Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement
2017, Vol. 49, No. 2, 78–88

0008-400X/17/$12.00 http://dx.doi.org/10.1037/cbs0000069

The Purpose and Practice of Exploratory and Confirmatory Factor Analysis in Psychological Research: Decisions for Scale Development and Validation

David B. Flora and Jessica K. Flake
York University

There are many high-quality resources available which describe best practices in the implementation of both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Yet, partly owing to the complexity of these procedures, confusion persists among psychologists with respect to the implementation of EFA and CFA. Primary among these misunderstandings is the very mathematical distinction between EFA and CFA. The current paper uses a brief example to illustrate the difference between the statistical models underlying EFA and CFA, both of which are particular instantiations of the more general common factor model. Next, important considerations for the implementation of EFA and CFA discussed in this paper include the need to account for the categorical nature of item-level observed variables in factor analyses, the use of factor analysis in studies of the psychometric properties of new tests or questionnaires and previously developed tests, decisions about whether to use EFA or CFA in these contexts, and the importance of replication of factor analytic models in the ongoing pursuit of validation.

Keywords: exploratory factor analysis, confirmatory factor analysis, best practices, validity, psychometric properties

Factor analysis is a valuable tool because psychology is a science of unobservable cognitive abilities and characteristics. Although it is easy to do factor analysis using modern statistical software, it is difficult to do factor analysis well. Many high-quality resources on best practices in EFA and CFA are available, and this paper reiterates some of their recommendations. For EFA, these sources include MacCallum (2009), Preacher and MacCallum (2003), and texts by Fabrigar and Wegener (2012) and Osborne (2014); and for CFA, Boomsma (2000), McDonald and Ho (2002), and texts by Brown (2015) and Kline (2016). Although we review the common factor model below, we still assume a basic familiarity with the central principles and procedures of EFA and CFA on a level with the references listed above.

Few resources go into detail regarding when it is appropriate to conduct exploratory or confirmatory factor analyses. In fact, there seems to be some confusion among psychologists about the very distinction between EFA and CFA, despite the ready availability of resources. Therefore, the main purposes of this article are to (a) clarify the distinction between EFA and CFA, (b) describe the appropriate uses of factor analysis in psychological research, and (c) discuss the choice between EFA and CFA, particularly with respect to the assessment of psychometric properties of tests and questionnaires. In describing the uses of EFA and CFA in psychometric research, we also review many common flaws in factor analytic applications and make recommendations for improved practice.

First, we present the common factor model to provide a statistical foundation and vernacular for the rest of the article. Although this presentation is nontechnical, having a conceptual understanding of the common factor model is essential for understanding the difference between EFA and CFA and can help avoid some of the common shortcomings of factor analytic practice.

The Common Factor Model

The primary statistical purpose of both EFA and CFA is to explain relations among a large set of observed variables using a small number of unobserved, or latent, variables called factors. In Thurstone's (1947) common factor model, the backbone of modern factor analysis, factors influence observed variables, accounting for their variation and covariation. Factor analysis traditionally proceeds by fitting models to bivariate associations among observed variables, with EFA most commonly using correlations and CFA most commonly using covariances.[1] Use of product-moment correlations or covariances follows from the fact that the common factor model specifies linear associations between factors and observed variables.

Specifically, Lawley and Maxwell (1963) showed that the common factor model is a linear regression model with observed variables as outcomes and factors as predictors such that

David B. Flora and Jessica K. Flake, Department of Psychology, York University.

Correspondence concerning this article should be addressed to David B. Flora, Department of Psychology, York University, 101 BSB, 4700 Keele Street, Toronto, ON M3J 1P3. E-mail: dflora@yorku.ca

[1] It is possible, though, to fit EFA models to covariances and to fit CFA models to correlations.

$$Y_p = \left(\sum_{m=1}^{M} \lambda_{pm}\eta_m\right) + \varepsilon_p \quad (1)$$

where Y_p is the pth observed variable from a set of P observed variables, η_m is the mth of M common factors, λ_{pm} is the regression coefficient, or *factor loading*, relating factor m to Y_p, and ε_p is the error term, or *unique factor*, for Y_p. The variance of ε for variable p is known as the variable's *uniqueness*, whereas 1 – VAR(ε) is that variable's *communality* and is equivalent to the regression R^2, here representing the proportion of variability in the observed variable explained by the common factors.

Thus, the common factor model is just a multiple regression model in which the values of the predictors (the factors) are unknown. Because there is a separate version of Equation 1 for each observed variable Y_p, these equations can be collected into a single matrix equation:

$$\mathbf{Y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

where $\mathbf{Y}$ is a vector of P observed variables, $\boldsymbol{\Lambda}$ is a $P \times M$ matrix of factor loadings (known as the *factor pattern matrix*), $\boldsymbol{\eta}$ is a vector of M common factors, and $\boldsymbol{\varepsilon}$ is a vector of unique factors. The estimated factor loading matrix represents the key output provided by software for EFA; CFA output also includes factor loading estimates, but they may not be explicitly organized in this matrix format.

The primary computational challenge of factor analysis is estimating the regression coefficients (the factor loadings in $\boldsymbol{\Lambda}$) despite that scores on the predictors (the common factors in $\boldsymbol{\eta}$) are unknown. It turns out that the factor loading parameters can be estimated according to the *covariance structure* (or *correlation structure*) among the observed variables in $\mathbf{Y}$ as implied by the linear form of the common factor model; this model-implied covariance structure is not a function of the unobserved factor scores in $\boldsymbol{\eta}$. Consequently, factor analysis proceeds as an analysis of the covariances (or correlations) among the observed variables (Bollen, 1989).

Statistical Distinction Between EFA and CFA

Given their typical implementation in software, EFA and CFA may seem quite different. Yet, the common factor model is the foundation of both: Jöreskog (1969) showed that the EFA model, or an *unrestricted solution* for the common factor model, can be constrained to produce the *restricted solution* that is commonly understood as the CFA model in modern research in the context of structural equation modelling (SEM). Specifically, with EFA, the elements of the factor loading matrix $\boldsymbol{\Lambda}$ are all freely estimated; that is, each of the M factors has an estimated relation (i.e., factor loading) with every observed variable; factor rotation is then used to aid interpretation by making some values in $\boldsymbol{\Lambda}$ large and others small. But with CFA, depending on a researcher's a priori hypotheses, many of the elements of $\boldsymbol{\Lambda}$ are fixed to equal zero, often so that each observed variable is determined by one and only one factor. That is, the CFA model is often constrained so that there are no so-called *cross-loadings*; this factor pattern is known as an *independent clusters solution*, which is more specific than Thurstone's (1947) original conception of *simple structure* (McDonald, 1999). Because of these a priori restrictions on $\boldsymbol{\Lambda}$, rotation is not an aspect of CFA.

To illustrate the distinction between the EFA and CFA models, Table 1 shows estimated factor loading matrices obtained by fitting a two-factor EFA model and a two-factor CFA model to the same data. (Here, using both EFA and CFA on the same data is only for demonstration; as discussed later, doing so in applied research is poor practice.) The observed variables are scores on eight subtests of the Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983) from $N = 200$ cases from the test's standardization sample for 10-year old children (these data are from Kline, 2016). Additionally, path diagrams of these models are in Figure 1.

The EFA results were obtained using unweighted least squares estimation with *oblimin* rotation whereas the CFA results were obtained by fitting the model specified in Figure 1 using maximum likelihood; the CFA factor loadings are standardized to place them on the same scale as the EFA factor loadings.[2] But at this point, the details of how these results were obtained are not important. Instead, it is important to recognise that with EFA, each observed variable has a nonzero association with both factors, even though factor rotation was used to approximate simple structure. For example, the *word order* scale (Y_1) has a relatively large loading (0.65) on the first factor (η_1, which might be termed *sequential processing* based on subjective interpretation of the overall factor pattern), but its loading on the second factor (η_2, termed *simultaneous processing*) is nonzero, equaling 0.15. In the CFA model, by contrast, the factor loading relating word order to η_2 is fixed to 0 exactly in the a priori model specification (i.e., according to the input commands of the software); in SEM jargon, this parameter is *fixed*, *constrained*, or *restricted* (to equal 0) rather than *free*. The same is true for each of the other factor loadings equaling 0 in the CFA factor pattern. Applying Equation 1, these results for word order imply:

$$Y_1 = 0.65\eta_1 + 0.15\eta_2 + e_1$$

for the EFA model and simply

$$Y_1 = 0.81\eta_1 + e_1$$

for the CFA model.

The EFA factor pattern in Table 1 also shows a cross-loading for *hand movements* (Y_3) in that although it has moderate loading on η_2 ($\hat{\lambda} = 0.40$), it also has a small to moderate loading on η_1 ($\hat{\lambda} = 0.25$, the cross-loading). Upon seeing these results, many researchers specifying a subsequent CFA model would allow hand movements to be an indicator of the CFA model's *simultaneous processing factor* (η_2) because it had a relatively large (or *salient*) loading on the corresponding factor in the EFA solution. In so doing, the CFA specification may or may not also allow hand movements to be an indicator of the CFA model's *sequential processing factor* (η_1). But here, consistent with the model specification in Kline (2016), the CFA model is based on an a priori rationale that performance on the hand movements subscale is determined by a child's sequential processing ability rather than

[2] Both the EFA and completely standardized CFA loadings are interpretable as standardized regression coefficients. A common mistake with EFA for models with ≥2 correlated factors is to interpret these pattern loadings as correlations. But a standardized CFA loading is also interpretable as a correlation *only* if it loads on a single factor.

Table 1
Factor Loading Matrices Obtained With EFA Two-Factor Model and With CFA Two-Factor Model

Observed variable	Variable label	EFA factors η_1	EFA factors η_2	CFA factors η_1	CFA factors η_2
Y_1	Word order	.65	.15	.81	0
Y_2	Number recall	.96	−.04	.81	0
Y_3	Hand movements	.25	.40	.50	0
Y_4	Gestalt closure	−.12	.57	0	.50
Y_5	Triangles	−.02	.73	0	.73
Y_6	Spatial memory	.04	.64	0	.66
Y_7	Matrix analogies	.12	.53	0	.59
Y_8	Photo series	−.01	.79	0	.78

Note. $N = 200$. Exploratory factor analysis (EFA) results obtained using unweighted least squares estimation with oblimin rotation; interfactor correlation = .44. Confirmatory factor analysis (CFA) results obtained using maximum likelihood estimation; interfactor correlation = .56. Tabled CFA values are completely standardized factor loadings; values exactly equal to 0 are fixed rather than freely estimated.

simultaneous processing ability; thus, in the CFA model specification in Figure 1, Y_3 has a free loading on η_1 and its loading on η_2 is fixed to 0.

In addition to needing to specify the pattern of fixed factor loadings in CFA, another distinction between EFA and CFA is that researchers using EFA often do not have a strong hypothesis about the optimal number of common factors M, whereas applications of CFA usually do begin with a strong hypothesis about the number of factors. But this is actually a procedural difference rather than a difference in the statistical model itself. In fact, with EFA, it is still necessary to specify how many factors a model has before it is estimated,[3] but then various criteria are available to determine an optimal value of M for the data at hand, which may not be the same as the researcher's initial guess, and subsequently models with differing numbers for M may be estimated. In CFA, conversely, researchers typically begin with a strong hypothesis about the number of factors. However, many CFA studies still undertake a model-comparison procedure by which models with differing numbers of factors are compared. Even though CFA is more of a strong hypothesis-testing framework, methodologists still recommend testing competing models against one another. Finally, model fit statistics which are usually associated with CFA can also be effectively used in EFA to determine the optimal number of common factors (Preacher, Zhang, Kim, & Mels, 2013).[4]

As mentioned earlier, rotation is used to facilitate interpretation of an EFA model with $M \geq 2$ factors, whereas rotation does not arise in CFA. A crucial decision in the context of EFA, then, is choice of rotation procedure. Rotations are either *orthogonal*, which forces factors to be uncorrelated with each other, or *oblique*, which leads to freely estimated interfactor correlations. Despite that the default rotation (i.e., *varimax*) in popular software (e.g., SPSS and SAS) is orthogonal, consensus among methodologists is that oblique rotations should be preferred in almost every situation[5] because it is unrealistic to expect the constructs that common factors purportedly represent to be perfectly uncorrelated. The EFA results above were based on an oblique rotation which led to a substantial interfactor correlation of $r = .44$. The extent of factor correlation should be considered an empirical question, but orthogonal rotation artificially fixes interfactor correlations to zero before the question is asked, which can distort interpretations of the factor loadings themselves.

Researchers often do not appreciate the fact that for any EFA model with $M \geq 2$ factors, there is an infinite number of factor patterns (i.e., rotations) which fit the observed data equally well; this property is known as *rotational indeterminacy*. Consequently, it is good practice to apply several different (oblique) rotations to verify that the overall interpretation of the factors is relatively consistent; use of *analytical rotation* (Browne, 2001) is especially helpful in this regard. For instance, using *oblimin* rotation, a weight parameter can be varied which affects the balance between *row parsimony* (the extent to which each variable loads strongly on one factor and near-zero on other factors) and *column parsimony* (the extent to which each factor has large loadings for some variables and small loadings for others); in turn, interfactor correlations are also adjusted. Concern over rotational indeterminacy was the primary reason for Jöreskog's (1969) development of CFA: estimation of a properly specified CFA model leads to a single, unique set of factor loading and interfactor correlation estimates.

Because the EFA and CFA models are both particular instantiations of a more general common factor model, the names *exploratory* and *confirmatory* factor analysis pertain more to alternative statistical models than to different research objectives (Brown, 2015; McArdle, 2011). Formal statistical inference is a hallmark of confirmatory research, but it is possible to carry out inference in the context of EFA (e.g., maximum likelihood estimation of an EFA model gives a significance test of the number of factors). Conversely, exploratory research often involves comparing alternative models (representing alternative conceptualisations of data) with respect to their fit to data and interpretative quality; this objective is often carried out using CFA (e.g., Boomsma, 2000). In practice, however, it may be difficult to determine whether a given research scenario calls for EFA or CFA. Later, we offer some guidance for reaching this decision. But first it is important to address factor analysis for individual items versus total test or scale scores.

Factor Analysis of Item-Level Variables

Factor analysis developed primarily as a procedure for uncovering the latent variable structure of a set of scores from an administration of many different tests or scales. Because test scores tend to be approximately continuous variables, factor analysis naturally evolved as a method of analysing the Pearson product–moment correlations among the observed variables, which led to the linear regression form of the common factor

[3] Some software (e.g., SPSS and SAS) defaults to determine the number of factors according to the number of eigenvalues of the observed correlation matrix greater than 1. This default criterion is well-known to be faulty (Preacher & MacCallum, 2003), and researchers should always consult other criteria to determine the optimal number of factors and estimate and interpret alternative models accordingly.

[4] Such fit statistics are available in the EFA output of the *psych* package for R (Revelle, 2016) as well as SEM software capable of executing EFA; unfortunately, they are not produced by EFA procedures in SPSS and SAS.

[5] An important exception is rotation to an exploratory *bifactor* solution in which a general factor is orthogonal to specific factors (Jennrich & Bentler, 2012); bifactor models are briefly described later in this paper.

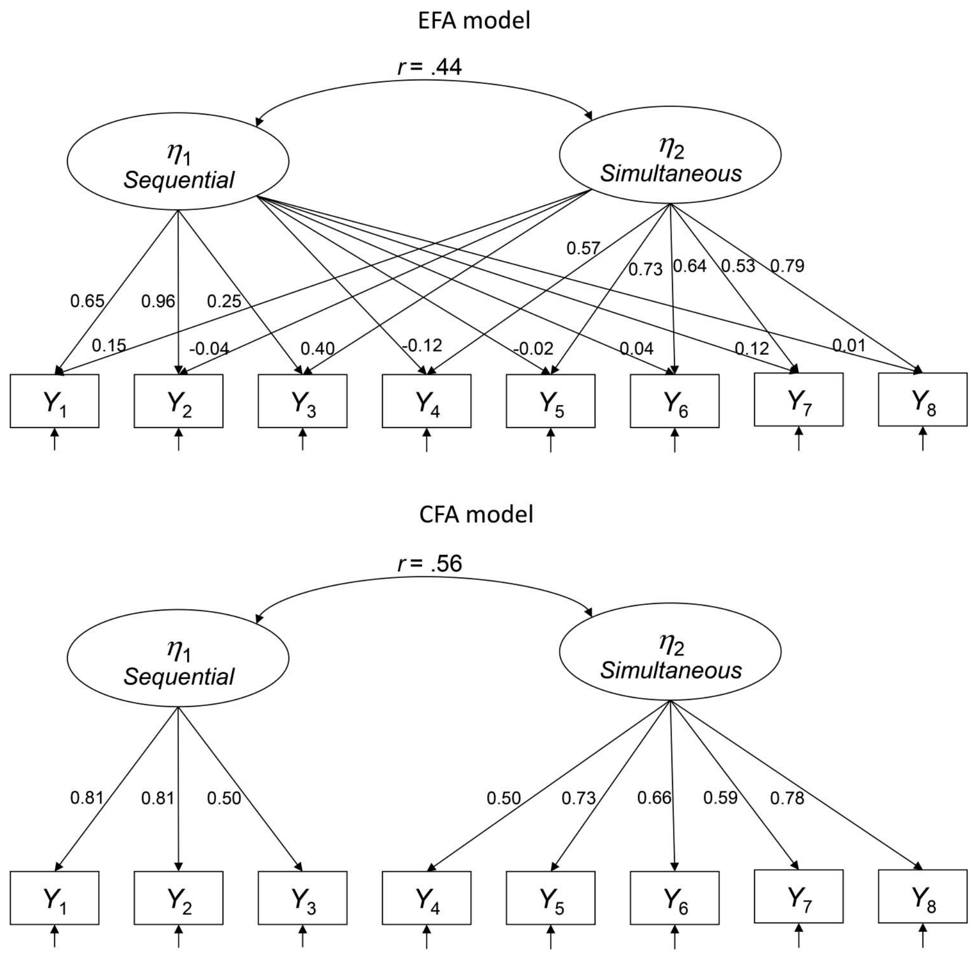

Figure 1. Path diagrams representing separate exploratory factor analysis (top) and confirmatory factor analysis (bottom) models for eight subtest variables.

model (Bartholomew, 2007). In modern research, however, factor analysis frequently involves analysis of individual items from a single test or questionnaire rather than analysis of total scores from multiple tests or subscales. Because items typically produce binary or ordered categorical variables (e.g., Likert-type items producing discrete, integer responses), the linear common factor model implied by analysis of product-moment correlations or covariances is clearly inappropriate for such variables. Just as a nonlinear procedure such as logistic regression is preferable to ordinary linear regression for a categorical outcome, a categorical variable methodology is often preferable for factor analyses of items. The potential consequences of factor analysing items as continuous variables include incorrect conclusions about the number of factors (or model fit more generally), biased factor loadings and interfactor correlations, and biased parameter standard errors, leading to incorrect significance tests and confidence intervals (for further explanation, see Flora, LaBrish, & Chalmers, 2012, and references therein). Unfortunately, countless publications present factor analyses which ignore the categorical nature of item-level variables.

Fortunately, modern software makes it is easy to carry out factor analyses which properly account for the categorical nature of items. A prominent approach is to analyse polychoric correlations rather than product-moment correlations[6]; this approach has been shown to perform well in numerous simulation studies (e.g., Flora & Curran, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Both EFA and CFA of polychoric correlations are available in multiple software packages (but not the base EFA procedures in SPSS and SAS). Wirth and Edwards (2007) give further description of this approach as well as alternative methods of item factor analysis.

Yet, not all items need to be analysed as categorical variables. For example, items using a visual analog scale produce approximately continuous variables. Furthermore, simulation studies (e.g., Rhemtulla et al., 2012) indicate that it may be preferable to treat items with more than five response categories as continuous variables for the purpose of factor analyses, as long as a method to account for their inherent nonnormality is used (such as the Satorra–Bentler procedure; Satorra & Bentler, 1994). But if the

[6] The tetrachoric correlation is a special case of the polychoric correlation obtained when both variables are binary. Karl Pearson was instrumental in the development of both the tetrachoric correlation and the product-moment correlation. Thus, it can be misleading to refer to only the latter as simply the Pearson correlation.

observed item response distribution is heavily skewed despite having five or more categories, then the polychoric correlation approach may still be optimal. For further details, see Finney and DiStefano (2013).

Applying Factor Analysis in Psychological Research

Many published factor analyses focus on the psychometric development and validation of individual scales, which is why factor analyses of items are so prevalent. Yet, factor analysis has also been an important tool for broad research questions that are not inherently tied to any one test or questionnaire. Perhaps most famously, factor analysis has a long history of contributing to theories of intelligence (see Horn & McArdle, 2007). Factor analysis is also prominent in studies on the structure of personality (e.g., Costa, Busch, Zonderman, & McCrae, 1986; but see Block, 1995, for a criticism of factor analyses in this area). More recently, factor analytic models have been used to represent the structure of executive functioning in neuropsychological research (e.g., Miyake et al., 2000) and the basic structure of psychopathology (e.g., Caspi et al., 2014). Aside from studies on broad psychological processes, factor analysis is most typically employed to help assess validity evidence for the use of a particular scale. The remainder of this paper focuses on this general purpose of factor analysis.

Validity Evidence for New Scales

Researchers presenting a new scale often use factor analysis to show that the internal structure (i.e., the dimensionality) of the items is consistent with expectations regarding the construct(s) that the scale is intended to measure. Specifically, the individual items are factor analysed, ideally using a large sample of participants from the primary population in which the scale is meant to be used in future applications. Although both EFA and CFA have been used for this purpose, EFA is preferable at the beginning phase of scale development because there may be unanticipated, but substantively meaningful, factors influencing subsets of items or unanticipated cross-loadings. For example, refer back to results for Y_3 (hand movements) in Table 1: Here, the CFA model placed this variable on the first factor, η_1, based on an a priori expectation, whereas the EFA results suggest that Y_3 may be an even stronger indicator of the second factor, η_2. Thus, although the CFA model fits the data adequately, EFA shows more clearly that certain predictions regarding factor structure could be incorrect.

This use of factor analysis is often presented as providing evidence for the *construct validity* of the use of scores from a new scale.[7] Although a correspondence between the expected dimensionality of a scale and the results of a factor analysis may be construed as evidence of construct validity of scale scores, this result does not comprehensively establish construct validity. Indeed, construct validity is a very broad concept and virtually all empirical evidence regarding a test's use contributes construct validity evidence (Cronbach & Meehl, 1955). Construct validation can be considered a program of research, broken into three major parts: substantive, structural, and external (Benson, 1998; Loevinger, 1957). In the substantive phase, researchers are primarily interested in the theory of the construct and developing a pool of test items which addresses all the theoretical aspects. The goal of the structural phase is to determine whether the items' empirical relations to one another are consistent with those theoretical aspects. In the final, external phase, researchers focus on how the construct is related to other constructs, known measures, and outcomes.

Factor analysis is squarely in the structural phase of construct validation and is one statistical vehicle for assessing whether the measurement theory is reflected in empirical data. Zumbo (2006) provides a more detailed discussion of the place of factor analysis (and latent variable models more generally) in the validation process; in particular, the factor analysis model can be used to represent a theoretical explanation for observed test scores, supporting a strong form of construct validity. The volume by Zumbo and Chan (2014) presents a comprehensive overview of validity theory and past and current practices of validation studies in the social, behavioural, and health sciences; these chapters cite a noticeable increase in the use of factor analysis in validity studies over the preceding decades.

If a researcher with extensive knowledge of a certain content domain expects an item pool to represent a certain number of factors, then either EFA or CFA can be used to assess whether empirical evidence is consistent with that expectation. If the prediction is not supported, depending on the pattern of results, the researcher may revise her theory about the measurement of the proposed construct(s), or she may simply decide that the questionnaire needs minor modifications. Such modifications might entail removal of one or more items based on low factor loadings or communality estimates. But a factor analysis does not give sufficient reason for removing an item: it may be that an item represents an important aspect of the conceptual content domain of a given construct. Hence, removing the item would be harmful for the *content validity* (Sireci, 1998) of scale scores intended to represent that construct. Furthermore, any factor analysis is influenced by sampling error, and thus items with weak factor loadings from one dataset might not have weak loadings when the analysis is repeated with a separate sample. Hence, replication of factor analytic models is particularly important for scale development.

Occasionally, researchers remove an item from a scale because it has a cross-loading in a rotated EFA factor pattern; that is, the item has a salient loading on more than one factor, even after different rotations have been explored. Similarly, in a CFA model, model fit may be improved if an item is allowed to have a nonzero, freely estimated loading on more than one factor. For example, an item may have a loading of .45 on one factor and .35 on another factor. Yet, because this item has a salient loading on the first factor, removing it would affect the reliability and content validity of (sub)scale scores calculated as measures of the construct represented by that factor (see McDonald, 1999, for the association between the common factor model and scale reliability). In addition, although the item has a weaker relation with the second factor, removal may also harm the reliability or content validity of scores representing that corresponding construct. There is nothing inherently problematic about allowing a given item to be used for

[7] Validity is not a property of a test or scale itself. Rather, validity is a property of a particular use (i.e., an inference or interpretation) of the scores obtained from a test or scale (see Messick, 1995).

more than one scale or subscale.[8] For example, an item asking about sleep difficulties in a general psychopathology questionnaire may be important for measuring both depression and anxiety. Overall, the desire to obtain a clean factor pattern should not dictate decisions about whether to remove an item from a test or questionnaire.

Validity Evidence for Extant Scales: Should I Do a Factor Analysis in *My* Study?

Because scale validation is an ongoing process (e.g., Zumbo, 2006), all empirical studies using a given scale contribute evidence for the validity of the use of that scale. For example, evidence for the scale's *criterion validity* is provided if a new study reports an association between the scale and an important outcome variable. Further construct validity evidence is given by any reliability estimates reported in the new study, given that reliability is an aspect of construct validity (Cronbach & Meehl, 1955). With the ongoing nature of validation in mind, each new study using a previously developed scale provides a new opportunity for an evaluation of the psychometric properties of that scale, which may very well include a new factor analysis of the scale's items or a factor analysis including the total scale score as an observed variable along with other the other scales in the study. There is also a good chance that the original scale development study was faulty in some way (see below), which, in turn, behooves researchers to reevaluate the factor model and general psychometric quality of the scale.

But to some extent, researchers using a previously developed test can rely on earlier validity studies to justify their own use of the test, especially if the test is well established. Additionally, it may be that the sample size of a new study is inadequate for in-depth analyses of the psychometric properties of an extant test. With factor analysis, there is a relatively large literature on the impact of sample size on estimate precision and (for CFA) statistical power. In brief, it is well-known that EFA and CFA are best considered as large-sample methods, with sample sizes in the hundreds typically needed to obtain accurate results (e.g., for EFA see MacCallum, Widaman, Preacher, & Hong, 2001; for CFA see Jackson, 2003). Yet, depending on features such as the number of variables per factor and the strength of factor loadings (or communalities), it may be reasonable to conduct factor analyses with smaller sample sizes (e.g., de Winter, Dodou, & Weiringa, 2009).

It is important to recognise that the internal structure of a test may differ across populations with respect to the number of factors or factor pattern (i.e., both the strength and pattern of zero and nonzero factor loadings). For instance, an item with the prompt "I do not do the same activities I used to" may be a strong indicator of negative affect in a young-adult clinical population but a weak indicator in a population of elderly respondents. Consequently, there is a possibility that results using the test in a different population from that used in the original scale development will confound true construct differences with measurement differences. Thus, it is prudent to reevaluate the psychometric properties of a scale for its use in a new population, including an investigation of the scale's factor structure. This investigation should begin with an EFA to investigate whether the same number of major factors underlies the scale in both populations. If that is the case, then ideally a follow-up study will also include a sample from the population in which the scale was originally developed to allow an investigation of *measurement invariance* in which the parameters of a CFA model can be formally tested for equivalence across populations. Specifically, a multigroup CFA model in which all parameters are set to be equal across groups (a complete invariance model) is statistically compared with a sequence of models in which subsets of parameters are allowed to vary freely across groups; see Millsap (2011) for a comprehensive overview. Although this CFA procedure is more popular, it is also possible to assess measurement invariance with EFA (Finch & French, 2008).

Other situations calling for a reevaluation of the factor structure of a test occur when the test has been modified in some way, such as when a short form is created (Smith, McCarthy, & Anderson, 2000), when a test is translated to a different language (e.g., from English to French), or when a test is modified to obtain a different reporter's perspective on a target person (e.g., a self-report test is modified to allow peer report). In essence, though, such modifications result in the creation of a *new* test, the use of which should then be validated as such.

Should I Use EFA or CFA?

In short, the choice between EFA and CFA depends on the particularities of the research scenario, and thus it is difficult to offer any rules of thumb. Because EFA and CFA are both rooted in the common factor model, one often can reach the same general conclusions about the factor structure of a set of variables using either an EFA or a CFA approach (McArdle, 2011). As stated earlier, EFA and CFA are not distinct methods; rather, the unrestricted factor pattern of the EFA model and the restricted, independent clusters pattern often tested with CFA should be considered two ends of the same continuum. In fact, it is possible to carry out EFA in the CFA (i.e., SEM) framework, which produces an unrestricted, EFA-type factor solution while providing features more easily obtained with CFA, such as statistical tests and confidence intervals for parameter estimates (Brown, 2015). Conversely, confirmatory features can be obtained from a more traditional EFA approach, such as formal tests of the number of common factors (e.g., the chi-square test of fit or, preferably, a test of *close fit* based on the root-mean-square error of approximation (RMSEA) statistic; Preacher et al., 2013) and *target rotation* toward a hypothesised factor pattern (Browne, 2001; MacCallum, 2009). Thus, the distinctions between the EFA model and the CFA model do not map directly onto the practicing researcher's more conceptual distinction between an "exploratory" analysis and a "confirmatory" analysis. Nevertheless, there are certain scenarios which clearly call for an EFA rather than CFA and vice versa. According to Brown (2015),

> CFA requires a strong empirical or conceptual foundation to guide the specification and evaluation of a factor model. Accordingly, CFA is typically used in later phases of scale development or construct validation after the underlying structure has been tentatively established by prior empirical analyses using EFA, as well as on theoretical grounds. (p. 41)

[8] It is important to keep in mind, though, that if an item is used for two separate scales, then those scales will necessarily be correlated to some degree by virtue of sharing the item.

Researchers developing an entirely new scale should use EFA to examine the dimensionality of the items; keep in mind that this analysis may demand a categorical-variable method to account for the discrete nature of item-level variables. It is likely that a new scale is expected to measure a single construct, but EFA may be used to determine whether the scale may be better considered multifactorial (even with one or two minor common factors which influence only a small subset of the items), and if so, whether there is a clear, conceptual interpretation of the obtained factors. If the factor pattern (in terms of the number of factors or the factor loadings) is not consistent with expectations, then one may revise the scale, which, in turn, calls for a new factor analysis of the revised set of items. Of course, any reanalysis of the same dataset essentially makes the analyses exploratory (in the sense that formal significance tests or confidence interval estimates are compromised), even if the initial goals were confirmatory.

As explained earlier, CFA should be used when researchers have strong a priori hypotheses about the factor pattern underlying a set of observed variables (which may be individual items in a single scale or a set of total scale or subscale scores). Such is likely the case when a previously developed test is being evaluated in a new sample drawn from the same population as that observed in the original scale development study. Thus, prior EFA results can be used to inform the specification of CFA models for further construct validity evidence (below we discuss replication of EFA models using CFA). Additionally, CFA may be used to verify that the original factor structure is retained when a test has been revised by adding or removing a few items. Keep in mind that CFA is still viable if one has hypothesised more than one potential model for the variables. For example, one hypothesised factor pattern may be a simple one-factor model, consistent with the expectation that a set of items measures a single construct, while an alternative factor pattern may be a *bifactor* model in which a single, general factor still underlies all items, but distinct groups of items share covariation beyond that explained by the general factor (Reise, 2012).[9] Using CFA, formal model comparisons can be used to test such hypothesised models against one another.

Furthermore, a potential advantage of CFA is that it allows the specification of error covariance parameters which can be used to capture method effects such as similar item phrasings, differential susceptibility to demand characteristics or social desirability, difficulty processing reverse-worded items, and so on (Brown, 2015). Ignoring such subtle method effects can produce inaccurate estimates of factor loadings and scale reliability (e.g., Raykov & Marcoulides, 2011). But it is important to recognise that error covariances should be specified judiciously; the goal is to represent plausible, conceptually justified method effects rather than simply to obtain good model fit. As with EFA, interpretational quality is essential for CFA.

Another advantage of CFA is that the model can be expanded into a larger SEM whereby key outcomes are regressed on the factors representing the construct(s) measured by the items; this expansion is useful for assessing a scale's criterion validity (Zumbo, 2006). Moving beyond factor analyses of the individual items in a given test or questionnaire, factor analytic models are also useful for assessing the *convergent* and *discriminant* (or *divergent*) *validity* of the scores from a new test with respect to associations with scores from other instruments. Convergent and discriminant validity represent important aspects of construct validity (Campbell & Fiske, 1959). In particular, convergent and discriminant validity can be assessed by fitting a CFA model to a *multitrait-multimethod* covariance matrix in which two or more constructs are each measured using two or more methods (one of which may be a new scale under development; Marsh & Grayson, 1995).

Model Modification

If no hypothesised CFA model adequately fits the sample data, it can be difficult to decide how to proceed. One option is simply to conclude that the hypotheses or expectations behind the model(s) were not empirically supported; such is the nature of science. Most often, though, researchers seek a revised model which does fit the data adequately; doing so necessarily moves the analysis from a confirmatory mode to an exploratory mode, even if the researcher continues to fit one or more revised CFA models to the same dataset.

In this situation, researchers may attempt to find an alternative model either by continuing to fit CFA models or by moving to EFA. In many cases, though, it is preferable to undertake a traditional EFA approach. For example, perhaps a scale is hypothesised to have a single underlying factor, but the one-factor model is rejected. At this point, a researcher might not have any strong rationale for the specification of a multifactorial model, and so an EFA may be used to determine both how many factors are needed to account for the associations among the observed variables and how strongly each variable is influenced by the factors. As another example, suppose a researcher specifies a two-factor CFA model which turns out not to fit the data adequately. If this researcher is stuck in a CFA perspective, she might not consider that the number of specified factors is incorrect, instead continually revising the two-factor CFA model by adding cross-loadings or freeing error covariance parameters, despite that an EFA could have revealed a well-fitting, conceptually plausible three-factor pattern.

Nonetheless, there are situations in which post hoc modifications of a CFA model are justifiable. For example, if a test which is intended to be unidimensional contains some positively worded items and some negatively worded items, then the one-factor model might not fit the data adequately, whereas residual covariance output indicates freeing error covariances among the negatively worded items would improve model fit (e.g., Brown, 2003, used CFA to assess this possibility with the Penn State Worry Questionnaire). Hence, it is reasonable to specify a revised one-factor model which includes these error covariances as freely estimated parameters (but it would likely be unnecessary to free error covariances both among positive items and among negative items; doing so could lead to underidentification). Although this new CFA model should be considered exploratory, the *essential unidimensionality* (Stout, 1987) of the item set is still supported if the revised one-factor model fits the data well and the factor loadings are uniformly large. Furthermore, rather than indicating that there should be free error covariances among a subset of items (which is impossible in EFA) in a one-factor model, an EFA would likely indicate that there are two distinct common factors under-

[9] Alternatively, a *higher-order factor* model may be specified, which is mathematically related to the bifactor model, but with a different conceptual interpretation (Yung, Thissen, & McLeod, 1999).

lying the items (one factor for positive items and one for negative items), which would not be consistent with essential unidimensionality. Overall though, unless revisions to a poorly fitting CFA model are justifiable in terms of psychological interpretation or conceptually plausible method effects, it is usually prudent to move from CFA into an explicitly exploratory EFA framework.

Replication of EFA With CFA

It is not logical to obtain a good-fitting factor structure using EFA and then seek to confirm that structure using CFA with the *same* dataset; doing so capitalizes on sample-specific, chance relationships and in no way verifies the EFA findings. Unfortunately, there are published examples of this poor practice. Instead, researchers may identify a plausible factor model for their observed variables using EFA with one dataset and then seek to confirm that factor pattern by fitting a CFA model to a *separate* dataset. Specifically, first a researcher uses EFA with one sample to determine the optimal number of common factors and rotates the factors to provide an interpretable solution. Next, using a separate dataset, a CFA model is estimated which has the same number of factors and with factor loadings specified such that observed variables that had relatively strong loadings on a given factor in the EFA solution are then allowed to load freely on the corresponding factor in the CFA model while relatively weak loadings in the EFA solution are fixed to zero in the CFA model.

To specify a CFA model based on EFA results, researchers are forced to make two arbitrary decisions. First, they must decide how large a factor loading in the EFA solution needs to be in order to conclude that the corresponding element of the factor loading matrix should be freely estimated in the CFA model. In particular, researchers often use a cutoff of 0.30 or 0.40 to conclude that a given factor loading is salient, while items with lower loadings on all factors are removed; however, there is no real statistical justification for such a cutoff (Gorsuch, 1983). Instead, decisions about item removal should be made along with other information regarding reliability and content validity. Second, a decision must be made regarding cross-loadings. For example, if a variable has an EFA loading of 0.25 on one factor and 0.40 on another factor, as is the case for Y_3 (hand movements) in Table 1, should the variable be allowed to have free loadings on just one or both factors in the CFA model? Sometimes researchers simply remove this variable from subsequent analyses, but again, that is problematic for reliability and validity. Alternatively, researchers might allow the variable to have a free loading on only one of the two factors, but doing so is likely to contribute to poor fit of the CFA model as explained next. Therefore, it is prudent to allow such a variable to cross-load on both factors in the CFA model.

Although the CFA model specified in this manner may fit the data adequately, researchers are frequently surprised if the CFA model does not fit. One explanation for this occurrence is that the EFA model and the subsequent CFA model are not equivalent; small but nonzero associations (i.e., weak factor loadings) found in EFA are fixed to exactly zero in CFA. Because the CFA model is more restricted than the EFA model, the fit of the CFA model to data is inevitably worse in an absolute sense (van Prooijen & van der Kloot, 2001).[10] For example, returning to Table 1, although Y_1 (word order) has a small loading on the second factor (η_2) in the EFA solution ($\hat{\lambda} = 0.15$), it is subject to sampling error and the population association between Y_1 and η_2 could be larger. Consequently, forcing this association to equal exactly zero in a CFA model could be a substantial enough misspecification to cause the CFA model to have inadequate fit in a separate dataset. Furthermore, as earlier, the EFA solution shows that Y_3 (hand movements) has notable loadings on both η_1 and η_2, suggesting that restricting Y_3 to be an indicator of only one factor in the CFA model is particularly likely to contribute to model misfit.

Upon finding that the CFA model specified based on an initial EFA does not fit in a separate dataset, the researcher is again put into an exploratory mode of data analysis. In this case, methodologists have recommended that EFA be used again in the second, separate sample (Floyd & Widaman, 1995; Osborne & Fitzpatrick, 2012). Criteria for assessing the extent to which replications of EFA models are successful are described in Osborne and Fitzpatrick (2012).

Common Mistakes With EFA

Many common errors in EFA are thoroughly reviewed in the "best practices" resources cited earlier; yet, they have persisted to some extent. These errors include mistaking principal components analysis (PCA) for factor analysis; PCA is not a type of factor analysis and can lead to incorrect conclusions if viewed as such. Next, it is a mistake to base the decision about the number of common factors on the criterion of the number of eigenvalues of the correlation matrix which are greater than 1 (also known as the Kaiser-Guttman rule), as mentioned earlier. Furthermore, oblique rotations should almost always be used rather than orthogonal rotation. Finally, when the observed variables are items, which tend to have ordered categorical distributions, then the EFA should adequately account for the categorical nature of the items; categorical variable methods have become more popular with CFA, but EFA applications seem to be lagging in this regard. Regrettably, these common mistakes continue to arise in large part because the base EFA procedures in popular software (SPSS and SAS) feature imprudent defaults (by default, PCA is the estimator, the Kaiser-Guttman rule determines the number of factors, and an orthogonal rotation, varimax, is implemented), and cannot readily implement methods to account for the categorical nature of items. Fortunately, many researchers are becoming more facile with alternative software which overcomes these limitations, such as Mplus (Muthén & Muthén, 2015) and the *psych* package in R (Revelle, 2016).

Another common mistake in EFA—one not addressed in great detail in previous resources—concerns the *naming fallacy*. In particular, when researchers estimate a model with too many factors, there is a temptation to ascribe a conceptual meaning to them, despite that such factors may be weakly defined based on only one or two observed variables or may be defined by variables which bear little conceptual relation to each other. In other situations, certain factors may be primarily determined by method effects (such as item wording) or an ambiguous combination of method effects and more meaningful constructs. Even statistically

[10] But model fit statistics which adjust for model parsimony (such as RMSEA and CFI) might indicate that the CFA model fit is as good or better than EFA because the CFA model is more parsimonious (i.e., it has larger degrees of freedom).

strong factors may lead to disagreements over their name and interpretation. In short, being able to name a factor does not imply that the factor represents a meaningful psychological phenomenon.

Common Mistakes With CFA

It is important to keep in mind that model fit is a matter of degree rather than a binary yes-or-no decision and that popular guidelines for determining adequate model fit do not generalise to all types of CFA models (e.g., Marsh, Hau, & Wen, 2004); consequently, the model comparison procedure advocated previously is particularly advantageous. Furthermore, researchers often do not recognise that popular model fit statistics (e.g., RMSEA, comparative fit index [CFI]) only summarise the overall, global fit of the model to data and that it is possible for *local* misfit to be present despite good global fit. That is, even when global fit is judged to be adequate, there may still be substantial residual correlations among certain variable pairs (Tomarken & Waller, 2003). In scale development, such residual correlations imply a degree of redundancy[11] between the items involved or method artifacts, as above. Yet, it often turns out that more than one model fits a given dataset reasonably well; in this case, one should present complete results for each well-fitting model and discuss their relative conceptual merit.

Unfortunately, researchers occasionally take good model fit as the only indication that the hypotheses that guided model specification were supported, without considering estimates of factor loadings, interfactor correlations, or other parameters. In scale development, one might conclude that good fit of a two-factor model indicates that two reliable subscales may be formed from the items. But if factor loadings are weak, despite good model fit, then the reliability of these subscales is compromised. Therefore, as with EFA, interpretation of individual parameter estimates is a critical aspect of CFA.

Upon finding that a hypothesised model does not adequately fit data, researchers often turn to *modification indices* to reveal which previously restricted parameters are most likely to improve model fit if they are instead freely estimated in a revised model. Similar information can also be obtained from the residual correlation or covariance matrix from a fitted CFA model. There are three problems with model revisions based on modification indices or residual covariances: First, the suggested free parameters might not be conceptually justifiable and could even be downright nonsensical. Second, freeing too many parameters (or the wrong parameters) could lead to an underidentified model.[12] Finally, such model revisions capitalise on chance relations specific to one particular sample which are then unlikely to replicate in independent samples (e.g., Chou & Huh, 2012; MacCallum, Roznowski, & Necowitz, 1992). Thus, even though one is using CFA methods, post hoc model modification leads to an exploratatively obtained model which should be cross-validated with an independent sample; following modification indices may improve model fit in the current sample, but does not necessarily lead one to the correct population model.

Summary

This paper reviewed the purpose and practice of factor analysis in psychological research, focusing on scale development and validation. The statistical distinction between EFA and CFA was clarified by placing each as a particular instantiation of the common factor model. In this general framework, it can be seen that EFA and CFA are not disparate statistical procedures; consequently, many of the same general conceptual conclusions about the underlying structure of a set of observed variables can be drawn using either EFA or CFA. Nonetheless, there are situations in which EFA is preferable to CFA and vice versa.

The past century has seen a large shift in the primary reason that psychologists use factor analysis. Specifically, factor analysis was originally employed for the purpose of basic theory development and evaluation, where the observed variables tended to be total scores from a battery of diverse psychological tests. But in modern research, factor analysis is much more commonly used to examine the internal structure of individual items from a single test or questionnaire. A critical consequence of this shift, which is often not recognised by practicing researchers or even textbook authors, is that observed variables in modern applications are frequently item-level variables which are characterised by ordered (or binary) categorical distributions. Fortunately, modern software makes it relatively easy to conduct factor analyses which explicitly account for the categorical nature of items by analysing polychoric correlations.

Factor analysis is now commonly used to assess validity evidence for the uses of newly developed scales, particularly regarding construct validity (although factor analysis alone is not sufficient for establishing construct validity). For instance, if a new scale is intended to measure a single construct, then a factor analysis should show that a single, dominant factor influences all items (but perhaps within a bifactor structure, as alluded to earlier). Although expectations about the internal structure of a scale may be precise enough to imply specification of a CFA model, the newness of the measure may be such that unexpected patterns among the variables are present, which implies that EFA may be appropriate. It is also important to interpret results from factor analyses along with other considerations, particularly content validity and reliability, when deciding whether to remove items from a scale. Ultimately, the value of a given factor may be judged based on whether the implied scale scores show evidence of criterion validity.

Owing to the ongoing nature of validation, it is also important for new studies to investigate the psychometric properties of extant tests which have been developed in previous studies, especially if the test is not well established, if the test is being used in a different population, or if the test has been revised in some way (e.g., in a short-form creation or translation to a different language); once again, factor analysis is a valuable tool in this endeavor. In these situations, although it may be that hypotheses are strong enough to allow the a priori specification of one or more CFA models, a more cautious approach would be to use EFA so that unexpected factor patterns can be uncovered. Either way, the researcher has moved

[11] This redundancy can cause the association between two variables to be even stronger than can be explained by the underlying factor(s).

[12] Underidentification is not only a matter of whether a hypothesized model has negative degrees of freedom (Bollen, 1989). For example, if a CFA model has too many free factor loadings, then rotational indeterminacy can come into play; this situation calls for an EFA for proper factor rotation.

into an exploratory mode of analysis and it is especially important to validate the final model using a separate sample.

Regardless of whether an optimal factor model for a given set of observed variables (or items) is obtained using EFA or CFA, one should seek to replicate that model. The continual nature of validation dictates that the factor pattern should be replicated in new studies with independent samples. If these new samples are drawn from a different population, then it is important to carry out formal tests of measurement invariance to investigate whether and how the factor model differs across populations. Recently, psychologists have been made more aware that replication is a critical feature of scientific progress. Even as a large-sample procedure, factor analysis is not immune to this concern.

Résumé

De nombreuses ressources de haute qualité existent pour décrire les meilleures pratiques en matière de mise en œuvre de l'analyse factorielle exploratoire (AFE) et de l'analyse factorielle confirmatoire (AFC). Or, en partie dû à la complexité de ces procédures, une certaine confusion persiste entre les psychologues quant à la mise en œuvre de l'AFE et de l'AFC. L'une des principales sources de ces malentendus réside dans la distinction mathématique entre l'AFE et l'AFC. Le présent article utilise un bref exemple pour illustrer la différence entre les modèles statistiques sous-jacents à l'AFE et l'AFC, lesquels sont tous deux des instanciations particulières du modèle factoriel plus général. Ensuite, d'importantes considérations relatives à la mise en œuvre de l'AFE et de l'AFC, abordées dans le présent article, incluent la nécessité de tenir compte de la nature catégorique de variables observées au niveau des items dans les analyses factorielles, l'utilisation de l'analyse factorielle dans l'étude de propriétés psychométriques de nouveaux tests ou questionnaires et de tests élaborés dans le passé, des décisions quant à la procédure la plus appropriée – soit l'AFE ou l'AFC – dans ces contextes et l'importance de la reproduction de modèles d'analyse factorielle dans la poursuite de la validation en cours.

Mots-clés : analyse factorielle exploratoire, analyse factorielle confirmatoire, meilleures pratiques, validité, propriétés psychométriques.

References

Bartholomew, D. J. (2007). Three faces of factor analysis. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 9–21). Mahwah, NJ: Erlbaum.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17,* 10–17. http://dx.doi.org/10.1111/j.1745-3992.1998.tb00616.x

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117,* 187–215. http://dx.doi.org/10.1037/0033-2909.117.2.187

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley. http://dx.doi.org/10.1002/9781118619179

Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling, 7,* 461–483.

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy, 41,* 1411–1426. http://dx.doi.org/10.1016/S0005-7967(03)00059-7

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

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

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105. http://dx.doi.org/10.1037/h0046016

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science, 2,* 119–137. http://dx.doi.org/10.1177/2167702613497473

Chou, C.-P., & Huh, J. (2012). Model modification in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 232–246). New York, NY: Guilford Press.

Costa, P. T. J., Jr., Busch, C. M., Zonderman, A. B., & McCrae, R. R. (1986). Correlations of MMPI factor scales with measures of the five factor model of personality. *Journal of Personality Assessment, 50,* 640–650. http://dx.doi.org/10.1207/s15327752jpa5004_10

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. http://dx.doi.org/10.1037/h0040957

de Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44,* 147–181. http://dx.doi.org/10.1080/00273170902794206

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

Finch, W. H., & French, B. F. (2008). Using exploratory factor analysis for locating invariant referents in factor invariance studies. *Journal of Modern Applied Statistical Methods, 7,* 223–233.

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *A second course in structural equation modeling* (2nd ed., pp. 439–492). Charlotte, NC: Information Age.

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

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 Quantitative Psychology and Measurement, 3,* 55.

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

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

Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 205–247). Mahwah, NJ: Erlbaum.

Jackson, D. L. (2003). Revisiting sample size and the number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10,* 128–141. http://dx.doi.org/10.1207/S15328007SEM1001_6

Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika, 77,* 442–454. http://dx.doi.org/10.1007/s11336-012-9269-1

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202. http://dx.doi.org/10.1007/BF02289343

Kaufman, A. S., & Kaufman, N. L. (1983). *K-ABC administration and scoring manual*. Circle Pines, MN: American Guidance Service.

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

Lawley, D. N., & Maxwell, A. E. (1963). *Factor analysis as a statistical method*. London, UK: Butterworth.
Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635–694. http://dx.doi.org/10.2466/pr0.1957.3.3.635
MacCallum, R. C. (2009). Factor analysis. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 123–147). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9780857020994.n6
MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111,* 490–504. http://dx.doi.org/10.1037/0033-2909.111.3.490
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. http://dx.doi.org/10.1207/S15327906MBR3604_06
Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Thousand Oaks, CA: Sage.
Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320–341. http://dx.doi.org/10.1207/s15328007sem1103_2
McArdle, J. J. (2011). Some ethical issues in factor analysis. In A. T. Panter & S. K. Sterba (Eds.), *Handbook of ethics in quantitative methodology* (pp. 313–339). New York, NY: Routledge.
McDonald, R. P. (1999). *Test theory: A unified treatment*. New York, NY: Routledge.
McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7,* 64–82. http://dx.doi.org/10.1037/1082-989X.7.1.64
Messick, S. (1995). Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749. http://dx.doi.org/10.1037/0003-066X.50.9.741
Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100. http://dx.doi.org/10.1006/cogp.1999.0734
Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.
Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. Louisville, KY: CreateSpace Independent Publishing Platform.
Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research & Evaluation, 17,* 2.
Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2,* 13–43. http://dx.doi.org/10.1207/S15328031US0201_02
Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research, 48,* 28–56. http://dx.doi.org/10.1080/00273171.2012.710386
Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Routledge.
Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. http://dx.doi.org/10.1080/00273171.2012.715555
Revelle, W. (2016). *Psych: Procedures for personality and psychological research* (Version 1.6.9) [Computer software]. Evanston, IL: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych
Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17,* 354–373. http://dx.doi.org/10.1037/a0029315
Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research, 45,* 83–117. http://dx.doi.org/10.1023/A:1006985528729
Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12,* 102–111. http://dx.doi.org/10.1037/1040-3590.12.1.102
Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52,* 589–617. http://dx.doi.org/10.1007/BF02294821
Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.
Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology, 112,* 578–598. http://dx.doi.org/10.1037/0021-843X.112.4.578
van Prooijen, J.-W., & van der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement, 61,* 777–792. http://dx.doi.org/10.1177/00131640121971518
Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12,* 58–79. http://dx.doi.org/10.1037/1082-989X.12.1.58
Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64,* 113–128. http://dx.doi.org/10.1007/BF02294531
Zumbo, B. D. (2006). Validity: Foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics* (Vol. 26, pp. 45–79). Amsterdam, the Netherlands: Elsevier. http://dx.doi.org/10.1016/S0169-7161(06)26003-6
Zumbo, B. D., & Chan, E. K. H. (2014). Setting the stage for validity and validation in social, behavioral, and health sciences: Trends in validation practices. In B. D. Zumbo & E. K. H. Chan (Eds.), *Validity and validation in social, behavioral, and health sciences* (pp. 3–8). New York, NY: Springer. http://dx.doi.org/10.1007/978-3-319-07794-9_1

Received June 10, 2016
Revision received January 23, 2017
Accepted January 25, 2017 ■