Examples Demonstrating How to Conduct Critical Reviews of Analyses from SEM Articles

Shenyang Guo, University of North Carolina at Chapel Hill

Supplemental Material for *Structural Equation Modeling*
Oxford University Press Pocket Guides to Social Work Research Methods, 2012

Critical review of findings from prior studies is important for all kind of statistical analyses, not just for research applying SEM. However, performing critical review of SEM applications exhibits its own challenges, requires extra care in terms of balancing three factors: research objective or questions, the nature of data, and the specific type of models employed. As such, it takes time to learn and practice. To help readers understand issues embedded in the critical review, this section presents three examples. We hope through these examples readers further understand requirements of SEM, challenges in its practice, and strategies for good practice.

Example 1: Recover Necessary Data

When conducting critical review, it's important to recover original data from published source and then replicate the author's original analysis as much similar as possible. This is in theory feasible, as long as the authors present correlations and standard deviations and modeling the mean structure is not an issue. However, it is challenging in practice. The following example illustrates how to get back the necessary data as much as one can to facilitate a replication study.

Conway et al. (2004) present a study to address the hypothesis that adults reporting sexual abuse in childhood are more likely to exhibit a general tendency to ruminate on sadness. The authors use SEM to test a theoretical model depicting relations between reported abuse, rumination on sadness, and dysphoria. To review the soundness of the study, we need first to replicate the original study.

In the publication, the authors report a correlation matrix for the study variables for both men and women (i.e., a correlation matrix based on the entire sample). However, they report standard deviations separately for men and women. To replicate the SEM based on the entire sample, we need input data of both the correlation matrix for the sample – which is known, and the standard deviations for the sample – which is known for men and women separately. Now the challenge is how to recover the sample standard deviations based on the men and women's separate standard deviations? The data complication is further added by the fact that the sample size n varies by variable due to different numbers of missingness.

Taking variables v6 and v9 as an example, the study reports that the sample correlation of v6 and v9 is 0.18 (n=201), in addition to the following information:

SD(v6) for men = .48, n for men (v6) = 95 (5 missing),
SD(v6) for women = 1.89, n for women (v6) = 97 (4 missing);
SD(v9) for men = 3.99, n for men (v9) = 100 (no missing),
SD(v9) for women = 3.82, n for women (v9) = 101 (no missing).

Given the data, how to calculate the sample standard deviations for v6 and v9? Note that standard deviations are required data when a correlation matrix is used.

The solution: we can use sum of squares (SS) to recover SD for the entire sample. Recall the formula calculating a sample SD:

$$SD(y) = \sqrt{\frac{\sum (y-\bar{y})^2}{n-1}}, \qquad \text{so} \qquad Var(y) = \frac{\sum (y-\bar{y})^2}{n-1}.$$

Thus, Var(y)*(n-1) = $\sum (y-\bar{y})^2$ = SS or sum of squares.

This suggests that we can Get SS(men) and SS(women) separately using the reported data, and then use the following formula to calculate SD for the entire sample:

$$\text{SD (Sample)} = \sqrt{\frac{SS_{men} + SS_{women}}{n_{men} + n_{women} - 1}}.$$

The calculation:

SD(v6) for men = 0.48, so Var(v6) for men = $(0.48)^2$=.2304
SS(v6) for men = (.2304)(95-1)=21.6576

SD(v6) for women = 1.89, so Var(v6) for women = $(1.89)^2$= 3.5721
SS(v6) for women = (3.5721)(97-1)=342.9216

$$\text{SD (v6) for the Sample} = \sqrt{\frac{SS_{men} + SS_{women}}{n_{men} + n_{women} - 1}} = \sqrt{\frac{21.6576 + 342.9216}{95 + 97 - 1}} = 1.3816.$$

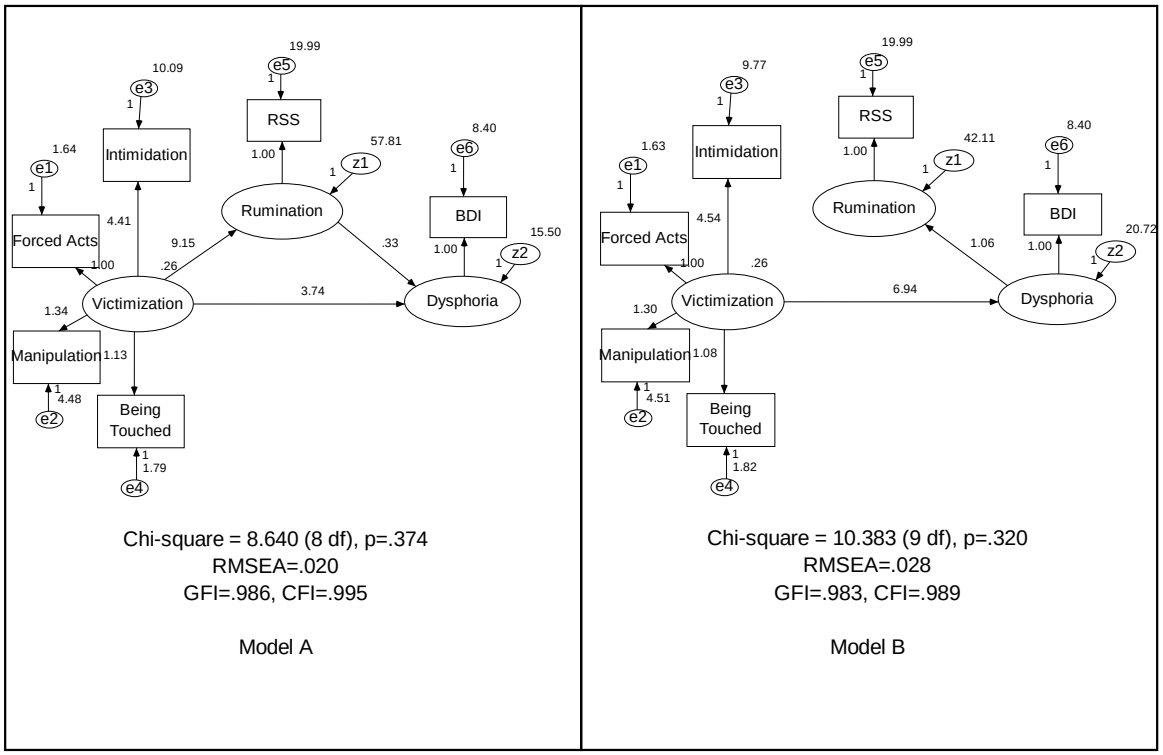
Replicate the above process for all variables, we obtain the required vector of standard deviations, and then we can input the vector in conjunction with the correlation matrix into SEM program.

Example 2: A Critical Review of Conway et al. (2004)

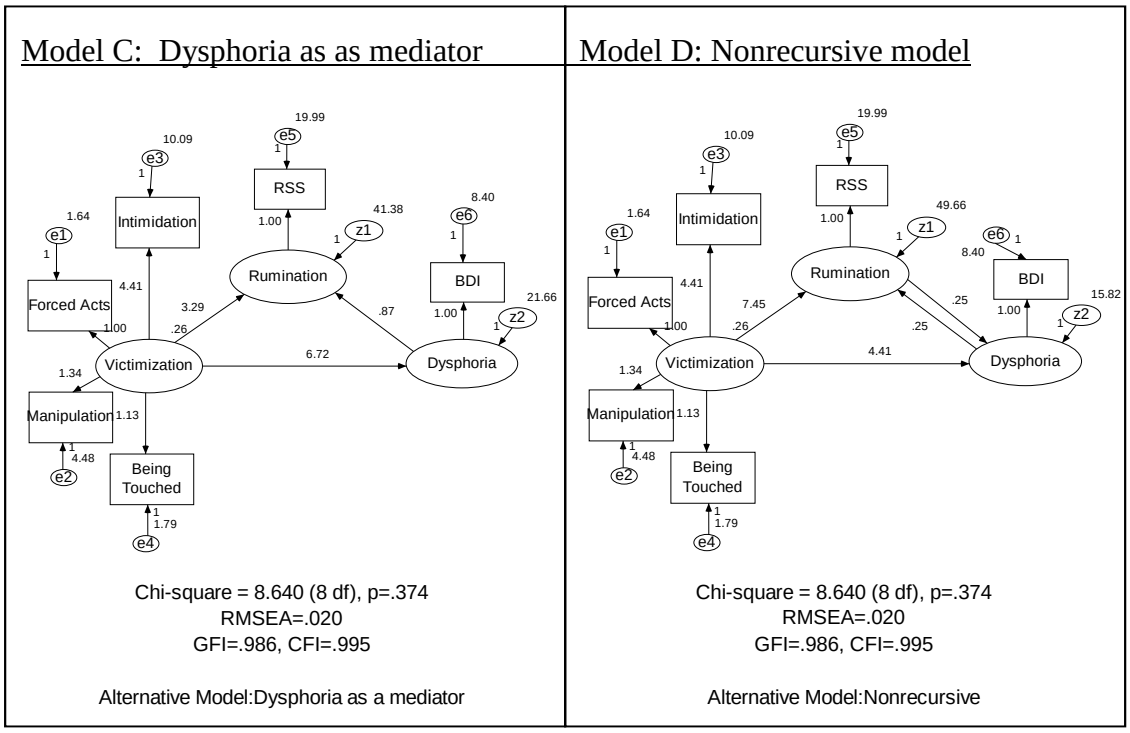
With the recovered data, the first task of a critical review is to replicate the authors' original study. Using the input data and Amos, we ran two original models specified by the authors (i.e., Models A & B). Results (Figure 7.1) show that the replicated models are very close to the authors' original model reported in their Table 3. This is especially true when comparing the recovered models with the original models on the model chi-square, RMSEA, GFI, and CFI. Thus, we concluded that we were at the same position as the authors' when they conducted the analysis. At this point, a number of new models can be tested, and interesting questions can be explored. For this illustration, we only focus on three questions: (a) What alternative models the

authors could have run? (b) Which models that might be equally important? And (c) do the new models support the authors' original findings?

Figure 7.1 Recovered Models A & B



Figures 7.2 Alternative Models



With regard to the first question, we ran several models. Examples of these models are shown in Figure 7.2 (i.e., Models C and D). As the fit indices indicate, both new models fit the data equally well. With regard to the second question, our reanalysis indicates these two alternative models are equally important, because they show alternative explanations to relationships between reported abuse, rumination on sadness, and dysphoria. With regard to the third question, we don't think the new models can fully support the authors' original conclusions. Specifically, conceptualizing dysphoria as a mediator and treating the interaction between rumination and dysphoria as a two-way nonrecursive relationship are important, and add implications to the study.

Example 3: A Critical Review of Colarossi & Eccles (2003)

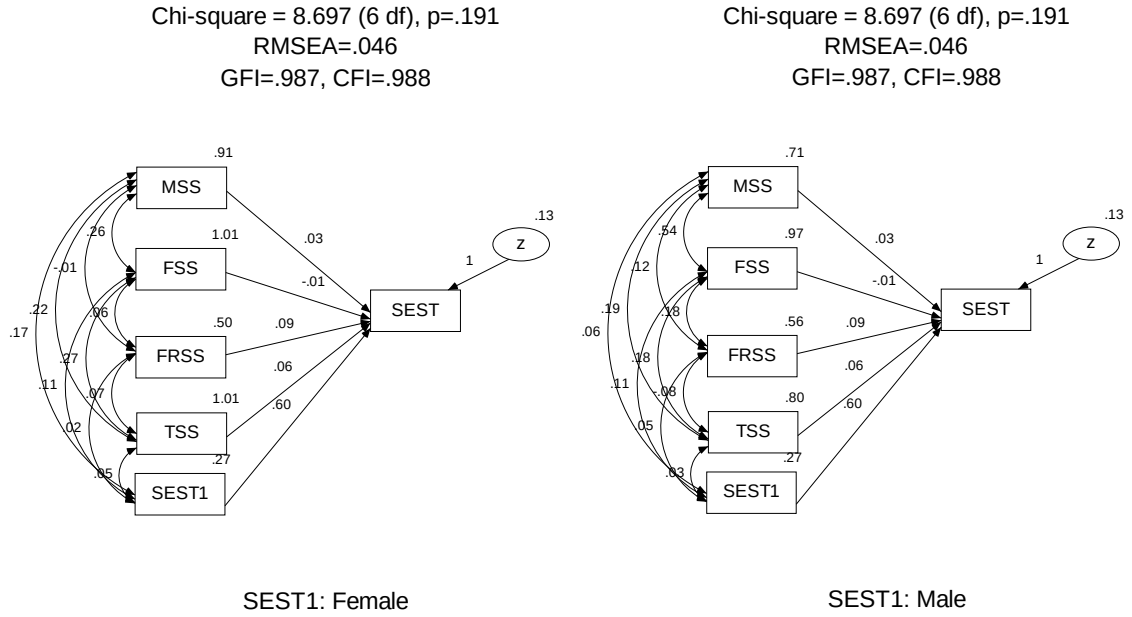
This study conducts a group comparison using SEM to evaluate differential effects of providers' social support on adolescents' mental health between females and males. The study groups are comprised of 125 females and 92 males, respectively. It analyzes two outcome measures: depression and self-esteem at time 2, while controlling for the same outcome variables at time 1. Four exogenous variables used in the study are mother's social support (SS), father SS, teacher SS, and friend SS.

The study presents data, study methods, and results in a clear, well-organized and succinct fashion. Tables 1 and 2 are especially good: not only the authors reported skewness and kurtosis to inform distributional properties of the study variables (Table 1), but also they presented correlations for each group efficiently (i.e., Table 2 presents correlations of each group using triangles above and below the diagonal).

In general, the authors made efforts to present the information sufficiently to allow replication. With the reported data, we first attempted to replicate the authors' original study. We now focus on one of the two outcomes (i.e., self-esteem at time 2) to show our replication and critical review.

Presented in Table 3 and Figure 1, the authors indicate that their study is a group comparison of "self-esteem at time 2" using five exogenous variables (the outcome at time 1 plus four support variables). Shown on page 24, the authors further constrained the five path coefficients and the disturbance variance to be the same between females and males. With this information in mind, we ran the same model by gender. The replicated models are presented in Figure 7.3. Comparing the magnitude of each path coefficient, p-value, R-square between the replicated models and the columns labeled "Equal" and the model-fit indices reported in the authors' Table 3, we are confident that the replication works well.

Figure 7.3: Replicated Self-esteem Model by Groups



A comment is worth making at this point. What model did the authors employ to run? Taking a careful look at the authors' Figure 1 and our replicated models, we find that the study model is a multiple regression analysis. In the paper, the authors stated that "MSEMs (i.e., multigroup structural equation models) were used to test hypotheses 3 and 4. MSEMS are preferable to multiple regression analyses because they can test a model for its applicability to different groups simultaneously by estimating group differences in path coefficients and model fit." No! The model the authors ran is exactly the same as a regression model using gender as a moderator and interacting it with other independent variables, one at a time. The model is no batter than a regression. Furthermore, the study does not take advantages of SEM (i.e., latent-variable strategy, direct and indirect effects, etc.). A more sophisticated analysis may treat self-esteem and depression as two indicators of a latent endogenous variable labeled as "mental health", and treat various types of support as a latent exogenous variable.

After obtaining similar model, the next step is to check whether or not the authors have followed the general guidelines for modeling. The crucial feature of the reported study is group comparison. So now we need to check steps for conducting a multiple group comparison. The reported model is a test of $H_{\Gamma\Psi}$ in a series of tests of hypotheses concerning equality of constrained coefficients between groups. As shown in Section 6.3, there are six such hypotheses need to be tested, and each should be performed in order. That is, if a same-form hypothesis cannot be rejected, then there is no need to test same-gamma, and so on. The paper does not show that the authors ever tested a same-form model, or followed the guidelines for testing a hierarchy of hypothesis in a sequential order.

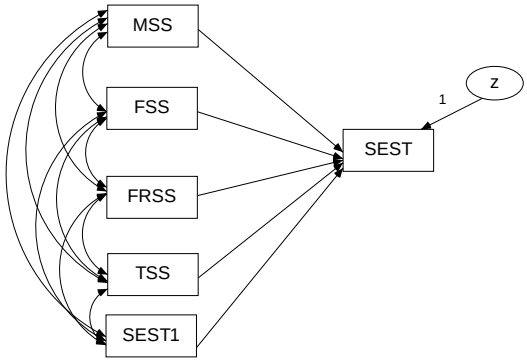
So naturally our next step is to test the same-form model. To run this model, we just need to delete all equality constraints in the previous run, and check the fit of the overall model. Results of our same-form model show that the model does not have a good fit to data. Specifically,

RMSEA for the same-form model is .187 (90% CI: .167, .209). Basically, we cannot accept the same-form hypothesis! This is not surprising, because the study's Table 3 reports that the model for Female has a significant path of "Friend SS", but the model for Male does not have such significant path. Precisely, the two groups have a different form shown by Figure 7.4.

Since the same-form hypothesis (i.e., H_{Form}) cannot be accepted, how could one accept a more restrictive hypothesis about equal gammas and psi (i.e., $H_{\Gamma\Psi}$)? The conclusion of "Hypothesis 4 was confirmed" made on page 24 of the paper cannot be supported. Note that the authors' Hypothesis 4 is "Despite gender differences in perceptions of support and levels of mental health symptoms, effects will not differ significantly for male and female adolescents." Based on the authors' Table 3 and the replicated model, we conclude that there is no evidence to support this hypothesis; at least the effect of Friend SS has a significant effect on self-esteem for females, but no such effect for males.

Following the above procedure, we ran similar models for the outcome depression, and obtained similar results. That is, the same-form hypothesis cannot be accepted, and therefore, equality constraints did not exist.

Figure 7.4 Differential Forms of Factors Affecting Self-esteem (SEST) between Females and Males: Results Suggested by Table 3 of Colarossi & Eccles (2003)



Female