Abstract

Structural Equation Modeling (SEM) techniques represent a powerful set of statistical methods for evaluating a variety of hypotheses in developmental research. The first purpose of this project was to determine the extent to which applied researchers are taking advantage of this analytic method. A review of the all published articles appearing in Child Development between 1993 to 1997 identified 41 articles (6%) that utilized SEM methods. The second purpose of this project was to determine how SEM methods were currently being utilized in Child Development. A summary of current reporting practices indicated that researchers need to communicate more information to the reader about their modeling strategy and rationale for evaluating model fit. The third purpose of this project was to examine whether a commonly used “rule of thumb” (i.e., fit indices > .90) regarding the adequacy of global model fit was empirically justified—particularly under conditions commonly encountered in developmental research. The results of a Monte Carlo simulation study strongly refute the adequacy of this criterion. We concluded that applied researchers should rely on multiple sources of information when evaluating and reporting model fit.

Copies of this poster can be downloaded from our project website:
http://www.duke.edu/~curran/remora/remora.html
**Recommended references**

**General Overviews of SEM**

**Non-normality**

**Global fit indices**

**Local fit indices**
Alternative Methods for Assessing the Fit of Structural Equation Models in Developmental Research

Michael T. Willoughby, B.S.

&

Patrick J. Curran, Ph.D.

Duke University
ABSTRACT

The first purpose of this project was to determine the extent to which applied researchers are taking advantage of structural equation modeling (SEM) methods. A review of the all published articles appearing in Child Development between 1993 to 1997 identified 41 articles (6%) that utilized SEM methods. The second purpose of this project was to determine how SEM methods were currently being utilized in Child Development. A summary of current reporting practices indicated that researchers need to communicate more information to the reader about their modeling strategy and rationale for evaluating model fit. The third purpose of this project was to examine whether a commonly used “rule of thumb” (i.e., fit indices > .90) regarding the adequacy of global model fit was empirically justified—particularly under conditions commonly encountered in developmental research. The results of a Monte Carlo simulation study strongly refute the adequacy of this criterion. We concluded that applied researchers should rely on multiple sources of information when evaluating and reporting model fit.
INTRODUCTION

Structural Equation Modeling (SEM) techniques represent a powerful set of statistical methods for evaluating a variety of hypotheses in developmental research. Some of the advantages of SEM methods include the examination of latent variable relationships that are corrected for measurement error, the estimation of both direct and indirect relationships, and the provision of indices of global model fit.

Over a decade ago, the journal Child Development published a special section on SEM methods to encourage developmental researchers to begin utilizing these methods. The current project was designed to answer three questions regarding the application of these methods in the Journal:

(1) How widely used are SEM methods in Child Development?
(2) How are SEM methods currently being utilized?
(3) Are current practices regarding the assessment of global model fit sensitive to issues that typically characterize developmental data (i.e., small samples, non-normal data).
METHODS

Questions 1 and 2 were addressed using a literature review. All published articles appearing in Child Development between 1993 to 1997 were examined. In order to be included in the current analysis, articles needed to report a measure of global model fit (i.e., $X^2$ test statistic, goodness of fit index). A variety of information was coded from the methods and results sections of articles that were included in the review.

Question 3 was addressed by way of a Monte Carlo simulation study. Specifically, a moderately misspecified oblique three-factor measurement model with three indicators per factor was estimated 200 times for each of four different sized samples (100, 200, 500, 1000) and two different distributions (normal, moderately non-normal). The model was considered misspecified because it omitted two factor loadings that existed in the population model. The sample size and distributional conditions reflected those typical in developmental research. The Comparative Fit Index (CFI) served as an exemplar fit index. Of interest was the proportion of misspecified models that were correctly rejected using CFI values that ranged from .90 to 1.00. This proportion is an estimate of empirical power.
RESULTS

**Question 1: How widely used are SEM methods in Child Development?**

A total of 41 articles were identified that utilized SEM methods. This represented 6% (41/659) of all published articles during this time period. Since behavioral-genetics studies utilize non-traditional SEM methods, five of these studies were excluded. Three additional studies were excluded because they utilized partial least squares estimation methods that do not provide relevant indices of global fit. In total, 33 studies were available for review.

**Question 2: How are SEM methods being utilized?**

Table 1. Model types

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Percent</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFA - only</td>
<td>27</td>
<td>9</td>
</tr>
<tr>
<td>SEM - only</td>
<td>52</td>
<td>17</td>
</tr>
<tr>
<td>CFA &amp; SEM</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Path Analysis – only</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

SEM refers to structural equation models, where latent variables influence each other. CFA refers to confirmatory factor analyses, where latent variables are allowed to correlate. Path analyses had to be estimated simultaneously and provide a relevant index of global fit to be included.
Table 2. Estimation methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Percent</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Likelihood (ML)</td>
<td>27</td>
<td>9</td>
</tr>
<tr>
<td>Robust ML</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>WLS or ADF</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Not reported</td>
<td>58</td>
<td>19</td>
</tr>
</tbody>
</table>

WLS = Weighted Least Squares; ADF = Asymptotic Distribution Free

Table 3. Information provided regarding global model fit

<table>
<thead>
<tr>
<th>Fit Index &amp; $X^2$ test statistic &amp; p value</th>
<th>Percent</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit Index &amp; $X^2$ test statistic – only</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>Fit Index – only</td>
<td>12</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 4. Information provided regarding local model fit

<table>
<thead>
<tr>
<th>Local Fit Measure</th>
<th>Percent</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagrange Multiplier (LM) test</td>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>LM &amp; Standardized Residual</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Information is not reported (or unclear)</td>
<td>70</td>
<td>23</td>
</tr>
</tbody>
</table>

The Lagrange Multiplier test is also referred to as a modification index in the LISREL software package.
Table 5. Criteria for global model fit

<table>
<thead>
<tr>
<th>Were criteria offered to imply good model fit?</th>
<th>Response</th>
<th>Percent</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td></td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>If so, what were these criteria?</td>
<td>FI &gt; .90</td>
<td>35</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Other criteria</td>
<td>35</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>FI &gt; .90 &amp; Other criteria</td>
<td>30</td>
<td>7</td>
</tr>
</tbody>
</table>

FI > .90 refers to the widely cited rule of thumb that fit indices that exceed .90 in value are indicative of good model fit. Other criteria included testing differences between nested models using the $X^2$ difference test and other rules of thumb (e.g., RMSEA < .05 or .08; $X^2$/df < 2).

Finally, only 18% of the studies that were reviewed reported examining distributional assumptions. However, 83% of these studies noted that some type of correction was necessary (e.g., power transformation, alternative estimation method). Given that the default estimation method for most software packages is maximum likelihood, which assumes multivariate normality, it is imperative that researchers examine the distributions of their data.
**Question 3: Are current practices regarding the assessment of global model fit sensitive to issues that characterize developmental data?**

Given that there is a strong reliance on fit indices > .90 to help distinguish good from poor fitting models, a simulation study was undertaken to empirically address this practice. Figure 1 depicts the target population model. This model can be considered an exemplar model, as 42% of the studies utilizing SEM methods over the past five years have included a confirmatory factor analysis. Figures 2 and 3 depict power curves for the Comparative Fit Index (CFI) under conditions commonly encountered in developmental research, namely moderately misspecified models with moderately non-normal data. Two important points emerge from these figures. First, the empirical power of the CFI to appropriately reject misspecified models clearly varies as a function of sample size and non-normality indicating that use of a “global” cutoff is not well informed. Second, the criterion of fit indices > .90 has exceedingly low power (80 – 90% of misspecified models were deemed to fit the data well).
Figure 1. Simulation model: 3 factor, 9 indicator CFA with 2 omitted cross loadings

Note: Dashed lines represent factor loadings that were omitted in the simulated model.
Figure 2. Power curves for the Comparative Fit Index (CFI) evaluating a **moderately misspecified** model with **normally distributed** indicator variables across four sample sizes.
Figure 3. Power curves for the Comparative Fit Index (CFI) evaluating a **moderately misspecified** model with **moderately non-normally distributed** indicator variables across four sample sizes.
SUMMARY & RECOMMENDATIONS

1. Only 6% of all published manuscripts in Child Development between 1993 – 1997 utilized SEM. Given the advantages of SEM: (A) the simultaneous estimation of equations, (B) the examination of latent variable relationships, (C) the estimation of both direct and indirect relationships, (D) the provision of indices of global model fit, and (E) the ability to estimate latent growth curves – we urge applied researchers to consider using these techniques in their own research.

2. Improvements need to be made in current reporting practices
   • 70% of articles did not clarify whether information about local model fit was evaluated and/or utilized in model respecification.
   • 30% of articles failed to report even minimally necessary information regarding global model fit (i.e., $X^2$, p value, GFI).
   • 30% of articles did not provide any explicit criteria to justify the adequacy of their model fit.

Better reporting practices will facilitate the ability of readers and reviewers to critically evaluate studies utilizing SEM methods.
3. Given that the maximum likelihood (ML) estimation method is among the most widely utilized and is the default method on most software packages, it is imperative that applied researchers evaluate whether their data meet the strict assumption of multivariate normality. Although only 18% of studies reported evaluating this assumption, 83% of these studies failed to meet this assumption and utilized corrective action. Researchers should consider utilizing corrective techniques in the analysis of non-normal data (e.g., robust ML estimation; boot-strapping).

4. Regarding global fit, 65% of studies explicitly mentioned relying on a fit index > .90. Simulation data demonstrated that the adequacy of this criterion is not empirically justifiable. The exemplar fit index varied as a function of sample size and the distribution of indicator variables. The determination of adequate, global model fit is not a decision that can be answered using any single source of information. Rather, researchers should consult multiple indices of global and local fit, test nested models, and evaluate the plausibility of estimated parameters to inform this decision.