Latent Variable and Longitudinal Modeling for Language Education Research

Steven J. Ross

Themes

- Part I Latent Variable Models
- Part II Multi-Level Modeling
- Part III Event History Analysis

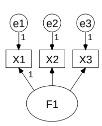
Rationale

- Cross-sectional (here and now) research methods are too sample-dependent.
- Educational policy-making usually requires generalizable causal models.
- Longitudinal research provides better basis for inferences to support policy.
- Education is about change. Growth models address change directly.

I Latent Variables Models

- Make traditional factor analysis models more explicit by including residuals
- Allow for path analysis using latent factors.
- Can be adapted to a wide range of empirical questions
- Are post-positivistic (seek to adjudicate empirical claims through model testing)

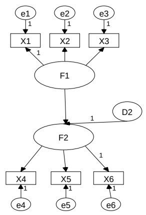
A Measurement Model



Variables

- Rectangles are measured variables hypothesized as indicators of a factor.
- Small circles are residuals
- Ovals are hypothesized LATENT variables or ‘factors’ in conventional factor analysis

Structural Equation



Latent Path Analysis

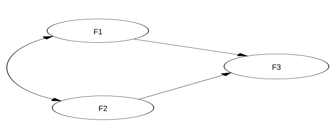
- Arrows from LATENT variables to Manifest variables indicate hypothesized covariance. X1-X3 Covary because they all indicate Factor 1.
- Arrow from F1 to F2 (LATENT to LATENT) indicate hypothesis that F1 ‘causes’ F2
- Example: Motivation → Classroom Achievement

Assumptions

- Theory preexists the data
- Model reflects the hypothesized relations a priori.
- Latent variables ‘compete’ with each other.
- An outcome is logically ‘caused’ by other latent variables
- Example: Proficiency is ‘caused’ by Motivation (F1) and/or Achievement (F2)

- Measurement error is included in a model
- Residuals are assumed to be independent
- Residuals are unexplained variances in Xs
- Disturbances are unexplained variances in Fs

Hypothesized Relations



- Straight unidirectional lines indicate hypothesized causal relation
- F1→F3 (Motivation *causes* Prof.)
- F2→F3 (Achievement *causes* Prof.)
- Curved arrows indicate non-zero covariance.
- F1←→F2 are correlated with each other

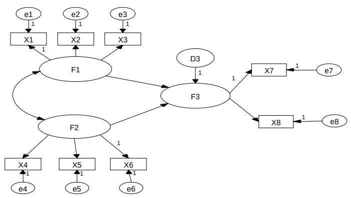
Assessing Fit

- Sample generalizability to Population assessment global Chi-square
- Hypothesized paths are tested with different fit indices; poor fit implies <u>missing</u> or <u>superfluous</u> paths in a model.
- Modification Indices diagnose missing paths and correlated residuals.

Potential Applications

- Theory testing through confirmatory factor analysis.
- Causal modeling with competing latent indicators in a latent path analysis.
- Multi-trait Multi-Method analysis with latent variables as competing traits and methods.
- Latent Growth Curve Modeling

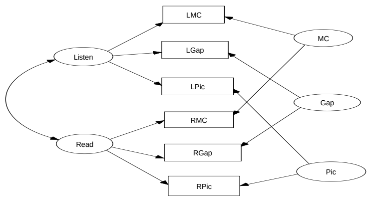
Confirmatory Factor Model



CFA Research Qs

- Do Instrumental motives influence school-based learning outcomes MORE THAN learning strategies influence outcomes?
- Instrumental motives and learning strategies <u>compete</u> in the confirmatory model.
- Larger standardized path coefficient indicate greater relative influence

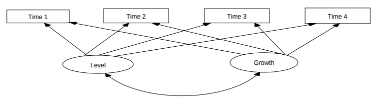
MTMM via SEM



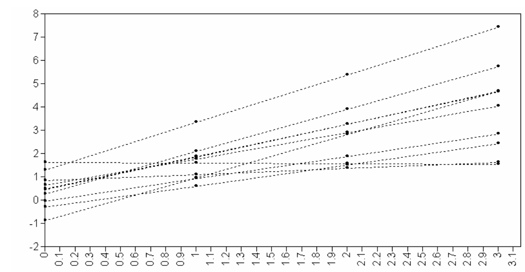
MTMM Research Qs

- How do methods of measurement contaminate construct-valid measures of Reading and Listening skills?
- Are some methods of measurement likely to reduce construct validity?
- How strong are method artifacts?

Latent Growth Curve



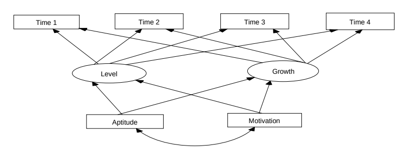
Variation in Individual Growth



LGC Research Qs

- How much variation is there in inter-individual learning?
- Is there regression/progression over time for high vs low starting ability learners?
- Is growth linear, flat, non-linear, or exponential in its shape?
- Is there a plateau effect? When?

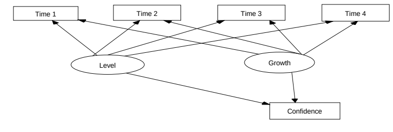
Covariates of Growth Curves



LGM Covariate Research Qs

- What static antecedent variables covary with <u>initial</u> individual differences?
- What static antecedent variables covary with <u>changes</u> over time?
- **Example**: Is Aptitude more influential than Motivation in understanding growth in language learning over time?

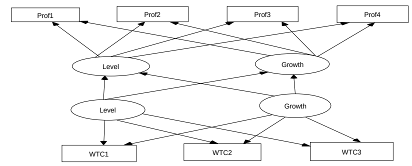
Sequelae of Growth



LGM Sequelae Research Qs

- Does growth cause other outcomes?
- Do initial individual differences better account for sequential outcomes?
- What are the long-term sustained effects of growth on other outcomes?
- Example consequences: employment, income, further study, confidence, motivation.

Cross-Domain Growth Curves



Parallel LGM Research Qs

- Does growth in achievement leverage growth in proficiency?
- Does change in motivation over time affect growth in achievement?
- Does change in perception of peer aspiration affect individual students' own growth in learning?

II The Analysis of Context

- Social-cultural theories put context at the apex of importance.
- Discourse-based methods focus on interaction to study context.
- Context can be analyzed quantitatively as well.

Multi-Level Models

- Individuals are <u>nested</u> within contexts
- School impacts are not exclusively attributable to individual differences.
- Contextual effects exert potentially large influences over individuals.
- Individuals and contexts require interactive modeling.

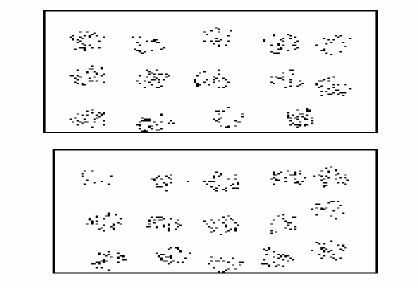
Multi-Levels

- Level 1: <u>Personal</u> attributes such as ability, experience, motivation, aptitude, strategy use, etc measured at the individual level.
- Level 2: <u>Collective</u> attributes such as average class ability, teacher experience, mean SES of class members, etc measured at the aggregate class level.
- Level 3: <u>Context</u> attributes of a whole school such as public or private, etc.

Nested Structure Example

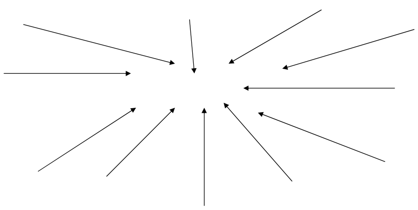
- Level 3 Sector (public vs private)
- Level 2 Schools (each within a community with a different level of social capital).
- Level 1 Students in schools (each person with unique ability, parental support, motivation, etc).

Nesting



Normative Environments

- Individuals tend to gravitate toward the norm within school contexts



Modeling Objectives

- Identify variables that co-vary with each level separately.
- Assess the impact of contextual variables as they moderate lower level variables.
- Test the effect of planned macro-level policy initiatives at the highest level.

Unconditional Model

- Intercepts alone are modeled first to assess the extent of variation between the individuals within and between the levels of schools.

$$Y_{ij} = \beta_0 + r$$

Adding level 1 predictors

- Level 1 variables added as in linear regression. Y_{ij} is the outcome of interest (e.g. achievement) for each student nested in a school.

$$Y_{ij} = \beta_0 + \beta(motivation) + r$$

Modeling Intercepts and Slopes

- The level 1 inter-individual differences and the influence of motivation on the individual students now are the *outcome* variables in the level 2 analysis. The between-school impact on student motivation is the focus:

$$\beta_0 = \gamma_{00} + \gamma_{01}(School) + u_0$$
$$\beta_1 = \gamma_{10} + \gamma_{11}(School) + u_1$$

Multi-level Advantages

- Individual students are not the only focus of analysis.
- Contextual variables (differences between school contexts, classrooms, etc) can be assessed.
- Value-added interventions can be assessed at the macro-level for policy analysis.

Applications

- Language learners (Level 1) nested in peer assessment groups (Level 2) doing cooperative learning tasks.
- Self-Assessing learners (Level 1) nested in classes (Level 2) with and without assessment training (Level 3).
- Interview candidates (Level 1) nested in interviewers (Level 2).

Normative Environment for Raters?

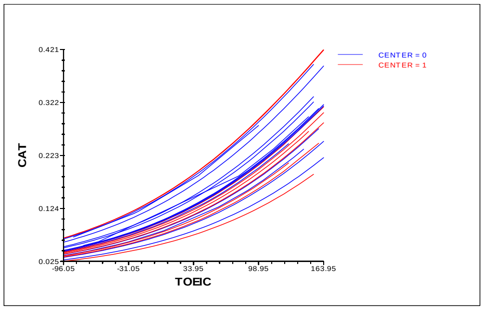
- RQ: Is there an interaction between a level one variable (TOEIC) and a tendency for oral proficiency raters to disagree?
- Does a Level 2 variable (community of practice ‘center’) covary with increased probability of disagreement?

Rater Disagreement Model

- $Log[P/(1-P)]=B_0+B_1(Prof)_{ij}+B_2(Sex)_{ij}$
- $B_0 = G_{00}+G_{01}(Liberal)+G_{02}(Center)+u_0$
- $B_1 = G_{10}$
- $B_2 = G_{20}$

Rater Severity and Location tested for their influence on the probability of rater disagreement on task based performance.

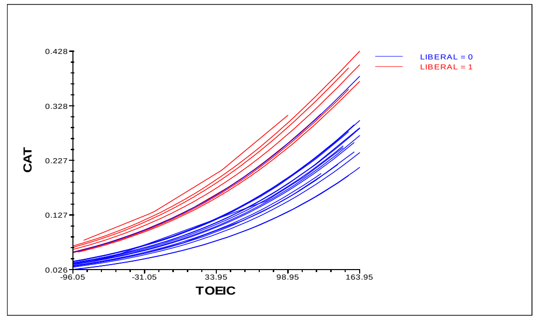
Disagreement across Centers



Rater Severity Risk Assessment

- RQ Does an OPI rater's previous history of lenient (liberal) rating patterns affect the probability that other raters will disagree with his/her ratings?
- Does the risk of disagreement increase through interaction with candidate proficiency (TOEIC)?

Rater Risk of Disagreement



More Applications

- Learners nested in classes (Level 1) who evaluate instructional quality (Level 2).
- Self-assessing learners nested in classes (Level 1) who estimate their classmates' mean motivation and aspiration (Level 2).
- Learners nested in institutions (Level 1) which experimentally employ summative or formative assessments (Level 2).

Peer Assessment Model

- Y_{ij}=B_0+B_1(prof.)+r
- B_0 = G_{00}+G_{01}(motivation)+u_0
- B_1 = G_{10}+G_{11}(motivation)+u_1

Test for Level 1 (prof) effects t>1.96 p<.05

and for Level 2 (motivation) of peer groups influencing Y (outcome). T>1.96 p<.05 for Level 2 diagnose peer group as 'context'

III Event History Analysis

- Events are <u>discrete changes of status</u> for an individual.
- Educational events:
- Students: certification, graduation, school-leaving, passing, continuing education, etc.
- Teachers: leaving the field, acquiring a post-graduate degree, getting tenure, etc.

- Events have a 'history' because they occur over a period of time.
- EHA is a longitudinal research method involving an event occurrence and a measure of time.
- Covariates or causes of the event can also be modeled.

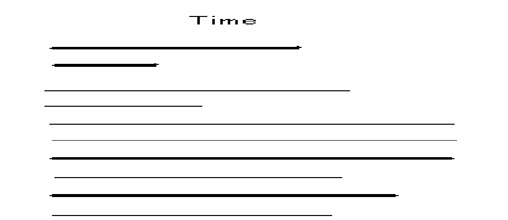
Baseline Model

- $\log E(t_{ij}) = \log(E_j)$
- The event (E) occurrence risk for each individual is the sole function of time.
- Basic data are a measure of time with a zero origin for each case and a discrete code for the event occurrence.

Data Structure

Case	Time	Event
001	25	0
002	32	1
003	44	1
004	15	0
005	65	1
006	22	0

Visualizing Events



Censored Data

- Censoring is when the event of interest does not occur for a case (student).
- Time continues on without a change of status for the case.
- The event may occur early, late, or not at all during the longitudinal study.
- EHA goal is to understand <u>when</u> and <u>why</u> events happen.

EHA with Covariates

- $\log E(t_{ij}) = \log(E_j) + B_1(X1)_i + B_2(X2)_I$
- Covariates X1 and X2 are hypothesized to explain why the event happens.
- Covariates can be continuous measures or categorical variables.

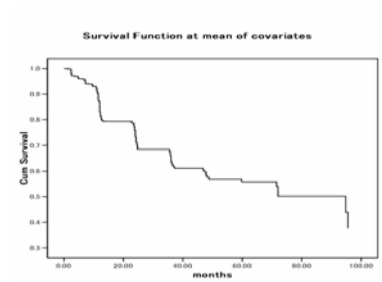
Data Structure with Covs

Case	Time	Event	X1	X2
001	25	0	14.1	1
002	32	1	15.2	0
003	44	1	16.6	1
004	15	0	13.7	0
005	65	1	19.0	1
006	22	0	17.1	1

Example: OPI Gain EHA

- RQ: How much is an observed gain in speaking proficiency (n=752) affected by differences in rater severity?
- How can apparent gains be distinguished from gains that are artifacts of rater differences?

Baseline time to Event

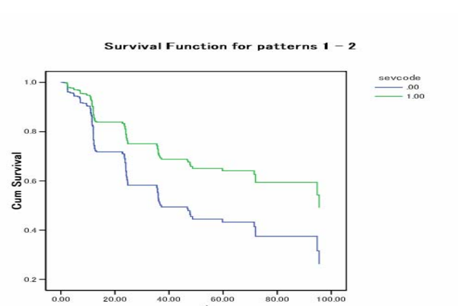


Baseline ‘Survival’

- With no information about rater differences, it takes about 70 months to reach a 50% chance of getting a higher OPI rating.
- Does a differences in severity between the raters covary with the gain event?
- $\log E(t_{ij})=\log(E_j)+B_1$(Severity Difference)

- B_1(Severity Difference) functions as the covariate.
- Difference between the earlier and latter rater severities (logits from MFRA)
- Effect-coded to denote higher than average severity by the latter rater.

Rater Severity Impact



Implications

- Severity difference highly significant
- After about 10 months, ‘gains’ may be artifacts of rater severity differences.
- Task based assessment is potentially contaminated by rater differences in applying the rating criteria!

Work in Progress

- How can rater-equating disambiguate authentic proficiency gain from rater artifacts?
- Anchoring designs
- Cumulative record 'mega matrix' equating for task based assessments

Summary

- I Latent Variable Models
    - Confirmatory factoring
    - Trait vs method analysis
    - Growth curve models (parallel, predicted, sequelae)
- II Multi-Level Models
    - Students in classes in schools
    - Candidates nested in raters within communities.
- III Event History Analysis
    - Persons x time influenced by covariates
    - TBA gains as possible rater differences

Take the Long View

- Longitudinal analyses affords many advantages over cross-sectional analyses.
- New methods are continuously invented to deal with data complexity
- Multi-level and longitudinal methods are converging and increasingly accessible