An Introduction to Structural Equation Modeling

Richard G. Lomax

1

A. What is SEM?

- It is a method to provide quantitative tests of theoretical models.
- The researcher hypothesizes certain **relationships** among a set of variables and then empirically tests those relationships.

2

- A few quick and very basic examples from education, marketing and health care:
- Home environment → achievement
- Consumer trust → product sales
- Diet →
  Exercise →
  Risk of heart attack

3

Bottom line question:

- Is the theoretical model supported by the sample data?
- The hypothesized model is based on theory and previous research.

4

B. Types of variables

- 1. **Latent variables** (constructs / factors): these are variables that are not directly observed or measured; they are indirectly observed or measured, and thus are inferred from a set of observed variables that are actually measured using tests, surveys, observations, interviews, etc.
- examples – intelligence (psychological construct), consumer confidence (economic construct), fitness (health-related construct).

5

- 2. **Observed variables** (indicators / measures): a set of measured variables that are used to define or infer the latent variables in a particular way.
- examples: WISC (measure of intelligence), Dow-Jones industrial average (measure of economy), blood pressure (measure of fitness).

6

- 3. **Independent variables**: variables that **are not** influenced by any other variable in the model.
- 4. **Dependent variables**: variables that **are** influenced by some other variable in the model.
- Home environment → achievement
- Consumer trust → product sales
- Diet →
  Exercise → Risk of heart attack

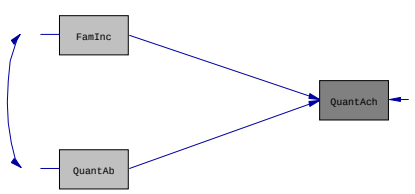
7

C. Most basic types of models

- 1. Regression models
- 2. Path models
- 3. Confirmatory factor analysis models
- 4. Structural equation analysis models
- 5. More complex models briefly described later

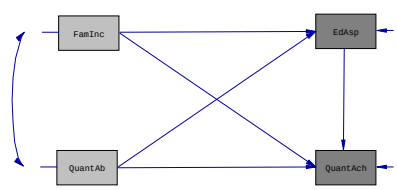
8

- 1. **Regression models** – an observed variable model with a single equation; thus one dependent variable is explained or predicted by one or more independent variables; only simple theoretical models are possible, but does not deal with measurement error.



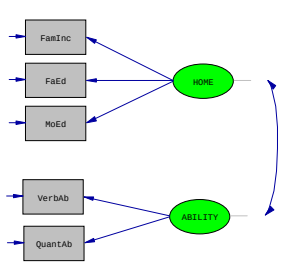
9

- 2. **Path models** – an observed variable model with multiple equations; thus allows for multiple independent variables and multiple dependent variables; thus more complex theoretical models are possible, but does not deal with measurement error.



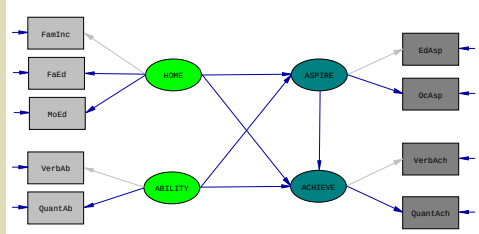
10

- 3. **Confirmatory factor analysis (CFA) models** – a latent variable model with multiple observed variables measuring each latent variable or factor; should lead to better assessment of the latent variables, as it can deal with measurement error.



11

- 4. **Structural equation analysis models** – a latent variable model which combines a confirmatory factor analysis model (latent variables assessed through multiple observed variables) and a path model (latent variables are related to one another); thus includes all 4 kinds of variables.



12

D. Why conduct SEM?

- 1. Researchers need to use multiple variables to better understand an area of inquiry. As theories have become more sophisticated, basic methods that are limited to a small number of variables are no longer sufficient. SEM has become the preferred method for confirming (or disconfirming) theoretical models in a quantitative fashion.
- 2. Recognizes the effect of measurement error (which SEM explicitly takes into account). Greater recognition is given to the validity and reliability of the observed variable scores obtained from measurement instruments.

13

- 3. More advanced SEM models are possible after 40 years of statistical and software development (e.g., models that involve multiple groups, multiple levels, and interactions).
- 4. SEM software more user-friendly than ever.

14

E. Most popular SEM software

- AMOS (add-on to SPSS).
- EQS (stand-alone program).
- LISREL (stand-alone program).
- Mplus (stand-alone program).

15

F. Steps in the SEM Process

1. Model Specification

- The researcher develops and hypotheses a theoretical model (or models) from available theory and research.
- This is the “hard part” of SEM.
- This is done prior to any data analysis (and preferably prior to data collection and even prior to the selection of measures).

16

- The components of model specification include:
- A) deciding which variables to include in the model and how these variables are related to one another.
- B) the sample covariance matrix is generated by by true population model (which is unknown); the researcher wants to determine which model actually generated this matrix and thus wants to find that model which best reproduces it.

17

- C) The tested model is properly specified when the true population model that generated this matrix is consistent with the model actually tested.
- D) The tested model is misspecified when the true population model that generated this matrix is not consistent with the model actually tested.
- E) Misspecified models can be due to errors of omission and/or inclusion of any variable and/or parameter.

18

- F) Misspecified models can result in biased parameter estimates (known as misspecification error); also the misspecified model may not fit the data according to global model fit indices.
- G) All models are misspecified to some degree.

19

2. Model Identification

- Often called solving "the identification problem."
- Given the sample covariance matrix and the theoretical model to be tested, can a unique set of parameter estimates be found?
- In other words, can we obtain a unique value for every parameter estimate? For example, $X + Y = 75$.

20

- Every potential parameter in a model is either:
- Free – unknown as you wish to estimate it;
- Fixed – not free, but fixed to a specific value (typically 0 or 1); or
- Constrained – unknown, but constrained to be equal to some other free parameter.

21

- For a model to be identified, every parameter must be identified.
- Otherwise the results cannot be trusted.
- This is why identification is important.

22

- Levels of identification can be assessed by looking at a models' degrees of freedom (*df*).
- A) under-identified model – not enough information (not identified); negative *df*.
- B) just-identified model – just enough information; zero *df*.
- C) over-identified model – more than enough information (preferred); positive *df*.

23

3. Model Estimation

- Some commonly used estimation methods:
- OLS (ordinary least squares);
- GLS (generalized least squares);
- ML (maximum likelihood).

24

4. Model Testing

- A) Use global fit indices to assess fit of an entire model.
- Unfortunately there is no single most powerful fit index (like the *F* in ANOVA or regression).
- Some global fit indices include chi-square, GFI (goodness-of-fix index), RMSEA (root mean square error of approximation), SRMR (standardized root-mean-square residual), and CFI (comparative fit index); it is recommended that you report multiple global fit indices.

25

- B) To assess the fit of individual parameters, examine…
- (1) *t* values: *t* = estimate / standard error;
- (2) significance of *t*;
- (3) sign of the estimate (in hypothesized direction);
- (4) whether the estimate makes sense (e.g., in expected range; no negative variances; no correlations beyond 1.0)

26

5. Model Modification

- When global model fit is not acceptable, need to look for ways to modify the initial theoretical model to achieve a better fit.
- Searching for a more properly specified model is known as a "specification search."

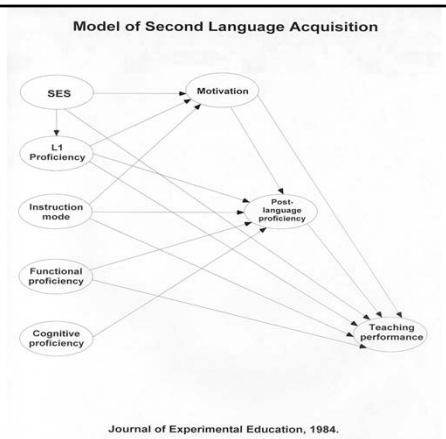
27

- True model ←----------→ Tested model
- The less consistent these models, the more the tested model is misspecified.
- **In any specification search, substantive knowledge must be the #1 priority.**
- **In other words, do not add additional parameters to achieve better model fit if they do not make substantive sense.**

28

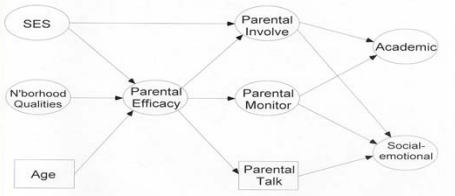
G. Some examples from my work

29



30

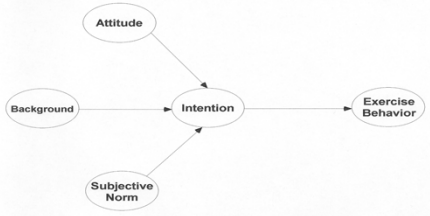
Model of Parental Efficacy



Parenting: Science & Practice, 2002

31

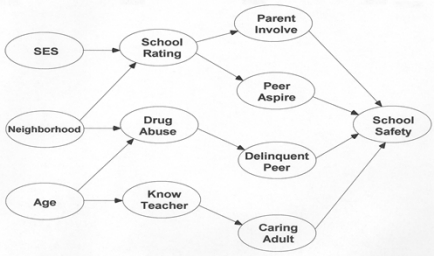
Model of Exercise Behavior



Research Quarterly for Exercise & Sport, 1990

32

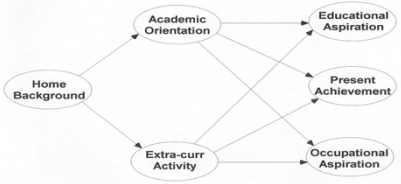
Model of School Safety



School Community Journal, 2001

33

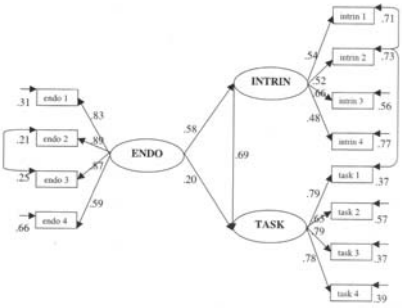
Model of Schooling



J Experimental Education, 1985 (multiple samples)

34

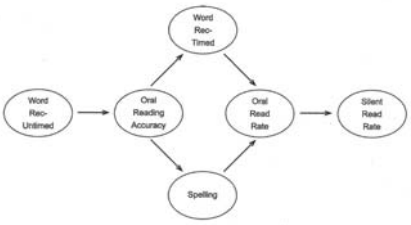
Model of Motivation



Contemporary Educational Psychology, 2004

35

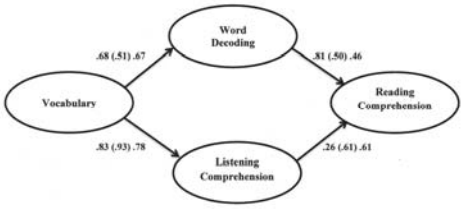
Model of Print Processing



Reading & Writing, 2012

36

Simple View of Reading



Reading Research Quarterly, in press

37

H. Advanced SEM Models

- Model validation (through multiple samples, cross-validation, or bootstrapping)
- Multiple group SEM (including structured means models)
- Second-order CFA
- Multitrait-multimethod models (MTMM)
- Multiple indicator-multiple cause models (MIMIC)

38

- Mixture models
- Multilevel models
- Interaction models
- Latent Growth models
- Monte Carlo methods

39

I. Further Information

- Schumacker, R. E. & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.) New York: Routledge.
- Spring course in SEM (ESQREM 8659, Wednesdays 12:45-3:25)

40