2026 ■ Vol. 22 ■ no. 1

A tutorial on evaluating confirmatory factor analysis with `lavaan` and `dynamic`: Integrating statistical and conceptual approaches

Thomas L. D. Boivin [a], Charles Veilleux [a], Maya Egerton-Graham [a], Pier-Olivier Caron [b] & Patrick Gaudreau [a]

[a]School of Psychology, University of Ottawa

[b]Département des Sciences Humaines, Lettres et Communication, Université TÉLUQ

Abstract ■ Confirmatory factor analysis (CFA) is a statistical method that allows researchers to quantify how well the hypothesized factorial structure of a psychological measure fits the associated data. Fit indices are statistical measures that evaluate the factorial structure by quantifying the degree of model/data fit or misspecification. Researchers generally use traditional or fixed cutoff values of fit indices to judge the appropriateness of their CFA models. However, many authors have discussed the limitations of using these cutoffs, given that they cannot be generalized to all models. Recently, McNeish and Wolf (2023) developed the Dynamic Fit Index (DFI) approach, which generates cutoff values of fit indices that are tailored to the characteristics of a specific model. In this tutorial, a CFA of the Attainment of School Achievement Goal Scale (A-SAGS) was conducted using the lavaan package in R. Next, cutoff values for the fit indices were generated using the dynamic package in R. The model fit of the A-SAGS was mostly satisfactory when interpretations were based on fixed cutoff values. However, the hypothesized model did not reach the dynamic cutoff values, prompting a more thorough investigation into the possible sources of model misspecification. The DFI approach is easily implemented, highly promising, and offers researchers valuable insights as they attempt to improve the construct validity of a given psychological measure.

Keywords ■ Confirmatory Factor Analysis, Fit evaluation, Fit indices, Dynamic fit. **Tools** ■ R, lavaan, dynamic.

✉ tboiv052@uottawa.ca

10.20982/tqmp.22.1.p009

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers ■ Two anonymous reviewers

Ce texte a reçu le prix SQRP-TQMP remis lors de la conférence annuelle de la Société québécoise pour la recherche en psychologie en 2026.

Introduction

Confirmatory factor analysis (CFA) is a statistical method used to evaluate whether the theoretically proposed structure of a measurement model fits the observed data (Jöreskog, 2007). Given that latent variables represent constructs that are not directly observable, their manifest indicators (Byrne, 2005) are expected to be an accurate representation of the construct (Weston & Gore, 2006). Furthermore, indicators, which are most often the items of a questionnaire, are expected to reflect their respective latent variable and to load onto it (i.e., correlate with the other items reflecting the latent variable). In other words, the correlation between the items is explained by a common, unobservable cause which is represented by a latent variable (Brown & Moore, 2012). When theory dictates it, a given item can be set to correspond to or "load onto" multiple variables, as opposed to just one. However, indicators are generally not permitted to cross-load onto multiple latent variables (McDonald, 1985; Thurstone, 1947). When testing a model using CFA, a priori hypotheses informed by theory dictate how each indicator is organized within that model (Jöreskog, 2007). It dictates which items will load onto or belong to a specific factor or latent variable. These hypotheses are then tested using observed data collected from participants.

CFA belongs to the factor analysis family, as does exploratory factor analysis (EFA). However, the two serve different purposes. EFA is a statistical method used to extract a factorial model from a dataset by determining the optimal number of factors to be extracted (Achim, 2017; Yong & Pearce, 2013). In EFA, items are organized into factors based on their intercorrelations, without considering where they might be expected to load. This allows researchers to determine how many factors will be necessary to accurately represent the correlations between indicators, as well as which indicators and factors should be linked (Jöreskog, 2007). In EFA, each item loads onto all factors, however, a given item is expected to load strongly (e.g., > .60) onto one factor and weakly (e.g., < .30) onto the others. These model specifications differ from those of a CFA, where hypotheses regarding which indicators load onto which latent factors are pre-determined based on theory (e.g., Rogers, 2024). In CFA, indicators are permitted to load onto their designated factor, while loadings of all other factors are fixed to zero (Byrne, 2005).

Fit Indices and Fit Index Cutoffs

When evaluating the fit of a CFA model, it is essential to use and report the χ^2 index (Jöreskog, 1969), a null-hypothesis significance test. The χ^2 index estimates the absolute fit of a model based on any incongruities between the implied covariance matrix (i.e., the model) and the observed covariance matrix (the data). Although the χ^2 index is the cornerstone of model fitting, it does have some drawbacks. For instance, it is quite sensitive to misspecification when sample sizes are increased (Marsh & Balla, 1994; Marsh et al., 1988). When it comes to CFA models, which require large sample sizes, the χ^2 value of a given model will almost certainly be statistically significant. However, a model not perfectly fitting the data does not necessarily indicate that it is worthless or should be discarded (Perry et al., 2015). A significant χ^2 simply indicates the presence of potential model misspecifications in need of scrutiny (e.g., Hayduk et al., 2007).

Researchers have developed a series of relative fit indices that complement the evaluation of CFA models (Bentler, 1990, 1995; Jöreskog, 1969; Steiger & Lind, 1980; Tucker & Lewis, 1973). The effect sizes that quantify the fit of a given model in relation to the data are referred to as fit indices and they belong to the "goodness" or "badness" of fit families (Marsh et al., 2005; Schermelleh-Engel et al., 2003). These indices provide researchers with an indication of the extent to which the data matches the hypothesized factorial model. The values of fit indices usually range from 0 to 1, although they can exceed 1 in certain circumstances (see West et al., 2012, 2023). Additionally, fit indices are interpreted subjectively, as there are no clearly defined values that would indicate a good or bad fit (McNeish & Wolf, 2023). Some of these indices, such as the Comparative Fit Index (CFI; Bentler, 1990) and the Tucker-Lewis Index (TLI; Tucker and Lewis, 1973), are part of the "goodness of fit" family. Obtaining higher values for these fit indices indicates a good fit between the model and data (West et al., 2012, 2023). Other indices, such as Root Mean Square Error of Approximation ($RMSEA$; Steiger and Lind, 1980) or the Standardized Root Mean Square Residual ($SRMR$; Bentler, 1995), are part of the "badness of fit" family. Obtaining lower values for these indices indicates a good fit between the model and data (West et al., 2012, 2023).

When evaluating the fit of CFA models, researchers often use fixed cutoffs that have been derived from previous significant methodological developments (Bentler & Bonett, 1980; Brown & Cudeck, 1993; Chen et al., 2008; Hu & Bentler, 1999; MacCallum et al., 1996). Therefore, some researchers have used and been more tolerant of values that could be seen as acceptable (e.g., Bentler & Bonett, 1980; Hu & Bentler, 1995; Jöreskog, 1993; Caron, 2018, 2023, $CFI \geq .90$, $TLI \geq .90$, $RMSEA \leq .08$, $SRMR \leq .08$) but nonetheless indicative of room for improvement (e.g., Bentler & Bonett, 1980; Bollen, 1989). A simulation study conducted by Hu and Bentler (1999) has become the golden rule upon which CFA models are being assessed. Based on this simulation, many researchers and authors of popular books (e.g., Brown, 2015; Byrne, 2013) have proposed that $SRMR$ values of .08 or lower, $RMSEA$ values of .06 or lower, and CFI or TLI values of .95 or higher could distinguish incorrect models from correct ones with a high rate of success. Since the publication of Hu and Bentler (1999), and despite recurrent calls for caution and nuances (e.g., Greiff & Heene, 2017; Hayduk et al., 2007; Marsh et al., 2004; Groskurth et al., 2023), the aforementioned values have become benchmarks stringently followed by researchers, reviewers, and editors when evaluating the fit of CFA models (Jackson et al., 2009).

Other studies have since shown that the criteria derived from the simulation studies of Hu and Bentler (1999) tend to generalize poorly to other types of models with a different number of factors, items, or characteristics (e.g., loadings, correlations between factors; McNeish, 2023; McNeish & Manapat, 2024; McNeish & Wolf, 2023, 2022; see McNeish, 2026, for a discussion of criticisms of these fixed cutoff values). Given the diversity of model characteristics encountered by empirical researchers, the rate at which this criterion correctly classifies models is unsatisfactory. This suggests that the derived cutoffs may be appropriate in instances where the structure of the model at hand matches that of the Hu and Bentler (1999) simulation. In fact, a recent simulation found that fit indices are highly influenced by the type of estimator used, the weight of factor loadings, the size of samples, and factor correlations (i.e., factors allowed to correlate or not; Groskurth et al., 2023). Most importantly, a lower proportion of unmodeled cross-loadings was associated with a worse fit when factors were uncorrelated, but a better fit when they were correlated. Groskurth et al. (2023) also found that when factors were uncorrelated, the Diagonally Weighted Least Squares (DWLS or WLSMV) estimator, a widely used estimator with ordered-categorical data, generally yielded the worst fit indices. However, when factors were correlated, the DWLS yielded most of the better fit indices. Groskurth et al. (2023) provide strong evidence indicating that using fixed index cutoffs (e.g., CFI and $TLI \geq .95$; $RMSEA \leq .06$; $SRMR \leq .08$) can lead researchers to make incomplete, biased, or erroneous decisions about their CFA models, indicating that the use of fixed cutoffs is not appropriate for all models.

Depending on the characteristics of a given model, the Hu and Bentler (1999) cutoffs are likely to either be too severe or too lenient (Groskurth et al., 2023). The cutoffs are likely to be too lenient for models with few items per factor (Marsh et al., 2004), high inter-item correlations, or low to null inter-factor correlations; which are more likely to generate fit index values with an almost perfect fit. If the cutoffs for a model are too lenient, they increase the odds of a misspecified model being incorrectly classified as correct (Type 1 error). Conversely, they may be too severe for more complex models with more diverse parameters or larger sample sizes; which can negatively impact the fit index values obtained (Kenny & McCoach, 2003; Marsh & Balla, 1994; Marsh et al., 1988; Shi et al., 2019). If the cutoffs for a model are too severe, they increase the odds of a model being incorrectly classified as misspecified when it is correct (Type 2 error). Therefore, for a fit index cutoff to inform any conclusions regarding the fit of CFA models, they need to be tailored to the specificities of each model. The use of such tailored fit indices helps ensure that models are correctly classified (i.e., correct models classified as correct and misspecified models classified as misspecified) as often as possible.

The Dynamic Fit Index Approach

Recognizing the limitations of using fixed benchmarks to evaluate model fit, some researchers have proposed using simulations to generate fit index cutoffs specific to each model being tested (e.g., Pornprasertmanit et al., 2012). However, McNeish and Wolf (2023) noted that many researchers lack the training necessary to properly conduct simulation studies. Additionally, it is not possible to conduct Monte Carlo simulations with all statistical software. McNeish and Wolf (2023) also pointed out that simulation studies tend to be time-consuming, which makes using the cutoffs derived from Hu and Bentler (1999) tempting. Not to mention, selecting a meaningful misspecification as a basis for conducting a simulation is no easy task (McNeish & Wolf, 2023). Therefore, for the reasons mentioned above, there is a need to develop a tool that makes simulating fit index cutoffs easy, automatic, and accessible, in order to reduce the burden that is currently placed on researchers.

The Dynamic Fit Index (DFI) approach (McNeish & Wolf, 2023; Wolf & McNeish, 2023) was developed to generate fit index cutoffs that properly account for misspecification in a model, based on that model's specific characteristics. The DFI approach is similar to that of Hu and Bentler (1999), who also generated cutoffs that could differentiate between correct and misspeficied models. However, there are some key differences between the two approaches. Hu and Bentler (1999) simulated cutoffs for correlated three-factor models that have five items per factor, sample sizes ranging from 150 to 5000, item loadings of .70, .75 or .80, and that had omitted cross-loadings. For example, they found that a CFI of .95 and a $RMSEA$ below .06 could generally differentiate between the proposed and alternative models with few specification errors. Notably, Hu and Bentler (1999) highly recommended the use of at least two different fit indices to evaluate the fit of a model.

The DFI approach also simulates cutoffs based on a given model, however, unlike the Hu and Bentler (1999) approach, it generates cutoffs for all kinds of models. The DFI approach accounts for the particularities of a given model (i.e., sample sizes, factor loadings, number of items, internal reliability of factors, degrees of freedom, etc.; McNeish & Wolf, 2023). For example, dynamic may find that a CFI value of .92 is sufficient to differentiate between the proposed and alternative models with minor misspecifications. To achieve this, McNeish and Wolf (2023) developed an `R` package (Wolf & McNeish, 2023) and a Shiny App (Wolf & McNeish, 2021) that can simulate DFIs without having to write a single line of code. Both the `R` package and the app, which implement a simulation algorithm, are based on `lavaan` (Rosseel, 2012) and `simsem` (Pornprasertmanit et al., 2012).

The algorithm used to calculate DFIs follows a series of steps (see McNeish & Wolf, 2023; McNeish & Manapat, 2024). Researchers start by running their CFA model with the software of their choice. Estimates of the model (e.g., factor loadings, inter-factor correlations) are the foundation that is used to run the subsequent DFIs. A data generation model is then created by adding a misspecified path to the tested model. To add this misspecified path, the DFI algorithm (McNeish & Wolf, 2023; McNeish & Manapat, 2024) replicated the "minor misspecification condition" used by Hu and Bentler (1999). The added misspecification is a standardized cross-loading of the item with the smallest target loading onto the factor with the highest construct reliability. The value of this cross-loading is identical to that of the lowest target factor loading. For example, assuming that in a questionnaire, Factor 1 has the highest construct reliability, Factor 3 has the lowest target loading (.450), and item 6 loads onto Factor 3. The misspecified path added at step 2 would cross-load item 6 onto Factor 1 with a value of .450. In an instance where the standardized factor loadings are high, the addition of a high magnitude cross-loading could result in a negative residual population variance. To avoid this issue, the dynamic algorithm determines 95% of the highest cross-loadings that could be added to the model without the residual population variance becoming negative (McNeish & Wolf, 2023).

Then, the tested model is individually fitted to 500 datasets that have been generated by the data generation model. This allows fit index values with minor misspecifications to be generated for the model at hand. By analyzing the distribution of said fit indices across these datasets, we can identify a value that correctly classifies at least 95% of misspecified models as misspecified. The same steps are then repeated, but the tested model is now used as the data generation model. Our goal here is to find a value that correctly classifies 95% of correct models as correct. The values obtained at this stage should be "closer to exact fit" (McNeish & Wolf, 2023), as the tested model is now the correct model. However, if it is not, we risk rejecting too many correctly specified models. [1] Once these steps are completed, DFI can determine, for example, that a CFI of .98 is required to determine if your model is preferable compared to a model with a small misspecification (i.e., a small non-target loading).

Another feature that differentiates this approach from the Hu and Bentler (1999) approach is the levels of misspecification that the DFI algorithm generates (McNeish & Wolf, 2023). The number of levels of misspecification in a model corresponds to the number of factors within that model, minus one. The use of these levels provides researchers with an indication as to the extent of misfit in their model by going beyond simply labeling a model as "correct" or "misspecified". If your model contains six factors, the proposed CFI cutoff corresponds to the CFI needed to differentiate your model from an alternative model that contains one (Level 1), two (Level 2), three (Level 3), four (Level 4) and five (Level 5) misspecifications. In cases where a model contains three factors, there would be two levels of misspecification (i.e., Level 1 and Level 2). The first level of misspecification would be a standardized cross-loading by the item with the smallest target loading to the factor with the highest level of construct reliability (McNeish & Wolf, 2023). The value of this cross-loading is the same as the target loading factor of this item, requiring that the population residual variance does not become negative. As for the second level, it would add a second misspecification which is a standardized cross-loading by the item with the second smallest target loading and the other factor with the highest level of construct reliability, so long as that factor is not the same as the one chosen for the level 1 misspecification (McNeish & Wolf, 2023). The value of this second cross-loading should be the same as the second-smallest target loading, requiring that the population residual variance does not become negative (McNeish & Wolf, 2023).

When using these levels of misspecification, one can compare the fit indices of a CFA model to the dynamic fit index (DFI) cutoffs generated by the DFI algorithm (McNeish & Wolf, 2023). If the cutoffs were generated for a three-factor model, one would evaluate whether model fit is equal to or smaller than a cross-loading of the level 1 value, whether the misspecification is between a cross-loading of level 1 value and two cross-loadings of level 2 value, or equal to or worse than two cross-loadings of level 2 value. This procedure allows for more nuanced fit evaluation.

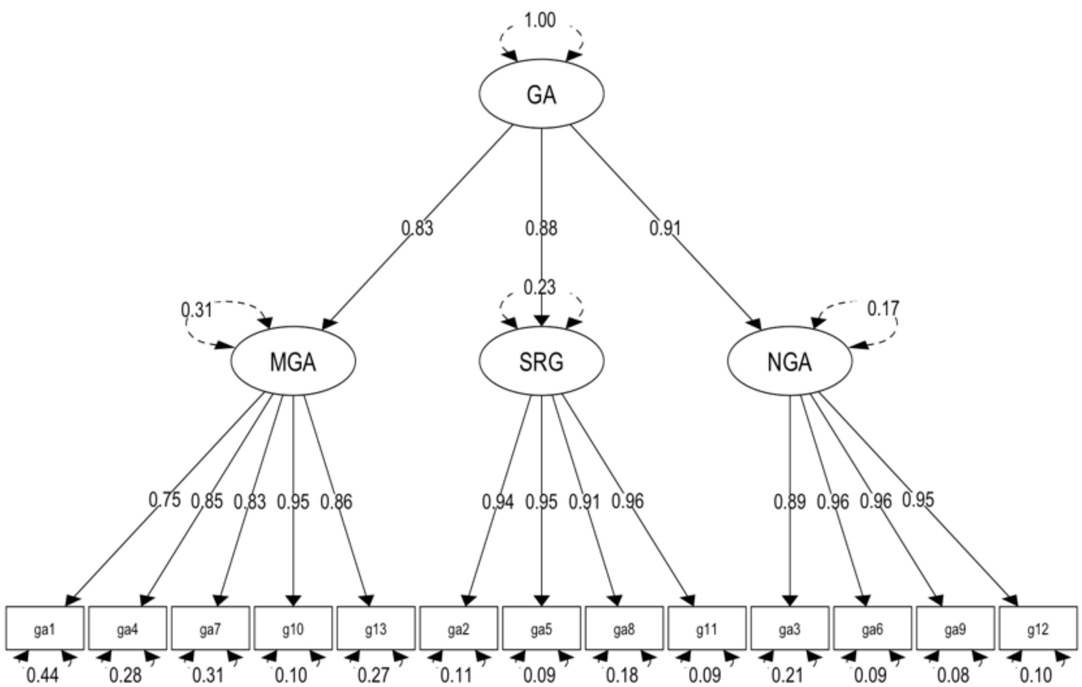
Objectives

In the following tutorial, there will be a discussion on how to conduct a CFA and how to generate DFI cutoffs in R (R Core Team, 2025) using the lavaan (Rosseel, 2012), and the dynamic (Wolf & McNeish, 2023) packages. R (R Core Team, 2025) was selected for this tutorial because it is freely available and widely accessible, making it an ideal tool for both students and researchers.

In our tutorial, we will evaluate the factor structure of the Attainment of School Achievement Goal Scale (A-SAGS; adapted from Amiot et al., 2004; Gaudreau & Blondin, 2004). We selected this measure for our tutorial because it contains three factors, each measured using 4 or 5 items, which is a similar factor structure to many other questionnaires in psychological sciences. The A-SAGS is built on the idea that people use three distinct but interrelated competence criteria when subjectively assessing their performance (e.g., Elliot & McGregor, 2001). First, they assess the extent to which they have mastered the absolute demands of the task (i.e., mastery attainment). Second, they evaluate the extent to which they performed better than their previous or habitual performances (i.e., self-referenced attainment). Third, they rate the extent to which they performed better than others (i.e., normative attainment). These three criteria can be organized in a hierarchical model with three first-order factors nested under a second-order dimension (see Figure 1).

[1] If the generated values are not closer to exact fit, it may still be possible to find fit index values that can classify at least 90% of misspecified models as misspecified. In some cases, no such values can be generated, meaning it is not possible to distinguish between correct and misspecified models due to the characteristics of the data.

Figure 1 ■ Factorial structure of the A-SAGS illustrated using `semPlot`. MGA = Mastery goal attainment; SRG = Self-referenced goal attainment; NGA = Normative goal attainment; GA = Goal attainment



Method

Participants

The data used in this tutorial have been analyzed and published in previous studies (Gaudreau, 2015; Kljajic et al., 2017). The sample comprised 527 undergraduate university students with a mean age of 19.16 years old (S.D. = 3.29). The majority of the sample being female (72.2%; Male = 27.8%). Most were first-year students (75.1%), 16.2% were second-year students, 6.3% were third-year students, and 2.5% were fourth-year students. Most of them were full-time students (97.9%), while 2.1% were part-time students. In terms of ethnic background, 60.3% are Caucasian, 12.1% are Asian, 6.8% are Arabic, 6.3% are African American, 0.9% are Hispanic-Latino, 0.9% are Aboriginal or Native, and 12.5% chose to identify as other (e.g., African, East Asian, Middle Eastern, South Asian, Mixed).

Measures

The Attainment of School Achievement Goal Scale. The Attainment of School Achievement Goal Scale (A-SAGS; adapted from Amiot et al., 2004; Gaudreau & Blondin, 2004), a 13-item scale, was used to measure student's perceptions of their own academic performance. [2] The students are instructed to answer the scale while thinking about the extent to which each item represents their performance in the previous academic semester on a scale from 1 (not at all) to 7 (totally). The A-SAGS contains three subscales, which are mastery-goal attainment (e.g., "I provided a quality effort"), self-referenced goal attainment (e.g., "I did better than my usual performances"), and normative goal attainment (e.g., "I showed that I am superior to other students"). Results of a hierarchical CFA on the sport version of the A-SAGS supported the proposed factor structure (Martinent et al., 2013). In this study, the internal reliability (McDonald's omega) of each subscale was excellent: mastery goal attainment ($\omega = .906$), self-referenced goal attainment ($\omega = .955$), normative goal attainment ($\omega = .954$).

Tutorial

The R code and the data which is described in this tutorial can be found in the Appendix or on OSF (https://osf.

[2] The A-SAGS is a 12-item scale (adapted from Amiot et al., 2004; Gaudreau & Blondin, 2004). However, there is an additional mastery goal-attainment item in case one of the original items does not behave as expected. We have decided to include this item (item 13) as an indicator of mastery goal-attainment in this tutorial to test for its properties.

io/8yp5m). The R output, generated using the rmarkdown package (Allaire et al., 2014), can be found under supplementary materials.

Step 1: Load R packages. As part of this tutorial, three R packages are used: `lavaan` (Rosseel, 2012), `dynamic` (Wolf & McNeish, 2023), and `semPlot` (Epskamp, 2013). First, each package needs to be downloaded using the `install.packages("name of package")` function. As for the `dynamic` package, it can be downloaded using the `install.github()` function from `devtools` (Wickham et al., 2011) and by providing the following argument to it `"melissagwolf/dynamic"`. Then, it is necessary to load each of them by using the `library(name of package)` function.

Step 2: Load the dataset. To conduct the analyses, the dataset, which will be referred to as `my_data`, needs to be loaded. As the dataset is in .csv format, the `read.csv()` function is used with the dataset's name in parentheses: `my_data <- read.csv("dataset.csv")`. It is important to note that it is possible to load datasets of other formats in R (R Core Team, 2025). There are different commands for each format. For more information about these commands, readers should consult Chapter 1 of Horton and Kleinman (2015) and Chapter 6 of Caron (2023).

Step 3: Quick look at the data before conducting the analyses. Consistent with Dion et al. (2021), the `summary(my_data)` function is used to have a quick look at the data (`my_data`) in R. This allows to look at the descriptive statistics for each item.

Step 4: CFA command. When conducting a CFA using lavaan, it is important to code which items load onto which factor. Each factor can be referred to by assigning it a number (e.g., 1, 2, or 3). Then, all items need to be referred to by their name in the dataset. For example, item 1 of the questionnaire is referred to as `ga1`. In addition, each item that loads onto a factor needs to be separated by a + (e.g., `Factor1 = ga1 + ga4 + ga7 + ga10 + ga13`). The `=` symbol needs to appear after the name of each factor. It means that the factor is made of the following items. If you want to set an equality constraint on a parameter such as a pair of factor loadings, you must add any label (e.g., `Equal`) and the symbol `*` before these parameters (e.g., `Equal*ga10`).

Step 5: Generate CFA results. To generate the CFA results, two lines of code are needed. One is adapted from Rogers (2024) and the other is adapted from Dion et al. (2021). First, the `fit <- cfa()` function is used to run the CFA (Rogers, 2024). In parentheses, the following was coded `(CFA, data = my_data, ordered = TRUE, std.lv = TRUE, estimator = "WLSMV")`." The `CFA` argument refers to the CFA model generated at Step 5, while `my_data` refers to the dataset. The `ordered = TRUE` argument tells `lavaan` that the data is ordered. As for the `std.lv = TRUE` argument, it generates standardized paths from the latent variables to each item. Finally, the `estimator = WLSMV` argument refers to the choice of estimator to handle missing data. When using Likert-type items, the mean- and variance-adjusted weighted least squares (WLSMV) is usually the better estimation method (see Brauer et al., 2023).

As for the second line of code, the "summary()" function is used to get summary results for the tested CFA model (Dion et al., 2021). The following was added in parentheses `(fit, fit.measures = TRUE, standardized = TRUE, rsquare = TRUE)`. The `fit` argument refers to the name of the CFA model. When using `fit.measures = TRUE`, `lavaan` generates fit measures for the CFA model. The `standardized = TRUE` argument generates standardized estimates. As for the `rsquare = TRUE` argument, it generates explained variance (R^2) values for each item. This gives an idea of the amount of variance of each item that is explained by their respective factors.

Step 6: Modification indices. To identify potential sources of model misspecification within this CFA model, one can look at modification indices. To do so, the `modindices()` function is used with the following in parentheses `(fit, sort = TRUE, maximum.number = 30)`. As for the `sort = TRUE` argument, it sorts modifications in order of the ones that would result in the largest improvement in fit. For example, to display the top 10 model misspecifications, `maximum.number = 10` is used.

Step 7: Comparing the CFA model to a plausible alternative. When estimating the fit of a CFA model, it is important to compare it to the fit of a plausible alternative model. To compare the fit of both models, we used the `anova()` function in `lavaan` with the following in parentheses `(fit1, fit2)`. As for `fit1`, it refers to our baseline CFA model, while `fit2` refers to our plausible alternative. By doing so, we obtain the results of a χ^2 difference test that compares the fit of both models.

Step 8: Command to calculate DFIs using dynamic. To generate DFI cutoffs for the model, two lines of code from Wolf and McNeish (2021), which are found on their *Dynamic fit app*, were adapted. As for the first line of code, an object called `DFIs` was created, which refers to the DFI cutoffs generated by the `dynamic` package. It was followed by the `hier1HB()` function and `(fit, reps = 500)`. The first part means that DFI cutoffs will be generated for the hierarchical CFA model. The cutoffs generated follow the same logic as Hu and Bentler (1999), referred to as `HB`. In parentheses, the `fit` object appears, which refers to the CFA model. As for `reps=500`, it means that `dynamic` will

Table 1 ■ Target loadings for the 13-item A-SAGS

	Mastery goal attainment	Self-referenced goal attainment	Normative goal attainment	Second-order factor
Mastery goal attainment	-	-	-	.833
Item 1	.748	-	-	-
Item 4	.848	-	-	-
Item 7	.833	-	-	-
Item 10	.950	-	-	-
Item 13	.856	-	-	-
Self-referenced goal attainment	-	-	-	.880
Item 2	-	.942	-	-
Item 5	-	.953	-	-
Item 8	-	.906	-	-
Item 11	-	.956	-	-
Normative goal attainment	-	-	-	.912
Item 3	-	-	.890	-
Item 6	-	-	.956	-
Item 9	-	-	.962	-

conduct 500 replications to generate the fit index cutoffs. `DFIs` was also entered, which allows one to list the DFIs generated for the model.

Step 9: Command to plot the CFA model using semPlot. It is possible to use `semPlot` (Epskamp, 2013) to plot the CFA model and its estimates. A line of code which is found on the `dynamic` fit app (Wolf & McNeish, 2021) was adapted. First, the name `ASAGSPlot` was assigned to the plot. The `semPaths()` function is used with the following in parentheses `(fit, thresholds = FALSE, intercepts = FALSE, whatLabels = "std", edge.label.cex = 1, edge.color = "black")`. The plot is created based on the CFA model referred to as `fit`. Thresholds (`Thresholds = FALSE`) and intercepts (`Intercepts = FALSE`) are not displayed to avoid displaying too much information. The last command (i.e., `whatLabels = "std"`) displays the standardized paths from the latent variables to each item in the plot. The `edge.label.cex` command helps choosing the size of the standardized estimates displayed in the model (i.e., `1`). Finally, the `edge.color` argument enables one to choose the color for each path in the model (i.e., `black`).

Results

Evaluating the Model Based on Fixed Cutoffs

Results of the hierarchical CFA with three first-order factors showed that the χ^2 test was statistically significant: $\chi^2(62) = 461.359, p < .001$. This implies that the model does not perfectly fit the data. The relative fit indices obtained for the model were as follows: $CFI = .975, TLI = .968, RMSEA = .171$, and $SRMR = .049$. By stringently following the cutoffs derived from Hu and Bentler (1999); [CFI and $TLI \geq .95$, $RMSEA \leq .06$, $SRMR \leq .08$], one would conclude that the fit of the model is mixed. One would reach the same conclusions using more liberal cutoffs (Bentler & Bonett, 1980; Bollen, 1989; Hu & Bentler, 1995; Jöreskog, 1993; , see also Caron, 2018, 2023, CFI and $TLI \geq .90$, $RMSEA \leq .08$, $SRMR \leq .08$). All fit indices were acceptable, but the $RMSEA$ largely exceeded recommended guidelines. However, we know that in models with high target loadings, few items per factor, and small to moderate degrees of freedom, $RMSEA$ values tend to falsely indicate a poor fitting model (Kenny et al., 2015; McNeish et al., 2018). In other words, based on the known limitations of the $RMSEA$, we can conclude that the proposed hierarchical model of the A-SAGS provides an acceptable fit when evaluating model fit with fixed cutoffs. Table 1 shows that each item loaded strongly onto their hypothesized factor. Mastery goal attainment loadings ranged from .748 to .950. Self-referenced goal attainment loadings ranged from .906 to .953. As for normative goal attainment, loadings ranged from .890 to .962. Secord-order factor loadings ranged from .833 to .912, so each of the three criteria contributed strongly to an overall subjective assessment of goal attainment.

One way to verify the fit of a factorial model is to compare it to a theoretically plausible alternative. Achievement Goal Theory (e.g., Elliot & McGregor, 2001; Nicholls, 1984)

Table 2 ■ Target loadings for the 2-factor A-SAGS

	Mastery and self-referenced goal attainment	Normative goal attainment
Mastery and self-referenced goal attainment	-	-
Item 1	.702	-
Item 2	.925	-
Item 4	.798	-
Item 5	.941	-
Item 7	.792	-
Item 8	.886	-
Item 10	.868	-
Item 11	.937	-
Item 13	.810	-
Normative goal attainment	-	-
Item 3	-	.890
Item 6	-	.955
Item 9	-	.962
Item 12	-	.951

has often subsumed mastery- and self-referenced goals under mastery goals. This means that a reasonable and theoretically plausible alternative would be for the items of factors 1 and 2 of the A-SAGS to load on one factor. The χ^2 test of this alternative model yielded a statistically significant result, $\chi^2(64) = 1395.253$, $p < .001$. We obtained the following fit indices: $CFI = .953$, $TLI = .943$, $RMSEA = .230$, $SRMR = .093$. Using the benchmarks derived from Hu and Bentler (1999), we observed that only the CFI value exceeded the cutoffs. Using more liberal cutoffs (Bentler & Bonett, 1980; Bollen, 1989; Hu & Bentler, 1995; Jöreskog, 1993; ; see also Caron, 2018, 2023), only the CFI and TLI values respected the cutoffs. Factor loadings are displayed in Table 2. We then decided to test whether the fit of this model was significantly different from our hierarchical model with three first-order factors. We obtained the following results: $\Delta\chi^2 = 260.46$, $\Delta\text{df} = 2$, $p < .001$. This means that our three-factor hierarchical model fitted significantly better than a two-factor model where mastery- and self-referenced goal attainment loaded onto a single factor. Thus, we retained our theoretically expected three-factor model.

Evaluating the Dynamic Fit Index Cutoffs

DFI cutoffs for the CFA model of the A-SAGS were generated using dynamic. First, a CFI value of .987, a $RMSEA$ value of .062, and a $SRMR$ value of .025 would indicate that misspecification in the model is equal to 95% of the largest allowable cross-loading (i.e., .290) that keeps the population residual variance positive (Level 1 misspecification). Second, a CFI value of .983, a $RMSEA$ value of .075, and a $SRMR$ value of .026 would indicate that misspecification in the model is equal to 95% of the two largest allowable cross loadings (i.e., .290 and .126) that keep the population residual variance positive (Level 2 misspecification). By looking at the observed fit indices of the A-SAGS ($CFI = .975$; $TLI = .968$; $RMSEA = .171$; and $SRMR = .049$), it can be deduced that none of the fit index values for our CFA model exceeded the levels of misspecification generated by dynamic. This indicates that misspecification in the model is equivalent to more than 95% of the two largest allowable cross loadings in the model (i.e., .290 and .126). Therefore, support for the factorial structure of the A-SAGS is unsatisfactory based on DFIs. Before using the A-SAGS in subsequent research, it would be important to investigate the potential sources of model misfit (i.e., small cross-loadings, small correlations between item residuals) to determine their impact on the overall fit and theoretical interpretation of the model.

Exploration of the Potential Sources of Model Misfit

Whenever the χ^2 of the CFA model is significant, it is generally advisable to explore the potential sources of misfit. However, this recommendation has rarely been widely implemented (e.g., Hayduk et al., 2007). We forward a similar cautionary note in recommending that modification indices (or alternative methods; Muthén & Asparouhov, 2012) be explored whenever a model fails to reach the model-tailored cutoffs from the DFI.

In the hierarchical CFA model with three first-order factors, the MIs indicated that item 10 of mastery goal attainment could cross-load on the other two first-order factors.

Table 3 ■ Target loadings for the A-SAGS with item 10 loading onto all three first-order factors

	Mastery goal attainment	Self-referenced goal attainment	Normative goal attainment	Second-order factor
Mastery goal attainment	-	-	-	.764
Item 1	.772	-	-	-
Item 4	.874	-	-	-
Item 7	.857	-	-	-
Item 10	.513	.112	.342	-
Item 13	.908	-	-	-
Self-referenced goal attainment	-	-	-	.894
Item 2	-	.942	-	-
Item 5	-	.954	-	-
Item 8	-	.906	-	-
Item 11	-	.956	-	-
Normative goal attainment	-	-	-	.898
Item 3	-	-	.891	-
Item 6	-	-	.956	-
Item 9	-	-	.960	-
Item 12	-	-	.951	-

This MI was the main contributor to misfit in our model. Having smaller cross-loadings compared to the target loading could be considered tolerable and trivial. In contrast, having cross-loadings as strong as the target loading could suggest the need to revise this item in subsequent versions of the A-SAGS. In order to estimate the value of the two cross-loadings, we tested a CFA model where item 10 was set to load on all three first-order factors. As shown in Table 3, the target loading of item 10 on mastery goal attainment was high (.513). It was higher than the non-target loadings on self-referenced goal attainment (.112) and normative goal attainment (.342) which were all statistically significant. The cross-loading on normative goal attainment can be considered substantial and sufficiently high to warrant further considerations (see discussion section).[3]

Discussion

Despite well-documented limitations and repeated calls for caution, researchers often rely on fixed cutoff values derived from Hu and Bentler (1999) to assess CFA model fit. In this tutorial, we aimed to address this limitation by leveraging a recent solution developed by McNeish and Wolf (2023), which illustrated how to generate cutoff values that are tailored to the specific characteristics of a given CFA model. Specifically, we utilized this approach to examine the factorial structure of a questionnaire (A-SAGS) designed to help students subjectively assess their school performance. We demonstrated how to conduct a CFA using the lavaan package (Rosseel, 2012) and how to obtain model-specific cutoff values with the dynamic package (Wolf & McNeish, 2023). Overall, our tutorial provided a practical framework for evaluating model fit based on recent methodological developments and updated recommendations.

We examined the factorial structure of a questionnaire designed to provide a subjective assessment of students' performance in school. Based on the thresholds derived from Hu and Bentler (1999) [CFI and $TLI \geq .95$, $RMSEA \leq .06$, $SRMR \leq .08$)], we would have concluded that the fit for A-SAGS was satisfactory ($CFI = .975$; $TLI = .968$; $RMSEA = .171$; $SRMR = .049$). However, when using cutoffs generated by dynamic at both levels (Level 1: $CFI = .987$, $RMSEA = .062$, $SRMR = .025$; Level 2: $CFI = .983$, $RMSEA = .075$, $SRMR = .026$), none of our fit indices met the model-specific cutoffs. This result indicated that the level of misspecification in the model was superior to 95% of the two largest allowable cross-loadings (i.e., .290 and .126) in the data. Using the DFI offered a clearer view of the extent to which the hypothesized factor structure of the A-SAGS may have been misspecified, illustrating the added value of evaluating model fit with model-specific rather than universal cutoffs.

[3] We also tested whether the non-target loading of item 10 on self-referenced goal-attainment (.112) was significantly lower than the one on normative goal attainment (.342) and whether the non-target loading on normative goal attainment was equivalent to the target loading on mastery goal attainment (.513). We did so by testing two models with an equality constraint on these parameters and compared them to the model where all loadings for item 10 are freely estimated. We obtained support for our first hypothesis only (see Supplementary materials).

What to do when DFI cutoffs are not met

The DFI cutoffs generated for our CFA of the A-SAGS may *appear* stricter. Yet, they are simply *more realistic* because they are calibrated based on the reality of our model. Researchers may wonder what their next steps should be if their model does not meet DFI cutoffs. Our tutorial relied on best practices in SEM. After rejecting the hypothesized model, we explored the modification indices (MI) to identify potential sources of model misspecification. MIs can help to locate the parameters that would lead to the greatest improvement in model fit if they were included in the model (Xiong et al., 2025). Prudence is warranted when using MIs because a model misspecification may be due to sampling variability and may not generalize to other samples (MacCallum et al., 1992). Thus, we encourage looking at modification indices only for the sake of taking note of the problems found in your data. We then suggest estimating those problematic parameters to get a sense of the extent of the problem in your data. In other words, we suggest not using modification indices to blindly improve your CFA model for the purpose of reaching a threshold of acceptable model fit.

The MIs for the A-SAGS indicated that the improper model fit is potentially due to one item that could potentially load onto all three factors. We tested a new model to assess and interpret these potential cross-loadings. We then concluded that one of these cross-loadings could be considered trivial and negligible, but the other seems substantial enough to warrant future research attention.[4] Overall, we concluded that this item violated the principle of simple factor structure, which states that each item should only load onto one factor (e.g., Thurstone, 1947). A popular solution would be to remove this item and run a new CFA without it. However, removing an item based on the modification indices of a single sample is questionable (Flake & Fried, 2020; MacCallum et al., 1992) and was therefore not attempted in this tutorial. In addition, it is important to consider the theory that is relevant to our measure before coming to any conclusions regarding model appropriateness (e.g., Bollen, 1989; West et al., 2012, 2023). Said theory should inform all decisions made at any stages of the questionnaire development, including when testing its factorial structure.

If it has been suggested that a model may not meet the DFI cutoffs, a Bayesian CFA is proposed as a better alternative to looking at the modification indices provided in the output of statistical packages (Muthén & Asparouhov, 2012; , see also Depaoli, 2021). Through the use of informed priors, a Bayesian CFA can allow cross-loadings and correlated residuals (correlations between the errors of items) to vary around a pre-specified interval. The posterior distribution of these parameters can determine whether misfit is due to the presence of many small, trivial, and tolerable cross-loadings or to larger, more problematic ones. The same logic is applicable when examining correlated residuals (Asparouhov et al., 2015; Muthén & Asparouhov, 2012). The Bayesian CFA framework offers a way forward, allowing for the exploration of sources of misspecification when rejecting a model based on traditional CFA constraining the cross-loadings and correlated residuals to zero (see Taylor, 2019, for an example).

Limitations and Future research

The CFA model (using traditional or DFI cutoff values) works under the assumption that all cross-loadings are fixed at zero. This assumption can be acceptable for some conceptual models and too strict for others. Meaning that the DFI approach may erroneously fail several CFA models, because it works under the assumption that non-freed cross-loadings are zero (e.g., Thurstone, 1947). Psychometric indicators are most often not entirely pure construct indicators. Assuming that cross-loadings are all zero renders the CFA model too limiting because it is nearly impossible for all the cross-loadings of a factor model to be null (Asparouhov et al., 2015; Muthén & Asparouhov, 2012; Marsh et al., 2014).

As Marsh et al. (2014) pointed out, questionnaire items are imperfect indicators of the constructs they are intended to measure. Thus, they often display small residual correlations with other variables. They demonstrated that omitting non-null cross-loadings can inflate inter-factor correlations and compromise the discriminant validity and predictive power of the scores. Alternative approaches, such as exploratory structural equation modeling (e.g., Asparouhov & Muthén, 2009; Marsh et al., 2014) and Bayesian structural equation models (e.g., Muthén & Asparouhov, 2012) have been deemed promising as a way to relax the overly stringent assumptions of the traditional CFA model. We believe future research should investigate whether these other frameworks can provide additional insight alongside DFIs. However, we do not argue that these frame-

[4] A one-size fits all approach should not be considered when evaluating target and non-target loadings (Knekta et al., 2019). However, based on a review of current practices, Howard (2016) recommended that each item should "(a) load onto their target factor above 0.40, (b) load onto alternative factors below 0.30, and (c) demonstrate a difference of 0.20 between their target and alternative factor loadings" (p. 55). Using this .40, .30, .20 rule, we concluded that one of the cross-loadings for item 10 was sufficiently small ($\lambda < .20$) to be considered trivial. The other cross-loading exceeded the .30 rule ($\lambda < .34$) and did not demonstrate a difference of at least .20 with the target factor loading (target $\lambda = .51$). We suggest that future research reevaluates whether this cross-loading replicates across samples before determining if this item should be rewritten, dropped, or tolerated as a complex indicator of two factors (see Supplemental materials for an explanation as to why this item wants to cross-load on factor 3).

works should substitute for them.

Conclusion

The fit of confirmatory factor analysis models has historically been evaluated using fixed cutoffs, which are known to generalize poorly to models with different characteristics. In this tutorial, we tested the DFI approach by McNeish and Wolf (2023) which generates cutoff values tailored to the characteristics of the model being tested. In this tutorial, we illustrate how to implement this approach by testing the factorial validity of the A-SAGS. Though fixed cutoffs would have supported the fit of the A-SAGS in our data, dynamic fit indices did not provide support for it. We then provide guidance on what to do if a CFA model does not meet the DFI cutoffs. We discuss how modification indices can give a preliminary idea of sources of misfit in our model, and why a Bayesian CFA can help one evaluate whether misfit is due to the sum of negligible problems or a few problematic ones. We also discuss the importance of theory and making informed decisions when evaluating a factorial model, as well as the overly optimistic assumptions of near zero cross-loadings or correlated residuals in CFA. The DFI is a strong alternative to fixed cutoffs, and we encourage researchers to try it out and implement it when evaluating CFA models.

Authors' note

This research was conducted while the fifth author was supported by a research grant from the Social Sciences and Humanities Research Council of Canada (SSHRC; 435-2022-0145) and a teaching release awarded by the Faculty of Social Sciences from the University of Ottawa.

References

Achim, A. (2017). Testing the number of required dimensions in exploratory factor analysis. *The Quantitative Methods for Psychology*, *13*(1), 64–74. doi: 10.20982/tqmp.13.1.p064.

Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2014). *Rmarkdown: Dynamic documents for R (version 2.30) [R package]*. doi: 10.32614/CRAN.package.rmarkdown.

Amiot, C. E., Gaudreau, P., & Blanchard, C. M. (2004). Self-determination, coping, and goal attainment in sport. *Journal of Sport & Exercise Psychology*, *26*(3), 396–411. doi: 10.1123/jsep.26.3.396.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. doi: 10.1080/10705510903008204.

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on stromeyer et al. *Journal of Management*, *41*(6), 1561–1577. doi: 10.1177/0149206315591075.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. doi: 10.1037/0033-2909.107.2.238.

Bentler, P. M. (1995). *Eqs structural equations program manual*. Multivariate software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. doi: 10.1037/0033-2909.88.3.588.

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley. doi: 10.1002/9781118619179.

Brauer, K., Ranger, J., & Ziegler, M. (2023). Confirmatory factor analyses in psychological test adaptation and development: A nontechnical discussion of the wlsmv estimator. *Psychological Test Adaptation and Development*, *4*(1), 4–12. doi: 10.1027/2698-1866/a000034.

Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 445–455). Sage Publications.

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

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361–379). The Guilford Press.

Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment*, *85*(1), 17–32. doi: 10.1207/s15327752jpa8501_02.

Byrne, B. M. (2013). *Structural equation modeling with mplus: Basic concepts, applications, and programming*. Routledge. doi: 10.4324/9780203807644.

Caron, P.-O. (2018). *La modélisation par équations structurelles avec mplus*. Presses de l'Université du Québec. doi: 10.2307/j.ctvt1sh9g.

Caron, P.-O. (2023). *Méthodes quantitatives avec R*. TÉLUQ. https://mqr.teluq.ca

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462–494. doi: 10.1177/0049124108314720.

Depaoli, S. (2021). *Bayesian structural equation modeling*. The Guilford Press.

Dion, K., Bodnaruc, A. M., Trudel, G., Lamarche, J., Ranger, V., Fobert, S., Church, K. A., Ntumba Mukunzi, J., & René, J.-L. (2021). La modélisation par équations structurelles—un guide d'accompagnement pour l'interface R. *The Quantitative Methods for Psychology*, *17*(3), 198–271. doi: 10.20982/tqmp.17.3.p198.

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501–519. doi: 10.1037/0022-3514.80.3.501.

Epskamp, S. (2013). *semPlot: Path diagrams and visual analysis of various SEM packages' output (version 1.1.7) [R package]*. doi: 10.32614/CRAN.package.semPlot.

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. doi: 10.1177/2515245920952393.

Gaudreau, P. (2015). Self-assessment of the four subtypes of perfectionism in the 2 x 2 model of perfectionism. *Personality and Individual Differences*, *84*, 52–62. doi: 10.1016/j.paid.2014.10.039.

Gaudreau, P., & Blondin, J.-P. (2004). Differential associations of dispositional optimism and pessimism with coping, goal attainment, and emotional adjustment during sport competition. *International Journal of Stress Management*, *11*(3), 245–269. doi: 10.1037/1072-5245.11.3.245.

Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about model fit. *European Journal of Psychological Assessment*, *33*(5), 313–317. doi: 10.1027/1015-5759/a000450.

Groskurth, K., Bluemke, M., & Lechner, C. M. (2023). Why we need to abandon fixed cutoffs for goodness-of-fit indices: An extensive simulation and possible solutions. *Behavior Research Methods*, *56*(4), 3891–3914. doi: 10.3758/s13428-023-02193-3.

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three – testing the theory in structural equation models! *Personality and Individual Differences*, *42*(5), 841–850. doi: 10.1016/j.paid.2006.10.001.

Horton, N. J., & Kleinman, K. (2015). *Using r and rstudio for data management, statistical analysis, and graphics* (2nd ed.). CRC Press. doi: 10.1201/b18151.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, *32*(1), 51–62. doi: 10.1080/10447318.2015.1087664.

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Sage Publications.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. doi: 10.1080/10705519909540118.

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6–23. doi: 10.1037/a0014694.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183–202. doi: 10.1007/BF02289343.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Sage Publications.

Jöreskog, K. G. (2007). Factor analysis and its extensions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 47–77). Lawrence Erlbaum Associates Publishers.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*(3), 486–507. doi: 10.1177/0049124114543236.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(3), 333–351. doi: 10.1207/S15328007SEM1003_1.

Kljajic, K., Gaudreau, P., & Franche, V. (2017). An investigation of the 2 × 2 model of perfectionism with burnout, engagement, self-regulation, and academic achievement. *Learning and Individual Differences*, *57*, 103–113. doi: 10.1016/j.lindif.2017.06.004.

Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, *18*(1), 1–17. doi: 10.1187/cbe.18-04-0064.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. doi: 10.1037/1082-989X.1.2.130.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*(3), 490–504. doi: 10.1037/0033-2909.111.3.490.

Marsh, H. W., & Balla, J. (1994). Goodness of fit in confirmatory factor analysis: The effects of sample size and model parsimony. *Quality & Quantity*, *28*(2), 185–217. doi: 10.1007/bf01102761.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391–410. doi: 10.1037/0033-2909.103.3.391.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydey-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for roderick p. mcdonald* (pp. 275–340). Lawrence Erlbaum Associates Publishers.

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: A Multidisciplinary Journal*, *11*(3), 320–341. doi: 10 . 1207 / s15328007sem1103_2.

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*(1), 85–110. doi: 10 . 1146 / annurev - clinpsy - 032813-153700.

Martinent, G., Nicolas, M., Gaudreau, P., & Campo, M. (2013). A cluster analysis of affective states before and during competition. *Journal of Sport and Exercise Psychology*, *35*(6), 600–611. doi: 10.1123/jsep.35.6.600.

McDonald, R. P. (1985). *Factor analysis and related methods*. Erlbaum.

McNeish, D. (2023). Dynamic fit index cutoffs for categorical factor analysis with likert-type, ordinal, or binary responses. *American Psychologist*, *78*(9), 1061–1075. doi: 10.1037/amp0001213.

McNeish, D. (2026). How do psychologists determine whether a measurement scale is good? A quarter-century of scale validation with Hu & Bentler (1999). *Annual Review of Psychology*, *77*, 567–591. doi: 10 . 1146/annurev-psych-121924-104021.

McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment*, *100*(1), 43–52. doi: 10 . 1080 / 00223891 . 2017.1281286.

McNeish, D., & Manapat, P. D. (2024). Dynamic fit index cutoffs for hierarchical and second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*(1), 27–47. doi: 10 . 1080 / 10705511 . 2023 . 2225132.

McNeish, D., & Wolf, M. G. (2022). Dynamic fit index cutoffs for one-factor models. *Behavior Research Methods*, *55*(3), 1157–1174. doi: 10.3758/s13428-022-01847-y.

McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, *28*(1), 61–88. doi: 10.1037/met0000425.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. doi: 10.1037/a0026802.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, *91*(3), 328–346. doi: 10.1037/0033-295X.91.3.328.

Perry, J. L., Nicholls, A. R., Clough, P. J., & Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. *Measurement in Physical Education and Exercise Science*, *19*(1), 12–21. doi: 10.1080/1091367X.2014.952370.

Pornprasertmanit, S., Miller, P., Schoemann, A., & Jorgensen, T. D. (2012). *simsem: SIMulated Structural Equation Modeling [R package]* (Version 0.5-17). doi: 10.32614/CRAN.package.simsem.

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.2) [R Foundation for Statistical Computing]. https://www.R-project.org

Rogers, P. (2024). Best practices for your confirmatory factor analysis: A JASP and lavaan tutorial. *Behavior Research Methods*, *56*(7), 6634–6654. doi: 10 . 3758 / s13428-024-02375-7.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. doi: 10.18637/jss.v048.i02.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*(2), 23–74. doi: 10.23668/psycharchives.12784.

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on sem fit indices. *Educational and Psychological Measurement*, *79*(2), 310–334. doi: 10.1177/0013164418783530.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors [paper presentation]. *Annual meeting of the Psychometric Society, Iowa City, Iowa, United States of America*.

Taylor, J. M. (2019). Overview and illustration of Bayesian confirmatory factor analysis with ordinal indicators. *Practical Assessment, Research & Evaluation*, *24*(1), 4. doi: 10.7275/vk6g-0075.

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

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. doi: 10.1007/BF02291170.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). The Guilford Press.

West, S. G., Wu, W., McNeish, D., & Savord, A. (2023). Model fit in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., pp. 184–205). The Guilford Press.

Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, *34*(5), 719–751. doi: 10.1177/0011000006286345.

Wickham, H., Hester, J., Chang, W., & Bryan, J. (2011). *Devtools: Tools to make developing R packages easier (version 2.4.6) [R package]*. doi: 10.32614/CRAN.package.devtools.

Wolf, M. G., & McNeish, D. (2021). *Dynamic model fit (version 1.1.0) [software]*. https://www.dynamicfit.app

Wolf, M. G., & McNeish, D. (2023). Dynamic: An R package for deriving dynamic fit index cutoffs for factor analysis. *Multivariate Behavioral Research*, *58*(1), 189–194. doi: 10.1080/00273171.2022.2163476.

Xiong, Z., Xia, H., Ni, J., & Hu, H. (2025). Basic assumptions, core connotations, and path methods of model modification—using confirmatory factor analysis as an example. *Frontiers in Education*, *10*, 1506415/1–9. doi: 10.3389/feduc.2025.1506415.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. doi: 10.20982/tqmp.09.2.p079.

Appendix A: R code generated using *rmarkdown* (Allaire et al., 2014)

```
#Listing 1 - Packages
install.packages("lavaan")

devtools::install_github("melissagwolf/dynamic")
install.packages("semPlot")

library(lavaan)
library(dynamic)
library(semPlot)

#Listing 2 - Loading data
my_data <- read.csv("dataset.csv")

#Listing 3 - Summary statistics
summary(my_data)

#Listing 4 - CFA
CFA <- 'MGA =~ ga1 + ga4 + ga7 + ga10 + ga13
SRGA =~ ga2 + ga5 + ga8 + ga11
NGA =~ ga3 + ga6 + ga9 + ga12
GA =~ MGA + SRGA + NGA'

#Listing 5 - Fit the CFA
fit <- cfa(CFA, data = my_data, ordered = TRUE, std.lv = TRUE, estimator = "WLSMV")
    #Rogers (2024)
summary(fit, fit.measures = TRUE, standardized = TRUE, rsquare = TRUE)
#Dion et al. (2021)
```

```
#Listing 6 - Modification indices
modindices(fit, sort = TRUE, maximum.number = 30)

#Listing 7 - Dynamic fit indices
DFIs <- hier1HB(fit, reps = 500)#Adapted from Wolf & McNeish (2021)
DFIs

#Listing 8 - Plotting the CFA model
ASAGSPlot <- semPaths(fit, thresholds = FALSE, intercepts = FALSE,
whatLabels = "std", edge.label.cex = 1, edge.color = "black") #Adapted from Wolf &
    McNeish (2021)

#Listing 9 - CFA - Alternative model
CFA2 <- 'Factor1 =~ ga1 + ga2 + ga4 + ga5 + ga7 + ga8 + ga10 + ga11 + ga13
Factor2 =~ ga3 + ga6 + ga9 + ga12
Factor1 ~~ Factor2'

#Listing 10 - Fit the CFA
fit2 <- cfa(CFA2, data = my_data, ordered = TRUE, std.lv = TRUE, estimator = "WLSMV
    ") #Rogers (2024)
summary(fit2, fit.measures = TRUE, standardized = TRUE, rsquare = TRUE)#Dion et al.
     (2021)

#Listing 11 - Difference in fit
anova(fit,fit2)

#Listing 12 - CFA - Estimating two non-target loadings
CrossLoad10 <- 'Factor1 =~ ga1 + ga4 + ga7 + ga10 + ga13
Factor2 =~ ga2 + ga5 + ga8 + ga10 + ga11
Factor3 =~ ga3 + ga6 + ga9 + ga10 + ga12
Factor4 =~ Factor1 + Factor2 + Factor3'

#Listing 13 - Fit the CFA
fitCrossLoad10 <- cfa(CrossLoad10, data = my_data, ordered = TRUE, std.lv = TRUE,
    estimator = "WLSMV") #Rogers (2024)
summary(fitCrossLoad10, fit.measures = TRUE, standardized = TRUE, rsquare = TRUE)#
    Dion et al. (2021)

#Listing 14 - CFA equality constraint #1
Equal.Constraint <- 'Factor1 =~ ga1 + ga4 + ga7 + ga10 + ga13
Factor2 =~ ga2 + ga5 + ga8 + Equal*ga10 + ga11
Factor3 =~ ga3 + ga6 + ga9 + Equal*ga10 + ga12
Factor4 =~ Factor1 + Factor2 + Factor3'

#Listing 15 - Fit the CFA
fitEqual.Constraint <- cfa(Equal.Constraint, data = my_data, ordered = TRUE, std.lv
     = TRUE, estimator = "WLSMV") #Rogers (2024)
summary(fitEqual.Constraint, fit.measures = TRUE, standardized = TRUE, rsquare =
    TRUE)#Dion et al. (2021)
```

```
#Listing 16 - Difference in fit
anova(fitCrossLoad10,fitEqual.Constraint)

#Listing 17 - CFA equality constraint #2
Equal.Constraint2 <- 'Factor1 =~ ga1 + ga4 + ga7 + Equal*ga10 + ga13
Factor2 =~ ga2 + ga5 + ga8 + ga10 + ga11
Factor3 =~ ga3 + ga6 + ga9 + Equal*ga10 + ga12
Factor4 =~ Factor1 + Factor2 + Factor3'

#Listing 18 - Fit the CFA
fitEqual.Constraint2 <- cfa(Equal.Constraint2, data = my_data, ordered = TRUE,
    std.lv = TRUE, estimator = "WLSMV") #Rogers (2024)
summary(fitEqual.Constraint2, fit.measures = TRUE, standardized = TRUE, rsquare =
    TRUE)#Dion et al. (2021)

#Listing 19 - Difference in fit
anova(fitCrossLoad10,fitEqual.Constraint2)
```

Open practices

The *Open Data* badge was earned because the data of the experiment(s) are available on osf.io/8yp5m/
The *Open Material* badge was earned because supplementary material(s) are available on osf.io/8yp5m/

Citation

Boivin, T. L. D., Veilleux, C., Egerton-Graham, M., Caron, P.-O., & Gaudreau, P. (2026). A tutorial on evaluating confirmatory factor analysis with lavaan and dynamic: Integrating statistical and conceptual approaches. *The Quantitative Methods for Psychology*, *22*(1), 9–24. doi: 10.20982/tqmp.22.1.p009.

Received: 16/02/2026 ∼ Accepted: 30/03/2026