: Implications for L2 Research

Reza Rezvani *   **Hessameddin Ghanbar** **
Kyle Perkins ***

Abstract

Second language (L2) researchers often rely on quantitative methods and measurement instruments like questionnaires and scales to explore latent constructs. They usually borrow sophisticated statistical techniques from disciplines like psychology and education to conduct quantitative analyses in L2 research. However, there are growing concerns about the inappropriate use and reporting practices of such statistical procedures like exploratory factor analysis (EFA). This study aimed to describe and evaluate the methodological issues in EFA research practices, examine the relationship between study features and outcomes, and improve future L2 research practices. This study presented the results of a methodological synthesis study on exploratory factor analysis use published in five universally reputable psychology-related journals since 2000. Specifically, the researchers identified 93 EFA studies and developed a coding scheme of key EFA considerations to analyze how EFA was used and reported in these articles. The article discussed the several results and provides essential recommendations for L2 researchers intending to employ EFA in their works.

Keywords: Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis, Methodological Synthesis, L2 Research

Since most language constructs of interest are latent in nature in L2 research, they cannot be directly measured (Alamer et al., 2024; Gass et al., 2020). Second language (L2) researchers often tend to quantify these constructs and employ various quantitative research and assessment methods (Loewen & Plonsky, 2016). To collect data, they also

* Received: 11/12/2023   Accepted: 01/03/2024
* Associate professor, Yasouj university/Shiraz University, Iran, rezvanireza@gmail.com
** Assistant professor, Department of Language and Linguistics, Islamic Azad University, Fereshtegaan International Branch, Tehran, Iran, hessam.ghanbar@gmail.com (Corresponding Author)
*** Professor, Florida International University, USA, perkins@fiu.edu

How to cite this article:
Rezvani, R., Ghanbar, H., & Perkins, K. (2024). Considerations and Praxis of Exploratory Factor Analysis: Implications for L2 Research. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills), 43*(1), 151-178. doi: 10.22099/tesl.2024.49564.3264

rely on measurement instruments such as questionnaires and scales (Dörnyei & Dewaele, 2022; Iwaniec, 2019; Ponto, 2015; Sudina, 2021, 2023). These instruments are commonly subject to rigorous psychometric examination to establish their reliability and validity (Flake et al., 2017; Kim, 2009).

The field of L2 is relatively new and still developing (De Bot, 2015; Plonsky & Gonulal, 2015). Despite the historical dominance of quantitative methods, the statistical techniques used in L2 have not originated from within the field itself (Loewen & Gass, 2009). Although L2 research has become "more sophisticated in its use of statistics" (Gass, 2009, p. 19), researchers in this area have relied on methodological practices from other disciplines (Selinker & Lakshmanan, 2001). Specifically, statistical techniques are often borrowed from sister disciplines such as psychology and education (Plonsky & Gonulal, 2015). Interestingly, many L2 scholars take statistics course from these departments (Loewen et al., 2014). This reliance becomes particularly evident when it comes to sophisticated statistical procedures like structural equation modeling and factor analysis (Gass, 2009; Ghanbar & Rezvani, 2023, & 2024; Lazaraton, 2000, 2005; Loewen et al., 2014, Plonsky & Gonulal, 2015).

Despite the common use of exploratory factor analysis (EFA) in L2 research, researchers often make questionable decisions and use inappropriate reporting practices when conducting these analyses. The appropriate use of statistics can significantly impact the quality of research and its reception within the L2 research community. Conversely, a lack of understanding and misuse of statistical procedures in quantitative L2 studies can threaten the credibility of research findings and weaken the legitimacy of the field (Loewen et al., 2020). This article explores EFA use and publication in educational psychology journals that routinely feature such research. The implications of the results are also discussed for L2 research. The study ended with some essential recommendations about how to effectively conduct EFA and report the findings in L2 research to improve methodological rigor and transparency.

Factor Analysis

Factor analysis (FA) entails the use of structure-analyzing procedures to determine the extent to which measures for different variables measure the same thing (i.e., measurement overlap). The procedure evaluates the pattern of variance between the multiple measured items (Wang et al., 2013). FA is "often used to explain a larger set of j measured variables with a smaller set of k latent constructs" (Henson & Roberts, 2006, p. 394). It is assumed that the relationships among the observed variables is due to the underlying latent variables. It is further assumed that the unobserved variables, the factors, account for the correlations of the observed measures because the underlying latent variable influenced them (Ding, 2013). The fact that the number of factors is smaller than the number of measured variables leads to a parsimonious account of the covariation among a set of variables (Brown, 2013).

In addition to identifying the latent relationships underlying multiple measured items, FA also produces factor loadings that provide an indication of the strength of the relationship between each item and the factor that is common to all the items in the scale (Wang et al., 2013). Another product of a FA is a quantitative dimensional representation of the data structure (Ding, 2013). FA utilizes either a correlation matrix or a covariance matrix to separate variance into two distinct components (Brown, 2013):

> (1) common variance or the variance accounted for by the latent variable(s), which is estimated on the basis of variance shared with other indicators in the analysis; and (2) unique variance, which is a combination of reliable variance specific to the indicator (i.e., systematic latent factors that influence only one indicator) and random error variance (i.e., measurement error or unreliability in the indicator). (p. 257).

Broadly, FA has been used for data reduction, hypothesis testing, test and survey development, test and survey validation and theory development (Tabachnick & Fidell, 2013). More specifically, it has been used to (1) explain a number of observed variables by a smaller set of latent variables, and (2) to determine (a) how many latent variables underlie a data set, (b) which variables are located on each factor, (c) if the factor structure is reliable, and (d) if the variables form a construct by a sufficient association with one another (Thorkildsen, 2005).

FA involves interpretation which entails at least three steps, according to Thorkildsen (2005). First, the researcher must verify that all the items are associated with the same construct in the initial (nonrotated) factor solution. The initial solution is supposed to provide the common factor variance among all the variables but does not identify meaningful dimensions. The statistical program squares and adds the loadings of variables that are highly correlated in each row of a factor matrix to estimate the shared variance. This phenomenon is called a *communality index* (Tabachnick & Fidell, 2013).

The second step involves statistical rotation, the purpose of which is to determine the number of dimensions in a particular construct and for the best position for the axes used to represent relationships among the variables in order to improve interpretation. It should be noted that different methods of rotation reflect different theoretical assumptions about the degree to which dimensions of a construct are independent or moderately correlated with one another. When dimensions are assumed to be independent, the axes along which factor loadings are rotated starts in an orthogonal position, at a 90-degree angle. When dimensions are assumed to be moderately correlated, an oblique rotation is used in which the angles between the axes are acute or obtuse (Thorkildsen, 2005, p. 89).

The third step entails determining which factors in the final rotated solution are significant. Researchers can use eigenvalues greater than 1 for this purpose in the following way: for each factor, square the factor loadings vertically for each column in the component matrix.

Thorkildsen (2005) offered the following guidelines for deciding whether dimensions are reliable. It should be mentioned that dimensions are reliable if they contain four or more indicants [variables] with loadings that are each greater than the absolute value of |.60|, regardless of sample size; contain three or more indicants with loadings that are each greater than |.80|, regardless of sample size; contain ten or more indicants with low loadings (near +/- .40) for sample sizes greater than 150.

Exploratory Factor Analysis (EFA)

There are two general types of factor analysis: exploratory factor analysis and confirmatory factor analysis. The works by Spearman (1904) and Thurstone (1935, 1947) are credited as the foundation of EFA. It was not until CFA emerged that the term 'exploratory' was widely added to Factor Analysis to differentiate it from CFA (Marsh & Alamer, 2024; Morin, 2023). According to Henson and Roberts (2006, p. 395) “EFA is an exploratory method used to generate theory” and to identify a smaller set of latent factors to represent a larger set of measured variables. Researchers using EFA have no hypotheses about the number of factors that underlie the input data nor about the factor loadings (i.e., the pattern of relationships between the common factors and the indicators). EFA is data-driven (Brown, 2013), and it is “a method of data reduction that provides an economical description of correlational data” (Haig, 2013, p. 21).

Other characteristics of EFA include the following: the loading for each item is estimated on all factors; chi-square and other fit statistics can be used for model fit; and loadings are estimated for all variables on all factors, even if they are very weak loadings (Wang et al., 2013). The latent factors postulated by EFA are referred to as common factors (Haig, 2013). Haig also described EFA as an abductive method of theory generation by generating "explanatory inference that leads back from presumed effects to understanding causes" (p. 21).

Confirmatory Factor Analysis (CFA)

Researchers generally use CFA to test theory (Henson & Roberts, 2006), because researchers develop theory-driven hypotheses concerning the underlying factor structure, the fit of the hypothesized factor structure to the data, the number of factors, which items load on which factors, and an orthogonal or an oblique relationship between the factors (Tabachnick & Fidell, 2013). These advantages missing in the EFA has sparked widespread interest in CFA among researchers to explore a priori hypothesized models (Marsh & Alamer, 2024).

Researchers use theoretical and empirical information in order to specify and to evaluate the factor model. For this reason, CFA is usually used in the later stages of construct validation and scale development. As a complement to EFA, CFA is conceived as a measurement model, which can be characterized as a statistical framework that elucidates the associations between a latent construct and the observable variables (e.g., items of a questionnaire) designed to measure it. In essence, it serves as a structured representation of how the underlying construct is reflected in the observed data, providing a means to assess the validity and reliability of the measurements employed in a study (Tabachnick & Fidell, 2019).

Thurstone's (1947) common factor model serves as the foundation for CFA and its purpose is to reproduce the observed relationships underlying a data set with a smaller set of latent variables. Frequently, a CFA follows an EFA. The input for a CFA is a variance-covariance matrix with variances on the diagonal and covariances in the off-diagonal. CFA results are parameter estimates (Brown, 2013).

Statistical Assumptions and Practical Considerations in EFA

There are several statistical assumptions and methodological considerations when conducting exploratory factor analysis (EFA). The first important consideration is the ratio of participants to variables. There is no consensus on the exact number of subjects or items per variable that are required, with estimates ranging from 3 to 20 (Gorsuch, 1990, 2003; Tabachnick & Fidell, 2013; Thompson, 2004). However, the most common suggestion is to have 10 to 15 subjects or items per variable (Field, 2009, Plonsky & Gonulal, 2015). Perhaps, the use of a power test to determine the number of subjects would ameliorate the rules of thumb. For example, Mertler and Vannatta (2013, p. 11) point out that "the required sample size for a study is a function of the level of significance or alpha level, power, and effect size. Because sample size has several relationships with these three factors, values for the factors must be set prior to the selection of a sample".

A second issue is the missing data. Data can go missing at random (MAR), missing completely at random (MCAR), or missing not at random (MNAR). In the first case, MAR, the missing data point may be related to the observed data. In the second case, MCAR, the missing data point has no relationship with other variables' values nor with its hypothetical value. In the last case, MNR, the phenomena are exactly opposite to those of MCAR.

There are deletion strategies to deal with missing values: deleting rows (listwise deletion), pairwise deletion, and deleting columns. Missing data imputation techniques include computing the overall mean, median, or mode and substituting that value for the missing data point(s)., which may lead to a reduction in variance in the data set. Regression can also be used for missing data imputation.

Regarding missing data, the variable with missing values becomes the dependent variable. Cases with complete data are used to developing this prediction equation. The equation is then used to predict missing cases. One disadvantage of regression is that the predicted scores are better than they actually would be (Mertler & Vannatta, 2013, p. 29).

A Markov Chain Monte Carlo simulation might be a better alternative. A third issue is normality. Skewness, kurtosis, and the Kolmogorov-Smirnov statistic, with Lilliefors significance level are typically used to assess univariate normality. Bivariate normality can be assessed by examining each pair of variables to determine if their scatterplot is elliptical. Deviant data should be transformed. A fourth issue is linearity. A straight-line relationship between two variables suggests linearity. Nonlinearity is also specified "through the examination of residual plots. Residuals are defined as the portions of scores not accounted for by multivariate analysis. If standardized residual values are plotted against the predicated values, nonlinearity will be indicated by a curved pattern to the points" (Mertler & Vannatta, 2013, p. 34).

Fifth in the list of issues concerns outliers. Outliers can be the result of recording mistakes or a participant's not being a member of the population under study. Outliers can distort the results of statistical procedures because many of them rely on squared deviations from the mean. Box plots can be used to identify univariate outliers, and Mahalanobis distance, i.e., the distance of a case from the centroid of the remaining cases, can be used to identify multivariate outliers (Tabachnick & Fidell, 2007). Liu et al., (2012) reported that very few studies discuss the effect of outliers on EFA performance.

Issue six is multicollinearity and singularity. Watson (2017) recommended the use of an interitem correlation matrix to determine if the data collected for an EFA are factorable. An interitem correlation matrix resulting in a majority of the correlation coefficients ranging between 0.20 and 0.80 is deemed to be factorable. Such a matrix can also identify occurrences of multicollinearity ($r > 0.80$) and singularity ($r = 1.00$). Items involved in multicollinearity and singularity should be removed from the data set before it is submitted to EFA. EFA researchers typically use the Kaiser-Meyer-Olkin (KMO) test for sampling adequacy and Bartlett's test of sphericity to estimate the extent to which the interitem correlation matrix is an identity matrix. Researchers generally use a threshold of > 0.60 for the KMO test and $p < .05$ for the Bartlett's test. It should be said that a major issue facing a psychologist intendent upon using EFA is the choice of a factor extraction method. Some of the options include principal component analysis, unweighted least squares, generalized least squares, principal axis factoring, alpha factoring, image factoring, and maximum likelihood. The choice of extraction method should reflect the goal of the proposed study and the research question(s) presented therein.

Principal component analysis (PCA) analyzes variance—all sources of error, shared, and unique variability for each variable, and "summarizes many variables into fewer components" (Henson & Roberts, 2006, p. 398). The procedure creates uncorrelated linear combinations of the observed variables. The first (principal) component has maximum variance, with each of the following, successive components explaining smaller portions of the variance. Unweighted least squares minimizes the sum of the squared differences between the observed and reproduced correlation matrices, ignoring the diagonals. Additionally, Generalized Least Squares (GLS) assumes that the observations are uncorrelated. Principal axis factoring (PAF) analyzes covariance, specifically focusing on the common variance among the items, thereby focusing on the latent factor(s). It is a preferred procedure when multivariate non-normality is an issue. The procedure extracts factors from the original correlation matrix. Squared multiple correlation coefficients are placed in the diagonals as estimates of the communalities. New communalities are estimated from these factor loadings, replacing the old communality estimates in the diagonals. This process continues until the changes in the communalities from one iteration to the next satisfy the convergence criterion for extraction. Another method, alpha factoring, considers the variables in the study to be a sample from the universe of potential variables. It maximizes the alpha reliability of the factors. Guttman developed image factoring which is based on image theory. The partial image, the common part of the variable, is defined as its linear regression on the remaining variables, rather than a function of hypothetical factors. The next estimation method, Maximum likelihood (ML), requires relatively normally distributed input data and produces parameter estimates that are likely to have produced the observed correlation matrix, if the sample input data were from a multivariate normal distribution (see Ghanbar & Rezvani, 2023 for more information on model estimation methods and its statistical considerations). ML has several desirable characteristics in that it "provides several indexes of goodness of fit, permits statistical significance testing of factor loadings, and computes intercorrelations among factors" (Watson, 2017, p. 233).

Rotation methods are a major decision issue facing the L2 researcher. Rotation makes a factor solution more interpretable without altering the underlying structure. Factors are rotated to more desirable positions "to maximize high loadings, minimize low loadings, and create the simplest factor structure" (Watson, 2017, p. 234). A researcher has two rotation options from which to choose, orthogonal and oblique. The choice of rotation depends upon the hypothesized relationship among the variables expected in the study based on compelling research and theory. The choice of a rotation method should be compatible with the goals of the study (Meyers et al., 2013).

If a researcher assumes that uncorrelated factors represent some unique aspect of the underlying structure, then s/he should use orthogonal rotation (Mertler & Vannatta, 2013). "The varimax option maximizes the variance across factors (Dimitrov, 2012) and is most easily interpreted (DeVellis, 2022). However, when the researcher suspects that the content being studied contains a single overall factor, quartimax and equamax might be more appropriate choices" (Watson, 2017, p. 234).

If a researcher, based on theory, research, and the goals of the study, expects minor to moderate interitem correlations, an oblique rotation should be used (see Tabachnick & Fidell, 2019). An oblique rotation method generates a factor pattern matrix which can be used to determine the extent to which a simple structure has been achieved, a factor structure matrix, and a factor correlation matrix (see Tabachnick & Fidell, 2019 for elaborated discussions on different types of rotation methods).

With the growing prevalence of studies utilizing scales, questionnaires, and EFA, there is a concerning lack of statistical literacy among L2 researchers (Loewen et al., 2014). Additionally, there is a notable absence of field-specific standards for the proper use and reporting of these techniques (Plonsky & Gonulal, 2015, p. 18). Our goal is to enhance statistical literacy and improve methodological rigor and transparency in the application of EFA within various areas of L2 research. Building upon the insights, we

aim to offer empirically supported recommendations and suggestions that can help L2 researchers effectively incorporate EFA practices into their studies. In order to accomplish these objectives, the study addresses the following research questions:
1-How and in what ways EFA is utilized in the field of educational psychology?
2-How do researchers adhere to the standards of rigor in conducting EFA?

The purpose of this paper was to present the results of a methodological synthesis study of exploratory factor analysis research published since 2000 in five psychology-related, peer-reviewed journals. According to Plonsky and Ghanbar (2018), a methodological synthesis "involves applying synthetic/meta-analytic techniques to a body of primary research. Unlike substantively-oriented syntheses, the focus of this type of research is not so much on the findings of studies but rather on their data analytic and reporting practices" (p. 5). Plonsky and Gonulal (2015) also pointed out that the goals of methodological synthetic studies involve describing and evaluating methodological issues (e.g., the purpose of such studies), examining the relationship between study features and outcomes, exploring the extent to which researchers used statistical methods in accordance with standards of methodological rigor, critically evaluating the statistical practices that researchers used, and identifying and explicating the major considerations in the application of statistical procedures. A methodological synthesist's critical evaluation provides a window into methodological culture with the intent of promoting more methodologically informed research practice and of improving future research.

There are well-established procedures for conducting methodological synthetic studies. First, a researcher must identify a set of parameters for inclusion and exclusion, followed by a systematic collection of peer-reviewed primary studies based on the aforementioned parameters (see Morea & Ghanbar, 2024 for a systematic approach in synthesizing a methodological technique). Step two involves the construction of a coding scheme or coding sheet which research will use to extract and to record the relevant information from each study. At this juncture, it should be noted that primary studies are viewed as participants which yield methodologically-oriented data. And finally, the researcher analyzes the data.

Method

Study Retrieval

We conducted an extensive search of the ERIC, PsycINFO, and JSTOR databases using the keywords "exploratory factor analysis" from the year 2000 onwards. We carefully reviewed the title, abstract, keywords, and occasionally the methodology sections of the studies to identify papers that utilized EFA as a primary method and reported the results of the analysis. Articles that mentioned EFA only in passing or as background information for new research were excluded from the study. After completing the search and screening process, we identified a total of 176 articles involving EFA. From this pool, we randomly selected 15 papers from each of the reputed journals of Educational and Psychological Measurement, Psychological Assessment, Journal of Educational Psychology, Measurement and Evaluation in Counseling and Development, and Journal of Counseling Psychology. Some papers in the sample (n = 75) included multiple instances of EFA, resulting in a final sample of 75 papers and 93 EFA studies for our analysis.

Data Collection

We developed a coding scheme which was used as the data collection instrument to give us an understanding of how EFA was used and reported in the educational psychology research community. Our coding scheme was based on recommendations, findings, and suggestions offered in statistical texts and previous studies (e.g., Field, 2013; Henson & Roberts, 2006; Mertler & Vannatta, 2013; Plonsky & Ghanbar, 2018; Plonsky & Gonulal, 2015; Riazi et al., 2023; Tabachnik & Fidell, 2013; Watson, 2017).

The coding scheme included seven sections: (1) article information, (2) statistical assumptions and considerations, (3) reporting practices, (4) extraction method issues, (5) rotation method issues, (6) factor retention criteria, and (7) CFA issues. In the first section of our coding, we focused on identifying key information such as the author(s)' name(s), the name of the journal, the year of publication, and the statistical software used in the study. Moving on to the second section, we examined details such as the number of participants, variables, participant to variable ratio, and whether crucial information like missing values, normality, linearity, outliers, multicollinearity, and singularity were reported in the study. Section three involved coding for the presence of a correlation matrix, KMO, Bartlett's test, reproduced matrix, cumulative percentages of variance for extracted factors, number of extracted factors, total variance explained, anti-image matrix, univariate descriptives, scree plot, factor scores, variance for retained factors, eigenvalues, factor loading matrix, communalities, loading magnitude, statistical power, and factor reliability. Moving on to section four, we focused on extraction method, principal component analysis, unweighted least squares generalized least squares, principal axis factoring, alpha factoring, image factoring, maximum likelihood, and extraction criteria. Section five concerned the rotation method issues, such as the use of single or multiple rotation methods, justification, orthogonal versus oblique techniques, and reporting of coefficients and delta values. In section six, factor retention criteria were examined, including the use of Kaiser-1, Joliffe's criterion, scree plot inspection, parallel analysis, and minimum average partial correlation. Finally, in section seven, we assessed whether a CFA was warranted, and if so, the rationale given for not using a CFA.

In developing our coding scheme, we were informed by various sources such as Riazi et al. (2023) and Ghanbar and Rezvani (2023). To maintain the integrity and precision of our data coding process, we initially coded a random selection of 10 studies, encompassing a total of 12 EFA analyses, independently. Subsequently, we engaged in discussions to address any discrepancies and challenges encountered during the coding process. Following this initial phase, each team member proceeded to independently code a random sample of 25 articles.

Data Analysis

We computed the median, mean, and standard deviation for the continuous variables, (e.g., sample size, number of variables factored, ratio of sample size to number of factored variables, number of factors extracted, and total variance explained). We also reported the minimum and maximum values for the continuous variables. For the categorical variables, we reported the frequencies and percentages.

Results and Discussion

EFA General Issues

Totally, we examined 93 uses of EFA (*k* = 75, representing the number of articles we reviewed) in the five targeted journals from 2000 to 2019. As it was mentioned elsewhere in this article, we believe that this number roughly portrayed a clear picture of EFA usage in these journals in the field of psychology. The first feature of EFA we investigated is the statistical software utilized in EFA implementation. The most widely used statistical software was SPSS, exploited in 39 EFAs (41.9%), and the least used ones were R (4, 4.3%) and SAS (1, 1.1%). Interestingly, the majority of EFA uses did not contain any information about the statistical program (42, 45.2%).

Table 1.

General Descriptive Results of EFA Reporting Practices

Variable	n	Median	M	SD	Minimum	Maximum
Sample size	93	429	810.36	1256.28	88	8912
Number of variables factored	93	25	32.78	22.98	6	115
Ratio of sample size to no. of factored variables	93	15	47.46	99.68	1.52	658.54
Number of factors extracted	93	4	3.9	2.42	1	20
Total variance explained	69	56%	56.46%	13.47%	27%	95%

Note: n = number of EFA containing that information
a. Indicates that there were 15 participants per one variable

As it was shown in Table 1, sample size is the first global EFA variable that we investigated in our work. EFA is a correlation-based technique; therefore, sample size is one of the prime determinants of a sound EFA application (Pituch & Stevens, 2016, Tabachnick & Fidell, 2013), boosting the reliability of the analysis. Our distribution of sample size was quite non-normal (highly positively skewed, with coefficient of skewness of 4.22), ranging from 88 to 8912 (median = 429, interquartile range = ۶۶۸). There exists a rich literature on the necessary sample size for EFA, proposing myriad of rules of thumb. For example, Field (2013) mentioned that the common rule of thumb is having at least 10-15 cases per variable, Hair et al., (2010) proposed a minimum of 100, or Tabachnick and Fidell (2013) recommended the maximum of 500. Other sources such as Comrey and Lee (1992) offered an explicit guideline of 50 as very poor, 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1000 as excellent. Additionally, Stevens (2009), from the standpoint of component saturation, suggested using a “subjects per variable” criterion of 5 to 20 participants per variable. In our study, the median ratio of cases to variables was 15:1, revealing that sample sizes were approximately large enough with regard to component saturation and showed an improvement of five units in comparison with the results of the previous review of EFA application in psychology literature (Henson & Roberts, 2006).

Nonetheless, as Henson and Roberts (2006) pointed out, the most important drawback of the aforementioned guidelines is that they failed to take into account the labyrinth of EFA implementation, as it depends on a variety of factors such as the size of the population correlation, the number of extracted factors, and the magnitude of communalities (e.g., greater than 0.6 needs lower sample size) and the loadings of items (e.g., loadings of more than .80 do not require more than 150 cases). Regarding investigating communalities, researchers have a tendency to not report checking them, with only 22.6% (k = 19) of EFAs containing the pertaining information. To conclude, given the multifaceted nature of choosing sample size, the wisest recommendation is to choose the largest possible sample size so as to guarantee the reliability of correlation coefficients.

The next important finding is related to the cumulative explained variance by a factor structure. On average, the extracted factors in our sample represented 59% of the total variance, which is less than “the minimum 70%” recommended by Stevens (2009, p. 329) and within a range of 55-65% proposed by Field (2013). Although there is an improvement in the explained variance in comparison with that of 52% found in Henson and Roberts’s (2006) review of EFAs, some studies (k = 20) accounted for less than 50 % of the total variance, and that average variance of 59% can also be considered moderate and at best. The potential grounds for this finding might be (a) a meaningful, uninterpretable factor structure was not extracted, (b) the items of instruments did not sufficiently represent the extracted factors, (c) the burdensome nature of working out a compromise between explaining the maximum variance and model parsimony, and (d) questionable reasonableness of benchmarks proposed by Field (2013) or Stevens (2009), as several underlying factors such as the nature of construct under investigation or sample size influence the variance extracted.

Henson and Roberts (2006) mentioned the relationship between variance extracted and the number of items, noting that, in their study, “the proportion of total variance explained tended to decrease as the total number of items factored increased” (p. 403). They explained that the retention of “items that add more unexplained than explained variance to the model” (p. 403) could be the cause of the problem. A potential solution might be to delete the items with the largest measurement error variance from the model. Using the jack-knife sampling procedure might also help to identify those item that are attenuating the correlation. We recommend that items with large measurement error variance and items that contribute more unexplained than explained variance to the model be deleted. Hopefully, the remaining items in the model will increase the cumulative explained variance by a factor structure closer to the 70 percent threshold recommended by Stevens (2009).

Table 2.

Frequencies and Percentages of Articles Reporting EFA Information

Feature	n	%
Statistical Software		
SPSS	39	41.9
R	4	4.3
SAS	1	1.1
Other	7	7.5
Not Reported	42	45.2
Checking Statistical Assumptions and Considerations		
Missing values	29	31.2
Normality	36	38.7
Linearity	9	9.7
Absence of Outliers Among Cases	15	16.1
Absence of Multicollinearity & Singularity	7	7.2
Reporting Practices		
Reported correlation Matrix	9	9.7
Reported KMO	39	41.9
Reported Bartlett	33	35.3
Reporduced matrix	0	0
Reported variance explained	61	65.6
Reported Anti-image matrix	2	2.2
Reported descriptives statistics	37	36.8
Reported factor scores	2	2.2
Reported EV for retained factors	33	35.5
Reported factor loading matrix	48	51.6
Reported communalities	21	22.6
Loading magnitude		
.3-.39	27	29
.4-.49	25	26.9
.50 or higher	11	11.8
Not reported	30	32.3
Reported statistical power	6	6.5
Reported reliability of variables	65	69.9
Extraction Method issues		
Reporting the exact name of method	84	90.3
Principal Components Analysis (PCA)	26	28
Principal Axis Factoring	31	33
Unweighted Least Squares	0	0
Generalized Least Squares	0	0
Alpha factoring	2	2.2
Image Factoring	0	0
Maximum Likelihood (ML)	27	29
Rotation Method Issues		
Are a single method or multiple methods used?		
Single method	67	72
Multiple method	15	16.2

Not reported	11	11.8
Rotation justification Reported	47	50.5
Rotation method type		
n= 82		
Orthogonal	18	21.95
Oblique	58	70.73
Both	6	7.32
If orthogonal, which type is used?		
Varimax	24	100
Quartimax	0	
Euamax	0	
Orthomax	0	
Parsimax	0	
Other	0	
If oblique, which type is used?		
Direct Oblimin	25	39.1
Direct Quartimin	0	0
Orthoblique	0	0
Promax	26	40.6
Procrustes	6	9.4
Geomin	2	3.1
Not mentioned	5	7.8
If direct oblimin, delta value is given?		
n= 25		
Yes	6	24
No	19	76
If promax, kappa value is given?		
n= 26		
Yes	11	42
No	15	58
If Oblique, coefficients reported?		
n= 64		
Pattern Matrix only	28	44
Structure Matrix only	3	5
Both	14	22
Not Reported	19	29
Factor Retention Decisions		
Are a single criterion or multiple criteria used?		
Single method	33	35.5
Multiple methods	50	53.8
Not reported	10	10.8
Type of retention method		
Kaiser's criterion	60	64.5
Joliffe's criterion	0	0
Scree test	38	40.9
Parallel analysis (PA)	46	49.5
Minimum average partial (MAP)	10	10.8

CFA Suggested			
	Yes, because it is not a new instrument	12	3.2
	No, because it is a new instrument	17	15.1
	Yes, both EFA and CFA done	64	68.8
If CFA warranted, reasons given for not using			
n= 12			
	sample size	2	16.6
	No strong Theory	3	25
	Other	0	0
	Not mentioned	7	58.4

Reporting Practices of EFA in Research

We also investigated the elements of EFA reports provided by researchers. As can be seen in Table 2, the first feature of EFA practices that we focused on is investigation of its underlying statistical assumptions and considerations by researchers. None of the practical issues related to EFA were checked routinely: absence of missing values (31.2%), normality (38.7%), linearity (9.7%), absence of outliers among cases (16.1%), (a lack of) multicollinearity (7.5%). It should be pointed out that not examining statistical assumptions of EFA could threaten the precision of the results of EFAs. For example, regarding the assumption of normality, items (variables) with similar levels of skewness and kurtosis (used for checking the shape of distributions of items) can form artificial factors, as the items that have similar distributions are more highly correlated, creating easy (negatively skewed) and difficult (positively skewed) items and resulting in factor solutions which are burdensome to interpret, and sometimes become very misleading (Bandalos & Finney, 2018). This problem, as well, would be exacerbated with the level of nonnormality. Researchers, thus, are highly recommended to examine skewness and kurtosis values of items to find those with nonnormal distributions (skewness and kurtosis values of 2.00 and 7.00, respectively, indicate nonnormality). We recommend that researchers provide a table of descriptive statistics of variables wherein the mean, standard deviation, in conjunction with skewness and kurtosis values are reported, information which was not found in 60% of our sample of EFAs. On a related note, as we found in our review, the majority of studies did not check for outliers despite the fact that outliers can adversely affect EFA results; hence, we recommend that researchers screen their data for multivariate outliers and report this procedure in their articles (see Pituch and Stevens (2016) and Plonsky and Ghanbar (2018) for more information on outliers and influential data points in correlation-based techniques).

In addition to assessing skewness and kurtosis, we recommend a best practice of checking the data for the general assumptions of normality, linearity, and homoscedasticity. Following Mertler and Vannatta (2013), we recommend the Kolmogorov-Smirnov statistic with Lilliefors significance level, verifying that the scatterplots for each pair of variables are elliptical, examining residual plots for nonlinearity, using Levene's test for univariate cases, using Box's M test for equality of variance-covariance matrices for multivariate cases, and Mahalanobis distance to identify outliers. A correlation matrix can also be useful for identifying multicollinearity and singularity

The next targeted element of EFA reports was correlation matrix, just reported in 9.7% of our sample. As EFA is implemented based on a Pearson Product-Moment (PPM) correlation matrix; therefore, reporting such a matrix, even in an appendix, would enhance the transparency of EFA results. The next two elements, related to factorability of correlations, were Kaiser-Meyer-Olkin (KMO) test of sampling adequacy and Bartlett's test of sphericity. The value of Kaiser-Meyer-Olkin (KMO) test of sampling adequacy was reported in 41.9% of the EFAs. It should be said that given that the KMO test presents a vivid, however approximate, picture of sampling adequacy for each variable and the factor structure (depending on the software used), the KMO value should be examined and, if it is below the acceptable value, sample size should be expanded, with value of .70 or greater signifying the adequate sample size (Kaiser, 1970, 1974). Bartlett's test of sphericity, also, was only reported in only 35.5% of EFAs and, as was discussed before, since EFA is based on PPM correlations, researchers should first examine whether variables are adequately correlated and then continue with the analysis, something which Bartlett's test (a significant test is preferred) can shed light on.

Another important component of EFA reports, which is provided by many statistical software programs, is a reproduced matrix, providing valuable information regarding the fit of the model. Unfortunately, none of the studies in our sample reported this matrix or examining it. As its elements illustrate the differences between the observed correlations and the correlations based on the factor structure, we suggest that researchers probe these residuals and be cautious if more than 50 percent of them are greater than .05 (Field, 2013), as it is a red signal of poor fit of the model. Furthermore, regarding factor quality, reliability estimates were reported in 65 EFAs (69.9%), with Cronbach alpha estimate of the reliability in all identified EFAs being reported. Several points are worthy of consideration apropos reliability of variables. First, as pointed out by Bandalos and Finney (2018), if a multidimensional solution is obtained, the reliability coefficients should be given for all the subscales separately. In addition, if the total scale is a higher-order factor, the reliability coefficient can be provided for the total scale. The second issue is related to the magnitude of reliability estimates in that the internal consistency values smaller than .7 should be discussed in studies (see Lance et al., (2006) for more information on the appropriate range of reliability estimates), and that the lower the estimate of internal consistency, the less stable and interpretable the factor solution is; hence, the researchers should pay careful attention to the underlying reasons for low internal consistency values, among which the number of variables representing a factor can be mentioned. Worthy of pursuit is squared multiple correlations (SMC) of factor scores resulting from the regression of factors scores on variable scores, and SMCs of .7 signify that variables represented a substantial amount variance in factors, illustrating a stable factor solution. We can thus argue that researchers would better to obtain factor scores and use them for investigating the factor stability, something which cannot be seen in our sample, with only two EFAs (2.2%) reporting factor scores in their results. As the final point regarding the type of reported reliability index, it should be said that Cronbach alpha, the most frequently reported reliability index, can only be used if a factor is conceived as a composite, that is, a sum of its representing variables in which all the related variables have the same weight; consequently, this is just a reliability of a composite, rather than that of a factor which is why nowadays other indexes such as *Coefficient H* (Hancock & Muller, 2001) is recommended, as it reflects the correlation the factor is predicted to have with itself over repeated measurements.

Factor extraction method is another feature we considered in our review. Because researchers have a wide range of extraction methods at hand, provided in all the statistical software programs, we analyzed their individual usage in our sample. First, regarding the explicit mentioning of the name of the extraction method, the majority of EFAs (90.3%) contained this type of information in their reports. Among those EFAs providing information about the method of extraction, 26 EFAs (28%) used PCA. Interestingly, this finding is not consistent with other reviews of EFA in different fields where PCA is the most preferable option among other methods (Henson et al., 2004, Henson & Roberts, 2006, Plonsky & Gonulal, 2015). In our study, we found that principal axis factoring (PAF) and maximum likelihood (ML), two factor analysis methods, were utilized in 31 EFAs (33%) and 27 EFAs (29%), respectively. Choosing between PCA and FA is one of the most important decisions made by researchers; therefore, they need to pay meticulous attention to the delicate difference between these two orientations toward factor extraction in EFA because the two types of analysis are confused. More specifically, although PCA is frequently utilized in scale development studies, if the researchers' orientation is more toward a theoretical solution (i.e., construing components as latent dimensions or factors) and not an empirical summary of variables, FA is a better choice. Within factor analysis methods, our findings, in keeping with Bandalos and Finney (2018), demonstrated that PAF and ML methods are the most commonly used methods, whereas other methods, such as generalized least squares, unweighted least squares, image factoring and alpha factoring are the least frequently used options by researchers. Because ML factor analysis is used routinely in our sample, it may have been used because it provides standard errors for model parameters and tests of the goodness of fit for factor solution. Researchers should be cautions that this method of factoring may not yield accurate pattern coefficient under the condition of having weak factors and or small sample size (Briggs & MacCallum, 2003). It should be noted that PAF may be popular amongst researchers because it has been described as the "classic factor analytic approach" (Pett et al., 2003, p. 103). It also "explicitly focuses on the common variance among the items and, therefore, focuses on the latent factor" (Henson & Roberts, 2006, p. 398). Additional desirable characteristics of PAF include the facts that it is preferred when multivariate normality is problematic, and it produces reliable solutions with high or low communalities (Watson, 2017).

Turning to decisions related to determining the number of components or factors to retain, the results reveal that in 50 EFAs (53.8%) multiple criteria are utilized, although 33 EFAs (35.5%) erroneously used a single yardstick, and 10 EFAs (10.8%) do not provide such information. Given the fact that determining the number of factors to retain is one of the major decisions in EFA studies, benefiting from several methods simultaneously enhances the quality of authors' decision, as factor retention's decision should never be based on one criterion (Bandalos & Finney, 2018; Thompson, 2004). More specifically, the EV>1 rule, a mathematically-based criterion, was the most recurrent method (64.5%). Albeit previous simulation studies illustrated that it consistently yields unreliable results (Cortina, 2002; Velicer et al., 2000), more recent simulation studies revealed that it is a powerful and promising factor retention method as long as sample size is large enough (Braeken & van Assen, 2017). It should be pointed out that Joliffe's criterion which proposes retaining factors with eigenvalues above 0.70 (Jolliffe, 2002) was never used in our sample, insinuating that researchers ignored it or at least were not aware of it. Two more statistically-oriented methods, Parallel Analysis and Minimum Average Partial (MAP) were not used in a similar way. Whereas Parallel Analysis was utilized frequently in EFAs (49.5%), just 10 uses (10.8%) of MAP were identified in our sample. This may be because of the mathematical complexity of MAP, with which researchers are not familiar, or, as Gorsuch (1983) pointed out, this method does not perform well if some factors have a few loading items; nonetheless, both Parallel Analysis and MAP function more effectively in comparison with the scree test, a more heuristic method, which was used in 38 EFAs (40.9%). More recently, Auerswald and Moshagen (2019) compared parallel analysis (PA), the efficiency of several methods such as Kaiser Criterion, sequential X^2 model tests (SMT), revised PA, comparison data (CD), the Hull method, and the Empirical Kaiser Criterion (EKC) and they found no significant difference among them, signifying the fact that a combination of methods should be used in an EFA to have more conclusive results. One interesting finding in this part is that in 19 EFAs (20.4%), it was mentioned that the number of factors to extract was determined in advance, based on an a priori theory or other similar instruments in previous studies. In line with Henson and Roberts (2006), and Kieffer (1999), we recommend not using this approach, as it is not an optimal option, and, instead, suggest using CFA, which is more robust and accurate. The bottom line in this section is that researchers should base their decisions about the adequacy of extraction and number of factors on both using the aforementioned techniques, and on the theory or related literature.

The findings, in some cases, from the factor retention discussions are a mixed bag. In a few reviewed studies, researchers reported using Kaiser's > 1, the scree plot, PA, and MAP, finding that each technique produced a different number of factors. In those cases, the researchers selected the solution that was the most interpretable. The Kaiser's > 1 criterion is not recommended to be used with PAF because it overfactors (Adelson & McCoach, 2011; Russell, 2002), and PA is also claimed to overfactor (For further discussion on the mixed results of the different factor retention criteria, see Henson and Roberts, 2006, p. 399).

Another main portion of our analysis was devoted to factor rotation methods. As it was shown in Table 2, depending on the existence of correlation among factors, one type of rotation, orthogonal or oblique, should be used. We recommend that researchers choose one oblique rotation method from the very beginning of analysis, as it produces a factor correlation matrix, and check whether there is an amount of correlation among factors (values greater than .3 signify a fair amount of correlation). If they find that factors are correlated, they continue their analysis, and if not, they will opt for an orthogonal method and this needs to be explicit in reports, something which was found to be missing in 50 % of our reviewed EFAs.

Turning to the type of factor rotation techniques, our analysis illustrated that the majority of EFAs used a single method (72%) and a small number of EFAs utilized several rotation methods (16.2%). As it was mentioned before, utilizing multiple methods has the virtue of boosting the precision of EFA and we recommend that prospective researchers use several methods to see which one provides them with more plausible results. Further, most of EFAs exploited an oblique rotation (70.73%), in contrast with two recent reviews of EFA in social sciences, Henson and Roberts (2006) and Plonsky and Gonulal (2015), which found that one type of orthogonal rotation (varimax) was the most frequently used technique. As it was argued by Fabrigar and Wegener (2012), Pedhazur and schmelkin (1991), Stevens (2009), and Tabachnick and Fidell (2013), assuming correlated factors has several conceptual advantages, something which manifests itself in our findings as well, notwithstanding the practical disadvantages pertaining to the complicated nature of factors' interpretations in this rotation type.

More specifically, we found that, among oblique techniques, promax (40.6%) and direct oblimin (39.1%) were two often-used techniques. This can be because of the fact that promax rotates orthogonal factors to their oblique positions which is fast and effective and direct oblimin simplifies the factors by minimization of cross-products of loadings. Few studies, however, reported kappa (42%) for promax and delta for direct oblimin (24%) which showed the amount of the exact amount of these values for improving the evaluation of results. Another rotation technique, procrustes, which is only available in the SAS program and was utilized only in 9.4 % of EFAs can be a useful option when researchers aim to use CFA. With regard to a detailed view of orthogonal rotation usage in our sample, as it was mentioned before, varimax was the only used orthogonal rotation. It should be mentioned that each orthogonal rotation method reported in Table 2 has its own merits and demerits. Quratimax, for instance, is a good option in case of cleaning up the variables; however, as Stevens (2008) mentioned, it triggers loading of variables on only one factor, culminating in a burdensome factors' interpretation. Pertaining to varimax, it can be inferred that because it is a default option in many statistical packages, it was used recurrently, but, nonetheless, we should note that varimax should only be used if the aim of EFA is to clean up the factors, facilitating the interpretation of factors which is why makes varimax the most popular rotation method in EFA (For more information on interpretation of factors in EFA see Gorsuch, 1983; Pituch & Stevens, 2016; Tabachnick & Fidell, 2013) for more comprehensive discussion on the internal mechanism of each factor rotation method). Our final note in this section pertains to the accompanying information which is needed to be reported with different sets of rotation methods. The reports of oblique rotation techniques, for instance, as recommended by Meyers et al., (2013), and Tabachnick and Fidell (2013), should be supplemented with both pattern matrix and structure matrix, as these two matrices contain different sets of coefficients, with each providing different types of information regarding the correlation between extracted factors and variables.

The last part of our analysis examined the relationship between EFA and CFA in the studies that we reviewed. As discussed before, CFA is utilized when the researchers aim to evaluate hypothesized structures of latent constructs or/and foster a better understanding regarding those structures. In EFA, however, the goal is to extract latent constructs or generating hypothesis. In our review, we examined whether CFA was considered as a complementary or potentially appropriate analysis in addition to EFA. It should be noted that CFA is used when there is at least one hypothesized structure postulated for latent constructs. Put it differently, when a new instrument is proposed, EFA should be used as the factor structure of items is not determined, and, gradually, as an instrument is used across different populations and contexts, several factor structures are uncovered, thereby using CFA can help researchers to test these rival structures (Kline, 2016) to check which one best represents the data. Our findings regarding the patterns of using EFA in research illustrated that 17 EFAs (15.1%) were used to extract factorial structure of new instruments, which is an indispensable function of EFA. Further, 12 EFAs (3.2%) used EFA in confirmatory manner, conducting it when an instrument was not new and they warranted CFA for further research. Delving more into these EFAs to find the underlying reasons for not implementing CFA revealed that in 2 EFAs (16.6%) sample size, and in 3 EFAs (25%) lacking a robust theory for underlying structure were mentioned as reasons for not conducting EFAs; however, in 7 (58.4%) EFAs no reasons were mentioned, indicating that authors somewhat lacked enough knowledge regarding the robustness of the theory underlying the investigated factorial structure. Eventually, pertaining to the collocation of EFA and CFA, we understood that most of EFAs (68.4%) were used in conjunction with EFAs. In line with Bandalos and Finney (2018), we recommend not using CFA to confirm EFA's results, a picture which is evident in our findings, given that using EFA then CFA on the same sample would result in inflation of amount of error, thereby resulting in spurious impression of validity. Nonetheless, L2 researchers can use EFA after CFA, if the results of CFA showed an extreme misfit of a model on data, a practice which is completely justified and recommended.

Conclusion and Recommendations

This study aimed to provide some essential recommendations for conducting EFA effectively and reporting the findings in L2 research. By adhering to these recommendations, researchers can enhance the rigor and quality of their L2 research studies utilizing EFA. This will ultimately lead to more reliable and valid findings in quantitative or quantitively driven mixed methods research in the field of L2 research. Based on the empirically grounded findings of our study, as well as conceptual recommendations from textbooks and reference guides, we have developed a concise set of specific recommendations for utilizing EFA in L2 research. These recommendations are aligned with the best EFA practices.

1. Largest possible sample size should be selected to ensure the reliability of the correlation coefficients upon which the EFA is based.
2. Since outliers can adversely affect EFA results, data sets should be screened for multivariate outliers and the results should be reported in the article.
3. Items that contribute more unexplained variance to the model than explained variance need to be excluded.
4. Items with large error variance should be excluded.
5. Skewness and kurtosis values of items should be examined to identify those with non-normal distributions. When identified these items should be either excluded or modified.
6. Descriptive statistics of variables are essential to present the mean and standard deviation, together with skewness and kurtosis values.
7. The data set for the general assumptions of normality, linearity, and homoscedasticity should always be screened.
8. A correlation matrix of the input variables should be provided in the study. In addition, it should be determined whether variables are adequately correlated (factorable) before proceeding with the analysis.
9. The reproduced matrix as evidence of a good fit (or misfit) of the model should be reported.
10. The reliability coefficients should be separately reported for all the subscales of a multidimensional solution. It is also recommended that the reliability coefficients for the total scale, if the scale is considered a higher-order factor, be reported.
11. Internal consistency reliability values smaller than 0.7 should be addressed and discussed in the article.
12. Factor scores need to be calculated and used to investigate factor stability.

13. Multiple criteria for factor retention decisions should be considered and reported in the article.
14. An oblique rotation method should be employed. However, if the factors are not correlated an orthogonal method should be opted for.
15. In order to determine which method provides more plausible, interpretable results, several factor rotation methods need to be exploited.
16. The kappa values for promax, if utilized, and the delta value for direct oblimin, if utilized, should be reported.
17. The pattern matrix and the structure matrix in reports of oblique rotation techniques should be included and reported.
18. If the CFA showed an extreme misfit of the model to the data, it is recommended that EFA be used with different samples after CFA.

Acknowledgments

We would like to thank the editorial team of TESL Quarterly for granting us the opportunity to submit and publish the current synthesis. We would also like to express our appreciation to the anonymous reviewers for their careful, detailed reading of our manuscript and their many insightful comments and suggestions.

Declaration of conflicting interests

The authors declare no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for this article's research, authorship, and/or publication.

References

Adelson, J. L., & McCoach, D. B. (2011). Development and psychometric properties of the Math and Me Survey: Measuring third through sixth graders' attitudes toward mathematics. *Measurement and Evaluation in Counseling and Development, 44*(4), 225-247.

Alamer A, Schuberth F, Henseler J. (2024). When and how to use confirmatory composite analysis (CCA) in second language research. *Studies in Second Language Acquisition*, 1-20. doi:10.1017/S0272263124000020

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*. Advance online publication.

Bandalos, D. L., & Finney, S. J. (2010). Factor Analysis. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93-114). Routledge.

Bandalos, D. L., & Finney, S. J. (2018). Factor analysis: Exploratory and confirmatory. In The reviewer's guide to quantitative methods in the social sciences (pp. 98-122). Routledge.

Bandalos, D. L., & Finney, S. J. (2019). Factor analysis: Exploratory and confirmatory. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd ed., pp. 98–122). Routledge. https://doi.org/10.4324/9781315755649-8

Braeken, J., van Assen, M. L. A. M. (2017). An empirical kaiser criterion. *Psychological Methods, 22*(3), 450-466.

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research, 38*, 25-56.

Brown, T. A. (2013). Latent variable measurement models. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 2, Statistical analysis) (pp.257-280). Oxford University Press.

Comrey, A. L., & Lee, H. B. (1992). *The first course in factor analysis* (2nd ed.). Lawrence Erlbaum.

Cortina, J. M. (2002). Big things have small beginnings: An assortment of "minor" methodological misunderstandings. *Journal of Management, 28*, 339-362.

De Bot, K. (2015). *A history of applied linguistics: From 1980 to the present*. Routledge.

DeVellis, R. F. (2022). *Scale development: Theory and applications* (3rd ed.). Sage.

Dimitrov, D. M. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields.* American Counseling Association.

Ding, C. S. (2013). Multidimensionality scaling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 2, Statistical analysis) (pp. 235-256). Oxford University Press.

Dörnyei, Z., & Dewaele, J.-M. (2022). Questionnaires in second language research: Construction, administration, and processing (3rd ed.). Routledge.

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

Field, A. (2009). Discovering statistics using SPSS (3rd ed.). Sage

Field, A. (2013). *Discovering statistics using SPSS* (4th ed.). Sage.

Flake, J., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science, 8*, 370–378. https://doi.org/10.1177/1948550617693063

Ghanbar, H., & Rezvani, R. (2023). Structural Equation Modeling in L2 Research: A Systematic Review. *International Journal of Language Testing, 13*, 79-108. 10.22034/ijlt.2023.381619.1224

Ghanbar, H. & Rezvani, R. (2024). A Methodological Synthesis of Research Articles in the English for Specific Purposes's Lifespan. *Language Related Research,15* (1), 119-143. 20.1001.1.23223081.1401.0.0.376.8

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

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate behavioral research, 25*(1), 33-39.

Gorsuch, R.L. (2003). Factor analysis. In A. Schinka & W.F. Velicer (Vol. Eds.), Handbook of psychology: Vol. 2. Research methods in psychology (pp. 143–164). Wiley

Haig, B. D. (2013). The philosophy of quantitative methods. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 1, Foundations) (pp. 7-31). Oxford University Press.

Hair, J. F., Black,W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking Construct Reliability within Latent Variable Systems. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural Equation Modeling: Present and Future—A Festschrift in Honor of Karl Joreskog* (pp. 195-216). Scientific Software International.

Henson, R. K., Capraro, R. M., & Capraro, M. M. (2004). Reporting practices and use of exploratory factor analyses in educational research journals: Errors and explanation. *Research in the Schools, 11*(2), 61-72.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416.

Iwaniec, J. (2019). Questionnaires: Implications for effective implementation. In J. McKinley & H. Rose (Eds.), The Routledge handbook of research methods in applied linguistics (pp. 324–335). Routledge.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer.

Kaiser, H. F. (1970). A second-generation Little Jiffy. *Psychometrika*, *35*, 401-415.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31-36.

Kieffer, K. M. (1999). An introductory primer on the appropriate use of exploratory and confirmatory factor analysis. *Research in the Schools*, *6*, 75-92

Kim, Y. (2009). Validation of psychometric research instruments: The case of information science. *Journal of the American Society for Information Science and Technology*, *60*, 1178–1191. https://doi.org/10.1002/asi.21066

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The Sources of Four Commonly Reported Cutoff Criteria. What Did They Really Say? *Organizational Research Methods*, *9*, 202-220.

Lazaraton, A. (2000). Current trends in research methodology and statistics in applied linguistics. *TESOL Quarterly*, *34*, 175–181.

Lazaraton, A. (2005). Quantitative research methods. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 109–224). Erlbaum.

Liu, Y., Zumbo, B. D., & Wu, A. D. (2012). A demonstration of the impact of outliers on the decisions about the number of factors in exploratory factor analysis. *Educational and Psychological Measurement*, *72*(2), 181-199.

Loewen, S., & Gass, S. (2009). Research timeline: The use of statistics in L2 acquisition research. *Language Teaching*, 42(2), 181–196. doi: 10.1017/S0261444808005624

Loewen, S., Gonülal, T., Isbell, D. R., Ballard, L., Crowther, D., Lim, J., et al. (2020). How knowledgeable are applied linguistics and SLA researchers about basic statistics? Data from North America and Europe. *Studies in Second Language Acquisition*, *42*(4), 871-890. https://doi.org/10.1017/S0272263119000548

Loewen, S., Lavolette, E., Spino, L. A., Papi, M., Schmidtke, J., Sterling, S., & Wolff, D. (2014). Statistical literacy among applied linguists and second language acquisition researchers. *TESOL Quarterly*, 48(2), 360–388. doi: 10.1002/tesq.128

Loewen, S., & Plonsky, L. (2016). *An A–Z of applied linguistics research methods*. Bloomsbury Publishing.

Marsh, H., & Alamer, A. (2024). When and how to use set-exploratory structural equation modelling to test structural models: A tutorial using the R package lavaan. *British Journal of Mathematical and Statistical Psychology*, 1–18. https://doi.org/10.1111/bmsp.12336

Mertler, C. A., & Vannatta, R. A. (2013). *Advanced and multivariate statistical methods. Practical application and interpretation* (5th ed.). Pyrczak Publishing.

Meyers, L. S., Gamst, G. C., & Guarino, A. J. (2013). *Performing data analysis using IBM SPSS*. Wiley.

Morea, N., & Ghanbar, H. (2024). Q methodology in applied linguistics: A systematic research synthesis. *System*, *120*, 103194.

Morin, A. J. (2023). Exploratory structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed.). Guilford.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Lawrence Erlbaum.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Sage.

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for social sciences* (6th ed.). Routledge.

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *The Modern Language Journal*, *102*, 713-731. DOI: 10.1111/modl.125090026-7902/18/1-19

Plonsky, L., & Gonulal, T. (2015). Methodological synthesis in quantitative L2 research: A review of reviews and a case study of exploratory factor analysis. *Language Learning*, 65(1), 9-36.

Ponto, J. (2015). Understanding and evaluating survey research. *Journal of the Advanced Practitioner in Oncology*, *6*, 168–171.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, *28*, 1626-1646.

Selinker, L., & Lakshmanan, U. (2001). How do we know what we know? Why do we believe what we believe? *Second Language Research*, *17*, 323–325.

Spearman, C. (1904). General intelligence' objectively determined and measured. *The American Journal of Psychology*, *5*, 201–293.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge.

Sudina, E. (2021). Study and scale quality in second language survey research, 2009–2019: The case of anxiety and motivation. *Language Learning*, *71*(4), 1149–1193. https://doi.org/10.1111/lang.12468

Sudina, E. (2023). Scale quality in second-language anxiety and WTC: A methodological synthesis. *Studies in Second Language Acquisition*, 1–29. https://doi.org/10.1017/S0272263122000560

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon.

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics (7th ed.)*. Pearson.

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

Thorkildsen, T. A. (2005). *Fundamentals of measurement in applied research.* Pearson Education, Inc.

Thurstone, L. L. (1935). *The vectors of the mind*. University of Chicago Press.

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

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41-71). Kluwer Academic/Plenum Publishers.

Wang, L. L., Watts, A. S., Anderson, R. A., & Little, T. D. (2013). Common fallacies in quantitative research methodology. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 2, Statistical analysis) (pp. 718-758). Oxford University Press.

Watson, J. C. (2017). Establishing evidence for internal structure using exploratory factor analysis. *Measurement and Evaluation in Counseling and Development*, *50*(4), 32-238.