Adequacy of Model Fit in Confirmatory Factor Analysis and Structural Equation Models: It Depends on What Software You Use

Susan R. Hutchinson University of Northern Colorado
Antonio Olmos   University of Denver
Eric Teman   University of Northern Colorado

American Evaluation Association Conference
Washington, DC October 2013

SEM and CFA in Evaluation

- Structural equation modeling (SEM) and confirmatory factor analysis (CFA) are long-term favorites among evaluators
  - However the requirement of large N (> 200) may be one reason these procedures have not been used more widely in evaluation

- Applications include:
  - Assessing validity of scores on self-report instruments
  - Testing models of factors affecting program outcomes
  - Evaluating patterns of change following interventions

Benefits of SEM and CFA

- CFA offers a more theory-based approach to examining the factor structure of self-report measures than exploratory factor analysis
  - Provides stronger validity evidence than EFA
  - Allows testing of equality of factor structure across groups

- SEM allows testing of factors or variables that both directly and indirectly affect outcomes
  - And SEM does a better job of accounting for measurement error than other statistical procedures → offers greater power

Assessing Model Fit

- Of major concern to those who use SEM/CFA is the extent to which their model "fits" the data

- Evidence of fit includes:

- Component fit
  - Related to plausibility of the solution such as the strength of factor loadings, the direction of path coefficients, and the statistical significance of hypothesized relationships

- Global fit
  - Indicates how well, overall, the model fits the data
  - Based on various fit indices

Research on Fit

- The research literature on global model fit is extensive
- Studies (mostly based on simulations) have addressed numerous factors affecting fit indices including:
  - sample size
  - distributional characteristics of the data
  - estimation procedures
  - model size
- One of the unresolved issues is the usefulness of specific cutoff criteria for judging model adequacy
  - Limited by sensitivity of fit indices to many factors

Motivation for the Presentation

- Informally observed differences in fit for models run in different software but based on identical models and data
  - In turn, raised questions about how widespread is this problem and across which SEM programs

- Relatively little is provided in the SEM research literature about potential discrepancies in model fit depending upon the software
  - Despite differences among software in estimation procedures and algorithms for computing fit indices

Purpose of the Presentation

- Illustrate differences (and similarities) in model fit:
  - across several different SEM software programs and versions
  - based on both CFA and full structural models
  - using both normal theory and robust estimators
- Describe salient software differences affecting the fit indices
- Discuss implications regarding choice of software and estimation procedure

Software Used

- EQS (version 6.2)
- LISREL (versions 8.7, 8.8, and 9.1)
- M*plus* (versions 5, 6, and 7)
- R lavaan (version .5-14, R 3.0.2)
- Originally planned to include SAS Proc Calis (SAS 9.3) but decided to drop it due to limited estimation options
- AMOS was not included

Models Examined

- Model 1: One-factor CFA (of faculty support) with 8 indicator variables (all ordinal)
- Model 2: Two-factor CFA (of students' experiences with fair/unfair treatment) with 15 indicator variables (all ordinal)
- Model 3: Full structural model (of students' enrollment satisfaction) with four latent variables and 15 indicator variables (mix of continuous and ordinal indicators)
- Model 4: Full structural model with four latent variables and 26 indicator variables (all ordinal)

Model Estimation

- Estimation procedures:
  - Standard maximum likelihood (ML) estimation (all models and all software)
  - Robust estimation for analysis of ordinal data (differed depending on availability by software)
    - **LISREL** (Robust ML with Satorra-Bentler rescaling)
    - **EQS** (Robust ML with Satorra-Bentler rescaling; but based on slightly different approach than LISREL's robust ML)
    - **M*plus*** and **R lavaan** (WLSMV which is a type of robust WLS)
- All analyses based on listwise deletion of missing data
- Only reported initial model fit; no post-hoc model modifications

Popular fit criteria

- Based on Hu & Bentler, 1999
- Values are considered as an indication of good fit:
- CFI: Values of 0.95 or greater
- RMSEA: Values of 0.05 or less
  - Include Confidence intervals with a lower bound of 0 and a higher bound no larger than 0.08
    - Size of the CI is also indicative of good fit
- TLI: Values of 0.95 or higher
- SRMR: Values of 0.05

Results – ML Estimation (Model 1)

	LISREL 8.7/8.8	LISREL 9.1	M*plus* 5	M*plus* 6/7	EQS 6.2	R lavaan
LR chi-squared (χ^2)	**524.393**	**524.631**	**524.636**	**524.636**	**524.399**	**524.636**

All based on N = 2203 and 20 degrees of freedom

Results – ML (Model 1) cont'd

	LISREL 8.7/8.8	LISREL 9.1	M*plus* 5	M*plus* 6/7	EQS 6.2	R lavaan
LR chi-squared	524.393	524.631	524.636	524.636	524.399	524.636
CFI	.947	.947	.905	.905	.906	.905
TLI	.926	.926	.868	.868	.868	.868
RMSEA	.113	.107	.107	.107	.107	.107
SRMR	.052	.052	.047	.047	.052	.052

Results – Robust Estimation (Model 1)

	LISREL 8.7/8.8	LISREL 9.1	*Mplus* 5	*Mplus* 6/7	EQS 6.2	R lavaan
Chi-squared	368.382	325.769	760.635	854.824	389.717	854.347
CFI	.967	.979	.896	.936	.965	.936
TLI	.977	.971	.942	.910	.951	.910
RMSEA	.089	.149	.145	.138	.092	.138
SRMR	.061	.061	-----	-----	-----	-----

Note: Degrees of freedom for M*plus* 5 = 16; all others = 20

Results – ML Estimation (Model 3)

	LISREL 8.7/8.8	LISREL 9.1	*Mplus* 5	*Mplus* 6/7	EQS 6.2	R lavaan
LR chi-squared	**500.213**	**500.618**	**500.618**	**500.618**	**500.215**	**500.618**
CFI	**.957**	**.957**	**.917**	**.917**	**.917**	**.917**
TLI	**.947**	**.947**	**.899**	**.899**	**.899**	**.899**
RMSEA	**.062**	**.063**	**.062**	**.062**	**.062**	**.062**
SRMR	**.051**	**.051**	**.048**	**.048**	**.051**	**.051**

Analyses based on N =1236 and 86 degrees of freedom

Results – ML Estimation (Model 3)

	LISREL 8.7/8.8	LISREL 9.1	*Mplus* 5	*Mplus* 6/7	EQS 6.2	R lavaan
LR chi-squared	**500.213**	**500.618**	**500.618**	**500.618**	**500.215**	**500.618**
CFI	**.957**	**.957**	**.917**	**.917**	**.917**	**.917**
TLI	**.947**	**.947**	**.899**	**.899**	**.899**	**.899**
RMSEA	**.062**	**.063**	**.062**	**.062**	**.062**	**.062**
SRMR	**.051**	**.051**	**.048**	**.048**	**.051**	**.051**

Analyses based on *N* =1236 and 86 degrees of freedom

Results – Robust Estimation (Model 3)

	LISREL 8.7/8.8	LISREL 9.1	M*plus* 5	M*plus* 6/7	EQS 6.2	R lavaan
Chi-squared	**160.287**	**163.780**	**549.414**	**760.305**	**115.339**	**759.391**
CFI	**.994**	**.994**	.877	.926	**.996**	.926
TLI	**.993**	**.992**	.946	.910	**.995**	.910
RMSEA	.026	.073	.092	.080	.017	.080
SRMR	.053	.053	----	----	----	----

M*plus* 5 was based on 48 degrees of freedom; all others = 86

Reasons for Differences

- For ML estimation, differences in fit are based on differences in how the fit indices are calculated
  - Although they all **use the same formulas** for indices such as the CFI, TLI, and RMSEA, they don't all use the same **type of χ^2** in the calculations

- For robust estimators, it gets worse…..
  - Differences in fit values are not only due to the particular type of χ^2 is used in computing fit indices but also differences between estimation procedures

For example, in ML estimation

- **RMSEA**
  - **LISREL 8.7/8.8** use the NTWLS χ^2
  - **LISREL 9.1, M*plus*, EQS,** and **R lavaan** use minimum fit function χ^2

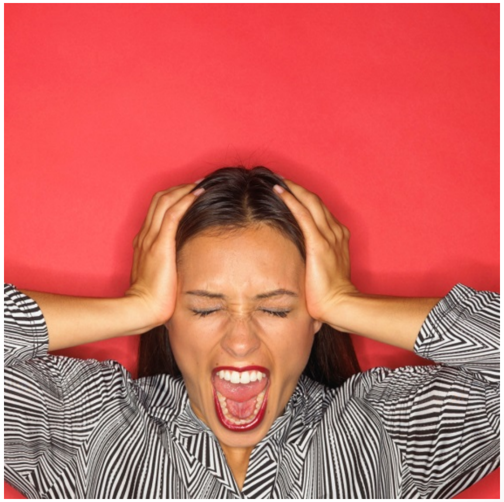
- **TLI and CFI**
  - **LISREL 8.7/8.8** use minimum fit function χ^2 from the tested model **but** normal theory weighted least squares (NTWLS) χ^2 for the independence null model
  - **LISREL 9.1** uses NTWLS χ^2 for both the tested model **and** the independence null model
  - **M*plus*, EQS, R lavaan** use minimum fit function χ^2 for both the tested model and the independence null model

In robust estimation…

- **RMSEA**
  - **LISREL 8.7/8.8** and **EQS** use the Satorra-Bentler rescaled χ^2
  - **LISREL 9.1** minimum fit function χ^2 but obtained from the robust ML results
  - **M*plus* 5, 6, 7** and **R lavaan** use NT robust WLS χ^2
- **TLI and CFI**
  - **LISREL 8.7/8.8/9.1** and **EQS** use Satorra-Bentler rescaled χ^2 for tested model but NT robust ML for independence null χ^2
  - **M*plus* 5, 6, 7** and **R lavaan** use robust WLS χ^2 for tested model and WLSMV-based χ^2 for independence null, with a different null model that also accounts for item thresholds when using ordinal data; **but** computation of χ^2 and degrees of freedom changed in version 6

And one more thing….

- For LISREL 9.1, results depend upon when you purchased or downloaded the software
- Results from October, 2012 to March, 2013 differ from those obtained after March, 2013



What's an evaluator to do?

- Learn the details about the SEM software you are using and how the various fit indices are computed
  - However, this is easier said than done
  - Finding this information can be like going on an archaeological dig

- Even when interpreting the findings of others, evaluators need to be aware of software differences that affect model fit

- Pay attention to component fit as well as global fit for consistencies versus inconsistencies
  - For example, do the results make sense even if global fit is good?

Implications

- For evaluation practice
  - Results of either CFA or tests of full structural models could reflect artifacts of the software used; which in turn, could affect conclusions drawn
  - Consider if "borderline" fit is based on software that tends to produce more versus less positive assessment of fit

- For evaluators attempting to publish, be sure to indicate software and version and if needed, supporting documentation of how the fit indices were produced

- For journal reviewers, keep up to date on SEM-based research as well as SEM software properties

Final thoughts

- Often choice of software is arbitrary
  - Cost, familiarity, availability, personal preference
- Software is generally never considered a critical factor in the fit of the model
- Researchers in the field have explored numerous factors that affect fit statistics, but software usually is not considered among those factors
- We encourage evaluators to include software among the factors to be considered

Questions?