*Article*

Journal of Black Psychology
2018, Vol. 44(3) 219–246

DOI: 10.1177/0095798418771807
journals.sagepub.com/home/jbp

# Exploratory Factor Analysis: A Guide to Best Practice

**Marley W. Watkins**[1]

**Abstract**

Exploratory factor analysis (EFA) is a multivariate statistical method that has become a fundamental tool in the development and validation of psychological theories and measurements. However, researchers must make several thoughtful and evidence-based methodological decisions while conducting an EFA, and there are a number of options available at each decision point, some better than others. Reviews of the professional literature have consistently found that many applications of EFA are marked by an injudicious choice of methods and incomplete reports. This article provides a systematic, evidence-based guide to the conduct of EFA studies that can be followed by researchers with modest statistical training, supplemented with an example to illustrate its application.

**Keywords**

exploratory factor analysis, EFA, measurement, validity, multivariate

Exploratory factor analysis (EFA) is one of a family of multivariate statistical methods that attempts to identify the smallest number of hypothetical constructs (also known as factors, dimensions, latent variables, synthetic variables, or internal attributes) that can parsimoniously explain the covariation

[1]Baylor University, Waco, TX, USA

**Corresponding Author:**
Marley W. Watkins, Department of Educational Psychology, Baylor University, One Bear Place, Waco, TX 76798-7301, USA.
Email: marley_watkins@baylor.edu

observed among a set of measured variables (also called observed variables, manifest variables, effect indicators, reflective indicators, or surface attributes). That is, to identify the common factors that explain the order and structure among measured variables. In the social and behavioral sciences, factors are assumed to be unobservable characteristics of people, which are manifested in differences in the scores attained by those people on the measured variables (Tucker & MacCallum, 1997). As described by Brown (2015),

> A factor is an unobservable variable that influences more than one observed measure and that accounts for the correlations among these observed measures. In other words, the observed measures are interrelated because they share a common cause (i.e., they are influenced by the same underlying construct); if the latent construct was partitioned out, the intercorrelations among the observed measures will be zero. (p. 10)

Founded on philosophical and statistical principles (Mulaik, 1987), EFA was first applied by Spearman (1904) and rapidly became a fundamental tool in the evaluation of theories and validation of measurement instruments (Haig, 2014; Henson & Roberts, 2006; Izquierdo, Olea, & Abad, 2014). As noted by Edwards and Bagozzi (2000), the relationships between constructs and their indicator variables are important because that knowledge allows unambiguous "mapping of theoretical constructs onto empirical phenomena" (p. 155) and, therefore, meaningful testing of theories (Loevinger, 1957; Meehl, 1990).

However, researchers must make several thoughtful and evidence-based methodological decisions while conducting an EFA (Henson & Roberts, 2006). There are a number of options available for each decision, some better than others (Lloret, Ferreres, Hernandez, & Tomas, 2017). Reviews of the professional literature have consistently found that many applications of EFA are marked by an injudicious choice of techniques and incomplete reports (Ford, MacCallum, & Tait, 1986). Although desktop software has made EFA readily accessible to all researchers, the quality of EFA practice does not seem to have improved (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Gaskin, & Happell, 2014; Henson & Roberts, 2006; Izquierdo et al., 2014; Lloret et al., 2017; Norris & Lecavalier, 2010). For instance, Fabrigar et al. (1999) judged that the quality of EFAs reported in psychological research is "routinely quite poor" (p. 295), and Norris and Lecavalier (2010) concluded that "many researchers continue to use suboptimal methodology" (p. 16).

There appears to be a constellation of reasons for the prevalence of poor quality EFA research (Fabrigar et al., 1999; Ford et al., 1986; Pett, Lackey, & Sullivan, 2003). First, researchers may receive relatively little formal training in EFA and are not familiar with the quantitatively complex literature on EFA methods. For example, Aiken, West, and Millsap (2008) found that only one-half of doctoral psychology programs included even half a semester/quarter class on factor analysis. Second, researchers tend to emulate existing publications that, as previously noted, are likely to be inadequate. Finally, researchers may rely on software that either defaults to unsound methods or fails to include optimal methods (Carroll, 1978; Fabrigar et al., 1999; Izquierdo et al., 2014; Lloret et al., 2017; Norris & Lecavalier, 2010).

Regardless of cause, suboptimal decisions in EFA can produce "distorted and potentially meaningless solutions" (Ford et al., 1986, p. 307) that can negatively affect the development and refinement of theories and measurement instruments (Bandalos & Gerstner, 2016; Fabrigar & Wegener, 2012; Haig, 2014; Henson & Roberts, 2006; Izquierdo et al., 2014; Lloret et al., 2017). To ameliorate this situation, a systematic, evidence-based guide to the conduct of EFA studies that can be followed by researchers with modest statistical training is needed. The goal of this article is to provide that guide, supplemented with an example to concretely illustrate its application. Depending on their statistical sophistication, readers may also want to consult relatively accessible (Child, 2006; Fabrigar & Wegener, 2012; Pett et al., 2003) or more demanding (Gorsuch, 1983; Tucker & MacCallum, 1997) texts.

## Software

EFA is included in many commercial statistical packages (e.g., SPSS, SAS, Stata) and tutorials on the use of EFA and supplemental statistical routines are accessible for those packages (e.g., Child, 2006; Fabrigar & Wegener, 2012; Garson, 2013; Lloret et al., 2017; Mehmetoglu & Jakobsen, 2017; O'Connor, 2000; O'Rourke & Hatcher, 2013; Pett et al., 2003; Price, 2017). There are also free statistical programs that include EFA. The most comprehensive is the R package (R Core Team, 2017), which operates on Windows, Macintosh, and Linux operating systems. Several tutorials on using R for EFA have been published (Beaujean, 2014; Finch & French, 2015; Revelle, 2016). Three other free EFA programs that operate on Windows systems are also available: (a) CEFA (Browne, Cudeck, Tateneni, & Mels, 2010), (b) FACTOR (Baglin, 2014; Ferrando & Lorenzo-Seva, 2017), and (c) Explorer (Fleming, 2017). Finally, a variety of free software packages that support EFA methods have been developed (see https://jasp-stats.org, http://folk.uio.no/ohammer/past, https://www.jamovi.org, http://www.macstats.org, and http://edpsychassociates.com.

## EFA Decisions

### *Variables to Include*

Measured variables are selected for their utility as indicators of anticipated factors. That is, their content, convergent, and discriminant validity (Izquierdo et al., 2014). Thus, measured variables should adequately represent the domains the factors are thought to tap and not include variables from unrelated domains (Tucker & MacCallum, 1997). For example, if the broad domain of reading is to be analyzed, multiple variables that measure each dimension of reading (e.g., phonemic awareness, phonics, vocabulary, comprehension) should be selected, but it would be inappropriate to include variables that measure addition and subtraction skills. In contrast, variables that tap blending, segmentation, rhyme, and deletion of phonemes would be appropriate if the narrow domain of phonemic awareness is to be investigated. Inadequate sampling of the domain may fail to uncover important common factors or produce spurious factors. Of course, investigations of existing measurement instruments have a ready-made list of variables, and the researcher's task is to evaluate the validity of those variables (structural, factorial, or construct validity).

At least three measured variables are needed for statistical identification of a factor although more indicators are preferable (Child, 2006; Fabrigar & Wegener, 2012; Izquierdo et al., 2014). Fabrigar et al. (1999) recommended four to six indicators per factor. In general, EFA tends to perform better when each factor is overdetermined (i.e., multiple measured variables are influenced by that factor). Statistical software will operate with fewer than three variables per factor but akin to locating a point in three-dimensional space with one, two, or three lines, the location of such factors will be imprecise. Regardless of quantity, variables that are dependent on each other should not be included in an EFA: for example, several subscores and a total score created by summing those same subscores.

In theory, there are situations where the observed variables are more properly treated as determinants rather than effects of constructs. These are called formative indicators (Edwards & Bagozzi, 2000). For example, education, income, and occupation may be formative indicators of socioeconomic status (SES). Elevated education and income and a high-status job cause higher SES; losing a job would result in lower SES, but this lowered SES would not change years of education. Formative indicator variables should be used with great caution in EFA (Bollen & Diamantopoulos, 2017).

## *Participants to Include*

Careful consideration must be given to which and how many participants to include in an EFA. Which participants is primarily a matter of logic and common sense. Does the sample of participants make sense given the constructs that are being measured? Is the sample representative of the population of interest? For example, selecting a group of male athletes for measurement of postpartum depression would not be the most sensible strategy. Child (2006) also warned against using samples collected from different populations because factors that are specific to a population might be obscured when pooled.

The number of participants needed to adequately reproduce population values has been intensively debated for decades. Early recommendations were typically based on some ratio of the number of variables to the number of factors such as 5:1 or 10:1 or some arbitrary number of participants such as 100 or 200 (Hair, Black, Babin, & Anderson, 2010). However, statistical simulations have revealed that an adequate sample size is complexly determined by the strength of the measured variables' relationships with the factors, factor overdetermination, and number of measured variables (Fabrigar et al., 1999; MacCallum, Widaman, Preacher, & Hong, 2001; Mundfrom & Shaw, 2005; Velicer & Fava, 1998). It appears that these three characteristics "interact in ways that permit compensation for weaknesses in one area by strengths in another area" (Velicer & Fava, 1998, p. 248). Handy tables of minimum sample sizes, based on these characteristics, were provided by Mundfrom and Shaw (2005). These tables clearly indicate that "factor analysis is a large-sample procedure" (Norman & Streiner, 2014, p. 223).

## *Distributional Properties of the Data*

The assumptions of EFA tend to be conceptual rather than statistical (Hair et al., 2010). For example, latent constructs are the source of covariation among measured variables, and those constructs exert a linear influence on the measured variables (Bandalos & Gerstner, 2016; Fabrigar & Wegener, 2012; Tucker & MacCallum, 1997). Departures from normality and linearity are important only because they affect the Pearson product-moment correlation coefficients ($r$) among measured variables used for computation of EFA results, which, in turn, "can result in misleading EFA findings" (Reise, Waller, & Comrey, 2000, p. 289). Therefore, it is important to investigate and report the distributional properties of the data that might affect the Pearson correlations (Goodwin & Leech, 2006). See Flora, LaBrish, and Chalmers (2012) for detailed examples of data screening.

*Variability.* Most research samples result from some type of selective sampling and are not identical to the population from which they are selected (Fabrigar & Wegener, 2012; Tucker & MacCallum, 1997). If a sample is more restrictive than the population then the variance of its variables will also be restricted, leading to attenuated $r$ coefficients (Fabrigar et al., 1999). This effect is strikingly illustrated by comparing the correlations between verbal and quantitative test scores for applicants to a service academy: $r$ was .50 for the total applicate pool ($N$ = 2,253) but only .12 for the 128 students who were admitted (Tucker & MacCallum, 1997). In such cases, it may be appropriate to correct for range restriction, being very careful to apply a suitable remedy (Hunter, Schmidt, & Le, 2006).

*Linearity.* The Pearson $r$ coefficient measures the linear relationship between two variables. If the actual relationship is not linear, then the value of $r$ will be reduced. Linearity can be subjectively judged by examination of scatterplots (Goodwin & Leech, 2006). If nonlinear, a more robust type of correlation coefficient might be used instead of $r$ (de Winter, Gosling, & Potter, 2016; Gorsuch, 1983; Lloret et al., 2017; Revelle, 2016).

*Normality.* The Pearson correlation coefficient assumes normality (Goodwin & Leech, 2006), but violations of normality appear to be common with real data sets (Cain, Zhang, & Yuan, 2017; Micceri, 1989). Skew and kurtosis are especially influential on $r$ and subsequent EFA results. Skew refers to the symmetry of the score distribution, whereas kurtosis is a measure of the height of the score distribution in relation to its width. In general, differences in distributions serve to decrease the size of the correlation coefficient, but two variables with extreme skew can produce artifactual factors (Bandalos & Gerstner, 2016). To reduce the possibility of skew affecting EFA results, all manifest variables should be scored in the same direction. That is, any negatively valenced variables should be reverse scored so that high scores on all the variables mean the same thing (Norman & Streiner, 2014).

Multivariate skew and kurtosis can be statistically evaluated with Mardia's (1970) estimates. Regardless of statistical significance, simulation studies have found that serious problems may exist when univariate skewness is $\geq$2.0 and kurtosis is $\geq$7.0 (Curran, West, & Finch, 1996). When normality is statistically improbable or when univariate skew and kurtosis are excessively elevated (Curran et al., 1996), $r$ is not the most appropriate input for EFA. More robust correlational methods (e.g., Spearman, phi, polychoric, tetrachoric) as well as judicious selection of EFA estimation methods would be advisable in that case (de Winter et al., 2016; Fabrigar et al., 1999; Fabrigar & Wegener, 2012; Flora et al., 2012; Gorsuch, 1983; Lloret et al., 2017; Revelle, 2016).

*Level of Measurement.* Pearson correlations assume that normally distributed variables are measured on interval or ratio scales of measurement—that is, essentially continuous data with equal intervals. In contrast, ordinal variables (categorical with an inherent order such as Likert-type items) and dichotomous (binary such as true-false items) variables will not meet these linearity and normality assumptions and will, consequently, negatively affect correlation coefficients and subsequent EFA results (Fabrigar & Wegener, 2012).

Given the ubiquity of ordinal data (Holgado-Tello, Chacon-Moscoso, Barbero-Garcia, & Vila-Abad, 2010), considerable research has been conducted to identify correlation estimates more robust to nonnormality than Pearson coefficients and situations in which their use would be advantageous. Polychoric correlations have received extensive consideration because they assume that an unobservable normally distributed continuous variable underlies each observed categorical variable and estimates the Pearson correlation between those underlying hypothetical variables. Results from both simulated and real data have converged to indicate that polychoric correlations are more likely to recover the true factor model than are Pearson correlations (Baglin, 2014; Barendse, Oort, & Timmerman, 2015; Flora et al., 2012; Holgado-Tello et al., 2010; Lee, Zhang, & Edwards, 2012) with the difference in methods becoming more extreme as the number of categories decreases. Given these results, methodologists have recommended that EFA be based on polychoric correlations if the ordinal variables are measured by fewer than five to seven categories or when distributions of the ordinal variables are asymmetrical (Bandalos & Gerstner, 2016; Fabrigar et al., 1999; Izquierdo et al., 2014; Lloret et al., 2017; Norris & Lecavalier, 2010).

*Missing Data.* Every study should report the quantity and nature of missing data as well as the rationale and methods used to deal with it. Obviously, the best strategy is to tightly control the experimental situation so that there is no missing data. Unfortunately, this approach is not always possible. Given this reality, considerable research has been conducted on missingness, and theory has been developed regarding its causes (see Little & Rubin, 2002, for a detailed account). Listwise and pairwise deletion of cases with missing data are the default methods in many statistical packages but are inefficient and typically not recommended (Baraldi & Enders, 2010). Alternatives include mean, regression, multiple, and maximum likelihood (ML) imputation methods (Baraldi & Enders, 2010). Studies of imputation methods with simulated and real data demonstrate that any method is probably effective when <5% of the data are missing, mean imputation is acceptable when <10% of the data are missing, and regression imputation is acceptable when <15% of the data are missing (Schumacker, 2015), but multiple imputation and ML methods are more accurate when larger proportions of data are missing (Baraldi & Enders, 2010).

*Outliers.* Methods to detect outliers include box plots and scatterplots for individual variables as well as Mahalanobis distance for multiple variables (Aguinis, Gottfredson, & Joo, 2013). One source of outliers is out-of-range or impossible values. These are likely the result of errors in data collection or input. Another source of outliers is failure to specify missing-value codes in software syntax so that missing-value indicators are read as real data. In either case, the erroneous data should be corrected. Alternatively, the outlier may be real but from a population that differs from the intended population. The later explanation can be addressed with deletion or transformation of the outlying value, but the desirability of these modifications is debatable, and many methodologists recommend against their application (Aguinis et al., 2013; Zijlstra, van der Ark, & Sijtsma, 2011). Use of more robust estimators, such as Spearman or polychoric correlations, might be more appropriate (de Winter et al., 2016; Gorsuch, 1983; Revelle, 2016). If the researcher deletes any data, the decision-making process as well as results from an analysis that did not delete or transform the outliers should be reported.

*Measurement Error.* Measurement error decreases the reliability of the variables, which, in turn, decreases the correlation between variables. Low reliability leaves little variance to be shared with other variables. Fabrigar et al. (1999) recommended that variables with reliabilities below .70 should be avoided in EFA. However, adhering to this reliability standard may not be possible when analyzing test items.

*Appropriateness of the Data for EFA.* Although great care may have been exercised in selecting the variables and participants, it is nevertheless important to verify that the measured variables are sufficiently intercorrelated to justify factor analysis. A subjective method is to examine the correlation matrix. A sizable number of correlations should exceed ±.30 or EFA may be inappropriate (Hair et al., 2010).

An objective test of the factorability of the correlation matrix is Bartlett's (1954) test of sphericity, which statistically tests the hypothesis that the correlation matrix contains ones on the diagonal and zeros on the off-diagonals. Hence, that it was generated by random data. This test should produce a statistically significant chi-square value to justify the application of EFA.

Large sample sizes make the Bartlett test sensitive to even trivial deviations from randomness, so its results should be supplemented with a measure of sampling adequacy. The Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) measure of sampling adequacy is the ratio of correlations and partial correlations that reflects the extent to which correlations are a function of the variance shared across all variables rather than the variance shared by particular pairs of variables. KMO values range from 0.00 to 1.00 and can be computed for the total correlation matrix as well as for each measured variable. Overall KMO values ≥.70 are desired (Hoelzle & Meyer, 2013; Lloret et al., 2017), but values less than .50 are generally considered unacceptable (Child, 2006; Hair et al., 2010; Kaiser, 1974), indicating that the correlation matrix is not factorable. As colorfully described by Kaiser (1974), KMO values "in the .90s, marvelous; in the 80s, meritorious; in the .70s, middling; in the .60s, mediocre; in the .50s, miserable; below .50, unacceptable" (p. 35).

## *Model of Factor Analysis*

The term EFA is often used rather loosely to refer to two models that differ in purpose and computation: specifically, principal components analysis (PCA) and common factor analysis (Fabrigar et al., 1999). PCA analyzes the entire correlation matrix (including the self-correlations of 1.00 found on the diagonal) and "is intended to reduce data while preserving as much information from the original data set as possible" (Norris & Lecavalier, 2010, p. 9). To do this, PCA computes linear combinations of the original measured variables that explain as much information as possible about those original variables. Called components, these new measured variables are parsimonious representations of the original measured variables but are not latent constructs (Cudeck, 2000). Instead, the measured variables influence the components. Accordingly, users of PCA should refer to these linear combinations as components, not factors.

Assuming that the measured variables are correlated because they are influenced by the same underlying latent construct, common factor analysis attempts to separate the total variance of the measured variables into the variance that is common to the measured variables (communality or $h^2$, similar to the familiar $R^2$ in regression) and variance that is unique ($u^2$). Unique variance is composed of variance that is reliable but not shared with other measured variables ($s^2$) plus unreliable measurement error ($e$). Thus, the total variance explained by common factors is equal to $h^2 + (s^2 + e)$ or $h^2 + u^2$. Common factor analysis cannot separate specific from error variance and lumps them both into the $u^2$ term (Tucker & MacCallum, 1997).

Common factor analysis partitions variance into $h^2$ and $u^2$ by analyzing a reduced correlation matrix with an estimate of the communality of each measured variable placed on the diagonal of the correlation matrix instead of the 1.00 values as in PCA. Fortunately, the squared multiple correlation of each measured variable with all other measured variables has been found to be a good initial estimate of the common variance (Guttman, 1956; Tucker & MacCallum, 1997) and is the default option in most EFA software.

Both PCA and common factor analysis produce estimates of communality, but only common factor analysis estimates the uniqueness ($u^2$) of each measured variable. By definition, $h^2$ is dependent on the reliable variance of the measured variables and is an indicator of variable importance useful for assessing the adequacy of measured variables and EFA results. For example, variables for which the common factors explain little variance (i.e., low communality) may distort EFA results (Fabrigar et al., 1999).

PCA components may enhance parsimony in other statistical analyses and may contribute information for decisions regarding the number of factors to retain for subsequent common factor analysis, but most methodologists recommend that common factor analysis be employed when the purpose is to identify latent constructs responsible for the variation of measured variables (Carroll, 1978; Fabrigar et al., 1999; Ford et al., 1986; Gaskin, & Happell, 2014; Gorsuch, 1983; Norman & Streiner, 2014; Norris & Lecavalier, 2010; Price, 2017; Tucker & MacCallum, 1997). For example, Fabrigar and Wegener (2012), recommended the following:

> When the goal of research is to identify latent constructs for theory building or to create measurement instruments in which the researcher wishes to make the case that the resulting measurement instrument reflects a meaningful underlying construct, we argue that common factor analysis (EFA) procedures are usually preferable. (p. 32)

However, this distinction may make little difference when there are ≥40 measured variables (Loehlin & Beaujean, 2017). Unfortunately, PCA is the default model in several statistical programs regardless of the number of variables.

## *Estimation Method*

After common factor analysis has been specified as the preferred model, the method used to estimate (extract) the common factor model must be selected. These mathematical procedures attempt to estimate the relationships between the measured variables and the factors (i.e., regression of measured variables on the common factors) that will replicate the observed correlation matrix as closely as possible (Finch & French, 2015).

A large number of estimation methods have been developed, but two methods that differ in assumptions are most common: ML and iterated principal axis (PA; also known as principal factors, MINRES, or OLS estimation). ML estimation derives from normal theory and is sensitive to multivariate normality and typically requires large sample sizes, whereas PA is a least-squares estimation method that makes no distributional assumptions (Cudeck, 2000). PA leverages the initial communality estimates into intermediate estimates that allow a more precise estimate of communalities that are iterated until a final solution is reached that best reproduces the measured correlation matrix (Norris & Lecavalier, 2010).

The algebraic decomposition of the correlation matrix can produce as many factors as measured variables. An eigenvalue is computed for each of the resulting factors to indicate the amount of variance accounted for by that factor independent of all other factors. Eigenvalues can be converted into proportions by dividing each eigenvalue by the total variance of the data (which is equal to the number of measured variables in the analysis). For example, if the first factor in a 10-variable analysis produces an eigenvalue of 4.0 then that factor would account for 40% of the total variance (4.0 ÷ 10 = .40).

Statistical simulations have found that PA outperforms ML when the relationships between measured variables and factors are relatively weak (≤.40), sample size is relatively small (≤300), multivariate normality is violated, or when the number of factors underlying the measured variables is misspecified (Briggs & MacCallum, 2003; Curran et al., 1996; MacCallum et al., 2001). Simulation results led Briggs and MacCallum (2003) to recommend use of PA "in exploratory factor analysis in practice to increase the likelihood that all major common factors are recovered" (p. 54). In contrast, ML estimation would be appropriate when factor-variable relationships are strong (>.40), sample size is large, multivariate normality is attained, and the number of factors is correctly specified (Fabrigar et al., 1999; Gaskin & Happell, 2014; Norman & Streiner, 2014).

Nonconvergence of ML or iterated PA estimates can often be traced to problems with the data. After ruling out data errors, the number of iterations can be restricted to two or three to reduce the possibility of improper solutions (Gorsuch, 1983; Loehlin & Beaujean, 2017). In any case, the researcher should also employ another estimation method to ensure that results replicate. The researcher should also understand that "a number of different common factors can be produced to fit the same pattern of correlations in the manifest variables" (Haig, 2014, p. 76). Called indeterminacy, this is a challenge common to all multivariate methods that rely on empirical evidence (Mulaik, 1987).

## *The Number of Factors to Retain*

Skillful application of EFA astutely balances parsimony and comprehensiveness. A compromise between these two extremes is to estimate a model that contains just enough factors to account for the important covariation among measured variables. This compromise necessitates a decision about the number of factors to retain in the model for further analysis. Helpfully, the process of model estimation assists in estimating the optimal model because, analogous to wringing water from a wet towel, the first factor extracts the most common variance with subsequent factors extracting successively smaller portions of variance. This process will result in a sequentially descending set of eigenvalues that can be used to estimate the optimum number of factors to retain.

The eigenvalues produced by a PCA have traditionally been used to estimate the number of factors to subsequently investigate in a common factor analysis (Carroll, 1978). A graphical method called the visual scree was developed by Cattell (1966), which entails plotting the magnitude of the component eigenvalues against the ordinal number of the components. Cattell (1966) speculated that trivial error factors would follow the "true" factors and be detected by an "elbow" or distinct break in the slope of the scree plot. Unfortunately, this is a subjective technique (Gorsuch, 1983), and researchers may disagree on the interpretation of scree plots (Child, 2006; Norman & Streiner, 2014).

Two empirical estimates of the estimated number of factors have also been developed: parallel analysis (Horn, 1965) and minimum average partials (MAP; Velicer, 1976). Parallel analysis statistically simulates a set of random data with the same number of variables and participants as the real data. That random data set is then submitted to PCA and the resulting eigenvalues saved. This process is repeated multiple times (100 at a minimum) and the resulting set of eigenvalues averaged and compared with the components extracted from the real data. The eigenvalues extracted from real data that exceed those extracted from random data indicate the number of factors to retain.

To compute MAP, a matrix of partial correlations is calculated after each principle component is extracted and the average of the squared partial off-diagonal correlations is calculated for each of these matrices. This quantity should reach a minimum when the correct number of components is extracted because the maximum common variance has been partialled out of the matrix. At the point where the common variance has been removed and only unique variance remains, the MAP criterion will begin to rise.

Although selection of the correct number of factors to retain is one of the most important decisions in EFA (Child, 2006; Fabrigar & Wegener, 2012; Gorsuch, 1983; Izquierdo et al., 2014; Norman & Streiner, 2014), the default method used by many statistical software programs (e.g., the "eigenvalue 1" rule) is usually wrong and should not be used (Fabrigar & Wegener, 2012; Izquierdo et al., 2014; Norris & Lecavalier, 2010). Measurement specialists have conducted simulation studies and concluded that parallel analysis and MAP are the most accurate empirical estimates of the number of factors to retain and that scree is a useful subjective adjunct to the empirical estimates (Velicer, Eaton, & Fava, 2000; Velicer & Fava, 1998). Unfortunately, no method has been found to be correct in all situations (Fabrigar et al., 1999; Gorsuch, 1983; Pett et al., 2003), so it is necessary to employ multiple methods and carefully judge each plausible solution to identify the most appropriate factor solution (Fabrigar & Wegener, 2012; Gorsuch, 1983; Hair et al., 2010; Henson & Roberts, 2006; Izquierdo et al., 2014; Lloret et al., 2017; Loehlin & Beaujean, 2017; Norris & Lecavalier, 2010; Pett et al., 2003). Of course, relevant theory and prior research must also be included as evidential criteria (Gorsuch, 1983). Consequently, a range of plausible factor solutions should be evaluated by selecting the smallest and largest number of factors suggested by these multiple criteria.

## *Rotation of Factors*

PA estimation can be conceptualized as a transformation from correlation space to factor space via a least-squares procedure. PA estimates the correlation (or loading) of each variable with each factor as illustrated in the top panel of Figure 1 where the $X$ and $Y$ axes represent the two dimensions in factor space. The three variables in the upper-right quadrant have moderately high loadings on Factor I (around .40-.60) and Factor II (approximately .75). The three variables in the lower-right quadrant have high loadings on Factor I (near .75) and low loadings on Factor II (roughly −.20). Thus, all six variables have high loadings on Factor I, making it difficult to interpret the underlying construct.

As this example illustrates, initial results are often difficult to understand because PA estimation concentrates on computational convenience without consideration of conceptual clarity (Fabrigar & Wegener, 2012). Factor rotation is designed to achieve a simpler and theoretically more meaningful solution by rotating the axes within factor space to bring them closer to the location of the variables. As demonstrated in the bottom panel of Figure 1, the factor axes are held at right angles to one another and rotated 24° about their origin to bring them nearer the variables, which are fixed in factor space. Following that rotation, three variables have low loadings on Factor I and high loadings on Factor II, while the other three variables have high loadings on Factor I and low loadings on Factor II. This rotation is called orthogonal because the factor axes were maintained at a 90° angle. The two factors could be made even more distinct by allowing an oblique rotation whereby each axis is allowed to rotate about its origin independent of the other axis. An oblique rotation is demonstrated in the bottom panel of Figure 1 (dashed lines). Following this rotation, each factor has three high loadings and three low loadings, making it easier to discern the underlying constructs being measured.

In practice, factor rotation is accomplished mathematically rather than geometrically. These analytic rotations adhere to the orthogonal versus oblique distinction as in the geometric perspective. Dozens of analytic rotation methods have been proposed (Gorsuch, 1983; Loehlin & Beaujean, 2017; Price, 2017), but varimax (Kaiser, 1958) is the most popular orthogonal rotation

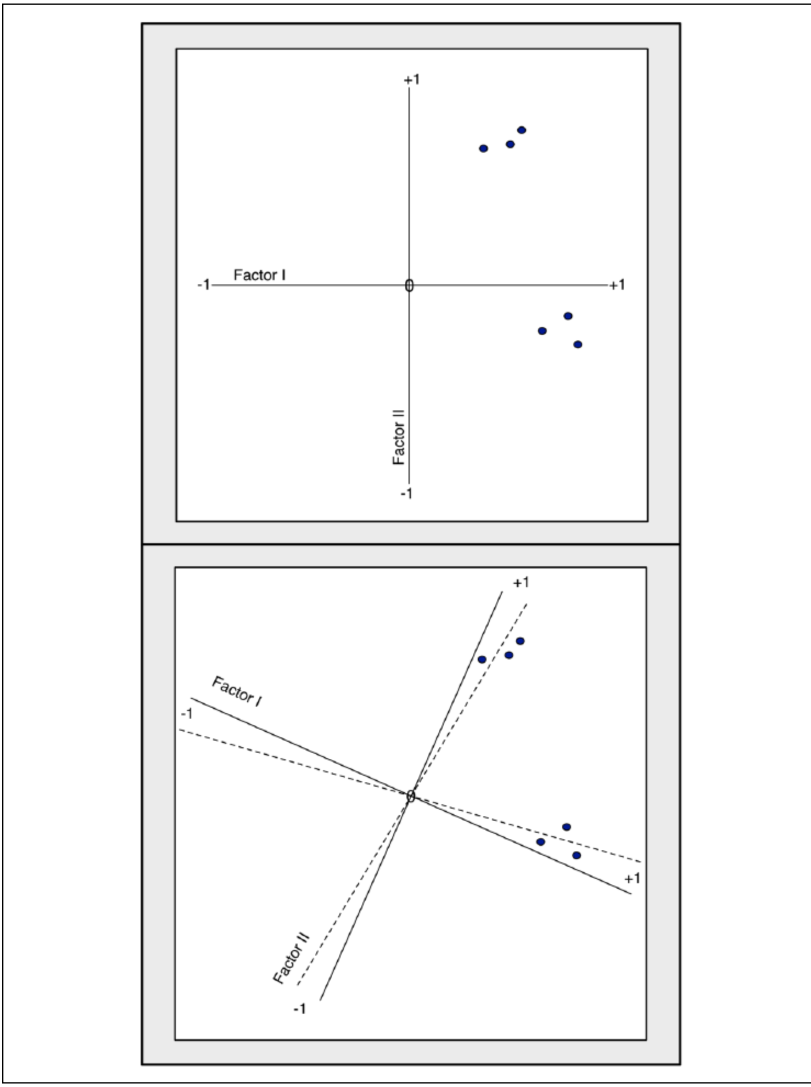


**Figure 1.** (Top panel) Conceptual illustration of two factors and six measured variables in two-factor space. (Bottom panel) Orthogonal (solid axis lines) and oblique (dashed axis lines) rotation about origin.

method, whereas promax (Hendrickson & White, 1964) and oblimin (Jenrich & Sampson, 1966) are the most popular oblique rotation methods. Both promax and oblimin allow the analyst to control the degree of interfactor correlation (via kappa and delta parameters, respectively, but the values automatically applied by most statistical software programs are usually adequate.

To honor the reality that almost everything measured in the social sciences is correlated to some degree (Meehl, 1990), measurement specialists typically recommend that an oblique rotation be applied to allow factor intercorrelations to emerge (Brown, 2015; Cudeck, 2000; Fabrigar et al., 1999; Flora et al., 2012; Gaskin, & Happell, 2014; Gorsuch, 1983; Norris & Lecavalier, 2010; Price, 2017). If there is no real relationship between the factors, then both promax and oblimin rotations will produce orthogonal results. As summarized by Fabrigar and Wegener (2012), "oblique rotations will often be a more realistic representation of the data, will provide a solution that allows for easier interpretation, and will give the researcher additional information not available in orthogonal rotations" (p. 78). Therefore, both promax and oblimin rotation are generally appropriate and either can be recommended (Carroll, 1978; Child, 2006; Gorsuch, 1983; Izquierdo et al., 2014; Loehlin & Beaujean, 2017; Norman & Streiner, 2014; Pett et al., 2003).

However, the superiority of oblique rotations comes at a cost. The initial estimation via PA produces an unrotated solution where each successive factor explains less variance than the prior factor, but rotation redistributes the variance among factors. Thus, the amount of variance accounted for by a factor will differ before and after rotation, but neither the total amount of explained variance nor the communalities of the measured variables will change.

Additionally, the interpretation of factor loadings differs between orthogonal and oblique rotations. For orthogonal solutions, the factor loadings can be interpreted as correlations between common factors and measured variables. These correlations range from −1.00 to +1.00 and the proportion of variance in a measured variable that was contributed by a common factor can be computed by squaring the factor loading. In contrast, oblique solutions produce two different types of factor loadings: structure and pattern coefficients. Structure coefficients can also be interpreted as correlations between common factors and the measured variables. In contrast, pattern coefficients are no longer simple factor-variable correlations; rather, they are similar to standardized partial regression coefficients. That is, they are correlations between common factors and measured variables after controlling for (partialing out) the influence of all other common factors. Accordingly, pattern coefficients might exceed |1.00| and cannot be squared to obtain the proportion of variance uniquely contributed by a common factor.

Unfortunately, rotation allows "an infinite number of alternative orientations of the common factors (or principal components) in multidimensional space" (Fabrigar & Wegener, 2012, p. 67). For example, the 24° orthogonal rotation illustrated in Figure 1 could have been 20° or 30° or 70.89°, etc. Thus, there is no unique rotational solution for any EFA model.

## Interpretation of Results

Interpretation of factors requires consideration of the measured variables and their relationships with the factors (Gorsuch, 1983). Both pattern and structure coefficients register variable-factor relatedness and will be similar if the factor intercorrelations are low, but dissimilar if the factor intercorrelations are high. Although not universally agreed (Gorsuch, 1983), pattern coefficients should be the first focus of interpretation in most analyses (Cudeck, 2000; Fabrigar & Wegener, 2012; Hair et al., 2010; Price, 2017). Structure coefficients should also be reviewed to ensure that anomalous results were not produced. For example, a large pattern coefficient but low structure coefficient might identify a variable with no direct overlap with the construct, whereas a small pattern coefficient but large structure coefficient might identify a variable that is also strongly influenced by other factors (Graham, Guthrie, & Thompson, 2003). Structure coefficients may also be useful for subsequent naming of factors. In any case, researchers should always identify the coefficient being interpreted (pattern or structure) rather than relying on the generic "factor loading" term that could apply to either coefficient.

Respecting the parsimony principle, weak pattern coefficients are not consequential. To be considered strong enough for interpretation (i.e., salient), pattern coefficients should be statistically significant (i.e., not likely due to chance) *and* large enough to be practically useful (Gorsuch, 1983). Norman and Streiner (2014) provided an approximate formula for the statistical significance of pattern coefficients at the 1% level: $\frac{5.152}{\sqrt{N-2}}$. If $p < .05$ is sufficient, then the numerator of 5.152 can be replaced by 3.92. The practical usefulness of pattern coefficients has often been judged to lie in the |.30| to |.40| range (Bandalos & Gerstner, 2016; Hair et al., 2010). That is, those variables with 9% to 16% of their variance explained by a factor after controlling for the influence of other factors.

First articulated by Thurstone in 1947, the concept of simple structure is the most common strategy to guide interpretation of EFA results (Gorsuch, 1983). Conceptually, simple structure is an attempt to find a solution where each factor is loaded by several salient variables and each variable has a salient loading on one factor and trivial loadings on the remaining factors (Brown, 2015). Simple structure recognizes "the purpose of science [which] is to uncover the relatively simple deep structure principles or causes that underlie the apparent complexity observed at the surface structure level" (Le, Schmidt, Harter, & Lauver, 2010, p. 112). However, simple structure must be combined with other criteria to identify acceptable EFA solutions. Specifically, (a) each factor should be saliently loaded by at least three variables (i.e., overdetermined), (b) each variable should load saliently on only one factor (no complex or cross-loadings), (c) each factor should demonstrate internal consistency reliability ≥.70, and (d) all factors should be theoretically meaningful (Bandalos & Gerstner, 2016; Fabrigar & Wegener, 2012; Ford et al., 1986; Gorsuch, 1983; Hair et al., 2010; Norman & Streiner, 2014; Velicer & Fava, 1998).

The interpretation process begins by conducting an EFA with the largest number of plausible factors identified by parallel analysis, MAP, scree, or theoretical convergence and evaluating its adequacy based on these four criteria of acceptability. Next, one fewer factor is extracted and its solution evaluated. This process continues until the complete range of plausible factor solutions has been evaluated (Fabrigar et al., 1999; Pett et al., 2003). With this approach, it is likely that some models will be overfactored (too many factors included) thereby introducing unwanted error variance and others will be underfactored (too few factors included), thereby leaving out useful common variance. It is generally agreed that overfactoring alters the solution less than underfactoring because the major factors will continue to appear in overfactoring, whereas they may be falsely combined into a single factor, altering true factor loadings, in underfactoring (Wood, Tataryn, & Gorsuch, 1996). A symptom of overfactoring is a factor saliently loaded by fewer than three variables (Gorsuch, 1983), whereas complex loadings may be a symptom of underfactoring (Bandalos & Gerstner, 2016). As described by Fabrigar and Wegener (2012), "often when too few common factors have been specified in the model, two or more factors will be collapsed onto the same factor, making it difficult to identify a single unifying theme among the measured variables" (p. 87). However, complex loadings may not be problematic if there is a clear theoretical reason to believe that the measured variable is influenced by more than one latent construct. A measured variable with high reliability and low communality might be an indicator of a specific factor that could be enhanced by including additional indicators (Child, 2006).

Researchers more familiar with the statistics of EFA may also want to review the residual matrix to identify specific misfitting parameters (Gorsuch, 1983; Tucker & MacCallum, 1997). The residual matrix represents the difference between the original correlation matrix and the correlation matrix that was implied by the factor solution. Given that the goal of EFA is to reproduce the correlation matrix, sizeable residuals (≥|.10| may indicate that there are more factors remaining to be extracted (Cudeck, 2000; Pett et al., 2003).

Researchers may also consider a higher-order factor solution if the interfactor correlations are substantial (Child, 2006; Gorsuch, 1983). If the higher-order factors explain a considerable portion of the variance (≥40%; Gorsuch, 1983), then the higher-order model may be more interesting than the oblique solution. Alternatively, it might be advantageous to omit one factor from the model if the interfactor correlations are *too* high (Le et al., 2010). As described by Brown (2015), "factor intercorrelations above .80 or .85 may imply poor discriminant validity, and suggest that a more parsimonious solution could be obtained" (p. 28). To this end, interfactor correlations that rival the reliability of the factors themselves should probably be questioned (Meehl, 1997).

Any interpretation of EFA results must keep in mind that factors are hypothetical constructs that cannot be measured directly; rather, they are inferred from their effects on manifest variables. Accordingly, factors are typically named by considering what their most salient manifest variables have in common. Both pattern and structure coefficients should be used for this purpose, but structure coefficients may be more useful because they reflect factor-variable correlations without the confounding effect of other factors. To reduce the possibility of confusion, factors should not be given the same names as manifest variables. Kline (2016) described three cautions regarding factor names: (a) they are solely for ease of verbal communication and may not mean that "the hypothetical construct is understood or even correctly labeled" (p. 300), (b) they should not be thought of as corresponding to real things (i.e., reification), and (c) it should not be assumed that if they have the same name that two factors are the same thing (jingle fallacy) or that if they have different names that they are different things (jangle fallacy). Rather, the value of factors should be judged by the meaningfulness of their relationships with external criteria and their replicability across samples, methods, and studies (Gorsuch, 1983; Tucker & MacCallum, 1997).

### *Report the Results*

As with all research reports, the EFA report should describe how the study was conducted and should present the results with sufficient detail, clarity, and coherence to support the validity of the results and should justify the conclusions of the study (Appelbaum, Cooper, Kline, Mayo-Wilson, Nezu, & Rao, 2018). Additionally, the EFA report should address each of the preceding decisions and include the information detailed in Table 1 (Ford et al., 1986; Pett et al., 2003).

## Example of an EFA Report

### *Method*

*Participants.* Participants were 197 secondary students (50.5% male) of African ancestry enrolled in 27 randomly selected classrooms (Forms 1-5) within 6 randomly selected secondary schools in the Republic of Trinidad and Tobago.

**Table 1.** Information to Include in an EFA Report.

- Justification of the measured variables included in the EFA
- Justification of type and number of participants included in the EFA
- Data characteristics (including descriptive statistics, normality, missing data, etc.)
- Appropriateness of data for EFA (Bartlett and KMO statistics)
- Computer program and version
- Correlation matrix analyzed (Pearson, polychoric, etc.)
- Factor model (principal components analysis vs. common factor analysis)
- Estimation method (iterated principal axis, maximum likelihood, etc.)
- Method of estimating communalities
- How number of factors to retain was determined
- Factor rotation method
- Strategy for interpreting factors (including how salience was defined)
- Percentage of variance accounted for by factors (specify before or after rotation)
- Complete pattern coefficient matrix (do not omit low coefficients)
- Interfactor correlations (for oblique rotations)
- Reliability estimates for the identified factors
- Complete structure coefficient matrix (when substantially different from pattern matrix)
- Eigenvalues for all factors if space permits
- Correlation matrix if space permits

Note: EFA = exploratory factor analysis; KMO = Kaiser-Meyer-Olkin.

Participants ranged in age from 12 to 17 years ($M$ = 14.1, $SD$ = 1.3). Based on the average factor loadings found in prior research (Gilman, Laughlin, & Huebner, 1999), it was anticipated that 197 participants would allow excellent recovery of the population factor solution (Mundfrom & Shaw, 2005).

*Instruments.* The Self-Description Questionnaire–II (SDQ) was published by Marsh in 1990 with an Australian normative sample. The full SDQ contains 102 items that address 11 areas of self-concept. However, only the math and verbal (reading and English) self-concept areas were included in this example for simplicity of presentation. Ten Likert-type items addressed each self-concept area with each item scored on a scale of 1 = *False*, 2 = *Mostly False*, 3 = *More False Than True*, 4 = *More True Than False*, 5 = *Mostly True*, to 6 = *True*. All negatively valenced items were reverse scored.

Reliability estimates for SDQ subscale scores have been reported in the .70 to .90 range (Marsh, 1990). In a comprehensive review of self-concept measures, Byrne (1996) strongly endorsed the construct validity of SDQ scores, and SDQ scores have subsequently been validated with a wide variety of participants (e.g., Gilman et al., 1999; Mucherah & Finch, 2010; Wastlund, Norlander, & Archer, 2001).

*Procedure.* A complete description of the research procedures employed in this study can be found in Worrell, Watkins, and Hall (2008).

*Analyses.* Analyses were conducted with the R statistical package (Macintosh Version 3.4.3; R Core Team, 2017) and its *psych* package (Version 1.7.8). Bartlett's test of sphericity (Bartlett, 1950) was used to ensure that the correlation matrix was not random and the KMO statistic (Kaiser, 1974) was required to be above a minimum of .50.

After confirming that the correlation matrix was factorable, it was submitted for EFA. Common factor analysis was selected over PCA because the intent was to identify a latent factor structure (Fabrigar et al., 1999). An iterated PA extraction method with initial communalities estimated by squared multiple correlations was employed because of its relative tolerance of nonnormality and demonstrated ability to recover weak factors (Briggs & MacCallum, 2003; Guttman, 1956). Following the advice of Velicer et al. (2000), parallel analysis (Horn, 1965), MAP (Velicer, 1976), and the visual scree test (Cattell, 1966) were used to determine the appropriate number of factors to retain. Parsimony and theoretical convergence were also considered. Due to the nature of the constructs, it was assumed that factors would be correlated. Therefore, an oblimin rotation (Jennrich & Sampson, 1966) was employed (Carroll, 1978; Child, 2006).

Criteria for determining factor adequacy were established a priori. Given the number of participants in this study, pattern coefficients ≥.37 were considered salient (i.e., both practically and statistically significant as per Norman & Streiner, 2014). Complex loadings that were salient on more than one factor were rejected to honor simple structure (Thurstone, 1947). Factors with a minimum of three salient pattern coefficients, internal consistency reliability ≥.70, and that were theoretically meaningful were considered adequate.

## Results

Of the 197 respondents, 20 were missing 1 or 2 item responses for a total of 25 missing data points. Given that this represented less than 1% of the data, mean imputation was employed (Schumacker, 2015). Descriptive statistics for the imputed data set are provided in Table 2. Although univariate skewness and kurtosis were not extreme (Curran et al., 1996), Mardia's multivariate skew and kurtosis were both statistically significant ($p < .001$). Given this nonnormality as well as the ordinal nature of SDQ items, a polychoric

**Table 2.** Descriptive Statistics and Pattern Coefficients for 197 Participants on the Math and Verbal Items of the Self-Description Questionnaire-II.

| | Descriptive Statistics | | | | Factors | | |
|---|---|---|---|---|---|---|---|
| Item | Mean | *SD* | Skew | Kurtosis | Math | Verbal | $h^2$ |
| Math 1 | 3.89 | 1.89 | −0.30 | −1.38 | **.83** | −.10 | .71 |
| Verbal 1 | 4.93 | 1.45 | −1.26 | 0.49 | .04 | **.68** | .47 |
| Math 2 | 3.68 | 1.92 | −0.19 | −1.50 | **.73** | .03 | .53 |
| Verbal 2 | 4.24 | 1.82 | −0.63 | −1.02 | −.01 | **.59** | .35 |
| Math 3 | 4.11 | 1.88 | −0.54 | −1.18 | **.63** | .09 | .40 |
| Verbal 3 | 4.85 | 1.67 | −1.23 | 0.10 | .06 | **.62** | .38 |
| Math 4 | 4.71 | 1.70 | −1.10 | −0.22 | **.64** | .12 | .41 |
| Verbal 4 | 4.61 | 1.49 | −0.97 | −0.04 | .05 | **.72** | .52 |
| Math 5 | 3.63 | 1.91 | −0.12 | −1.45 | **.76** | .00 | .58 |
| Verbal 5 | 4.98 | 1.55 | −1.34 | 0.49 | .06 | **.59** | .35 |
| Math 6 | 3.96 | 1.90 | −0.38 | −1.36 | **.83** | −.04 | .69 |
| Verbal 6 | 4.44 | 1.83 | −0.81 | −0.82 | −.19 | **.73** | .59 |
| Math 7 | 3.88 | 1.89 | −0.31 | −1.41 | **.82** | −.02 | .67 |
| Verbal 7 | 5.03 | 1.53 | −1.54 | 1.15 | .02 | **.57** | .32 |
| Math 8 | 4.50 | 1.85 | −0.84 | −0.85 | **.70** | .12 | .49 |
| Verbal 8 | 4.69 | 1.57 | −1.06 | 0.06 | .00 | **.74** | .54 |
| Math 9 | 3.71 | 1.88 | −0.26 | −1.42 | **.71** | −.04 | .51 |
| Verbal 9 | 4.09 | 1.89 | −0.42 | −1.38 | .07 | **.41** | .17 |
| Math 10 | 4.38 | 1.89 | −0.71 | −1.04 | **.83** | .00 | .68 |
| Verbal 10 | 4.77 | 1.55 | −1.27 | 0.60 | .01 | **.83** | .69 |

Note: $h^2$ = communality. Salient pattern coefficients ≥.37 in boldface. Structure coefficients were almost identical to pattern coefficients due to factor correlation of −.08.

correlation matrix was deemed to be appropriate input for EFA (Bandalos & Gerstner, 2016; Fabrigar et al., 1999; Lloret et al., 2017).

The results of Bartlett's test of sphericity (Bartlett, 1954) indicated that the correlation matrix was not random, $\chi^2(190) = 2{,}648$, $p < .001$, and the KMO statistic (Kaiser, 1974) was .80, well above the minimum standard for conducting factor analysis. Therefore, it was determined that the correlation matrix was appropriate for factor analysis.

Parallel analysis, MAP, and scree all suggested that three factors should be retained but theory (Marsh, 1990) indicated that only two factors were required. Therefore, the three- and two-factor solutions were sequentially examined. The three-factor solution was inadequate: Although four items saliently loaded on the third factor, all four items complexly loaded on the first two factors, and the third factor's internal consistency (alpha) reliability was only .57 (95% CI = .48-.67).

The two-factor solution was next examined for adequacy. Each factor was saliently loaded by 10 items (see Table 2) with math and verbal items cohering as specified by Marsh (1990). Following rotation, the math factor accounted for 28% of the total variance and 56% of the common variance while the verbal factor accounted for 22% of the total variance and 44% of the common variance. Coefficient alpha was .90 (95% CI = .88-.92) for the math factor and .82 (95% CI = .79-.86) for the verbal factor. The math and verbal factors correlated at −.08, producing almost identical pattern and structure coefficients. Given these results, the two-factor solution was accepted as the most adequate structural representation of the SDQ with these participants and was subsequently found to be robust across alternative extraction and rotation methods as well as when missing data were deleted.

## Discussion

Using evidence-based EFA methods, this study found that the 20 math, reading, and English items of the SDQ coalesced into relatively orthogonal math and verbal self-concept factors in a sample of students of African ancestry enrolled in secondary schools in the Republic of Trinidad and Tobago. These results are broadly consistent with those found for adolescents in Africa, Asia, Australia, Europe, and North American (Gilman et al., 1999; Marsh, 1990; Mucherah & Finch, 2010; Nishikawa, Norlander, Fransson, & Sundbom, 2007; Wastlund et al., 2001; Yeung & Lee, 1999). However, scrutiny of Table 2 reveals that several verbal items contributed relatively little to the analysis (especially Verbal Item 9 with a communality of .17) making them prime suspects for future revision of the SDQ (Child, 2006). This result is compatible with recommendations based on item response theory analyses to shorten SDQ scales because of redundancy among the items (Flannery, Reise, & Widaman, 1995). Regardless, the value of SDQ factors should be judged by the meaningfulness of their relationships with external criteria and their replicability across samples, methods, and studies (Gorsuch, 1983; Tucker & MacCallum, 1997).

### Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Marley W. Watkins https://orcid.org/0000-0001-6352-7174

## References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*, 270-301. doi:10.1177/1094428112470848

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology. *American Psychologist*, *63*, 32-50. doi:10.1037/0003-066X.63.1.32

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist*, *73*, 3-25. doi:10.1037/amp0000191

Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation*, *19*(5), 1-15.

Bandalos, D. L., & Gerstner, J. J. (2016), Using factor analysis in test construction. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advances* (pp. 26-51). Boston, MA: Hogrefe.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*, 5-37. doi:10.1016/j.jsp.2009.10.001

Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using exploratory factor analysis to determine the dimensionality of discrete responses. *Structural Equation Modeling*, *22*, 87-101. doi:10.1080/10705511.2014.934850

Bartlett, M. S. (1954). A further note on the multiplying factors for various chi-square approximations in factor analysis. *Journal of the Royal Statistical Society, Series B*, *16*, 296-298.

Beaujean, A. A. (2014). *Latent variable modeling using R*. New York, NY: Routledge.

Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, *22*, 581-596. doi:10.1037/met0000056

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. doi:10.1207/S15327906MBR3801_2

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (2010). CEFA: Comprehensive exploratory factor analysis [Computer software]. Retrieved from http://faculty.psy.ohio-state.edu/browne/download.htm

Byrne, B. M. (1996) *Measuring self-concept across the life span*. Washington, DC: American Psychological Association.

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, *49*, 1716-1735. doi:10.3758/s13428-016-0814-1

Carroll, J. B. (1978). How shall we study individual differences in cognitive abilities? Methodological and theoretical perspectives. *Intelligence*, *2*, 87-115. doi:10.1016/0160-2896(78)90002-8

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

Child, D. (2006). *The essentials of factor analysis* (3rd ed.). New York, NY: Continuum.

Cudeck, R. (2000). Exploratory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 265-296). New York, NY: Academic Press.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29. doi:10.1037/1082-989X.1.1.16

de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, *21*, 273-290. doi:10.1037/met0000079

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*, 155-174.

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

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

Ferrando, P. J., & Lorenzo-Seva, U. (2017). FACTOR [Computer software]. Retrieved from http://psico.fcep.urv.es/utilitats/factor

Finch, W. H., & French, B. F. (2015). *Latent variable modeling with R*. New York, NY: Routledge.

Flannery, W. P., Reise, S. P., & Widaman, K. F. (1995). An item response theory analysis of the general and academic scales of the Self-Description Questionnaire II. *Journal of Research in Personality*, *29*, 168-188. doi:10.1006/jrpe.1995.1010

Fleming, J. S. (2017). Explorer: A program for common factor analysis and related models for data analysis [Computer software]. Retrieved from http://swppr.org/EXPLORER.html

Flora, D. B., LaBrish, C., & Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology*, *3*(55), 1-21. doi:10.3389/fpsyg.2012.00055

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

Garson, G. D. (2013). *Factor analysis*. Asheboro, NC: Statistical Associates.

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*, 511-521. doi:10.1016/j.ijnurstu.2013.10.005

Gilman, R., Laughlin, J. E., & Huebner, E. S. (1999). Validation of the Self-Description Questionnaire-II with an American sample. *School Psychology International*, *20*, 300-307. doi:10.1177/0143034399203005

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of *r*. *Journal of Experimental Education*, *74*, 251-266. doi:10.3200/JEXE.74.3.249-266

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

Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling*, *10*, 142-153. doi:10.1207/S15328007SEM1001_7

Guttman, L. (1956). "Best possible" systematic estimates of communalities. *Psychometrika*, *21*, 273-285. doi:10.1007/BF02289137

Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral sciences*. Cambridge: MIT Press.

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

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Mathematical Psychology*, *17*, 65-70.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement*, *66*, 393-416. doi:10.1177/0013164405282485

Hoelzle, J. B., & Meyer, G. J. (2013). Exploratory factor analysis: Basics and beyond. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, 2nd ed., pp. 164-188). Hoboken, NJ: Wiley.

Holgado-Tello, F. P., Chacon-Moscoso, S., Barbero-Garcia, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, *44*, 153-166. doi:10.1007/s11135-008-9190-y

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

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, *91*, 594-612. doi:10.1037/0021-9010.91.3.594

Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema*, *26*, 395-400. doi:10.7334/psicothema2013.349.

Jenrich, R. L., & Sampson, P. F. (1966). Rotation for simple loading. *Psychometrika*, *31*, 313-323. doi:10.1007/BF02289465

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

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31-36. doi:10.1007/BF02291575

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

Le, H., Schmidt, F. L., Harter, J. K., & Lauver, K. J. (2010). The problem of empirical redundancy of constructs in organizational research: An empirical investigation. *Organizational Behavior and Human Decision Processes*, *112*, 112-125. doi:10.1016/j.obhdp.2010.02.003

Lee, C.-T., Zhang, G., & Edwards, M. C. (2012). Ordinary least squares estimation of parameters in exploratory factor analysis with ordinal data. *Multivariate Behavioral Research*, *47*, 314-339. doi:10.1080/00273171.2012.658340

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.

Lloret, S., Ferreres, A., Hernandez, A., & Tomas, I. (2017). The exploratory factor analysis of items: Guided analysis based on empirical data and software. *Anales de Psicologia*, *33*, 417-432. doi:10.6018/analesps.33.2.270211

Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural equation analysis* (5th ed.). New York, NY: Routledge.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635-694. doi:10.2466/PR0.3.7.635-694

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*, 611-637. doi:10.1207/S15327906MBR3604_06

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519-530. doi:10.1093/biomet/57.3.519

Marsh, H. W. (1990). *Self-Description Questionnaire–II manual*. Campbelltown, New South Wales, Australia: University of Western Sydney, Macarthur.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195-244. doi:10.2466/pr0.1990.66.1.195

Meehl, P. E. (1991). Four factors about factor reality. *History and Philosophy of Psychology Bulletin*, *3*, 16-19.

Mehmetoglu, M., & Jakobsen, T. G. (2017). *Applied statistics using Stata: A guide for the social sciences*. Thousand Oaks, CA: Sage.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156-166. doi:10.1037/0033-2909.105.1.156

Mucherah, W., & Finch, H. (2010). The construct validity of the Self-Description Questionnaire on high school students in Kenya. *International Journal of Testing*, *10*, 166-184. doi:10.1080/15305051003739904

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research*, *22*, 267-305. doi:10.1207/s15327906mbr2203_3

Mundfrom, D. J., & Shaw, D. G. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, *5*, 159-168. doi:10.1207/s15327574ijt0502_4

Nishikawa, S., Norlander, T., Fransson, P., & Sundbom, E. (2007). A cross-cultural validation of adolescent self-concept in two cultures: Japan and Sweden. *Social Behavior and Personality*, *35*, 269-286. doi:10.2224/sbp.2007.35.2.269

Norman, G. R., & Streiner, D. L. (2014). *Biostatistics: The bare essentials* (4th ed.). Shelton, CT: People's Medical Publishing.

Norris, M., & Lecavalier, L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychological research. *Journal of Autism and Developmental Disorders*, *40*, 8-20. doi:10.1007/s10803-009-0816-2

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *38*, 88-91. doi:10.3758/BF03200807

O'Rourke, N., & Hatcher, L. (2013). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.

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*. Thousand Oaks, CA: Sage.

Price, L. R. (2017). *Psychometric methods: Theory into practice*. New York, NY: Guilford Press.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: Author. Retrieved from http://www.r-project.org

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*, 287-297. doi:10.1037/1040-3590.12.3.287

Revelle, W. (2016). *An introduction to psychometric theory with applications in R*. Retrieved from http://personality-project.org/r/book/

Schumacker, R. E. (2015). *Learning statistics using R*. Thousand Oaks, CA: Sage.

Spearman, C. E. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychiatry*, *15*, 201-293.

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

Tucker, L. R., & MacCallum, R. C. (1997). *Exploratory factor analysis*. Retrieved from https://www.unc.edu/~rcm/book/factor.pdf

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321-327. doi:10.1007/BF02293557

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). Boston, MA: Kluwer Academic.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*, 231-251. doi:10.1037/1082-989X.3.2.231

Wastlund, E., Norlander, T., & Archer, T. (2001). Exploring cross-cultural differences in self-concept: A meta-analysis of the Self-Description Questionnaire-1. *Journal of Comparative Social Science*, *35*, 280-302. doi:10.1177/106939710103500302

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, 1, 254-265. doi:10.1037/1082–989X.1.4.354

Worrell, F. C., Watkins, M. W., & Hall, T. E. (2008). Reliability and validity of self-concept scores in secondary school students in Trinidad and Tobago. *School Psychology International*, *29*, 466-480. doi:10.1177/0143034308096435

Yeung, A. S., & Lee, F. L. (1999). Self-concept of high school students in China: Confirmatory factor analysis of longitudinal data. *Educational and Psychological Measurement*, *59*, 431-450. doi:10.1177/00131649921969965

Zijlstra, W. P., van der Ark, L. A., & Sijtsma, K. (2011). Outliers in questionnaire data: Can they be detected and should they be removed? *Journal of Educational and Behavioral Statistics*, *36*, 186-212. doi:10.3102/1076998610366263