Quantitative Analysis for Research in Education and Communication

Dr. Wilfred Lau
Faculty of Education
The Chinese University of Hong Kong

Dr. Alex Chan
Hong Kong Shue Yan University

Presentation at the HKAECT-AECT 2017 Summer International Research Symposium
15-17 June 2017, The University of Hong Kong

Today's Topic

- Quantitative Methods in Social Sciences
- Multi-level Modeling
- Longitudinal Data Analysis
- Structural Equation Modeling
- Latent Class Analysis
- Questions

Quantitative Methods in Social Sciences

Chi-Square
Two-Samples t-test
Paired Sample t-test
ANOVA/MANOVA
ANCOVA/MANCOVA
Repeated Measures
Multiple Regression
Logistic Regression
EFA/CFA/Cluster

Multi-level Analysis
Longitudinal Data Analysis
Mixed-Effects Model
Structural Equation Model (SEM)
Latent Class Analysis
Mixture Modeling
Profile Pattern Analysis
Latent Trait Analysis

Statistical Inference
H0, H1, p-value

Model-based Inference

Multi-level Modeling

Level 3
School

Level 2
Classrooms

Level 1
Students

Why Multi-level Modeling?

1) Many data have a nested/clustered/hierarchical structure (e.g. students within classrooms within schools; workers within teams within departments; faculty within departments within universities …) **Multi-level modeling can capture the dependencies**

2) Single-level analyses (e.g. student-level/school-level in multiple regression analysis) fail to model dependency at level 2 or induce aggregation bias can affect the parameter estimations and incorrect inferences could be made (e.g. model misspecification)

Logic of Multi-level Modeling

1) Regression-based techniques (assumptions – homogeneity of variance and normality of error distribution at different levels are key assumption
2) Modeling an unconditional model to partition the variance-covariance components (error terms) at different levels
3) Putting in fixed, random (varying), and cross-interaction coefficients to explain the variance components at different levels
4) Theory-based and iterative model building process

(Suggested Reading: Raudenbush & Bryk, 2002, Chapters 4 & 9)

Examples of Research Questions

- Educational Research (e.g. how do student, teacher, and school characteristics explain the variation in student achievement?)
- Organizational Research (e.g. how do individual, team, and organizational characteristics relate to the differences in staffs' work satisfaction/performance?)
- Sociological Research (e.g. how do neighborhood segregation account for the individual aggressive behaviors?)
- Psychological Research (e.g. how do home and school contexts relate to child development?)
- Communication Research (e.g. how do the types of internet platform and individual personality explain variation in commitment to knowledge sharing?)

Illustration of Multi-level Modeling : Neighborhood Poverty & Achievement

Research Questions:

Do students living in high poverty-concentrated have lower math achievement scores than their peers living in low poverty-concentrated neighborhoods?

Data: 15,684 students enrolled in an urban school district (2000-2001) from 80 neighborhoods (census tract)

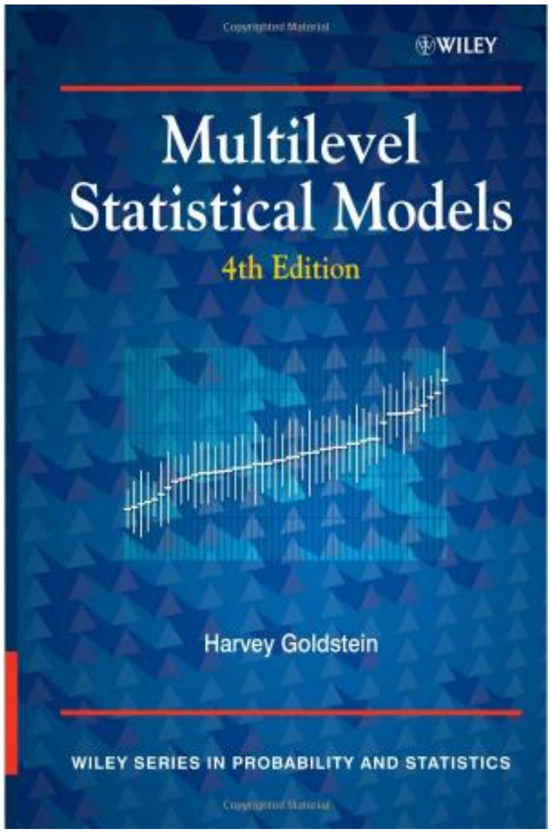
Source: Chan, C-K. & Maruyama,G. (2002). *Relations of disparities in housing and neighborhood poverty with achievement*. Society for the Psychological Study of Social Issues (SPSSI) 4^{th} Biennial Convention.

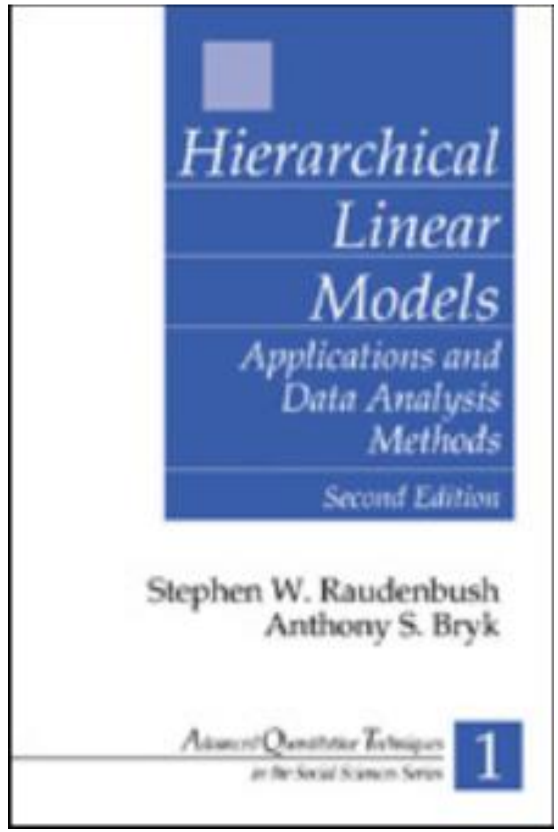
	Predictor	Base Model Coefficient	Base Model se	Neighborhood Model Coefficient	Neighborhood Model se
	Student Level				
Level 1	Intercept	52.27 ***	1.23	60.76 ***	0.69
	Public housing – Family Development			- 8.89 ***	1.33
	Public housing – Scattered Sites			-12.57 ***	1.54
	Section 8			-11.36 ***	1.32
	Non-subsidized housing low-SES			-11.22 ***	0.59
	Asian			- 2.57 ***	0.61
	Black			-10.41 ***	0.67
	Hispanic			- 8.19 ***	0.94
	Gender			- 1.00 ***	0.27
	Special education			-18.46 ***	0.56
	Elementary school			2.80 ***	0.43
	Middle school			0.64	0.43
	Average School Moves			- 6.36 ***	0.66
Level 2	Neighborhood Level				
	Neighborhood Poverty			-27.44 ***	1.86
Levels 1 x 2	Student x Neighborhood Level				
	Public housing – Scattered Sites			24.35 ***	6.72
	Section 8			12.22 *	5.33
	Non-subsidized housing low-SES			13.28 ***	1.91
	Asian			2.11	2.43
	Black			6.63 **	2.53
	Hispanic			9.31 *	3.98
	Elementary			2.32 *	1.09

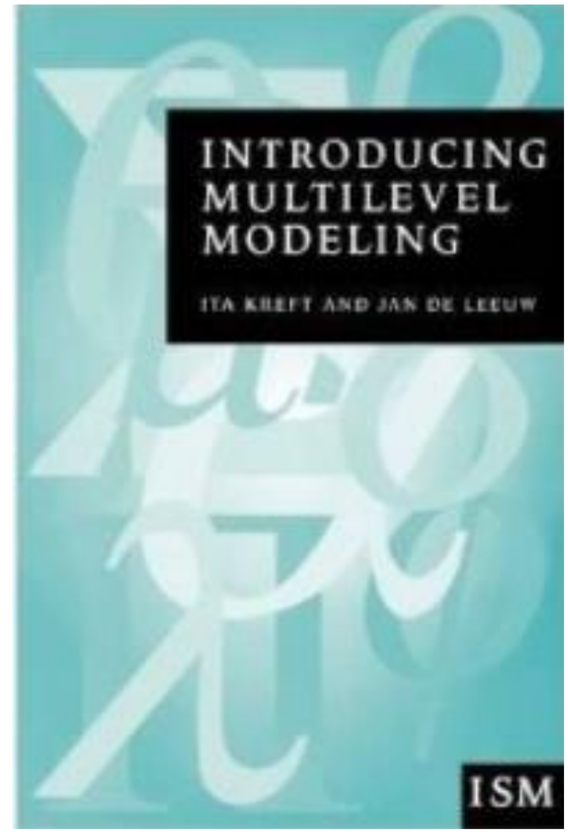
	Base Model	Neighborhood Model	
Variance Components			
Between neighborhood variance	162.29	17.67	
Percent of variance explained			89.1%
Within neighborhood variance	393.57	301.49	
Percent of variance explained			23.4%
Model Comparison Statistics			
Deviance	136590.67	132364.39	
Number of Parameters Estimated	3	42	
χ^2		4226.28 *** (df = 39)	

136590.67-132364.39

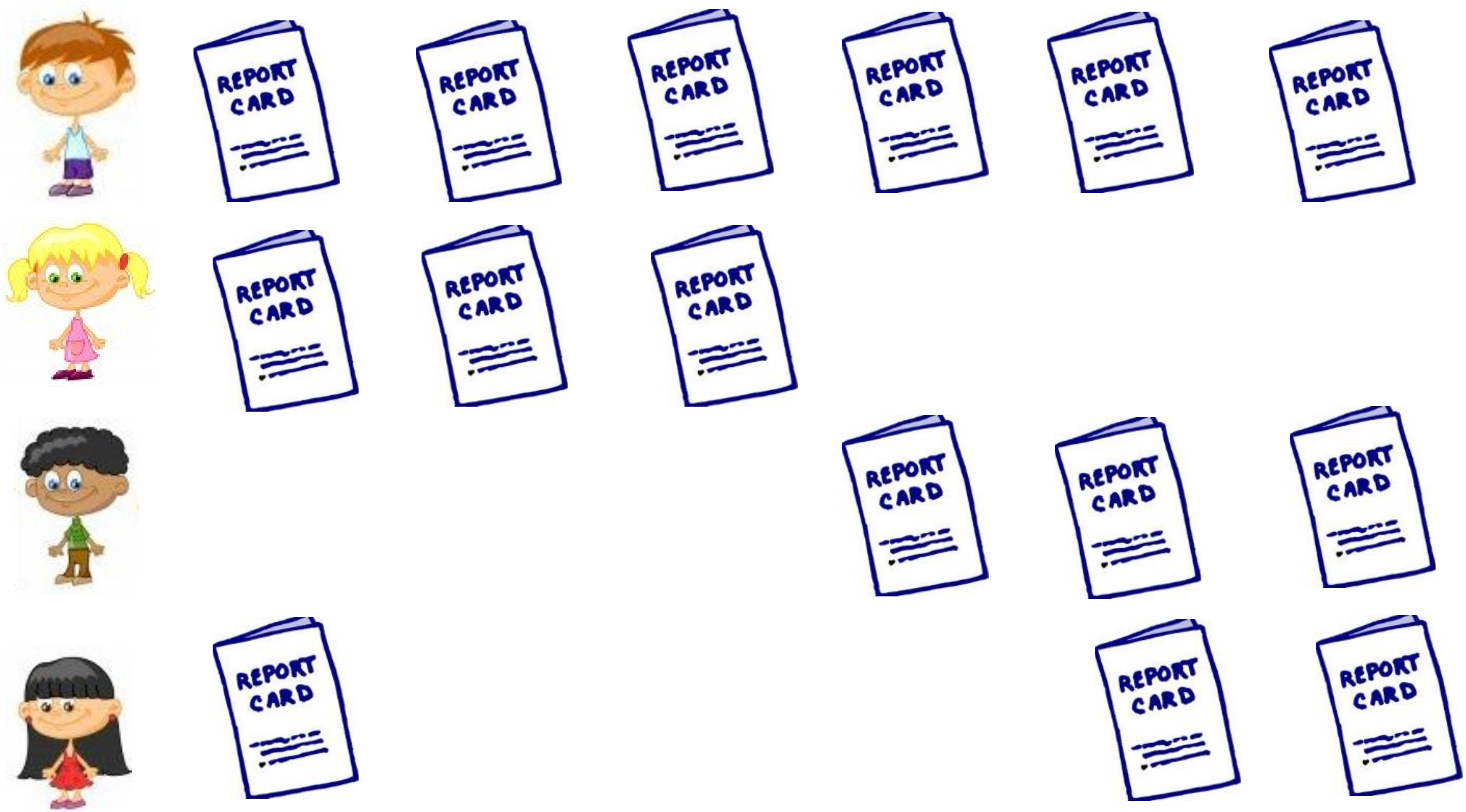
References: Multi-Level Modeling







Longitudinal Data Analysis

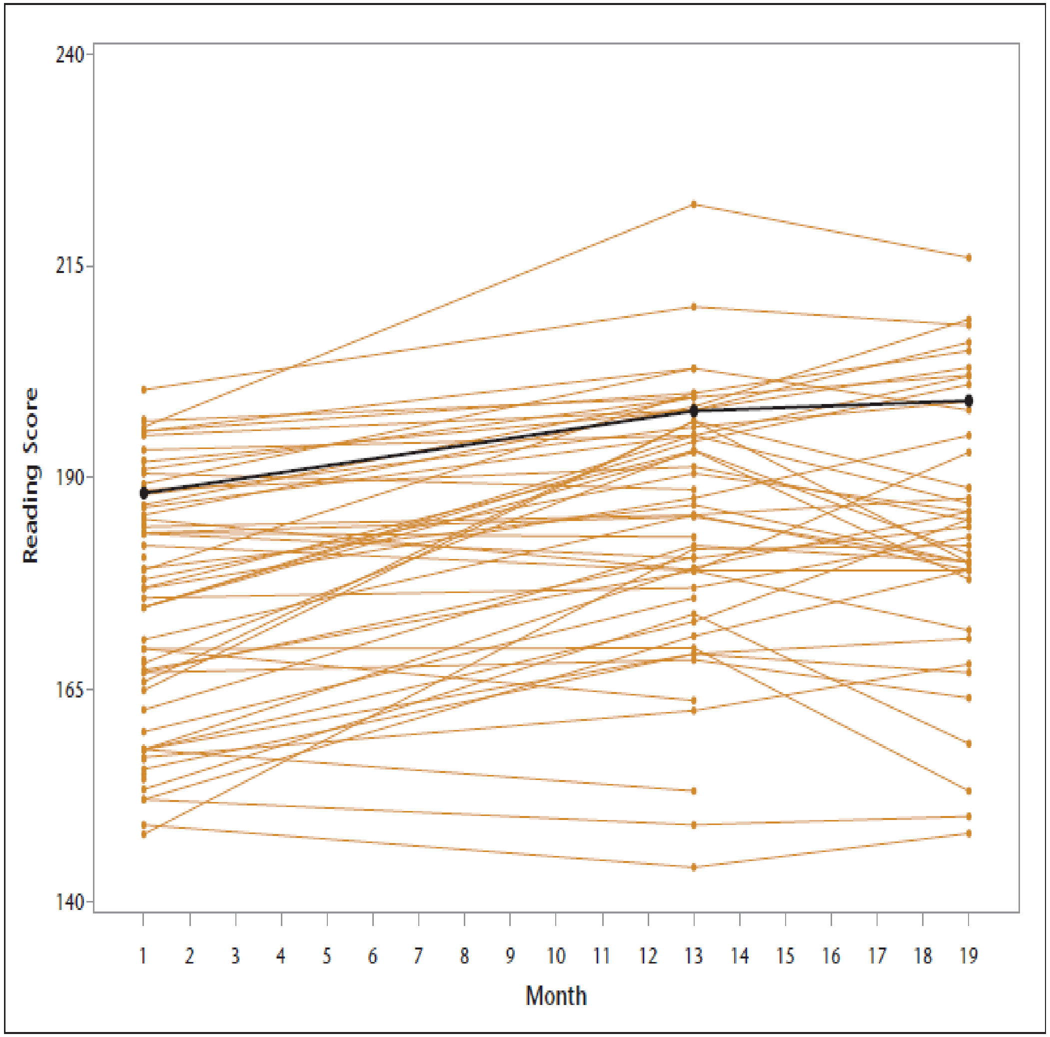


P1	P2	P3	P4	P5	P6

Why Longitudinal Data Analysis

1) Extension of multi-level modeling (e.g. measurements within individuals) - longitudinal data analysis decomposes inter- and intra-individual variability
2) Able to handle static and dynamic covariates
3) Able to handle categorical and continuous covariates
4) Accommodation of missing data
5) Accommodation of unequal spacing of time

Logic of Longitudinal Data Analysis



What is the shape of the mean growth curve (linear or not)?

Is there any difference in the intercept across individuals (groups)?

Is there any difference in the growth/change (slopes) across individuals (groups)

What variables can explain the individual differences in the intercept and growth/change?

Examples of Research Questions

- Educational Research (e.g. What is the relationship between teacher support and school engagement of students over time?)
- Organizational Research (e.g. What is the relationship between team leadership & staff performance across a year)
- Sociological Research (e.g. How does neighborhood segregation affect the mental health of individuals over the lifespan?)
- Psychological Research (e.g. How do parenting styles relate to the children's development of executive function skills)
- Communication Research (e.g. How do parent-child relationship relate to the youth's online risky behaviors from primary to secondary schools)

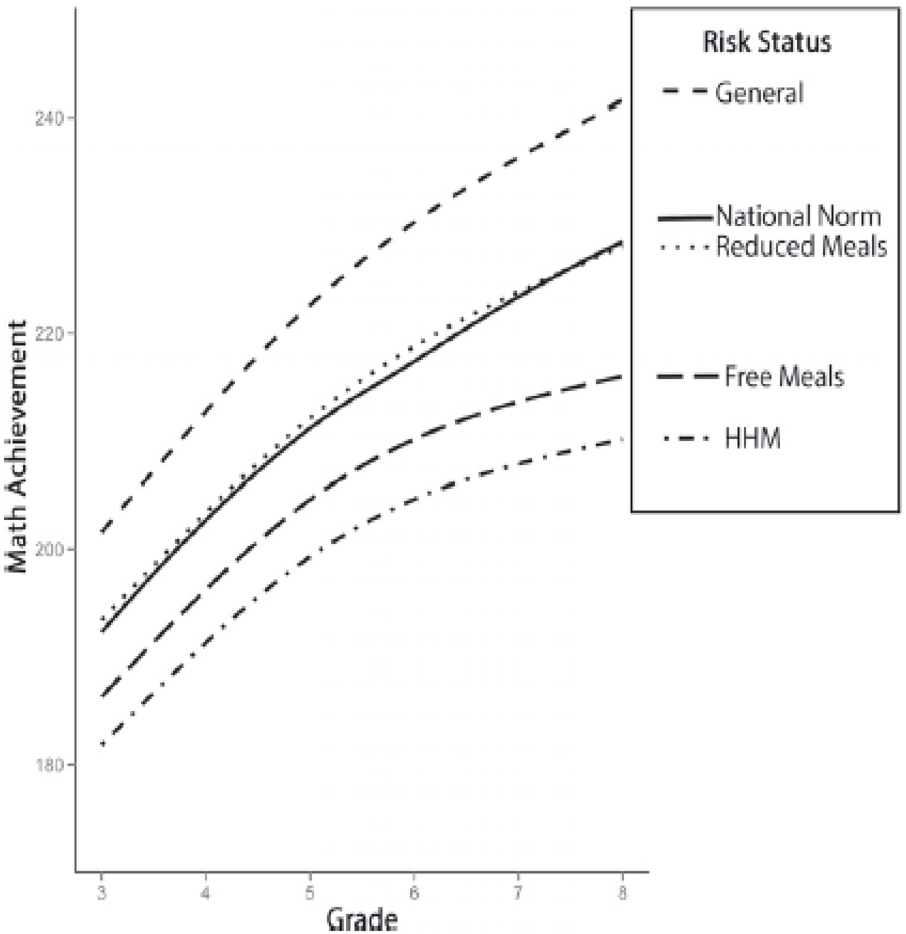
Illustration of Multi-level Modeling : Risk & Resilience of Homeless Children

Research Questions:

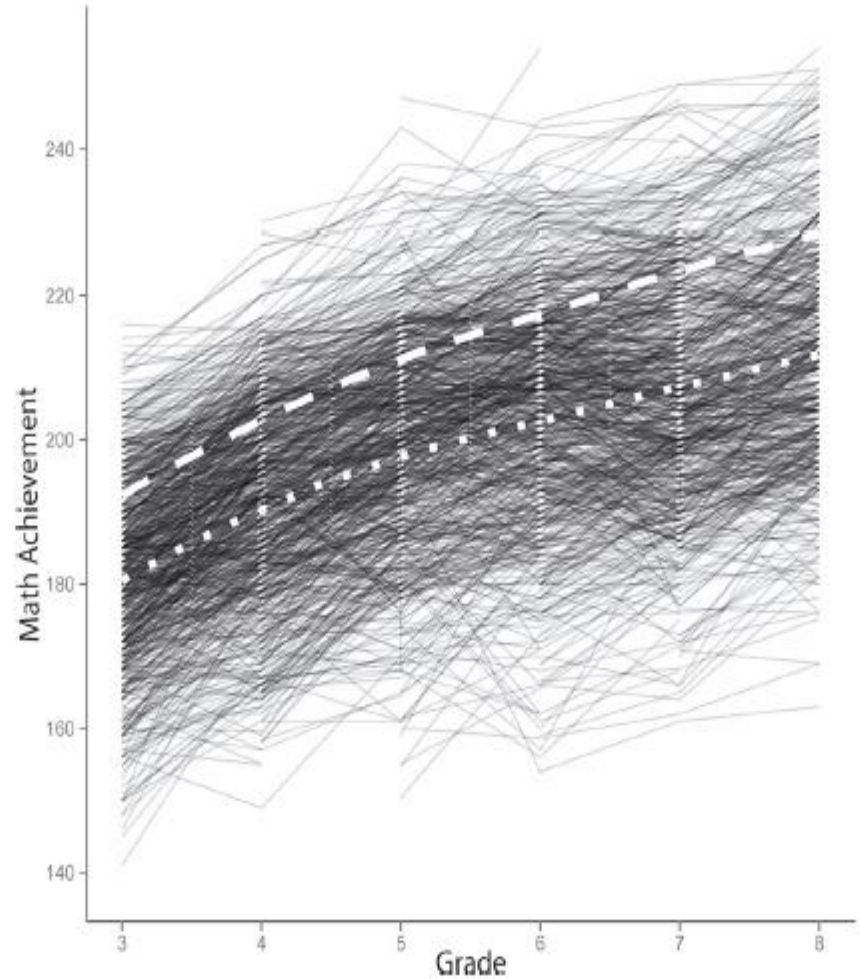
Are children experiencing homelessness more likely to be at-risk for their math achievement over time?

Data: 26,474 students (grades 3-8) and 13.8% of the sample experienced homelessness

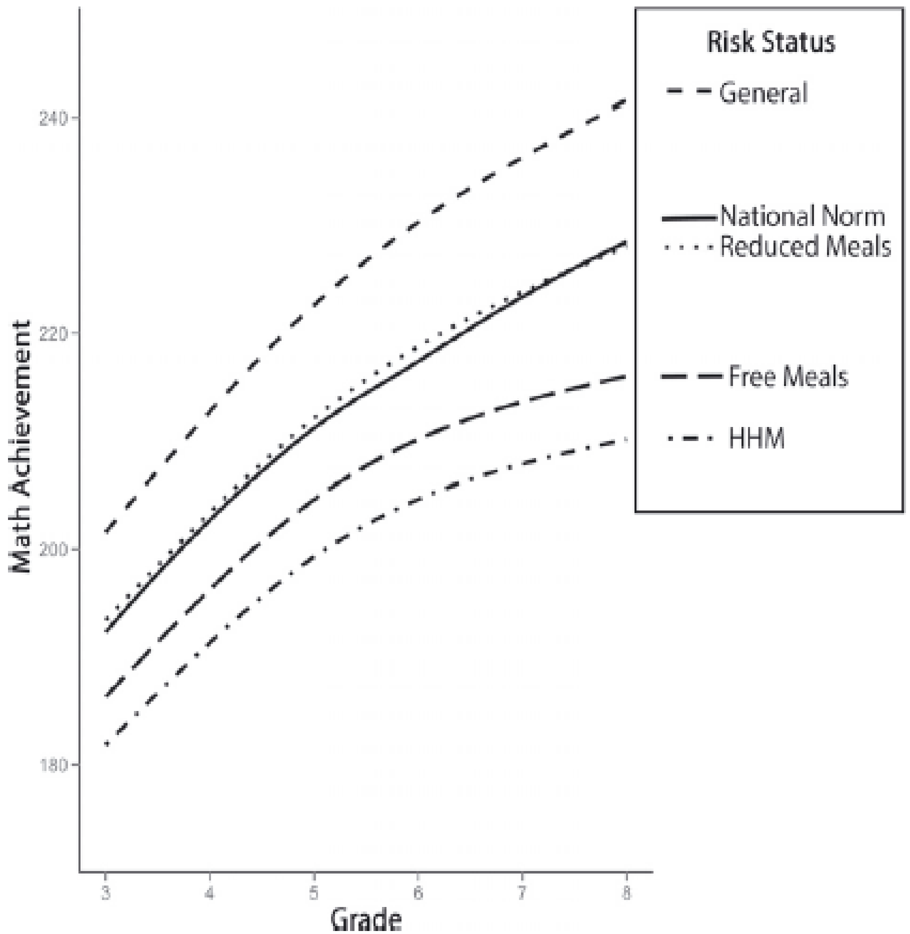
Source: Cutuli, J. J., Desjardins, C. D., Herbers, J. E., Long, J. D., Heistad, D., Chan, C-K., Hinz, E., & Masten, A. (2013). Academic achievement trajectories of homeless and highly mobile students: Resilience in the context of chronic and acute risk. *Child Development*, *84* (3), 841-857.

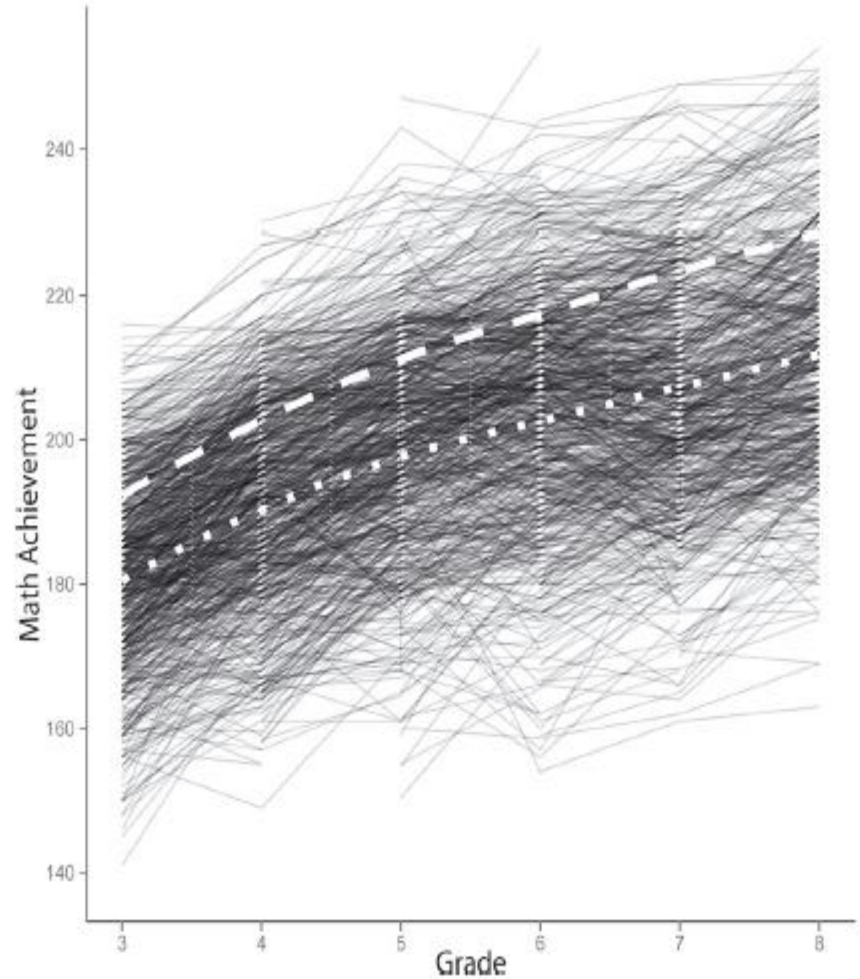


Risk Status
General
National Norm
Reduced Meals
Free Meals
HHM
Math Achievement
Grade
180
200
220
240
3
4
5
6
7
8



Math Achievement
Grade
140
160
180
200
220
240
3
4
5
6
7
8

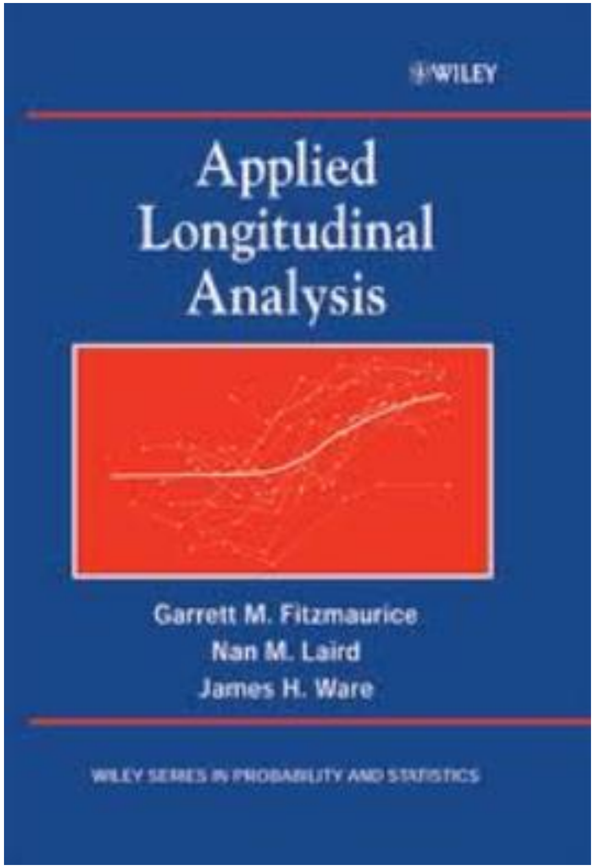

Risk Status
General
National Norm
Reduced Meals
Free Meals
HHM
Math Achievement
Grade
180
200
220
240
3
4
5
6
7
8

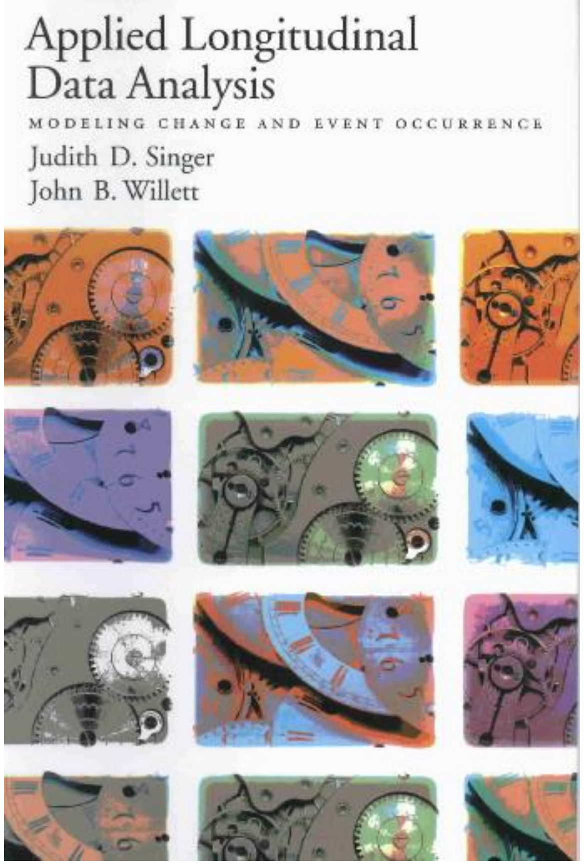

Math Achievement
Grade
140
160
180
200
220
240
3
4
5
6
7
8

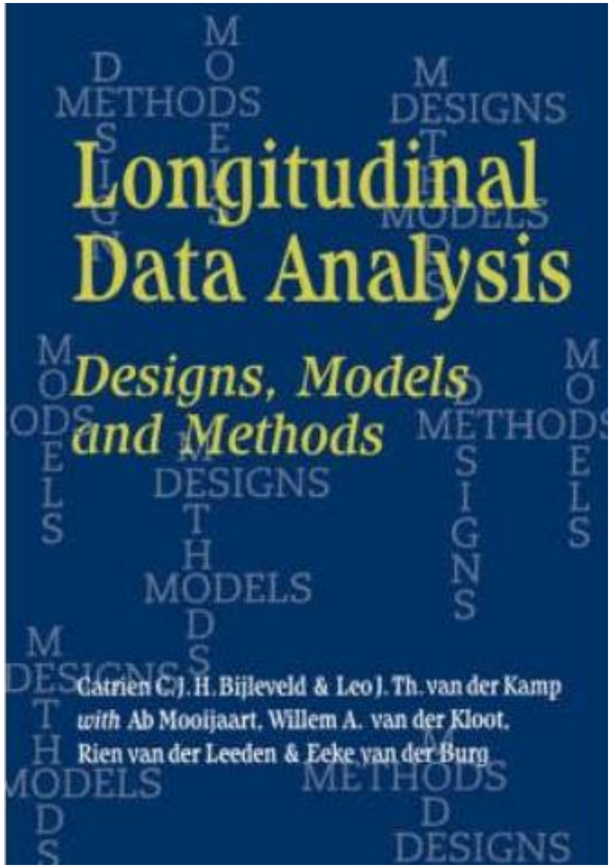
Risk effect	Curve	Math achievement AIC	ΔAIC	Weight
Static-risk models				
None	Linear	451,450	5,329	<.01
None	Log	447,976	1,855	<.01
None	Quadratic	447,662	1,541	<.01
Intercept	Linear	450,132	4,011	<.01
Intercept	Log	446,592	471	<.01
Intercept	Quadratic	446,266	145	<.01
Intercept, trajectory	Linear	449,968	3,847	<.01
Intercept, trajectory	Log	446,465	344	<.01
Intercept, trajectory	Quadratic	**446,121**	0	>.99

	Math achievement		
	Fixed effects		
	Intercept	Linear slope	Quadratic trajectory
Risk			
HHM vs. general	9.60 (0.39)	0.54 (0.27)	0.09 (0.05)
HHM vs. reduced	5.70 (0.56)	−0.16 (0.40)	0.09 (0.08)
HHM vs. free	2.80 (0.31)	−0.06 (0.22)	−0.01 (0.04)
Free vs. general[a]	6.80 (0.28)	0.60 (0.20)	0.10 (0.04)
Free vs. reduced[a]	2.90 (0.50)	−0.10 (0.35)	0.09 (0.07)
Reduced vs. general[a]	3.90 (0.51)	0.70 (0.36)	0.00 (0.07)
Ethnicity (White vs. . . .)			
American Indian	−6.66 (0.49)	−0.03 (0.34)	−0.08 (0.07)
African American	−8.61 (0.29)	−0.61 (0 20)	−0.07 (0 04)
Asian	−3.06 (0.42)	0.24 (0.29)	0.01 (0.06)
Hispanic	−5.13 (0.38)	0.45 (0.26)	−0.20 (0.05)
Sex (male vs. female)	−1.31 (0.19)	−0.32 (0.13)	0.03 (0.03)
ELL (no vs. yes)	−6.21 (0.31)	−0.99 (0.22)	0.15 (0.04)
Special ed. (no vs. yes)	−8.98 (0.24)	−0.92 (0.17)	−0.05 (0.03)
Attendance[b]	37.50 (2.48)	−2.79 (1.83)	1.32 (0.34)
Reference	159.06 (2.32)	13.53 (1 72)	−2.01 (0.32)
	Variance components		
Intercept (*SD*)	111.82 (10.57)		
Linear slope (*SD*)	8.55 (2.92)		
Quadratic slope (*SD*)	0.23 (0.48)		
Intercept, quadratic slope covar.	0.01		
σ^2	28.75 (5.36)		
	Model fit		
Akaike's information criterion	446,121		

References: Longitudinal Data Analysis







What is Structural Equation Modeling?

- Structural equation modeling (SEM) is a multivariate statistical modeling technique used specifically to examine relationships between a set of independent variables and dependent variables.
- In SEM, the hypothesized model needs to fit the covariance matrix of the data in order to obtain a good fit and is thus sometimes known as a covariance-based SEM technique.

Advantages of Structural Equation Modeling

- SEM offers the following advantages over other statistical techniques like multiple linear regression (MLR)
  - Measurement errors in variables;
  - Correlations between disturbance terms;
  - Recursive relations between variables;
  - Fit indicators and modification indicators for the construction of structural models.

Applications of Structural Equation Modeling

- Major applications of SEM include
  - Causal modeling or path analysis;
  - Confirmatory factor analysis (CFA);
  - Second order factor analysis;
  - Covariance structure models;
  - Correlation structure models.

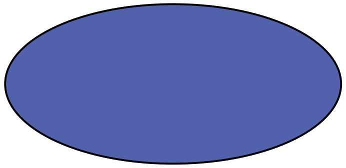
A Four-Stage General Modeling Process

- Model specification.
- Model estimation.
- Model evaluation.
- Model modification.

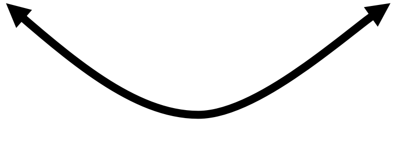
Terminology and Symbols

- Observed/manifest variable

- Latent variable



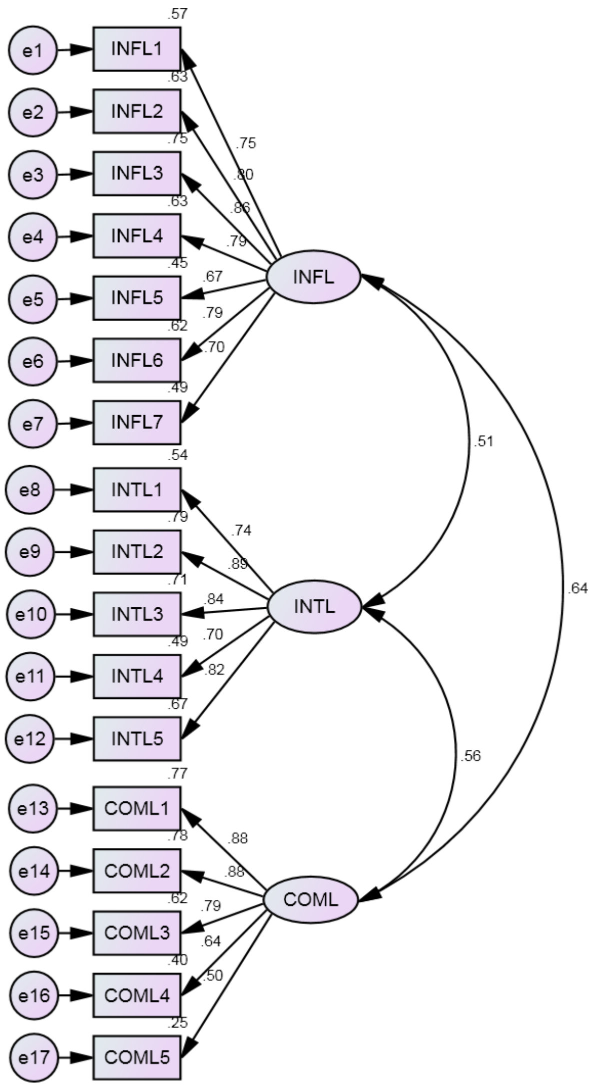
Terminology and Symbols

- Exogenous variable — Independent variable
- Endogenous variable — Dependent variable
- Direct effects 
- Reciprocal effects 
- Correlation or covariance 

Terminology and Symbols

- Measurement model
  - Relationship between latent variables and indicators.
- Structural model
  - Relationship between exogenous and endogenous variables in the model.

An example of SEM: Confirmatory Factor Analysis (CFA)



Source: Lau, W. W. F., & Yuen, A. H. K. (2014). Developing and validating of a perceived ICT literacy scale for junior secondary school students: Pedagogical and educational contributions. *Computers & Education, 78*, 1-9.

This study developed and validated a three-factor, 17-item perceived ICT literacy scale (3F-PICTLS) assessing information literacy (information), internet literacy (communication), and computer literacy (technology) for junior secondary school students in Hong Kong.

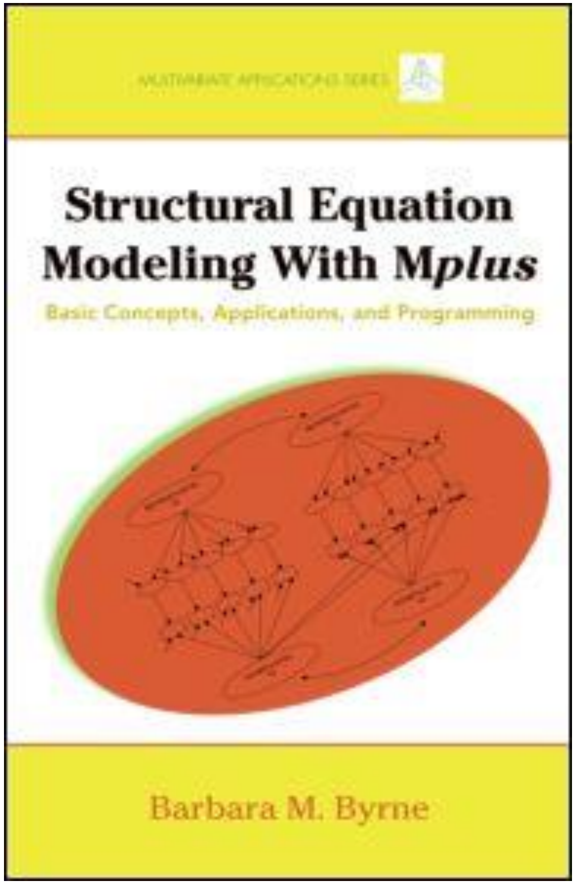
Goodness-of-Fit Indexes:
χ^2/df = 2.244, CFI = .964,
TLI = .958, and RMSEA = .057

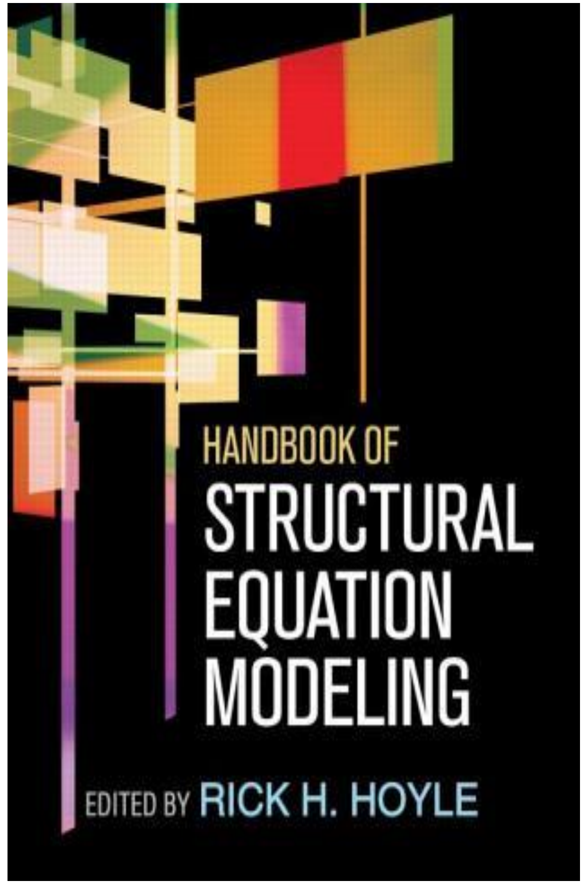
An example of SEM: Confirmatory Factor Analysis (CFA)

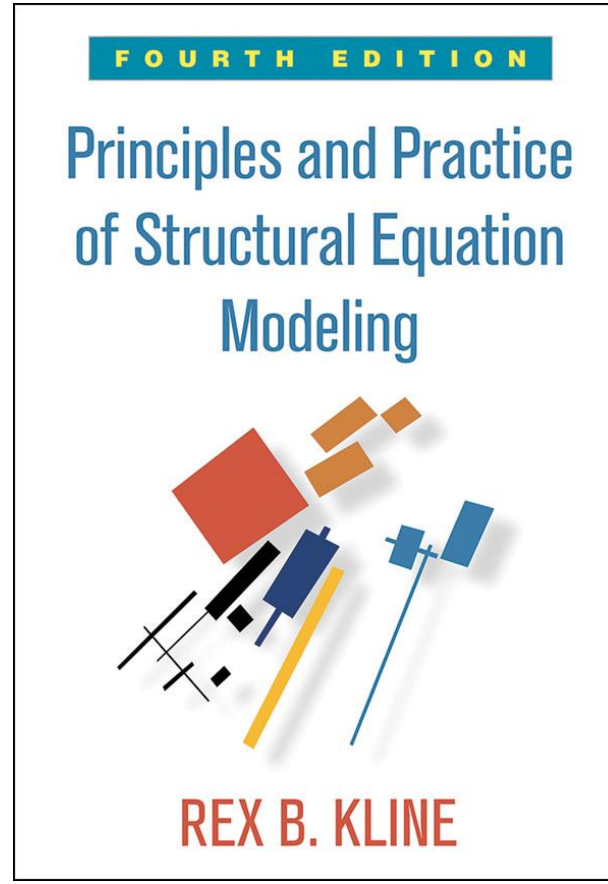
Results of the CFA (n= 413) of the 17-item perceived ICT literacy scale.
a this value was fixed at 1.00 for model identification purpose and thus no critical ratio was calculated.

Item	Unstandardized estimate	Standardized estimate	t-value	R^2	α
INFL					.908
INFL1	1	0.752	a	.566	
INFL2	1.229	0.795	16.002	.632	
INFL3	1.267	0.863	17.549	.745	
INFL4	1.162	0.793	15.961	.629	
INFL5	1.05	0.667	13.165	.445	
INFL6	1.202	0.79	15.878	.624	
INFL7	1.018	0.701	13.899	.491	
INTL					.890
INTL1	1	0.737	a	.544	
INTL2	0.951	0.89	17.411	.792	
INTL3	1.065	0.84	16.446	.706	
INTL4	1.021	0.703	13.607	.494	
INTL5	1.075	0.816	15.947	.665	
COML					.844
COML1	1	0.88	a	.774	
COML2	1.01	0.881	22.688	.776	
COML3	0.95	0.788	19.015	.621	
COML4	0.812	0.636	13.865	.404	
COML5	0.767	0.503	10.309	.253	

References: Structural Equation Modeling







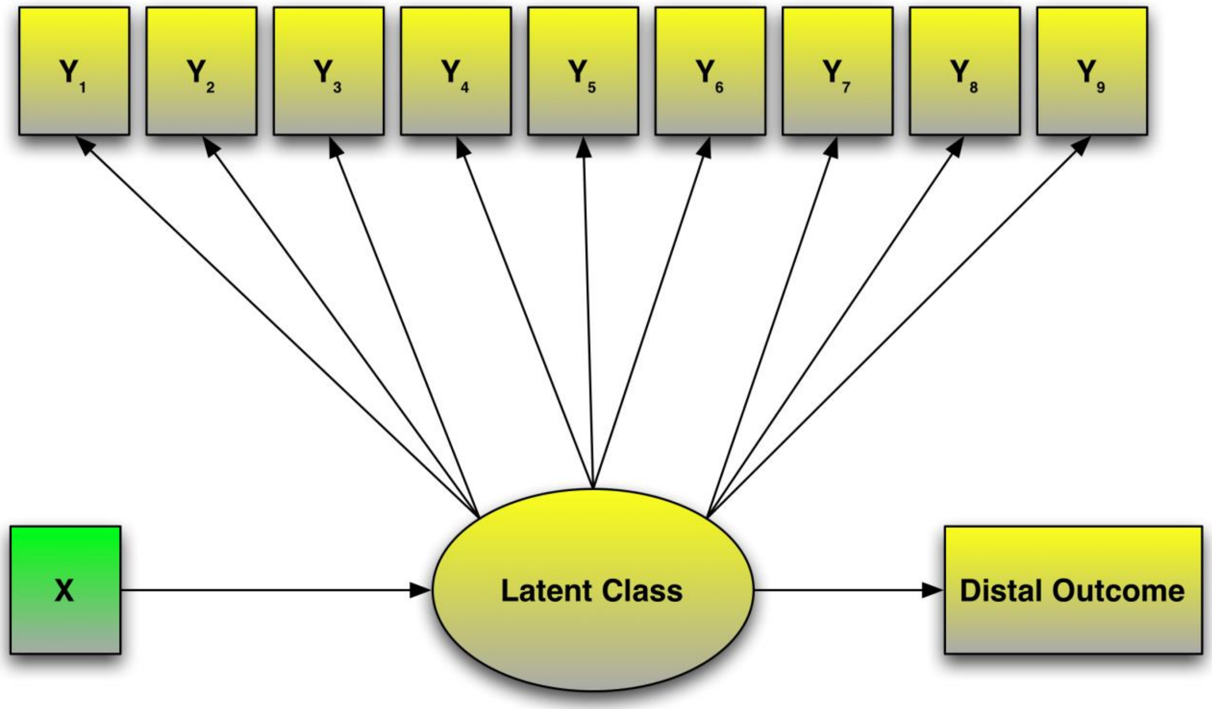
What is Latent Class Analysis?

- Latent class analysis (LCA) allows researchers to determine whether there are unobserved meaningful latent classes of individuals based on their responses to the items in an inventory.
- LCA is an example of statistical procedures using the person-centred approach.

Variable-centred versus Person-centred Approaches

- The variable-centred approach analyses the relationships between variables with the assumption that such relationships are generalizable to a homogenous population.
- The person-centred approach takes the view that there are individual differences to the variables under consideration. It assumes that unobserved subgroups of population exist (heterogeneous population) and that findings can only be generalized to certain class or cluster in a population.

A General Latent Class Analysis Model



- Indicators(Y_i)
- Covariates(X_i)

Steps in Latent Class Analysis

- Determine the number of classes of individuals.
- Identify the characteristics of individuals within a class.
- Estimate the prevalence of the classes.
- Classify individuals into classes.

An Example of Latent Class Analysis

- Source: Lau, W. W. F., Yuen, A. H. K., & Chan, A. (2015). Variable-centered and person-centered approaches to studying the VARK learning style inventory. In W. W. K. Ma, A. H. K. Yuen, J. Park, W. W. F. Lau, & L. Deng (Eds.), *New Media, Knowledge Practices and Multiliteracies – HKAECT 2014 International Conference (pp. 207-216).* Springer-Verlag: Berlin Heidelberg.
- The study attempted to use LCA to identify any unobserved meaningful latent subgroups of adolescents based on their response patterns to the items in a learning style inventory called VARK (visual, aural, read/write, and kinesthetic).

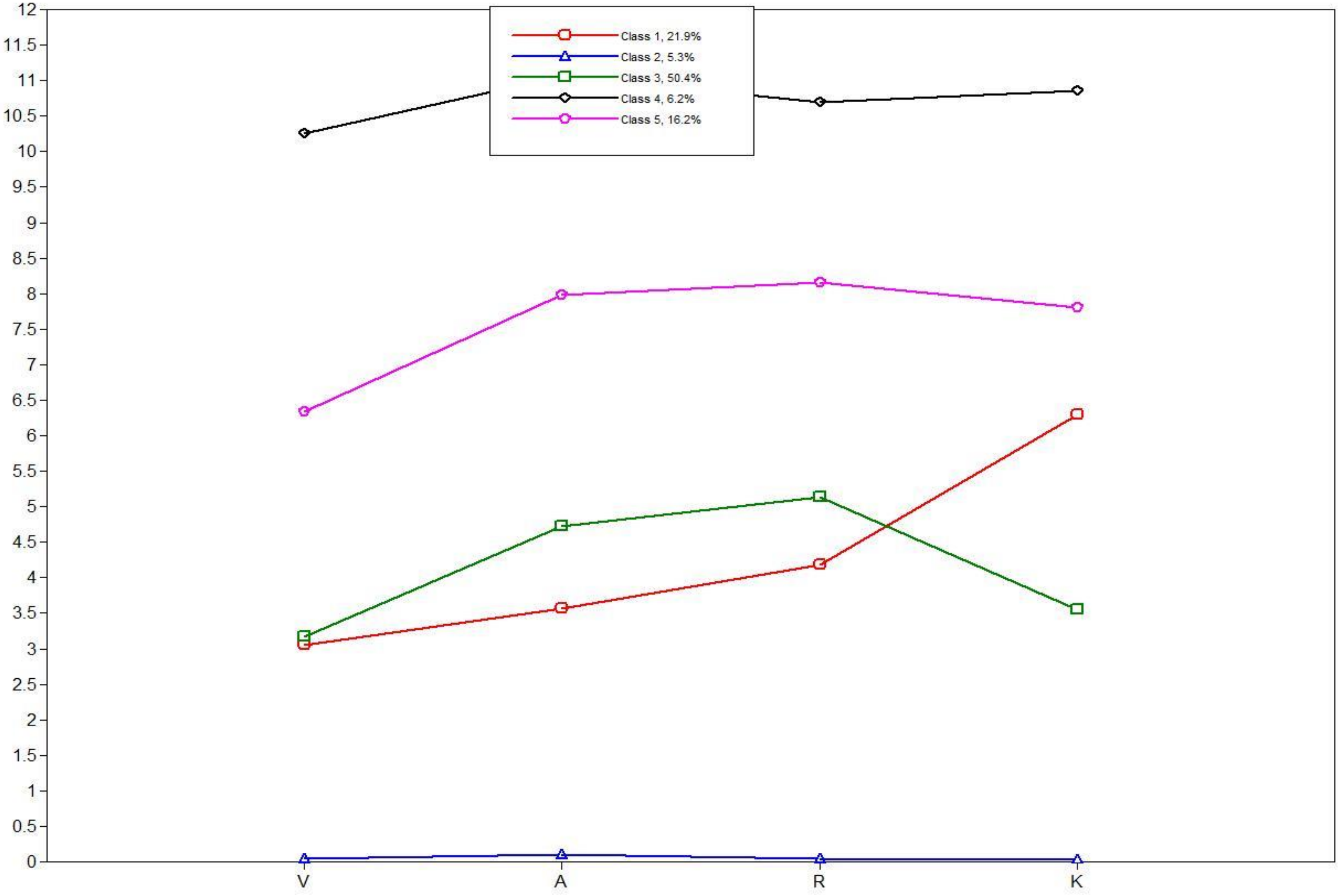
Model fit indexes for the 1-, 2-, 3-, 4-, 5-, and 6-class solutions

	1-class	2-class	3-class	4-class	5-class	6-class
AIC	16074.780	15040.016	14712.290	14557.275	14512.161	14470.451
BIC	16112.326	15101.029	14796.769	14665.222	14643.574	14625.331
Sample-Size Adjusted BIC	16086.922	15059.747	14739.609	14592.183	14554.658	14520.536
VLMRLRT	n/a	0.0000	0.0000	0.0013	0.0212	0.2633
LMRALRT	n/a	0.0000	0.0000	0.0015	0.0240	0.2738

AIC, Akaike Information Criterion; BIC, sample size-adjusted Bayesian Information Criterion;
VLMRLRT, Vuong-Lo-Mendell-Rubin Likelihood Ratio Test; LMRALRT, Lo-Mendell-Rubin Adjusted Likelihood Ratio Test

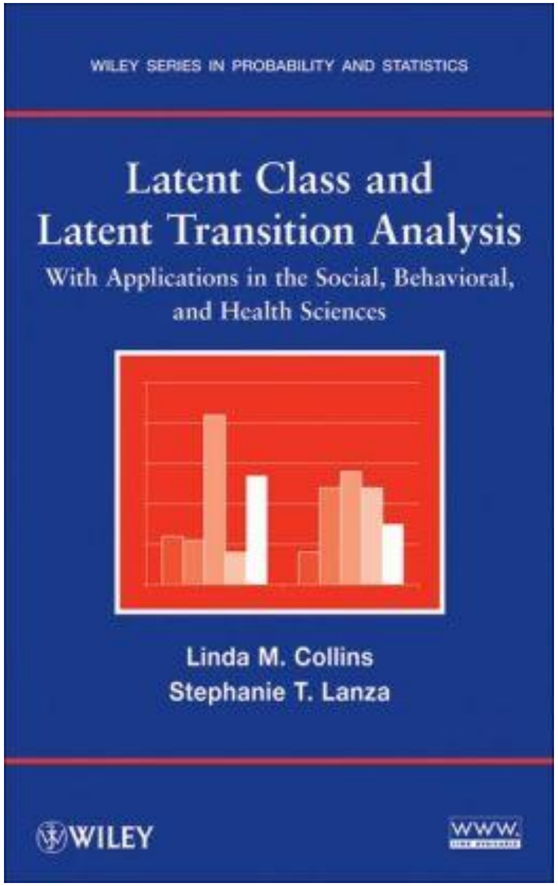
An Example of Latent Class Analysis

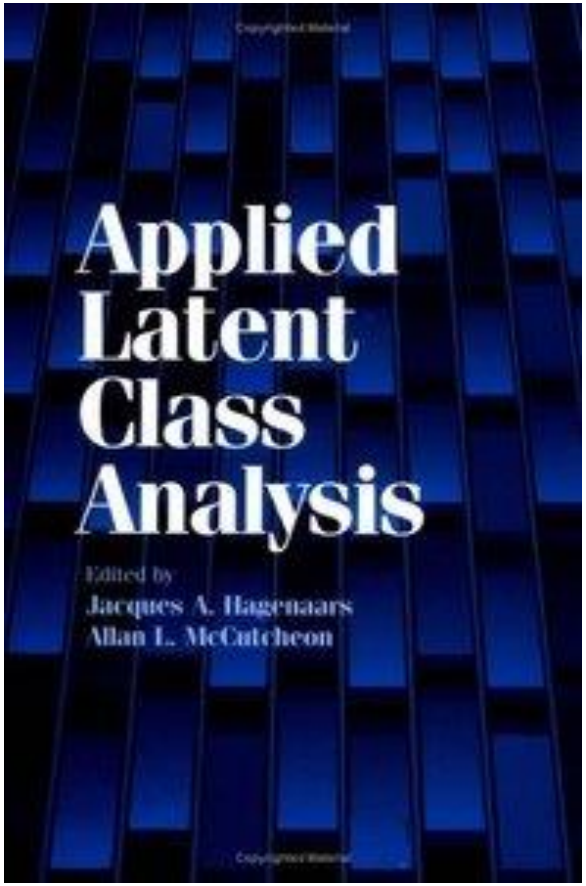
Latent class solution with five classes

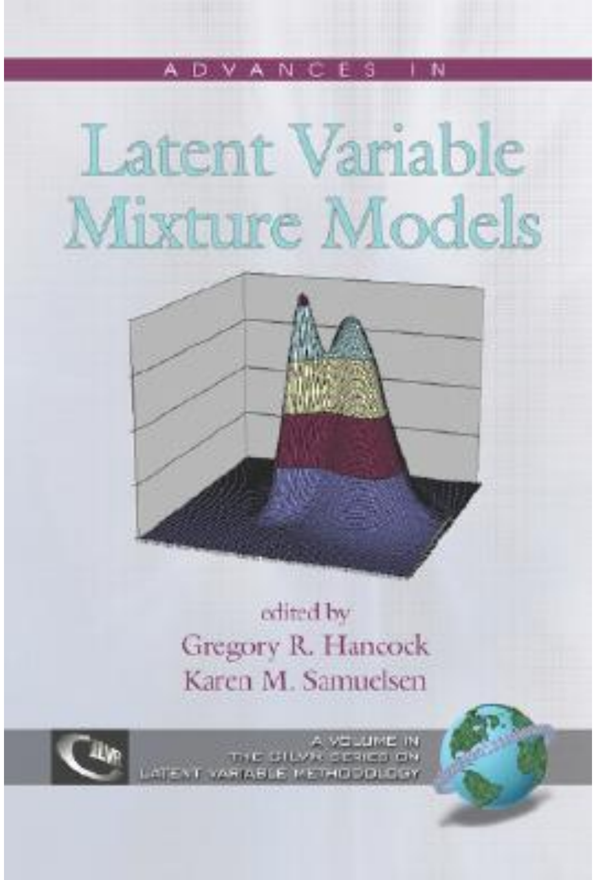


Class 1: kinesthetic oriented
Class 2: uninvolved
Class 3: read oriented
Class 4: all rounded
Class 5: mediocre

References: Latent Class Analysis







Q & A Session