Chapter 12

How to Fool Yourself with SEM

When we are self-indulgent and uncritical,
when we confuse hopes and facts,
we slide into pseudoscience and superstition.

—Carl Sagan

Overview

- Eyes wide open
- Tripping at the starting line: Specification
- Improper care and feeding: Data
- Checking critical judgment at the door: Analysis and respecification
- The garden path: Interpretation

Eyes wide open

- The technique of SEM is a very flexible analytical tool
- But as with any complex statistical procedure, its use must be guided by reason
- Potential pitfalls to avoid are considered under four categories:

    1. Specification
    2. Data
    3. Analysis and respecification
    4. Interpretation

Eyes wide open

- These categories are not mutually exclusive, but they correspond to the usual sequence in which researchers should address these issues
- This list of ways to take leave of your senses in SEM is not exhaustive, but it contains many of the more common mistakes
- See also Blalock (1991), Freedman (1991), and Mason (1991)

Tripping at the starting line: Specification

1. *Specify the model after the data are collected rather than before*

   This case concerns the specification of a model for an archival data set. Potential problems include the realization that key variables are omitted or that the model is not identified

2. *Omit causes that are correlated with other variables in a structural model.*

   If an omitted cause is correlated with variables already in the structural model, then estimates of the direct effects may be incorrect

Tripping at the starting line: Specification

3. *Fail to have sufficient numbers of indicators of latent variables*

   Having only two indicators per factor may lead to problems such as nonconvergence of iterative estimation and empirical underidentification. It may also be difficult to estimate measurement error correlations for factors with only two indicators

   Kenny's (1979) rule of thumb about indicators is apropos: "Two *might* be fine, three is better, four is best, and anything more is gravy" (p. 143)

Tripping at the starting line: Specification

4. *Use psychometrically inadequate measures*

   The analysis of variables with a lot of measurement error in the scores can lead to inaccurate results. Estimates about effects of latent variables are also more precise when the indicators are psychometrically sound

5. *Fail to give careful consideration to the question of directionality*

   If solid reasons cannot be provided for the specification of directionality, then either use another type of statistical procedure (e.g., multiple regression) or test alternative models with different causal sequences, but some of the latter may be equivalent models

Tripping at the starting line: Specification

6. *Specify feedback effects in structural models (e.g., $Y_1 \rightleftarrows Y_2$) as a way to mask uncertainty about directionality*

   Feedback relations have their own assumptions (e.g., equilibrium) and their presence makes a structural model nonrecursive, which is more difficult to analyze than a recursive model

7. *Overfit the model (i.e., forget the goal of parsimony)*

   Any model made sufficiently complex will explain the data, but such models may test no particular hypothesis (e.g., when $df_M = 0$)

Tripping at the starting line: Specification

8. *Add disturbance or measurement error correlations without substantive reason*

   Adding such unanalyzed associations can be a way to improve fit simply by making a model more complex. Note in some disciplines, such as econometrics, the specification of correlated residuals is routine

9. *Specify that indicators load on more than one factor without substantive reason*

   This specification may be appropriate if you really believe that an indicator measures more than one construct, but adding factor loadings makes the model less parsimonious

Improper Care and Feeding: Data

10. *Don’t check the accuracy of data input or coding*

    Data entry mistakes are easy to make, and even machine-based data entry is not always perfect

11. *Ignore whether the pattern of missing data loss is random or systematic*

    Failure to use an appropriate method for dealing with missing observations may lead to incorrect results (e.g., they do not generalize to the intended population)

Improper Care and Feeding: Data

12. *Fail to examine distributional characteristics*

If the distributions of continuous endogenous variables are severely nonnormal, then use an estimation method that does not assume normality or use corrected statistics (e.g., robust standard errors) when normal theory methods, such as ML estimation, are used

If the indicators are discrete with a small number of categories, then use an appropriate estimation method for this type of data

Improper Care and Feeding: Data

13. *Don't screen for outliers*

    Even a few extreme scores in a relatively small sample can distort the results. If it is unclear whether outlier cases are from a different population, the analysis can be run with and without these cases in the sample

14. *Assume that all relations are linear without checking*

    A standard assumption in SEM is that variable relations are linear. Curvilinear or interactive relations can be represented with product terms, but such variables must be created by the researcher and then included in the model (chap. 13)

Improper Care and Feeding: Data

15. *Ignore lack of independence among the observations*

    This problem may arise in two contexts: The observations are from a repeated measures variable or a hierarchical data set where the variables may not be repeated measures variables (chap. 13)

Checking critical judgment: Analysis and respecification

16. *Respecify a model based entirely on statistical criteria*

    A specification search guided entirely by statistical criteria, such as modification indexes, is unlikely to lead to the correct model

17. *Fail to check the accuracy of computer syntax*

    It is easy to make an error in computer syntax that specifies the model or data. Although SEM computer programs have become easier to use, they still cannot generally detect a mistake that is a logical error rather than a syntax error

Checking critical judgment: Analysis and respecification

18. *Fail to carefully inspect the solution for admissibility*

    The presence of a Heywood case or other kinds of illogical results indicates a problem in the analysis, that is, the solution should not be trusted

19. *Report only standardized estimates*

    Always report the unstandardized estimates in a primary analysis. Otherwise, it may be difficult to compare the results to those from later studies where either the same or a similar model is estimated in different samples

Checking critical judgment: Analysis and respecification

20. *Analyze a correlation matrix when it is clearly inappropriate*

    These situations include the analysis of a model across independent samples with different variabilities, longitudinal data characterized by changes in variances over time, or a type of SEM that requires the analysis of means

21. *Estimate a covariance structure with a correlation matrix without using appropriate methods*

    Appropriate methods, such as constrained estimation, should be used to analyze a correlation matrix in situations where it is not inappropriate to do

Checking critical judgment: Analysis and respecification

22. *Fail to check for constraint interaction when testing for equality of loadings across different factors or of direct effects on different endogenous variables*

    If there is constraint interaction, then it may make sense to analyze the correlation matrix using the method of constrained estimation, assuming it is appropriate to analyze standardized variables

23. *Analyze variables so highly correlated that a solution is unstable*

    Extreme multicollinearity (e.g., $r > .85$) may cause the results to be statistically unstable or estimation to fail

Checking critical judgment: Analysis and respecification

24. *Estimate a complex model in a small sample*

    As the ratio of cases to the number of parameters is smaller, the statistical stability of the estimates becomes more doubtful. Cases-to-parameter ratios less than 10:1 may be cause for concern, as are sample sizes less than 100

25. *Set scales for latent variables inappropriately*

    This includes analyzing standardized variables in multiple-sample SEM or in single-sample analyses when means are expected to change

Checking critical judgment: Analysis and respecification

26. *Ignore the problem of start values or provide grossly inaccurate ones*

    Iterative estimation may fail to converge because of poor initial estimates, which is more likely with complex models or nonrecursive models

27. *When identification status is uncertain, fail to conduct tests of solution uniqueness*

    If it is unknown whether a model is theoretically identified but a SEM computer program yields a converged, admissible solution, then the researcher should conduct empirical tests of the solution's uniqueness

Checking critical judgment: Analysis and respecification

28. *Fail to recognize empirical underidentification*

    Estimation of models that are identified can nevertheless fail because of data-related problems, such as multicollinearity or estimates of key parameters that are close to zero or equal to one another

29. *Fail to separately evaluate the measurement and structural portions of a structural regression model*

    Two-step (or four-step) estimation of structural regression models can help determine whether the source of poor fit of the whole model lies in the measurement component or in the structural component

Checking critical judgment: Analysis and respecification

30. *Estimate relative group mean or intercept differences on latent variables without establishing at least partial measurement invariance*

    If the observed variables do not at least have the same basic factor structure across all groups, then it makes little sense to evaluate relative group mean differences on latent variables

31. *Analyze parcels of categorical items as continuous indicators without checking to see whether items in each parcel are unidimensional*

    If a set of items assigned to the same parcel do not measure one common domain, analysis of the total score across the items may not be very meaningful

The garden path: Interpretation

32. *Look only at indexes of overall model fit; ignore other types of information about fit*

    This refers to "fit index tunnel vision," a disorder that is fortunately curable by looking through the entire output. A related mistake is selective reporting of fit indexes

33. *Interpret good fit as meaning that the model is "proved"*

    In general, SEM is more useful for rejecting false models than for somehow "proving" whether a given model is in fact true

The garden path: Interpretation

34. *Interpret good fit as meaning that the endogenous variables are strongly predicted*

    Indexes of overall fit indicate whether the model can reproduce the observed correlations or covariances, not whether substantial proportions of the variance of the endogenous variables are explained

35. *Rely solely on statistical criteria in model evaluation*

    Other important considerations include model generality, parsimony, and theoretical plausibility—see Kaplan (2000, chap. 9), Robert and Pashler (2000), and Sikström (2001)

The garden path: Interpretation

36. *Rely too much on statistical tests*

    This includes interpreting statistical significance as evidence for effect size and forgetting that results of statistical tests are affected by distributional characteristics

37. *Interpret the standardized solution in inappropriate ways*

    This includes comparing standardized estimates across groups that differ in their variabilities or when the unstandardized counterparts have been constrained to be equal

The garden path: Interpretation

38. *Fail to consider equivalent models*

    Essentially all structural equation models have equivalent versions that generate the same predicted correlations or covariances. Researchers should offer reasons why their models are to be preferred over some obvious equivalent versions of it

39. *Fail to consider (nonequivalent) alternative models*

    If the overall fits of some of these alternative models that are not equivalent models are comparable, then the researcher must explain why a particular model is to be preferred

The garden path: Interpretation

40. *Reify the factors*

    Believe that constructs represented in your model *must* correspond to things in the real world. Perhaps they do, but do not assume so

41. *Believe that naming a factor means that it is understood or correctly named (i.e., commit the naming fallacy)*

    Meaningful factor names are conveniences, not explanations

The garden path: Interpretation

42. *Believe that a strong analytical method like SEM can compensate for poor study design or slipshod ideas*

    No statistical procedure, SEM or otherwise, can make up for inherent logical or design flaws

43. *As the researcher, fail to report enough information so that your readers can reproduce your results*

    There are still too many reports in the literature where SEM was used in which the authors do not give sufficient information for readers to re-create the original analyses or evaluate models not considered by the authors

The garden path: Interpretation

44. *Interpret estimates of relatively large direct effects from a structural model as "proof" of causality*

    It is better to view structural models as "as if" models of causality that may or may not correspond to causal sequences in the real world

References

Blalock, H. M. (1991). Are there really any constructive alternatives to causal modeling? *Sociological Methodology*, *21*, 325-335.

Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology*, *21*, 291-313.

Kaplan, D. (2000). *Structural equation modeling*. Thousand Oaks, CA: Sage.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.

Mason, W. M. (1991). Freedman is right as far as he goes, but there is more, and it's worse. Statisticians could help. *Sociological Methodology*, *21*, 337-351.

Robert, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing in psychology. *Psychological Review*, *107*, 358–367.

Sikström, S. (2001). Forgetting curves: Implications for connectionist models. *Cognitive Psychology*, *45*, 95–152.