

# DFI Cutoffs: A Tutorial

Melissa G Wolf      Dan McNeish

2023-01-22



# Contents

<b>Introduction</b>	<b>5</b>
<b>1 What are the different types of model fit?</b>	<b>7</b>
<b>2 What are the weaknesses of the current approaches to evaluate model fit?</b>	<b>9</b>
<b>3 Why should I use DFI cutoffs instead?</b>	<b>11</b>
<b>4 How does the DFI algorithm work?</b>	<b>13</b>
4.1 Multi-Factor Models . . . . .	13
4.2 One-Factor Models . . . . .	14
<b>5 How do I calculate DFI cutoffs?</b>	<b>15</b>
5.1 Applied Example . . . . .	15
<b>6 What does a model statement look like?</b>	<b>17</b>
6.1 Mplus . . . . .	17
6.2 SPSS Amos . . . . .	17
6.3 Model Statement . . . . .	19
<b>7 Why are there different levels and what do they mean?</b>	<b>21</b>
7.1 Applied Example . . . . .	22
<b>8 What does 95/5 and 90/10 mean, and how should I interpret the plots?</b>	<b>25</b>

<b>9</b>	<b>What does NONE mean, and what should I do if I see it?</b>	<b>27</b>
<b>10</b>	<b>How do I include DFI cutoffs in a manuscript?</b>	<b>29</b>
10.1	Applied Example: Standard Reporting . . . . .	29
<b>11</b>	<b>What should I do if DFI cutoffs don't exist for my model type?</b>	<b>31</b>
<b>12</b>	<b>What are the limitations of DFI cutoffs?</b>	<b>33</b>

# Introduction

Confirmatory factor analysis (CFA) is a commonly used statistical method in the social sciences. Although these models have been used for over a century, debate remains about how to evaluate the fit of factor models. Recently, we proposed the use dynamic fit index (DFI) cutoffs to evaluate model fit (McNeish and Wolf, 2021, 2022) and introduced a corresponding Shiny application to facilitate their use (Wolf and McNeish, 2020).

This book was written as a tutorial for applied psychologists that are interested in using the DFI Shiny App to compute model specific cutoffs for their CFA models. We wrote this to make DFI cutoffs more accessible to everyone, especially those that use SPSS Amos and Mplus.

In this book, we will walk through 12 commonly asked questions about DFI cutoffs and use an applied example to demonstrate how to use the Shiny app to calculate them. For R users, DFI cutoffs are also available on CRAN under the package `dynamic` (Wolf and McNeish, 2022).

```
install.packages("dynamic")  
library(dynamic)
```



# Chapter 1

## What are the different types of model fit?

There are two types of global model fit in CFA: exact fit and approximate fit. Exact fit is a test of model fit in that it compares a test statistic to a probability distribution to calculate a  $p$ -value, while an approximate fit index can be thought of as an effect size measure that quantifies the degree of misfit in the model. Both are derived from the amount of overall misfit in the model, where misfit is defined as the difference between the model-implied variance-covariance matrix (e.g., the user's path diagram) and the data-generated variance-covariance matrix (i.e., the observed relationships in the user's data).

The most commonly used test of **exact fit** is the  $\chi^2$  test, although others can be used as well (see McNeish, 2020). Tests of exact fit are concerned with the presence of misfit (of any kind) anywhere in the variance-covariance matrices. The  $\chi^2$  test is a test of exact fit because the null hypothesis is that the model-implied variance-covariance matrix exactly matches the data-generated variance-covariance matrix. If the  $\chi^2$  test is significant, researchers have evidence to suggest that the model does not exactly fit the data. As such, the  $\chi^2$  test is the strictest way to evaluate model fit and test psychological theory (Hayduk et al., 2007).

The most used **approximate fit** indices are the standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA), and the comparative fit index (CFI). There are no probability distributions or  $p$ -values associated with these indices, and thus they are not tests of fit but rather a way to evaluate the amount of misfit in the model (e.g., to determine if the misfit is trivial or substantial). In other words, approximate fit indices are different from tests of exact fit because they are concerned with quantifying the degree of misfit in the model. The SRMR is derived from the residual correlation matrix (i.e., the difference between the model-implied variance-covariance matrix and

the data-generated variance-covariance matrix) and can be thought of as the average magnitude of the residuals. The RMSEA is derived from the  $\chi^2$  statistic but differs in that it includes a parsimony correction (i.e., it rewards simpler models with stronger theory that restrict more paths to 0). The CFI is the ratio of the model-reported  $\chi^2$  and the baseline  $\chi^2$  and can thus be thought of as the relative improvement in model fit. Lower values of SRMR and RMSEA are indicative of better fit while higher values of the CFI are indicative of better fit (Brown, 2015).

The last type of model fit is localized area of fit. In contrast to an overall evaluation of global fit, **local fit** is an investigation of each cell of the variance covariance matrix to diagnose any areas of strain in the model. This is often done to probe the source of misfit after finding a significant  $\chi^2$  statistic or a value of an approximate fit index that is indicative of substantial misfit. Still, good global fit can mask local misfit and thus it may be prudent to check regardless of global model fit. Although local fit is discussed at length in introductory textbooks for factor analysis (Brown, 2015; Kline, 2011), it is often not presented in journal articles.

Briefly, local fit can be investigated by probing the residual correlation matrix for extreme cases or by consulting modification indices for modifications that would substantially reduce the  $\chi^2$  statistic (Kline, 2011). Even if an extreme value is found, the model should not be changed unless the revisions are supported theoretically. There are several reasons for this. The first is that the modifications suggested by the software are derived statistically and may substantially alter the theory behind the initial model. Secondly, the misfit may be sample-specific and thus may not generalize to other samples. Third, there is no guarantee that a modification will resolve misfit and could instead lead a researcher down a rabbit hole akin to p-hacking. Fourth, it is difficult to know if misfit is due to an issue with the theory about the internal structure or one or more of the items. As such, any modifications should be justified theoretically and qualitatively.



## Chapter 2

# What are the weaknesses of the current approaches to evaluate model fit?

In terms of **exact fit**, the  $\chi^2$  test does not perform as well in small samples and may also be overly sensitive to minor misspecifications at large sample sizes (Browne and Cudeck, 1992; Hu et al., 1992). Additionally, with extremely small samples and smaller loadings, the  $\chi^2$  test can be underpowered and unable to detect misfit (McNeish et al., 2018). Some models (such as bifactor models) inherently have a higher “fit propensity”, meaning that they are more likely to fit the data regardless of the true nature of the data generating model (Bonifay and Cai, 2017; Bonifay et al., 2017; Preacher, 2006). Further, even if one does find a non-significant p-value for a  $\chi^2$  test with a reasonable sample size, this does not guarantee that the true model has been recovered as there are often multiple equivalent models that can fit the data (Hayduk, 2014). As such, all modeling decisions should have strong theoretical grounding.

Unlike the  $\chi^2$  test, **approximate fit** indices do not have a corresponding p-value. Given that these indices essentially function as effect size measures that capture the degree of misfit in the model, the difficulty lies in how to interpret them and which cutoff values to use (if any). Researchers often rely on a set of fixed cutoff values derived from a simulation study conducted by Hu and Bentler (1999), which has over 96,000 citations as of 2022. This simulation study produced the well-known cutoff values of SRMR < .08, RMSEA < .06, and CFI > .95. However, interpretations of the results from simulations studies are limited to the conditions sampled in the simulation study. Hu and Bentler manipulated the sample size (from 250 – 5000), the number and type of misspecifications (omitted crossloadings and omitted factor covariances), and the normality of the factors and errors. However, they did not manipulate the

number of factors (3), the number of items (15), the magnitude of the factor loadings (.7 - .8), or the model type (single level CFA estimated using maximum likelihood estimation).

Several studies have demonstrated that these fixed cutoff values cannot reliably be extrapolated to other model subspaces (e.g., one-factor models, multi-factor models with stronger or weaker loadings, or models with fewer or greater numbers of items or factors). In other words, if a researcher evaluates the fit of a single-level CFA model that does not have 15 items, 3 factors, and a sample size between 250 – 5000, the cutoff values derived from Hu and Bentler’s study cannot be used to reliably determine if there is substantial misfit in the model (McNeish and Wolf, 2021). Most concerning is the “reliability paradox” which stipulates that lower loadings (e.g., a smaller reliability coefficient) are associated with “better” values of approximate fit indices (Hancock and Mueller, 2011; Heene et al., 2011; Marsh et al., 2004; McNeish et al., 2018; Saris et al., 2009). In other words, holding all else equal, as factor loadings decrease, the SRMR and the RMSEA will also decrease, mistakenly leading researchers to conclude that less reliable models fit the data better (when compared to a set of fixed cutoff values). If Hu and Bentler had varied the factor loadings in their original simulation study, it is possible that they would have been unable to recommend a set of fixed cutoff variables since fit indices are so sensitive to loading magnitude.

## Chapter 3

# Why should I use DFI cutoffs instead?

Hu and Bentler’s approach to quantifying the degree of misspecification in a model was sensible; the problem lies with the interpretation of the fixed cutoff values (Millsap, 2007; Pornprasertmanit et al., 2013). The first problem is the extrapolation of this fixed set of cutoffs to conditions outside of the ones sampled in their original simulation. Secondly, because Hu and Bentler presented a single set of cutoffs, researchers were inadvertently encouraged to incorrectly treat misfit as a binary decision akin to a test of model fit (e.g., either the model fits well, or it does not). However, the only test of model fit is a test of exact fit; approximate fit indices are useful primarily because they can help researchers judge the extent to which the misfit in their model may be trivial or substantial.

DFI cutoffs are an improvement over the traditional fixed cutoff values because they address both of the issues raised above. DFI cutoffs are tailored to the user’s specific model, which alleviates the first problem of improper extrapolation. Researchers can think of DFI cutoffs as “if Hu and Bentler had used my exact model for their simulation study, these are the cutoff values that they would have published”. Unlike the traditional fixed cutoffs, DFI cutoffs (when available) are accurate for the user’s model and can reliably distinguish between a correctly specified model and an incorrectly specified model. In this sense, DFI cutoffs can be thought of as analogous to a custom power analysis (albeit one that is quite simple to conduct).

Secondly, the DFI algorithm is written to return a series of custom cutoff values that range from trivial misfit to substantial misfit. This addresses the second problem of improper interpretation because it encourages researchers to treat misfit as a continuum or a spectrum rather than a binary decision of “good” or “bad”. Because there is less finality associated with an interpretation of model fit when a series of cutoffs are used, the DFI approach also encourages researchers to

properly reconceptualize model fit as only one type of validity evidence rather than the crux of validity (American Educational Research Association et al., 2014). As such, if a researcher found evidence of trivial misfit according to the DFI cutoffs, they could still potentially defend the use of their scale if they have other sources of evidence of validity in support of its use (while still acknowledging that if the  $\chi^2$  test is significant then the model does not exactly fit the data).

## Chapter 4

# How does the DFI algorithm work?

The DFI algorithm follows Hu and Bentler’s (1999) approach to model misspecification in that it simulates a distribution of fit indices from a correctly specified model and a misspecified model and then chooses a cutoff value that distinguishes between the two distributions. Hu and Bentler created their misspecified models by omitting one or two cross-loadings from a three-factor model (the “complex” condition; see Figure 4.1) and by omitting one or two factor correlations from the same three-factor model (the “simple” condition). The DFI algorithm mimics this approach by making the user’s model both the data generating model and the analytic model (this is the true condition). In the misspecified condition(s), the analytic model remains the user’s model, but a series of misspecifications are added to the data generating model (in line with the conventions established by Hu and Bentler). Readers interested in understanding the DFI algorithm in depth should consult McNeish and Wolf (2021; 2022).

### 4.1 Multi-Factor Models

For multi-factor models, the DFI algorithm creates  $f-1$  levels of misspecifications, where  $f$  is the number of factors in the model. As such, a two-factor model will have one level of misspecification, while a six-factor model will have five levels of misspecification. The misspecification follows Hu and Bentler’s approach of omitting a cross-loading with a magnitude equivalent to the lowest loading in the model from the factor with the highest reliability. For example, the Level 1 misspecification will omit one cross-loading with a magnitude equivalent to the lowest loading, while the Level 2 misspecification will omit both the

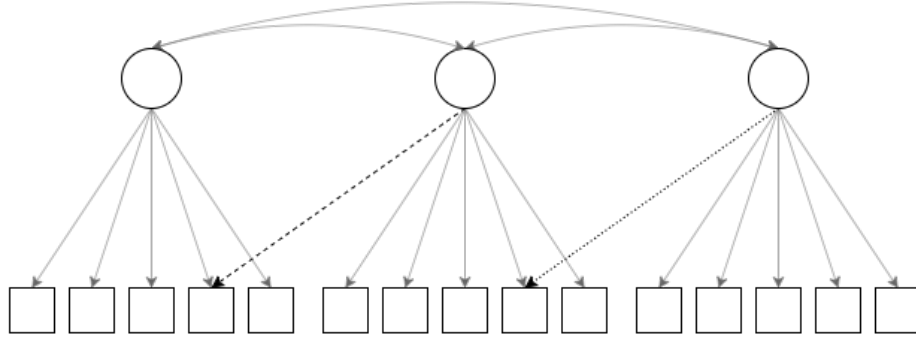


Figure 4.1: A path diagram of the model misspecifications used by Hu and Bentler (1999). The true data generating model has all of the loadings. The minor misspecification condition omits the loading with the dashed line from the data generating model, and the major misspecification condition omits both the dashed line and the dotted line from the data generating model. The analytic model is the same for all conditions.

Level 1 cross-loading and an additional cross-loading with a magnitude equivalent to the second lowest loading in the model. The magnitude of the loading that is omitted can be found in the Info tab of the app and will be returned by default in the R package.

## 4.2 One-Factor Models

The algorithm for one-factor models is unique in that it cannot exactly follow the approach established by Hu and Benter (1999) because it is impossible to omit factor correlations or cross-loadings in a one-factor model. As such, the DFI algorithm employs an approach inspired by Shi & Maydeu-Olivares (2020) of omitting residual correlations to create a misspecified model. One-factor models have three levels of misspecification, where Level 1 has approximately 1/3rd of items with an omitted residual correlation of .3, Level 2 has approximately 2/3rds of items with an omitted residual correlation of .3, and Level 3 has omitted residual correlations of .3 from all items. As such, the DFI algorithm for the one-factor model is standardized such that the number of items with an omitted residual correlation is proportional to the total number of items in the model, making it easier to compare degree of misfit across models.

## Chapter 5

# How do I calculate DFI cutoffs?

DFI cutoffs can easily be computed using the free, open source, web-based Shiny application, accessible at [www.dynamicfit.app](http://www.dynamicfit.app). The app has a simple, user-friendly, point-and-click interface, which requires no knowledge of coding to operate. The user need only enter their model statement with standardized loadings (see Chapter 6 for more details) and their sample size. Behind the scenes, the Shiny app will use R to run a series of Monte Carlo simulations to return a continuum of cutoff values tailored to the user’s individual model. R users who wish to bypass the app can instead make use of the corresponding R package `dynamic`, available on `CRAN`, which will return the same results as the web application.

### 5.1 Applied Example

Computing DFI cutoffs is a post-hoc endeavor; in other words, users must first run their CFA model to get some of the information that is necessary to calculate custom fit index cutoffs. Thus, to make this tutorial easier to follow, we introduce an applied example which will be used throughout the rest of the paper. The data comes from a popular personality assessment commonly referred to as the “Big Five”, which was provided by the Open Source Psychometrics Project (Goldberg, 1992). We will use the 10-item “extraversion” factor to compute DFI cutoffs for a one-factor model (see Figure 5.1).

To use the Shiny application, researchers should visit the website and select the app that corresponds to their model type. In this case, we will select the one-factor CFA application (see Figure 5.2). The app description states that only two pieces of information are needed: (1) the user’s standardized loadings from

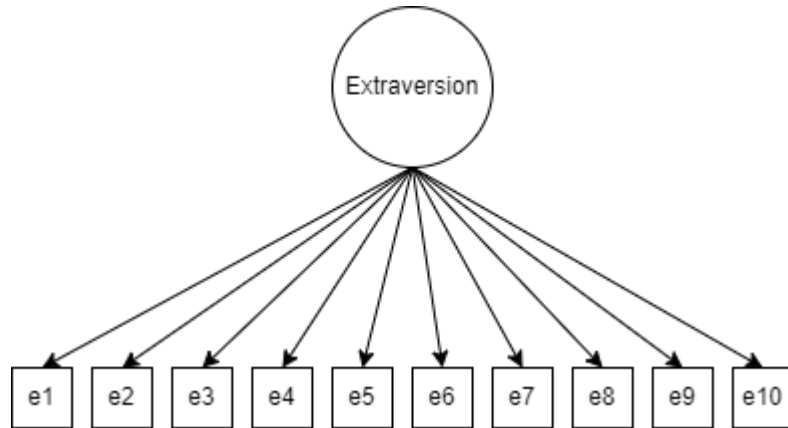


Figure 5.1: The one-factor model used for the demonstration in this tutorial ( $n = 1,222$ ).

the fitted model, and (2) the sample size. The standardized loadings will be used to create the model statement which will be uploaded to the app to compute the custom DFI cutoffs (see Chapter 6). The function in the R package `dynamic` that corresponds to the one-factor CFA app is `cfaOne`.

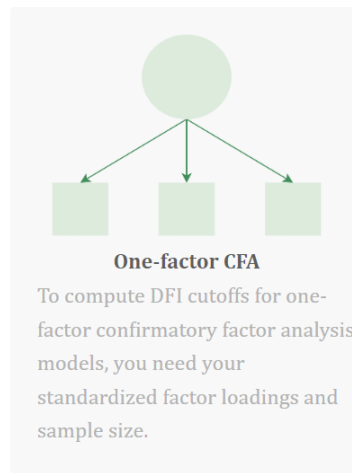


Figure 5.2: The one-factor CFA application on [www.dynamicfit.app](http://www.dynamicfit.app).



## Chapter 6

# What does a model statement look like?

After opening the one-factor CFA app, users are prompted to enter two pieces of information: (1) their sample size, and (2) their model statement (see Figure 6.1). The model statement is created using the standardized loadings from the user's fitted model (i.e., the results from the CFA model that the user wants to calculate DFI cutoffs for). These will be found in the software output that was used to run the original CFA model. We will walk through an example from Mplus and Amos using the Extraversion factor from the Big 5 dataset provided.

### 6.1 Mplus

To get the standardized loadings from Mplus (current as of version 8.7), we add the following argument to the end of the input file: `OUTPUT: STDYX; .` The standardized loadings will be found under the section of the output titled `STDYX STANDARDIZATION` (see Figure 6.2). The magnitude of the standardized loading for each indicator is under the header titled Estimate. For example, the standardized loading for E1 is .671.

### 6.2 SPSS Amos

To get the standardized loadings from SPSS Amos (current as of version 28), users should select the “Analysis Properties” icon, and check the “Standardized estimates” box under the “Output” tab. After running the model, the standardized loadings can be found in the “Parameter Formats” box by deselecting “Unstandardized estimates” and selecting “Standardized estimates”. They will

This app uses Monte Carlo simulations to generate dynamic fit index cutoff values for one-factor models.

Input Sample Size

Input Model Statement

Browse... .txt file

This may take a few minutes. Please only press submit once.

Submit

Figure 6.1: The required inputs for the one-factor CFA app.

STDYX Standardization

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
E	BY				
	E1	0.671	0.018	38.165	0.000
	E2	-0.705	0.016	-43.147	0.000
	E3	0.704	0.016	43.064	0.000
	E4	-0.702	0.016	-42.615	0.000
	E5	0.750	0.015	51.506	0.000
	E6	-0.572	0.021	-27.147	0.000
	E7	0.744	0.015	50.347	0.000
	E8	-0.514	0.023	-22.454	0.000
	E9	0.605	0.020	30.275	0.000
	E10	-0.703	0.016	-42.941	0.000

Figure 6.2: The standardized loadings from Mplus.

appear on the path diagram and can easily be copied from the syntax. The syntax can be accessed by toggling to the “Syntax” tab underneath the path diagram (see Figure 6.3).

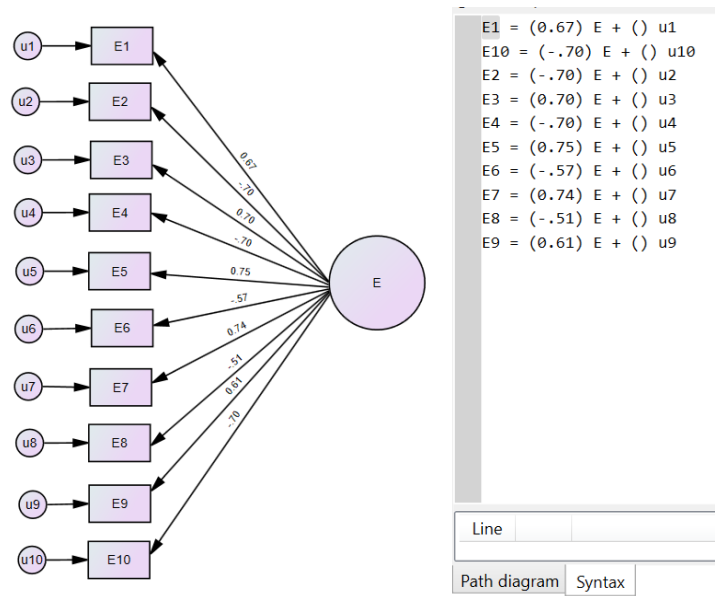


Figure 6.3: A path diagram with standardized loadings from Amos, and the standardized loadings as seen in the Amos Syntax tab.

## 6.3 Model Statement

We will use the standardized loadings from Mplus and Amos to write out the model statement. Note that the model statement must be saved in a .txt file to be uploaded to the app. The easiest way to create a .txt file on a PC is in Notepad, while the easiest way on a Mac is in TextEdit (make sure to save your TextEdit file as plain text).

The model statement should be written in lavaan style syntax . The regression relationship between the factor and any items will use the syntax  $=\sim$ , while any correlational relationships (e.g., between factors or items) will use the syntax  $\sim\sim$ . The model statement for this one-factor CFA model will be written as:

$$Extraversion = .671 * E1 + -.705 * E2 + .704 * E3 + -.702 * E4 + .750 * E5 + -.572 * E6 + .744 * E7 + -.514 * E8 + .605 * E9 +$$

As seen above, the model statement follows the following format: Factor = item loading magnitude \* item name. Because the first loading had a magnitude

of .671, it is written as .671\*E1. The magnitude of the loading will always come before the name of the item. The factors and items can have any name (e.g., orange =~ .671\*apple would work), however the name cannot start with a number (e.g., 123orange would not be permissible). Note that negative loadings still need to have a + sign as a link in the model statement (e.g., .671\*E1 + -.705\*E2). This model statement would then be saved as a .txt file and uploaded to the app along with the sample size (in this case, the sample size is 1,222). Users would then press submit to begin the simulation.

## Chapter 7

# Why are there different levels and what do they mean?

After submitting the model statement and the sample size, the app will compute several Monte Carlo simulations (each with 500 replications) to return the DFI cutoffs. Once the busy bar finishes running, the DFI cutoffs will be found under the “Results” tab. For most models, there will be a series of cutoff values beginning with “Level 1”. One-factor models will typically have three levels of cutoff values. Figure 7.1 displays the three levels of DFI cutoffs for the model used in this tutorial.

Instructions	Results	Plots	References
These are the dynamic fit index cutoff values for your model:			
	SRMR	RMSEA	CFI
Level 1: 95/5	.025	.04	.986
Level 1: 90/10	--	--	--
Level 2: 95/5	.037	.075	.953
Level 2: 90/10	--	--	--
Level 3: 95/5	.043	.095	.928
Level 3: 90/10	--	--	--

Figure 7.1: The DFI cutoffs for the model used in this tutorial.

The levels correspond to increasing degrees of misspecification in the fit of the model, enabling researchers to reconceptualize misfit as a continuum analogous to an effect size measure. Misspecifications are cumulative such that higher levels are equivalent to more egregious model misspecifications. The Level 1 cutoff can therefore be thought of as the strictest fit criteria because it is consistent

with the smallest misspecification from the misfit continuum. Thus, the ideal outcome would be if a user's fitted values for their SRMR, RMSEA, and CFI were all below the Level 1 cutoff, because this would indicate that the observed fitted values were more similar to a model that was correctly specified. In other words, if the user's model fit was below the Level 1 cutoffs, that would mean it fit better than a model that had only a minor misspecification. As the levels increase, the cutoff values will become more lenient because they will correspond to models that are more misspecified.

If the fitted values for a one-factor model were above the Level 1 cutoff but below the Level 2 cutoff, one might conclude that their model was consistent with modest misspecifications and could potentially argue that the misspecifications were not a substantial threat to the validity of their inferences. If the fitted values were above the Level 2 cutoff but below the Level 3 cutoff, the model fit might be categorized as moderately misspecified, while fitted values above the Level 3 cutoff might be described as substantially misspecified. It is also possible to observe fitted values for each of the indices that are classified at differing cutoff levels. In this case, the researcher might report that the fit indices were consistent with different degrees of misspecification, and potentially attempt to diagnose the inconsistencies (e.g., by investigating local areas of strain). As always, it is up to the researcher to present multiple types of evidence in defense of the validity of their assessment (e.g., finding fit consistent with a minor misspecification should not be the entirety of the claim for evidence of validity). See Chapter 4 for more discussion about levels and model misspecification.

## 7.1 Applied Example

To determine how well their model fits their data, researchers should compare the fit of their model (derived from their software of choice) to the DFI cutoffs derived from the app or the R package. In this case, the fit index values for the Extraversion empirical model are  $\chi^2 = 436.55$  ( $df = 35$ ,  $p < .001$ ), SRMR = .048, RMSEA = .097 [.089, .105], CFI = .921. Compared to the traditional fixed cutoff values from Hu and Bentler (1999), the SRMR would be indicative of good fit while the RMSEA and the CFI would be indicative of poor fit.

There are two drawbacks to using the fixed cutoffs. The first is that the fixed cutoffs are derived from a different model subspace (a 3-factor model with 15 items) that does not generalize to our model (a one-factor model with 10 items). The second is that because there is only one set of cutoff values, we cannot infer the degree of misspecification present, forcing us to make a binary decision about a continuum of misspecification. Alternatively, when comparing our fit index values to the DFI cutoffs that are tailored to our empirical model, we see that all three indices are consistent with a Level 3 misspecification (see Figure 7.1). This is because the SRMR and RMSEA are greater than the Level 3 DFI cutoff, while the CFI is less than the Level 3 DFI cutoff. As such, we

can infer that the model is substantially misspecified, rendering it difficult to defend as valid especially without other types of validity evidence. We might follow up by investigating local fit such as consulting the modification indices in our software of choice. In doing so, we see that adding a residual correlation between E8 and E9 would reduce the  $\chi^2$  by 156.158, or 36%. The next steps might involve qualitatively investigating the cause of the relationship between these two items, removing one of them if they are deemed redundant, adding a residual correlation between the two of them if it is theoretically justified, or something else.





## Chapter 8

# What does 95/5 and 90/10 mean, and how should I interpret the plots?

The 95/5 and 9/10 thresholds are derived from Hu and Bentler’s approach to minimizing the classification error rates. In Chapter 4, we mentioned that the DFI algorithm simulates a distribution of fit index values from a correctly specified model and a misspecified model and then selects a cutoff value that distinguishes between the two distributions. In the app, these distributions are visualized under the “Plots” tab (see Figure 8.1 for the Level 3 distributions from the applied example). But how is that value selected, especially when the distributions might overlap leaving researchers unsure as to whether the fit index value that they observed is consistent with one that is derived from a misspecified model or a correctly specified model?

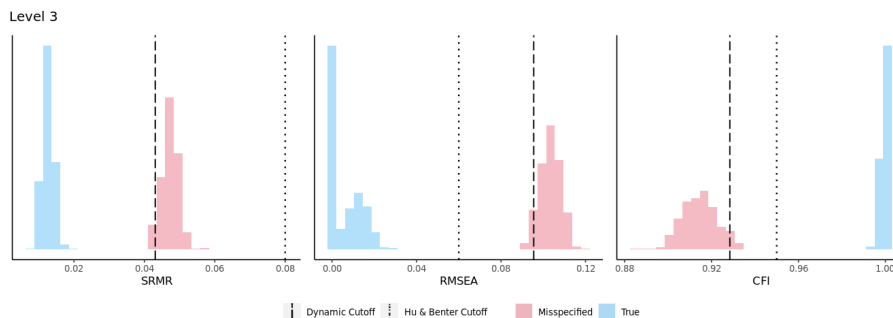


Figure 8.1: The Level 1 distributions from the DFI algorithm for the Extraversion scale.

To avoid ambiguity, the DFI algorithm consistently selects the cutoff value from the misspecified distribution. Specifically, the magnitude of the cutoff value corresponds to the 5th percentile of the misspecified distribution for the SRMR and RMSEA and the 95th percentile of the CFI (this is because low values of SRMR and RMSEA are indicative of better fit while high values of CFI are indicative of better fit). These values correspond to the dashed lines in Figure 8.1 (the Hu and Bentler cutoffs are also presented for comparison as dotted lines). Since the fitted value for the SRMR in the Extraversion example was .048, it is more likely that this value would come from a distribution of misspecified fit indices because the SRMR misspecified distribution (in red) ranges from .041 to .056, while the SRMR correctly specified distribution (in blue) ranges from .008 to .019. As such, the conclusion is that the fitted model is likely misspecified in a way that is consistent with a Level 1 misspecification.

In Figure 8.1, the two distributions are distinct and clearly separated. However, sometimes the misspecified and the correctly specified distributions will overlap. Since the distributions can overlap, the rule that the DFI algorithm uses is to select the 5th percentile of the misspecified distribution (for the SRMR and RMSEA) so long as the value that is returned is also greater than the 95th percentile from the correctly specified distribution. This check is put in place to safeguard against mistakenly choosing a cutoff value that could just as easily come from a correctly specified distribution. If the distributions overlap too much, (such that the 5th percentile of the misspecified distribution is less than the 95th percentile of the correctly specified distribution), then the DFI algorithm will attempt to return the 10th percentile of the misspecified distribution so long as the value is greater than the 90th percentile of the correctly specified distribution. Conceptually, this is like changing the alpha from .05 to .10 in standard null hypothesis significance testing. As such, the probability of misclassification is higher with the 90/10 rule than with the 95/5 rule, but the overall likelihood of making an error is still reasonably low (McNeish and Wolf, 2022).

## Chapter 9

# What does NONE mean, and what should I do if I see it?

In Chapter 8, we spoke about how the cutoff values are derived from the 5th percentile of the misspecified distributions of fit indices (for SRMR and RMSEA; the 95th percentile for CFI), and the problems that begin to arise if the misspecified and correctly specified distributions overlap. If the misspecified and correctly specified distributions of fit indices overlap substantially, then we become unsure as to whether an observed fitted value would be more likely to be found in a distribution of fit indices that were derived from a correctly specified model or a misspecified model. An example of this can be seen in Figure 9.1. When this happens, the DFI algorithm will return the word “NONE” for that fit index.

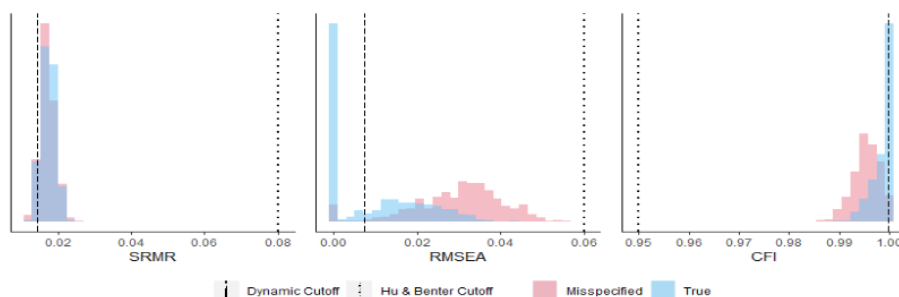


Figure 9.1: An example of overlapping distributions that would result in a NONE outcome.

When the word “NONE” is returned, that means that there are no cutoff values for that level of misfit that can reliably distinguish between a correctly specified model and a misspecified model. This can be verified visually in the “Plots” tab of the app. This is more likely to happen when sample sizes are small and loadings are low (Hancock and Mueller, 2011; Heene et al., 2011; McNeish et al., 2018). If there are DFI cutoff values available for other indices or other levels, they can and should still be used. If there are no DFI cutoff values available for any indices or any levels, the solution is not to rely on the traditional fixed cutoff values from Hu and Bentler as they similarly cannot distinguish between a correct and misspecified model. Instead, users can attempt to collect more data to increase their sample size, rely on the  $\chi^2$  test, or investigate local fit.

## Chapter 10

# How do I include DFI cutoffs in a manuscript?

In our experience, it is easiest to include DFI cutoffs in a manuscript by putting them in a table and then referencing the table in the text of the article. Researchers should report the fit of their model as they normally would, and then reference the likely magnitude of misspecification by comparing each approximate fit index to the table of DFI cutoff values. An example of a write up using the model from this tutorial is presented in the next section. Because tailored cutoffs (specifically, DFI cutoffs) are a relatively new development, it may be worthwhile to mention that they are used to quantify the degree of misfit in the model or include a sentence about the limitations of fixed cutoff values for readers who are not familiar with existing literature.

In this tutorial, the model was substantially misspecified. When this happens, researchers may be interested in modifying the model to attempt to improve the fit (although, note that model modification is only potential resolution and modifications should not be made without strong theoretical justifications). It is not clear to us quite yet how to proceed with DFI cutoffs when using modifying the model (i.e., we are not sure if DFI cutoffs should be updated for the new model or not). This is discussed more in Chapter 12. We are currently working on resolving this conundrum so that we can make clear recommendations for researchers.

### 10.1 Applied Example: Standard Reporting

The test of exact fit was statistically significant ( $\chi^2 = 436.55$ ,  $df = 35$ ,  $p < .001$ ) indicating that the model did not exactly fit the data. The approximate fit indices for the model were SRMR = .048, RMSEA = .097 [90% CI (.089,

Table 10.1: The DFI cutoffs used to quantify the degree of misfit in the model.

	SRMR	RMSEA	CFI
Level 1	0.025	0.040	0.986
Level 2	0.037	0.075	0.953
Level 3	0.043	0.095	0.928

.105)], and  $CFI = .921$ . These indices are essentially effect size measure for the magnitude of misfit. To quantify the degree of misfit reflected in these indices is, we compare the fit indices to a series of dynamic fit index (DFI) cutoffs (McNeish and Wolf, 2022) calculated by the one-factor DFI Shiny app version 1.1.0 (Wolf and McNeish, 2020). A table with the resulting cutoffs derived for this model are shown below. The SRMR and RMSEA from the model were above the Level-3 DFI cutoff and the CFI was below the Level-3 DFI cutoff, indicating that the fit of the model is consistent with a substantial misspecification.

## Chapter 11

# What should I do if DFI cutoffs don't exist for my model type?

As of this writing, the DFI method supports CFA models with continuous indicators only and the simulation component of the software assumes multivariate normality. It takes some time to work out generalizations for other types of models because a general method for identifying relevant misspecifications must be done model by model. For instance, potential misspecification that are relevant to latent growth models would likely be very different than a confirmatory factor analysis because latent growth models are typically interested in aspects like the function form of growth being correct or whether the correlation between repeated measures is reasonable rather than things like omitted cross-loadings that are relevant to confirmatory factor analysis. Our current work is focused on extending the method to higher-order models, categorical indicators, non-normality, missing data, and measurement invariance; so we expect those or related extensions will be the next to be added to the DFI method.

In the meantime, researchers should rely on the chi-square test and investigate local areas of strain to look for obvious misfit (e.g., viewing the standardized residual covariance matrix). Researchers can also use the Exact Fit application in the Shiny App or the `exactFit` function in the R package to return the 95th or 99th percentile of the distribution of cutoff values for a correctly specified model (for researchers that used one of the currently available apps, this can also be found in the Level 0 tab). Because this is a distribution of cutoff values for the true model, the values that are returned are the strictest way to evaluate approximate model fit. If the researchers fit index values fall below these values, this indicates that the fit of their model is consistent with a model that is correctly specified. These cutoff values can be computed for any model with

continuous outcomes (e.g., models estimated using ML or MLR).



## Chapter 12

# What are the limitations of DFI cutoffs?

In addition to only being available for a limited number of model types, DFI cutoffs have three other notable limitations. Currently, the cutoff values are derived by simulating data that is multivariate normal, which may not be consistent with the researcher's real data. We are currently working on switching to a bootstrapping approach which would sample the researcher's data and account for any non-normality anywhere in the model, resulting in cutoff values that are more accurate. Implementing this would require researchers to upload their data to the app, but it would mean that the model statement was simpler to write (e.g., it would no longer be necessary to include the magnitude of the standardized loadings in the model statement).

Additionally, the misspecifications for the one-factor model are currently standardized and thus somewhat comparable across models regardless of the number of items, but the misspecifications for the multi-factor models are not. This is because the multi-factor model replicates Hu and Bentler's approach to misspecification which involves adding one cross-loading for each  $f-1$  factor in the model with a magnitude equivalent to the item with the lowest loading in the model. Meanwhile, the one-factor model simply adds residual correlations with magnitudes of .3 proportional to the total number of items in the model. We are also working on introducing a similar standardized approach to model misspecification for multi-factor models.

Lastly, it is not clear to us quite yet if DFI cutoffs should be recomputed every time a model is modified. It is possible that cutoff values could change considerably for multi-factor models if the magnitude of the lowest factor-loading changes substantially (e.g., from .3 to .7). This will likely be resolved by standardizing the approach to misspecification for multi-factor models, which will make it easier to determine if the cutoff values should be recalculated. At this

point, we hypothesize that it may not be necessary to recompute DFI cutoffs for small modifications to a model (i.e., adding a residual correlation) but it may be necessary to recompute DFI cutoffs for larger modifications (e.g., switching from a one-factor model to a two-factor model).

# Bibliography

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association, Washington, D.C. OCLC: ocn826867074.
- Bonifay, W. and Cai, L. (2017). On the Complexity of Item Response Theory Models. *Multivariate Behavioral Research*, 52(4). Publisher: Routledge  
\_eprint: <https://doi.org/10.1080/00273171.2017.1309262>.
- Bonifay, W., Lane, S. P., and Reise, S. P. (2017). Three Concerns With Applying a Bifactor Model as a Structure of Psychopathology. *Clinical Psychological Science*, 5(1):184–186. Publisher: SAGE Publications Inc.
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. The Guilford Press, 2nd edition.
- Browne, M. W. and Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2).
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1):26–42. Place: US Publisher: American Psychological Association.
- Hancock, G. R. and Mueller, R. O. (2011). The Reliability Paradox in Assessing Structural Relations Within Covariance Structure Models. *Educational and Psychological Measurement*, 71(2):306–324.
- Hayduk, L. (2014). Seeing Perfectly Fitting Factor Models That Are Causally Misspecified: Understanding That Close-Fitting Models Can Be Worse. *Educational and Psychological Measurement*, 74(6):905–926.
- Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., and Boulianne, S. (2007). Testing! testing! one, two, three – Testing the theory in structural equation models! *Personality and Individual Differences*, 42(5):841–850.
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., and Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A

- cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, 16(3):319–336.
- Hu, L. and 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.
- Hu, L., Bentler, P. M., and Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112:351–362. Place: US Publisher: American Psychological Association.
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*. The Guilford Press, 3rd edition.
- Marsh, H. W., Hau, K.-T., and 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.
- McNeish, D. (2020). Should we use F-tests for model fit instead of chi-square in overidentified structural equation models? *Organizational Research Methods*, 23(3):487–510. Publisher: SAGE Publications Inc.
- McNeish, D., An, J., and 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.
- McNeish, D. and Wolf, M. G. (2021). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*. Place: US Publisher: American Psychological Association.
- McNeish, D. and Wolf, M. G. (2022). Dynamic fit index cutoffs for one-factor models. *Behavior Research Methods*.
- Millsap, R. E. (2007). Structural equation modeling made difficult. *Personality and Individual Differences*, 42(5):875–881.
- Pornprasertmanit, S., Wu, W., and Little, T. D. (2013). A Monte Carlo Approach for Nested Model Comparisons in Structural Equation Modeling. In Millsap, R. E., van der Ark, L. A., Bolt, D. M., and Woods, C. M., editors, *New Developments in Quantitative Psychology*, pages 187–197, New York, NY. Springer.
- Preacher, K. J. (2006). Quantifying Parsimony in Structural Equation Modeling. *Multivariate Behavioral Research*, 41(3):227–259.
- Saris, W. E., Satorra, A., and van der Veld, W. M. (2009). Testing Structural Equation Models or Detection of Misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4):561–582.

- Shi, D. and Maydeu-Olivares, A. (2020). The effect of estimation methods on SEM fit indices. *Educational and Psychological Measurement*, 80(3):421–445.
- Wolf, M. G. and McNeish, D. (2020). Dynamic Model Fit.
- Wolf, M. G. and McNeish, D. (2022). *dynamic: DFI Cutoffs for Latent Variable Models*. R package version 1.1.0.