Structural Equation Modelling: Guidelines for Determining Model Fit

Daire Hooper[1], Joseph Coughlan[1], and Michael R. Mullen[2]
[1]Dublin Institute of Technology, Dublin, Republic of Ireland
[2]Florida Atlantic University, Florida, USA
daire.hooper@dit.ie
joseph.coughlan@dit.ie
mullen@fau.edu

Abstract: The following paper presents current thinking and research on fit indices for structural equation modelling. The paper presents a selection of fit indices that are widely regarded as the most informative indices available to researchers. As well as outlining each of these indices, guidelines are presented on their use. The paper also provides reporting strategies of these indices and concludes with a discussion on the future of fit indices.

Keywords: Structural equation modelling, fit indices, covariance structure modelling, reporting structural equation modelling, model fit

1. Introduction

Structural Equation Modelling (SEM) has become one of the techniques of choice for researchers across disciplines and increasingly is a 'must' for researchers in the social sciences. However the issue of how the model that best represents the data reflects underlying theory, known as model fit, is by no means agreed. With the abundance of fit indices available to the researcher and the wide disparity in agreement on not only which indices to report but also what the cut-offs for various indices actually are, it is possible that researchers can become overwhelmed by the conflicting information available. It is essential that researchers using the technique are comfortable with the area since assessing whether a specified model 'fits' the data is one of the most important steps in structural equation modelling (Yuan, 2005). This has spurred decades of intense debate and research dedicated to this pertinent area. Indeed, ever since structural equation modelling was first developed, statisticians have sought and developed new and improved indices that reflect some facet of model fit previously not accounted for. Having such a collection of indices entices a researcher to select those that indicate good model fit. This practice should be desisted at all costs as it masks underlying problems that suggest possible misspecifications within the model.

This paper seeks to introduce a variety of fit indices which can be used as a guideline for prospective structural equation modellers to help them avoid making such errors. To clarify matters to users of SEM, the most widely respected and reported fit indices are covered here and their interpretive value in assessing model fit is examined. In addition to this, best practice on reporting structural equation modelling is discussed and suggests some ways in which model fit can be improved. In recent years the very area of fit indices has come under serious scrutiny with some authors calling for their complete abolishment (Barrett, 2007). While such a drastic change is unlikely to occur any time soon, the shortcomings of having stringent thresholds is becoming more topical within the field of structural equation modelling (Kenny and McCoach, 2003; Marsh et al, 2004).

2. Absolute fit indices

Absolute fit indices determine how well an a priori model fits the sample data (McDonald and Ho, 2002) and demonstrates which proposed model has the most superior fit. These measures provide the most fundamental indication of how well the proposed theory fits the data. Unlike incremental fit indices, their calculation does not rely on comparison with a baseline model but is instead a measure of how well the model fits in comparison to no model at all (Jöreskog and Sörbom, 1993). Included in this category are the Chi-Squared test, RMSEA, GFI, AGFI, the RMR and the SRMR.

2.1 Model chi-square (χ^2)

The Chi-Square value is the traditional measure for evaluating overall model fit and, 'assesses the magnitude of discrepancy between the sample and fitted covariances matrices' (Hu and Bentler, 1999: 2). A good model fit would provide an insignificant result at a 0.05 threshold (Barrett, 2007), thus the Chi-Square statistic is often referred to as either a 'badness of fit' (Kline, 2005) or a 'lack of fit' (Mulaik et al, 1989) measure. While the Chi-Squared test retains its popularity as a fit statistic, there exist a number of severe

Reference this paper as:
Hooper, D., Coughlan, J. and Mullen, M. R. "Structural Equation Modelling: Guidelines for Determining Model Fit." *The Electronic Journal of Business Research Methods* Volume 6 Issue 1 2008, pp. 53 - 60, available online at www.ejbrm.com

limitations in its use. Firstly, this test assumes multivariate normality and severe deviations from normality may result in model rejections even when the model is properly specified (McIntosh, 2006). Secondly, because the Chi-Square statistic is in essence a statistical significance test it is sensitive to sample size which means that the Chi-Square statistic nearly always rejects the model when large samples are used (Bentler and Bonnet, 1980; Jöreskog and Sörbom, 1993). On the other hand, where small samples are used, the Chi-Square statistic lacks power and because of this may not discriminate between good fitting models and poor fitting models (Kenny and McCoach, 2003). Due to the restrictiveness of the Model Chi-Square, researchers have sought alternative indices to assess model fit. One example of a statistic that minimises the impact of sample size on the Model Chi-Square is Wheaton et al's (1977) relative/normed chi-square ($\chi 2/df$). Although there is no consensus regarding an acceptable ratio for this statistic, recommendations range from as high as 5.0 (Wheaton et al, 1977) to as low as 2.0 (Tabachnick and Fidell, 2007).

2.2 Root mean square error of approximation (RMSEA)

The RMSEA is the second fit statistic reported in the LISREL program and was first developed by Steiger and Lind (1980, cited in Steiger, 1990). The RMSEA tells us how well the model, with unknown but optimally chosen parameter estimates would fit the populations covariance matrix (Byrne, 1998). In recent years it has become regarded as 'one of the most informative fit indices' (Diamantopoulos and Siguaw, 2000: 85) due to its sensitivity to the number of estimated parameters in the model. In other words, the RMSEA favours parsimony in that it will choose the model with the lesser number of parameters. Recommendations for RMSEA cut-off points have been reduced considerably in the last fifteen years. Up until the early nineties, an RMSEA in the range of 0.05 to 0.10 was considered an indication of fair fit and values above 0.10 indicated poor fit (MacCallum et al, 1996). It was then thought that an RMSEA of between 0.08 to 0.10 provides a mediocre fit and below 0.08 shows a good fit (MacCallum et al, 1996). However, more recently, a cut-off value close to .06 (Hu and Bentler, 1999) or a stringent upper limit of 0.07 (Steiger, 2007) seems to be the general consensus amongst authorities in this area.

One of the greatest advantages of the RMSEA is its ability for a confidence interval to be calculated around its value (MacCallum et al, 1996). This is possible due to the known distribution values of the statistic and subsequently allows for the null hypothesis (poor fit) to be tested more precisely (McQuitty, 2004). It is generally reported in conjunction with the RMSEA and in a well-fitting model the lower limit is close to 0 while the upper limit should be less than 0.08.

2.3 Goodness-of-fit statistic (GFI) and the adjusted goodness-of-fit statistic (AGFI)

The Goodness-of-Fit statistic (GFI) was created by Jöreskog and Sorbom as an alternative to the Chi-Square test and calculates the proportion of variance that is accounted for by the estimated population covariance (Tabachnick and Fidell, 2007). By looking at the variances and covariances accounted for by the model it shows how closely the model comes to replicating the observed covariance matrix (Diamantopoulos and Siguaw, 2000). This statistic ranges from 0 to 1 with larger samples increasing its value. When there are a large number of degrees of freedom in comparison to sample size, the GFI has a downward bias (Sharma et al, 2005). In addition, it has also been found that the GFI increases as the number of parameters increases (MacCallum and Hong, 1997) and also has an upward bias with large samples (Bollen, 1990; Miles and Shevlin, 1998). Traditionally an omnibus cut-off point of 0.90 has been recommended for the GFI however, simulation studies have shown that when factor loadings and sample sizes are low a higher cut-off of 0.95 is more appropriate (Miles and Shevlin, 1998). Given the sensitivity of this index, it has become less popular in recent years and it has even been recommended that this index should not be used (Sharma et al, 2005). Related to the GFI is the AGFI which adjusts the GFI based upon degrees of freedom, with more saturated models reducing fit (Tabachnick and Fidell, 2007). Thus, more parsimonious models are preferred while penalised for complicated models. In addition to this, AGFI tends to increase with sample size. As with the GFI, values for the AGFI also range between 0 and 1 and it is generally accepted that values of 0.90 or greater indicate well fitting models. Given the often detrimental effect of sample size on these two fit indices they are not relied upon as a stand alone index, however given their historical importance they are often reported in covariance structure analyses.

2.4 Root mean square residual (RMR) and standardised root mean square residual (SRMR)

The RMR and the SRMR are the square root of the difference between the residuals of the sample covariance matrix and the hypothesised covariance model. The range of the RMR is calculated based upon the scales of each indicator, therefore, if a questionnaire contains items with varying levels (some items may range from 1 – 5 while others range from 1 – 7) the RMR becomes difficult to interpret (Kline, 2005). The

standardised RMR (SRMR) resolves this problem and is therefore much more meaningful to interpret. Values for the SRMR range from zero to 1.0 with well fitting models obtaining values less than .05 (Byrne, 1998; Diamantopoulos and Siguaw, 2000), however values as high as 0.08 are deemed acceptable (Hu and Bentler, 1999). An SRMR of 0 indicates perfect fit but it must be noted that SRMR will be lower when there is a high number of parameters in the model and in models based on large sample sizes.

3. Incremental fit indices

Incremental fit indices, also known as comparative (Miles and Shevlin, 2007) or relative fit indices (McDonald and Ho, 2002), are a group of indices that do not use the chi-square in its raw form but compare the chi-square value to a baseline model. For these models the null hypothesis is that all variables are uncorrelated (McDonald and Ho, 2002).

3.1 Normed-fit index (NFI)

The first of these indices to appear in LISREL output is the Normed Fit Index (NFI: Bentler and Bonnet, 1980). This statistic assesses the model by comparing the $\chi 2$ value of the model to the $\chi 2$ of the null model. The null/independence model is the worst case scenario as it specifies that all measured variables are uncorrelated. Values for this statistic range between 0 and 1 with Bentler and Bonnet (1980) recommending values greater than 0.90 indicating a good fit. More recent suggestions state that the cut-off criteria should be NFI ≥ .95 (Hu and Bentler, 1999). A major drawback to this index is that it is sensitive to sample size, underestimating fit for samples less than 200 (Mulaik et al, 1989; Bentler, 1990), and is thus not recommended to be solely relied on (Kline, 2005). This problem was rectified by the Non-Normed Fit Index (NNFI, also known as the Tucker-Lewis index), an index that prefers simpler models. However in situations were small samples are used, the value of the NNFI can indicate poor fit despite other statistics pointing towards good fit (Bentler, 1990; Kline, 2005; Tabachnick and Fidell, 2007). A final problem with the NNFI is that due to its non-normed nature, values can go above 1.0 and can thus be difficult to interpret (Byrne, 1998). Recommendations as low as 0.80 as a cutoff have been proferred however Bentler and Hu (1999) have suggested NNFI ≥ 0.95 as the threshold.

3.2 CFI (Comparative fit index)

The Comparative Fit Index (CFI: Bentler, 1990) is a revised form of the NFI which takes into account sample size (Byrne, 1998) that performs well even when sample size is small (Tabachnick and Fidell, 2007). This index was first introduced by Bentler (1990) and subsequently included as part of the fit indices in his EQS program (Kline, 2005). Like the NFI, this statistic assumes that all latent variables are uncorrelated (null/independence model) and compares the sample covariance matrix with this null model. As with the NFI, values for this statistic range between 0.0 and 1.0 with values closer to 1.0 indicating good fit. A cut-off criterion of CFI ≥ 0.90 was initially advanced however, recent studies have shown that a value greater than 0.90 is needed in order to ensure that misspecified models are not accepted (Hu and Bentler, 1999). From this, a value of CFI ≥ 0.95 is presently recognised as indicative of good fit (Hu and Bentler, 1999). Today this index is included in all SEM programs and is one of the most popularly reported fit indices due to being one of the measures least effected by sample size (Fan et al, 1999).

4. Parsimony fit indices

Having a nearly saturated, complex model means that the estimation process is dependent on the sample data. This results in a less rigourous theoretical model that paradoxically produces better fit indices (Mulaik et al, 1989; Crowley and Fan, 1997). To overcome this problem, Mulaik et al (1989) have developed two parsimony of fit indices; the Parsimony Goodness-of-Fit Index (PGFI) and the Parsimonious Normed Fit Index (PNFI). The PGFI is based upon the GFI by adjusting for loss of degrees of freedom. The PNFI also adjusts for degrees of freedom however it is based on the NFI (Mulaik et al 1989). Both of these indices seriously penalise for model complexity which results in parsimony fit index values that are considerably lower than other goodness of fit indices. While no threshold levels have been recommended for these indices, Mulaik et al (1989) do note that it is possible to obtain parsimony fit indices within the .50 region while other goodness of fit indices achieve values over .90 (Mulaik et al 1989). The authors strongly recommend the use of parsimony fit indices in tandem with other measures of goodness-of-fit however, because no threshold levels for these statistics have been recommended it has made them more difficult to interpret.

A second form of parsimony fit index are those that are also known as 'information criteria' indices. Probably the best known of these indices is the Akaike Information Criterion (AIC) or the Consistent Version of AIC (CAIC) which adjusts for sample size (Akaike, 1974). These statistics are generally used when comparing non-nested or non-hierarchical models estimated with the same data and indicates to the researcher which of the models is the most parsimonious. Smaller values suggest a good fitting, parsimonious model however because these indices are not normed to a 0-1 scale it is difficult to suggest a cut-off other than that the model that produces the lowest value is the most superior. It is also worth noting that these statistics need a sample size of 200 to make their use reliable (Diamantopoulos and Siguaw, 2000).

5. Reporting fit indices

With regards to which indices should be reported, it is not necessary or realistic to include every index included in the program's output as it will burden both a reader and a reviewer. Given the plethora of fit indices, it becomes a temptation to choose those fit indices that indicate the best fit (see Appendix A for a summary of some key indices discussed herein). This should be avoided at all costs as it is essentially sweeping important information under the carpet. In a review by McDonald and Ho (2002) it was found that the most commonly reported fit indices are the CFI, GFI, NFI and the NNFI. When deciding what indices to report, going by what is most frequently used is not necessarily good practice as some of these statistics (such as the GFI discussed above) are often relied on purely for historical reasons, rather than for their sophistication. While there are no golden rules for assessment of model fit, reporting a variety of indices is necessary (Crowley and Fan 1997) because different indices reflect a different aspect of model fit. Although the Model Chi-Square has many problems associated with it, it is still essential that this statistic, along with its degrees of freedom and associated p value, should at all times reported (Kline, 2005; Hayduk et al, 2007). Threshold levels were recently assessed by Hu and Bentler (1999) who suggested a two-index presentation format. This always includes the SRMR with the NNFI (TLI), RMSEA or the CFI. The various combinations are summarised in Appendix B below. Kline (2005) speaks strongly about which indices to include and advocates the use of the Chi-Square test, the RMSEA, the CFI and the SRMR. Boomsma (2000) has similar recommendations but also advises for the squared multiple correlations of each equation to be reported. Based on these authors guidelines and the above review it is sensible to include the Chi-Square statistic, its degrees of freedom and p value, the RMSEA and its associated confidence interval, the SRMR, the CFI and one parsimony fit index such as the PNFI. These indices have been chosen over other indices as they have been found to be the most insensitive to sample size, model misspecification and parameter estimates.

6. How to improve model fit

Given the complexity of structural equation modelling, it is not uncommon to find that the fit of a proposed model is poor. Allowing modification indices to drive the process is a dangerous game, however, some modifications can be made locally that can substantially improve results. It is good practice to assess the fit of each construct and its items individually to determine whether there are any items that are particularly weak. Items with low multiple r2 (less than .20) should be removed from the analysis as this is an indication of very high levels of error. Following this, each construct should be modelled in conjunction with every other construct in the model to determine whether discriminant validity has been achieved. The Phi (φ) value between two constructs is akin to their covariance, therefore a Phi of 1.0 indicates that the two constructs are measuring the same thing. One test which is useful to determine whether constructs are significantly different is Bagozzi et al's (1991) discriminant validity test. The formula for this is: parameter estimate (phi value) ±1.96 * standard error. If the value is greater than 1.0 discriminant validity has not been achieved and further inspections of item cross-loadings need to be made. Items with high Lambda-Y modification indices are possible candidates for deletion and are likely to be causing the discriminant validity problem. By deleting indiscriminant items fit is likely to improve and is advantageous in that it is unlikely to have any major theoretical repercussions.

A further way in which fit can be improved is through the correlation of error terms. This practice is generally frowned upon (Gerbing and Anderson, 1984) as it means that there is some other issue that is not specified within the model that is causing the covariation. If a researcher decides to correlate error terms there needs to be strong theoretical justification behind such a move (Jöreskog and Long, 1993). Correlating within-factor error is easier to justify than across latent variable correlations, however it is essential that the statistical and substantive impact are clearly discussed. If a researcher feels they can substantiate this decision, correlated error terms is acceptable, however it is a step that should be taken with caution.

7. The future for fit indices

As a final point it must be noted that while fit indices are a useful guide, a structural model should also be examined with respect to substantive theory. By allowing model fit to drive the research process it moves away from the original, theory-testing purpose of structural equation modelling. In addition, fit indices may point to a well-fitting model when in actual fact, parts of the model may fit poorly (Jöreskog and Sörbom, 1996; Tomarken and Waller, 2003; Reisinger and Mavondo, 2006). Indeed, the area of fit indices 'rules of thumb' is highly topical at the moment with some experts in the area calling for a complete abandonment of fit indices altogether (Barrett, 2007). Others are less hesitant to denounce their usefulness but do agree that strictly adhering to recommended cutoff values can lead to instances of Type I error (the incorrect rejection of an acceptable model) (Marsh et al, 2004). Although the debate is ongoing, it is unlikely that fit indices will become defunct any time soon and for this reason the above literature review remains pertinent to those using SEM.

References

Akaike, H. (1974), "A New Look at the Statistical Model Identification," IEE Transactions on Automatic Control, 19 (6), 716-23.

Bagozzi, R.P., Yi, Y., and Phillips, L.W. (1991), "Assessing Construct Validity in Organizational Research," Adminstrative Science Quarterly, 36 (3), 421-58.

Barrett, P. (2007), "Structural Equation Modelling: Adjudging Model Fit," Personality and Individual Differences, 42 (5), 815-24.

Bentler, P.M. (1990), "Comparative Fit Indexes in Structural Models," Psychological Bulletin, 107 (2), 238-46.

Bentler, P.M. and Bonnet, D.C. (1980), "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures," Psychological Bulletin, 88 (3), 588-606.

Bollen, K.A. (1990), "Overall Fit in Covariance Structure Models: Two Types of Sample Size Effects," Psychological Bulletin, 107 (2), 256-59.

Boomsma, A. (2000), "Reporting Analyses of Covariance Structures," Structural Equation Modeling, 7 (3), 461-83.

Byrne, B.M. (1998), Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Crowley, S.L. and Fan, X. (1997), "Structural Equation Modeling: Basic Concepts and Applications in Personality Assessment Research," Journal of Personality Assessment, 68 (3), 508-31.

Diamantopoulos, A. and Siguaw, J.A. (2000), Introducing LISREL. London: Sage Publications.

Fan, X., Thompson, B., and Wang, L. (1999), "Effects of Sample Size, Estimation Methods, and Model Specification on Structural Equation Modeling Fit Indexes," Structural Equation Modeling, 6 (1), 56-83.

Gerbing, D.W. and Anderson, J.C. (1984), "On the Meaning of Within-Factor Correlated Measurement Errors," Journal of Consumer Research, 11 (June), 572-80.

Hayduk, L., Cummings, G.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 (2), 841-50.

Hu, L.T. and Bentler, P.M. (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives," Structural Equation Modeling, 6 (1), 1-55.

Jöreskog, K. and Long, J.S. (1993), "Introduction," in Testing Structural Equation Models, Kenneth A. Bollen and J. Scott Long, Eds. Newbury Park, CA: Sage.

Jöreskog, K. and Sörbom, D. (1993), LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Chicago, IL: Scientific Software International Inc.

Jöreskog, K. and Sörbom, D. (1996), LISREL 8: User's Reference Guide. Chicago, IL: Scientific Software International Inc.

Kenny, D.A. and McCoach, D.B. (2003), "Effect of the Number of Variables on Measures of Fit in Structural Equation Modeling," Structural Equation Modeling, 10 (3), 333-51.

Kline, R.B. (2005), Principles and Practice of Structural Equation Modeling (2nd Edition ed.). New York: The Guilford Press.

MacCallum, R.C., Browne, M.W., and Sugawara, H., M. (1996), "Power Analysis and Determination of Sample Size for Covariance Structure Modeling," Psychological Methods, 1 (2), 130-49.

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 Findings " Structural Equation Modeling, 11 (3), 320-41.

McDonald, R.P. and Ho, M.-H.R. (2002), "Principles and Practice in Reporting Statistical Equation Analyses," Psychological Methods, 7 (1), 64-82.

McIntosh, C. (2006), "Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007)," Personality and Individual Differences, 42 (5), 859-67.

McQuitty, S. (2004), "Statistical power and structural equation models in business research," Journal of Business Research, 57 (2), 175-83.

Miles, J. and Shevlin, M. (1998), "Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis," Personality and Individual Differences, 25, 85-90.

Miles, J. and Shevlin, M. (2007), "A time and a place for incremental fit indices," Personality and Individual Differences, 42 (5), 869-74.

Mulaik, S.A., James, L.R., Van Alstine, J., Bennet, N., Lind, S., and Stilwell, C.D. (1989), "Evaluation of Goodness-of-Fit Indices for Structural Equation Models," Psychological Bulletin, 105 (3), 430-45.
Reisinger, Y. and Mavondo, F. (2006), "Structural Equation Modeling: Critical Issues and New Developments," Journal of Travel and Tourism Marketing, 21 (4), 41-71.
Sharma, S., Mukherjee, S., Kumar, A., and Dillon, W.R. (2005), "A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models," Journal of Business Research, 58 (1), 935-43.
Steiger, J.H. (1990), "Structural model evaluation and modification," Multivariate Behavioral Research, 25, 214-12.
Steiger, J.H. (2007), "Understanding the limitations of global fit assessment in structural equation modeling," Personality and Individual Differences, 42 (5), 893-98.
Tabachnick, B.G. and Fidell, L.S. (2007), Using Multivariate Statistics (5th ed.). New York: Allyn and Bacon.
Tomarken, A.J. and Waller, N.G. (2003), "Potential Problems With "Well Fitting" Models," Journal of Abnormal Psychology, 112 (4), 578-98.
Wheaton, B., Muthen, B., Alwin, D., F., and Summers, G. (1977), "Assessing Reliability and Stability in Panel Models," Sociological Methodology, 8 (1), 84-136.
Yuan, K.H. (2005), "Fit Indices Versus Test Statistics," Multivariate Behavioral Research, 40 (1), 115-48.

Appendix A

Table 1: Fit indices and their acceptable thresholds

Fit Index	Acceptable Threshold Levels	Description
Absolute Fit Indices		
Chi-Square χ^2	Low χ^2 relative to degrees of freedom with an insignificant *p* value ($p > 0.05$)	
Relative χ^2 (χ^2/df)	2:1 (Tabachnik and Fidell, 2007) 3:1 (Kline, 2005)	Adjusts for sample size.
Root Mean Square Error of Approximation (RMSEA)	Values less than 0.07 (Steiger, 2007)	Has a known distribution. Favours parsimony. Values less than 0.03 represent excellent fit.
GFI	Values greater than 0.95	Scaled between 0 and 1, with higher values indicating better model fit. This statistic should be used with caution.
AGFI	Values greater than 0.95	Adjusts the GFI based on the number of parameters in the model. Values can fall outside the 0-1.0 range.
RMR	Good models have small RMR (Tabachnik and Fidell, 2007)	Residual based. The average squared differences between the residuals of the sample covariances and the residuals of the estimated covariances. Unstandardised.
SRMR	SRMR less than 0.08 (Hu and Bentler, 1999)	Standardised version of the RMR. Easier to interpret due to its standardised nature.
Incremental Fit Indices		
NFI	Values greater than 0.95	Assesses fit relative to a baseline model which assumes no covariances between the observed variables. Has a tendency to overestimate fit in small samples.
NNFI (TLI)	Values greater than 0.95	Non-normed, values can fall outside the 0-1 range. Favours parsimony. Performs well in simulation studies (Sharma et al, 2005; McDonald and Marsh, 1990)
CFI	Values greater than 0.95	Normed, 0-1 range.

Appendix B

Table 2: Hu and Bentler's Two-Index Presentation Strategy (1999)

Fit Index Combination	Combinational Rules
NNFI (TLI) and SRMR	NNFI of 0.96 or higher and an SRMR of .09 or lower
RMSEA and SRMR	RMSEA of 0.06 or lower and a SRMR of 0.09 or lower
CFI and SRMR	CFI of .96 or higher and a SRMR of 0.09 or lower

