Mplus Short Courses
Day 5B

Multilevel Modeling With Latent Variables Using Mplus

Linda K. Muthén
Bengt Muthén

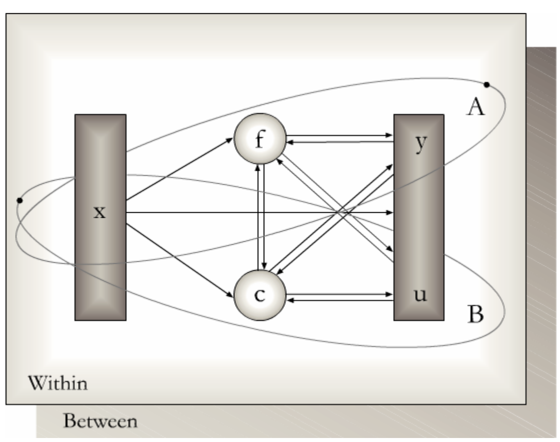
Table Of Contents

Mplus Background

- Inefficient dissemination of statistical methods:
  - Many good methods contributions from biostatistics, psychometrics, etc are underutilized in practice
- Fragmented presentation of methods:
  - Technical descriptions in many different journals
  - Many different pieces of limited software
- Mplus: Integration of methods in one framework
  - Easy to use: Simple, non-technical language, graphics
  - Powerful: General modeling capabilities
- Mplus versions
  - V1: November 1998
  - V2: February 2001
  - V3: March 2004
  - V4: February 2006
- Mplus team: Linda & Bengt Muthén, Thuy Nguyen, Tihomir Asparouhov, Michelle Conn

3

General Latent Variable Modeling Framework



4

Mplus

Several programs in one

- Structural equation modeling
- Item response theory analysis
- Latent class analysis
- Latent transition analysis
- Survival analysis
- Multilevel analysis
- Complex survey data analysis
- Monte Carlo simulation

Fully integrated in the general latent variable framework

5

Overview

Single-Level Analysis

	Cross-Sectional	Longitudinal
Continuous Observed And Latent Variables	***Day 1*** Regression Analysis Path Analysis Exploratory Factor Analysis Confirmatory Factor Analysis Structural Equation Modeling	***Day 2*** Growth Analysis
Adding Categorical Observed And Latent Variables	***Day 3*** Regression Analysis Path Analysis Exploratory Factor Analysis Confirmatory Factor Analysis Structural Equation Modeling Latent Class Analysis Factor Mixture Analysis Structural Equation Mixture Modeling	***Day 4*** Latent Transition Analysis Latent Class Growth Analysis Growth Analysis Growth Mixture Modeling Discrete-Time Survival Mixture Analysis Missing Data Analysis

6

Overview (Continued)

Multilevel Analysis

	Cross-Sectional	Longitudinal
Continuous Observed And Latent Variables	***Day 5*** Regression Analysis Path Analysis Exploratory Factor Analysis Confirmatory Factor Analysis Structural Equation Modeling	***Day 5*** Growth Analysis
Adding Categorical Observed And Latent Variables	***Day 5*** Latent Class Analysis Factor Mixture Analysis	***Day 5*** Growth Mixture Modeling

7

Regression Mixture Analysis

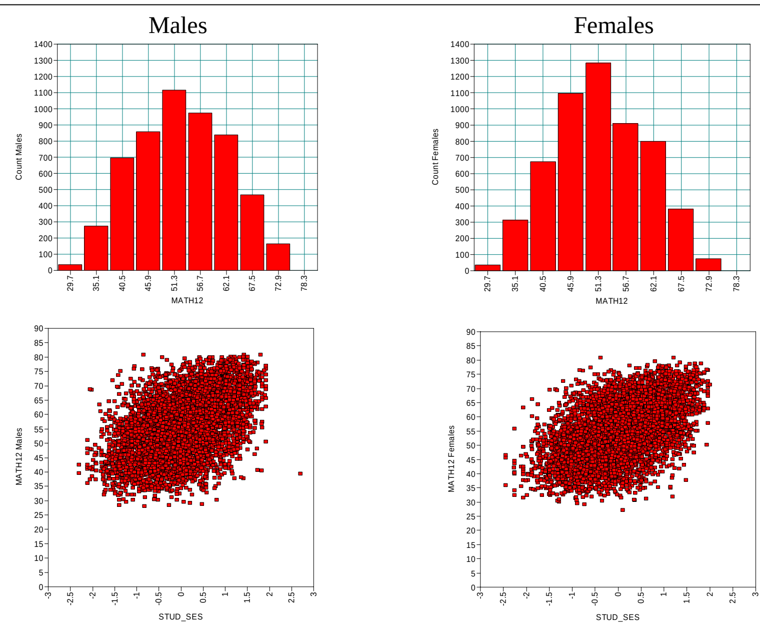
8

Two-Level Data

- Education studies of students within schools
  - LSAY (3,000 students in 54 schools, grades 7-12)
  - NELS (14,000 students in 900 schools, grades 8-12),
  - ECLS (22,000 students in 1,000 schools, K- grade 8)
- Public health studies of patients within hospitals, individuals within counties

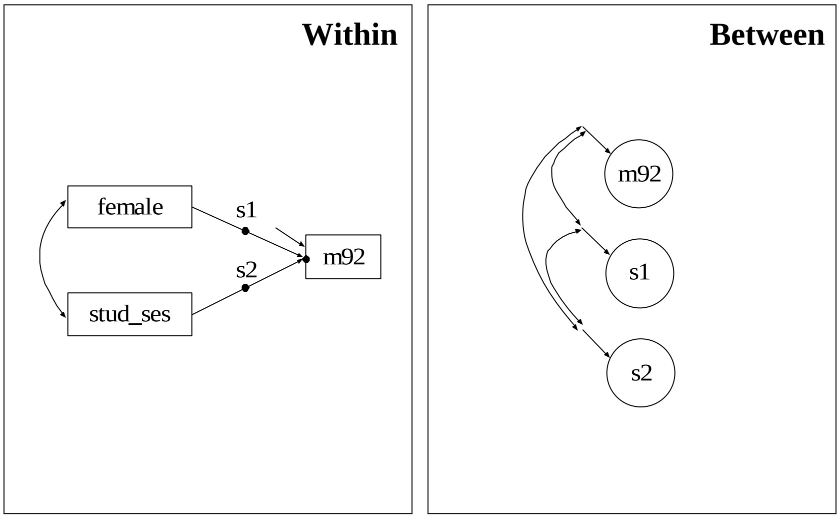
9

NELS Data: Grade 12 Math Related To Gender And SES



10

NELS Two-Level Math Achievement Regression



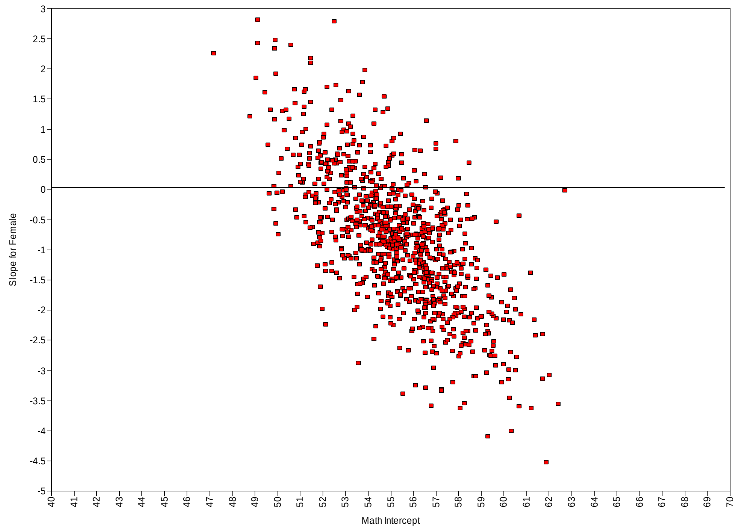
11

Output Excerpts NELS Two-Level Regression

	Estimates	S.E.	Est./S.E.
Between Level			
Means			
M92	55.279	0.174	317.706
S_FEMALE	-0.850	0.188	-4.507
S_SES	5.450	0.132	41.228
Variances			
M92	11.814	1.197	9.870
S_FEMALE	5.762	1.426	4.041
S_SES	0.905	0.538	1.682
S_FEMALE WITH			
M92	-4.936	1.071	-4.610
S_SES	0.068	0.635	0.107
S_SES WITH			
M92	1.314	0.541	2.431

12

Random Effect Estimates For Each School: Slopes For Female Versus Intercepts For Math



13

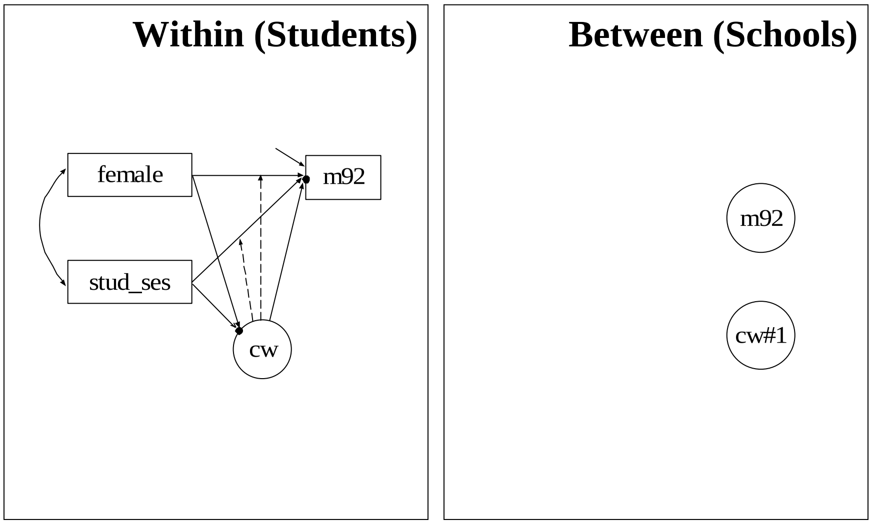
Is The Conventional Two-Level Regression Model Sufficient?

- Conventional Two-Level Regression of Math Score Related to Gender and Student SES
  - Loglikelihood = -39,512, number of parameters = 10, BIC = 79,117
- New Model
  - Loglikelihood = -39,368, number of parameters = 12, BIC = 78,848

  - Which model would you choose?

14

Two-Level Regression With Latent Classes For Students



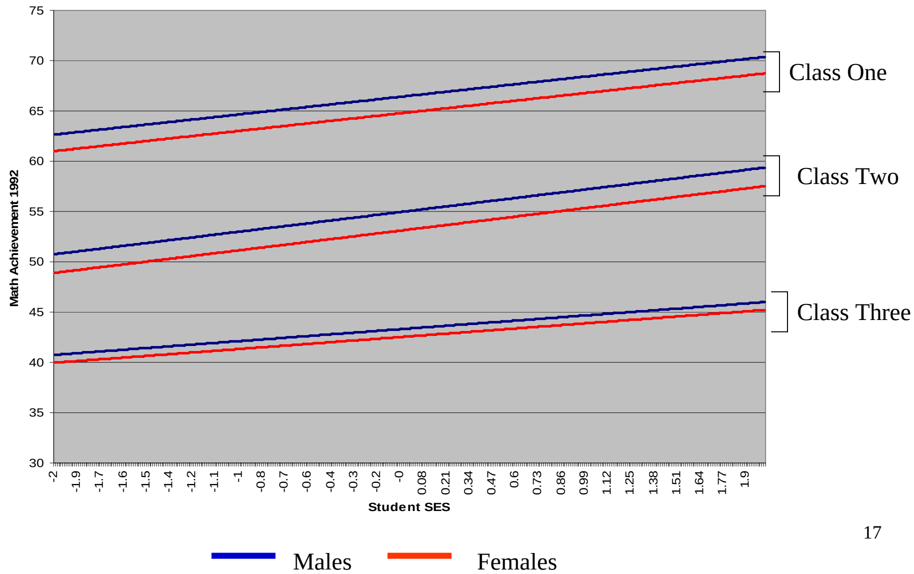
15

Model Results For NELS Two-Level Regression Of Math Score Related To Gender And Student SES

Model	Loglikelihood	# parameters	BIC
(1) Conventional 2-level regression with random intercepts and random slopes	-39,512	10	79,117
(2) Two-level regression mixture, 2 latent classes for students	-39,368	12	78,848
(3) Two-level regression mixture, 3 latent classes for students	**-39,280**	**19**	**78,736**

16

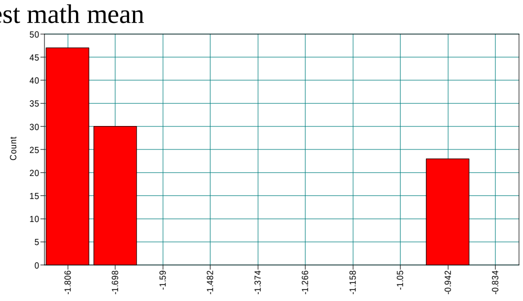
NELS: Estimated Three-Class Model



17

Estimated Two-Level Regression Mixture With 3 Latent Classes For Students

- Estimated Female slope means for the 3 latent classes for students do not include positive values.
- The class with the least Female disadvantage (right-most bar) has the lowest math mean



- Significant between-level variation in cw (the random mean of the latent class variable for students): Schools have a significant effect on latent class membership for students

18

Input For Two-Level Regression With Latent Classes For Students

```
TITLE:     NELS 2-level regression
DATA:      FILE = comp.dat;
           FORMAT = 2f7.0 f11.4 13f5.2 79f8.2 f11.7;
VARIABLE:
           NAMES = school m92 female stud_ses;
           CLUSTER = school;
           USEV = m92 female stud_ses;
           WITHIN = female stud_ses;
           CENTERING = GRANDMEAN(stud_ses);
           CLASSES = cw(3);
ANALYSIS:
           TYPE = TWOLEVEL MIXTURE MISSING;
           PROCESS = 2;
           INTERACTIVE = control.dat;
           !STARTS = 1000 100;
           STARTS = 0;
```

19

Input For Two-Level Regression With Latent Classes For Students (Continued)

```
MODEL:
           %WITHIN%
           %OVERALL%
           m92 ON female stud_ses;
           cw#1-cw#2 ON female stud_ses;
   ! [m92] class-varying by default
           %cw#1%
           m92 ON female stud_ses;
           %cw#2%
           m92 ON female stud_ses;
           %cw#3%
           m92 ON female stud_ses;
           %BETWEEN%
           %OVERALL%
           f BY cw#1 cw#2;
```

20

Cluster-Randomized Trials And NonCompliance

21

Randomized Trials With NonCompliance

- Tx group (compliance status observed)
  - Compliers
  - Noncompliers
- Control group (compliance status unobserved)
  - Compliers
  - NonCompliers

Compliers and Noncompliers are typically not randomly equivalent subgroups.

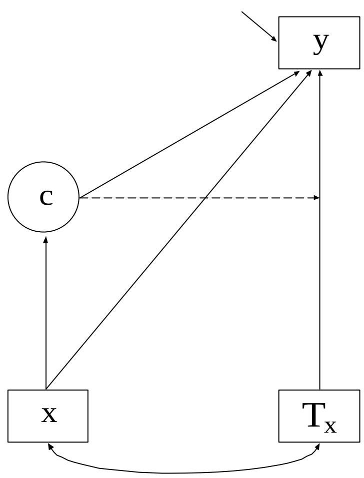
Four approaches to estimating treatment effects:

1. Tx versus Control (Intent-To-Treat; ITT)
2. Tx Compliers versus Control (Per Protocol)
3. Tx Compliers versus Tx NonCompliers + Control (As-Treated)
4. Mixture analysis (Complier Average Causal Effect; CACE):
   - Tx Compliers versus Control Compliers
   - Tx NonCompliers versus Control NonCompliers

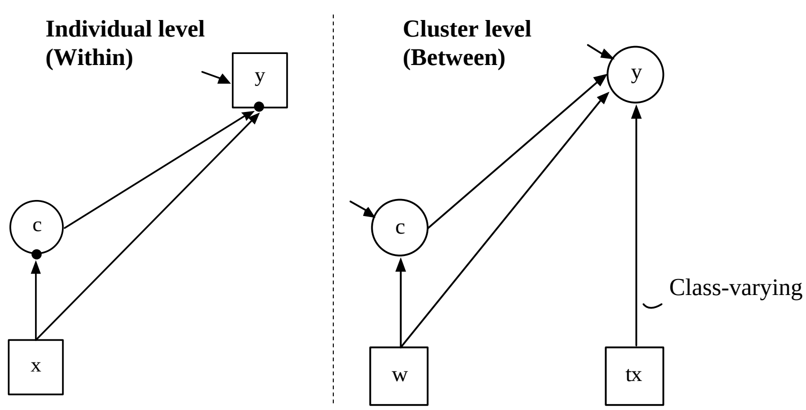
CACE: Little & Yau (1998) in Psychological Methods

22

Randomized Trials with NonCompliance: Complier Average Causal Effect (CACE) Estimation



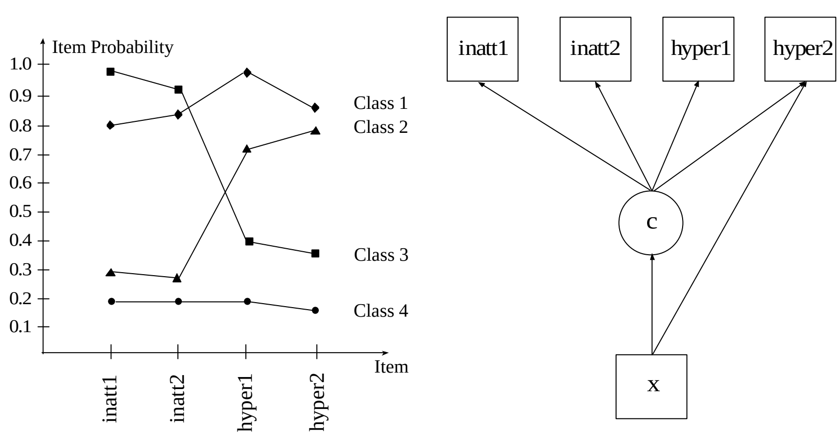
Two-Level Regression Mixture Modeling: Cluster-Randomized CACE



Latent Class Analysis

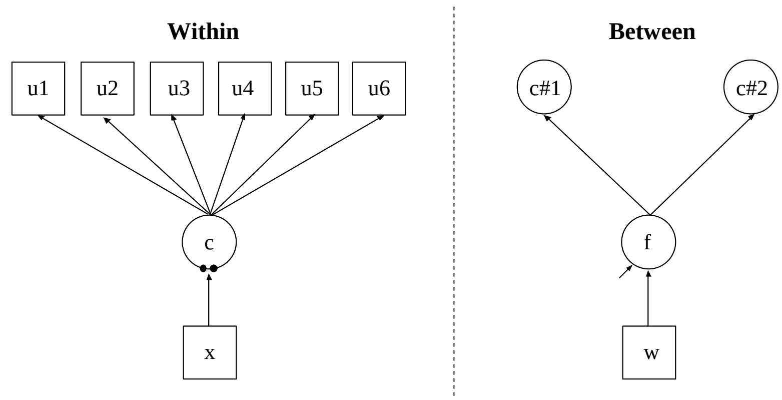
25

Latent Class Analysis



26

Two-Level Latent Class Analysis



Input For Two-Level Latent Class Analysis

```
TITLE:      this is an example of a two-level LCA with
            categorical latent class indicators
DATA:       FILE IS ex10.3.dat;
VARIABLE:   NAMES ARE u1-u6 x w c clus;
            USEVARIABLES = u1-u6 x w;
            CATEGORICAL = u1-u6;
            CLASSES = c (3);
            WITHIN = x;
            BETWEEN = w;
            CLUSTER = clus;
ANALYSIS:   TYPE = TWOLEVEL MIXTURE;
```

Input For Two-Level Latent Class Analysis (Continued)

```
MODEL:      %WITHIN%
            %OVERALL%
            c#1 c#2 ON x;

            %BETWEEN%
            %OVERALL%
            f BY c#1 c#2;
            f ON w;
OUTPUT:     TECH1 TECH8;
```

29

Latent Transition Analysis

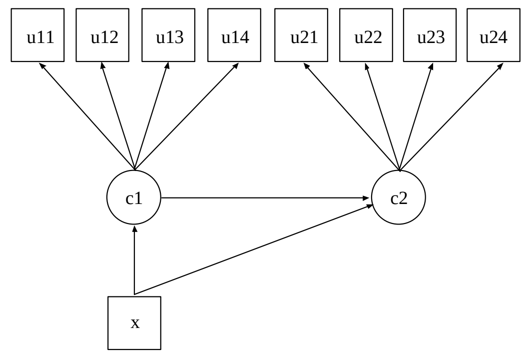
30

Latent Transition Analysis

Transition Probabilities

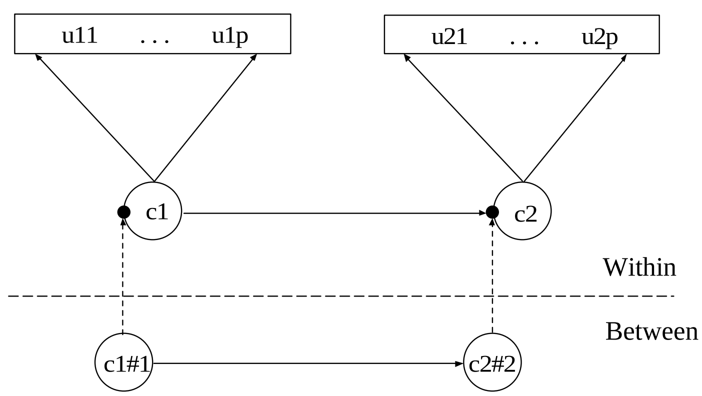
		c2: 1	c2: 2
c1	1	0.8	0.2
	2	0.4	0.6

Time Point 1 Time Point 2



31

Two-Level Latent Transition Analysis



Asparouhov, T. & Muthen, B. (2006). Multilevel mixture models. Forthcoming in Hancock, G. R., & Samuelsen, K. M. (Eds.). (2007). Advances in latent variable mixture models. Charlotte, NC: Information Age Publishing, Inc.

32

Input For Two-Level LTA

```
        CLUSTER = classrm;
        USEVAR = stub1f bkrule1f bkthin1f-tease1f athort1f
                 stub1s bkrule1s bkthin1s-tease1s athort1s;
        CATEGORICAL =  stub1f-athort1s;
        MISSING = all (999);
        CLASSES = cf(2) cs(2);
DEFINE:
        CUT stub1f-athort1s(1.5);
ANALYSIS:
        TYPE = TWOLEVEL MIXTURE MISSING;
        PROCESS = 2;
MODEL:
        %WITHIN%
        %OVERALL%
        cs#1 ON cf#1;
        %BETWEEN%
        OVERALL%
        cs#1 ON cf#1;
        cs#1*1 cf#1*1;
```


Input For Two-Level LTA (Continued)

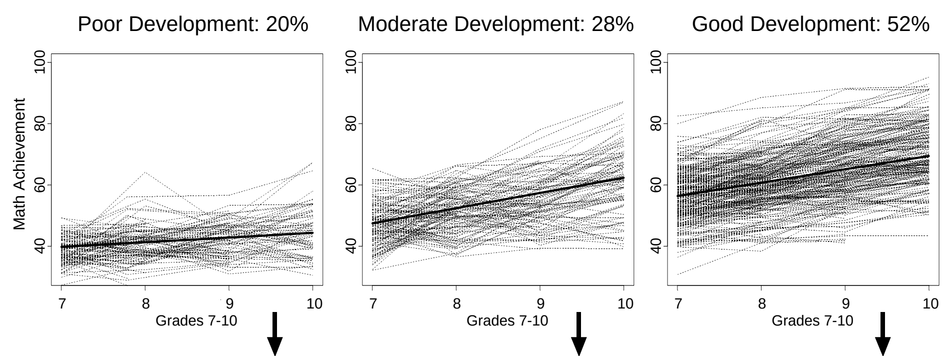
```
MODEL cf:
   %BETWEEN%
        %cf#1%
        [stub1f$1-athort1f$1] (1-9);
        %cf#2%
        [stub1f$1-athort1f$1] (11-19);
MODEL cs:
   %BETWEEN%
        %cs#1%
        [stub1s$1-athort1s$1] (1-9);
        %cs#2%
        [stub1s$1-athort1s$1] (11-19);
OUTPUT:
        TECH1 TECH8;
PLOT:
        TYPE = PLOT3;
        SERIES = stub1f-athort1f(*);
```

Output Excerpts Two-Level LTA

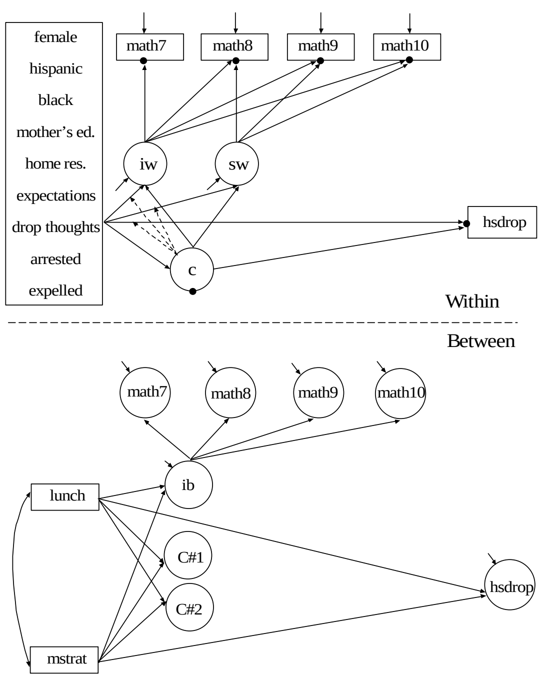
```
Categorical Latent Variables
Within Level
  CS#1     ON
      CF#1                  3.938      0.407      9.669
Means
      CF#1                 -0.126      0.189     -0.664
      CS#1                 -1.514      0.221     -6.838
Between Level
  CS#1     ON
      CF#1                  0.411       0.15      2.735
Variances
      CF#1                  2.062      0.672      3.067
Residual Variances
      CS#1                  0.469      0.237      1.984
```


Multilevel Growth Mixture Modeling

Growth Mixture Modeling: LSAY Math Achievement Trajectory Classes And The Prediction Of High School Dropout



Dropout:	69%	8%	1%



Input For A Multilevel Growth Mixture Model For LSAY Math Achievement

```
TITLE:      multilevel growth mixture model for LSAY math
            achievement

DATA:       FILE = lsayfull_Dropout.dat;

VARIABLE:   NAMES = female mothed homeres math7 math8 math9 math10
            expel arrest hisp black hsdrop expect lunch mstrat
            droptht7 schcode;
            !lunch = % of students eligible for full lunch
            !assistance (9th)
            !mstrat = ratio of students to full time math
            !teachers (9th)
            MISSING = ALL (9999);
            CATEGORICAL = hsdrop;
            CLASSES = c (3);
            CLUSTER = schcode;
            WITHIN = female mothed homeres expect droptht7 expel
            arrest hisp black;
            BETWEEN = lunch mstrat;
```

39

Input For A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
DEFINE:     lunch = lunch/100;
            mstrat = mstrat/1000;

ANALYSIS:   TYPE = MIXTURE TWOLEVEL MISSING;
            ALGORITHM = INTEGRATION;

OUTPUT:     SAMPSTAT STANDARDIZED TECH1 TECH8;

PLOT:       TYPE = PLOT3;
            SERIES = math7-math10 (s);
```

40

Input For A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
MODEL:      %WITHIN%

            %OVERALL%

            i s | math7@0 math8@1 math9@2 math10@3;

            i s ON female hisp black mothed homeres expect
            droptht7 expel arrest;

            c#1 c#2 ON female hisp black mothed homeres expect
            droptht7 expel arrest;

            hsdrop ON female hisp black mothed homeres expect
            droptht7 expel arrest;
```

41

Input For A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
%c#1%
[i*40 s*1];
math7-math10*20;
i*13 s*3;

%c#2%
[i*40 s*5];
math7-math10*30;
i*8 s*3;
i s ON female hisp black mothed homeres expect
droptht7 expel arrest;

%c#3%
[i*45 s*3];
math7-math10*10;
i*34 s*2;
i s ON female hisp black mothed homeres expect
droptht7 expel arrest;
```

42

Input For A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
%BETWEEN%

%OVERALL%

ib | math7-math10@1; [ib@0];

ib*1; hsdrop*1; ib WITH hsdrop;
math7-math10@0;

ib ON lunch mstrat;

c#1 c#2 ON lunch mstrat;

hsdrop ON lunch mstrat;

%c#1%
[hsdrop$1*-.3];

%c#2%
[hsdrop$1*.9];

%c#3%
[hsdrop$1*1.2];
```

43

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement

Summary of Data

```
     Number of patterns           13
     Number of y patterns         13
     Number of u patterns          1
     Number of clusters           44

  Size (s)    Cluster ID with Size s

     12           304
     13           305
     38           112
     39           109
     40           138
     42           120
     43           307
     44           303
     45           143       146
```

44

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

46	101					
48	144	106				
51	102	308				
52	136	118	133	111		
53	140	142	108	131	122	124
54	301	117	127	137	126	
55	103	141	123			
56	110					
57	147					
58	121	105	145	135		
59	119					
73	104					
89	302					
94	309					
118	115					

45

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

```
MAXIMUM LOG-LIKELIHOOD VALUE FOR THE UNRESTRICTED (H1) MODEL IS
-36393.088

THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE
TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE
FIRST-ORDER DERIVATIVE PRODUCT MATRIX. THIS MAY BE DUE TO THE
STARTING VALUES BUT MAY ALSO BE AN INDICATION OF MODEL
NONIDENTIFICATION. THE CONDITION NUMBER IS -0.758D-16. PROBLEM
INVOLVING PARAMETER 54.

THE NONIDENTIFICATION IS MOST LIKELY DUE TO HAVING MORE
PARAMETERS THAN THE NUMBER OF CLUSTERS. REDUCE THE NUMBER OF
PARAMETERS.

THE MODEL ESTIMATION TERMINATED NORMALLY
```

46

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Tests Of Model Fit

```
Loglikelihood
             H0 Value                        -26247.205
Information Criteria
             Number of Free Parameters              122
             Akaike (AIC)                     52738.409
             Bayesian (BIC)                   53441.082
             Sample-Size Adjusted BIC         53053.464
                (n* = (n + 2) / 24)
             Entropy                              0.632

FINAL  CLASS COUNTS AND PROPORTIONS OF TOTAL SAMPLE SIZE BASED
ON ESTIMATED POSTERIOR PROBABILITIES

 Class 1        686.43905          0.29285
 Class 2        430.83877          0.18380
 Class 3       1226.72218          0.52335
```


Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Model Results

```
                     Estimates     S.E.   Est./S.E.    Std      StdYX

 Between Level
 CLASS 1
  IB        ON
     LUNCH            -1.805      1.310     -1.378    -0.851    -0.176
     MSTRAT          -13.365      3.086     -4.331    -6.299    -0.448

  HSDROP    ON
     LUNCH             1.087      0.543      2.004     1.087     0.290
     MSTRAT           -0.178      1.478     -0.120    -0.178    -0.016

  IB        WITH
     HSDROP           -0.416      0.328     -1.267    -0.196    -0.253
```

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

Intercepts					
MATH7	0.000	0.000	0.000	0.000	0.000
MATH8	0.000	0.000	0.000	0.000	0.000
MATH9	0.000	0.000	0.000	0.000	0.000
MATH10	0.000	0.000	0.000	0.000	0.000
IB	0.000	0.000	0.000	0.000	0.000
Residual Variances					
HSDROP	0.550	0.216	2.542	0.550	0.915
MATH7	0.000	0.000	0.000	0.000	0.000
MATH8	0.000	0.000	0.000	0.000	0.000
MATH9	0.000	0.000	0.000	0.000	0.000
MATH10	0.000	0.000	0.000	0.000	0.000
IB	3.456	1.010	3.422	0.768	0.768

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

	Estimates	S.E.	Est./S.E.
Within Level			
C#1 ON			
FEMALE	-0.751	0.188	-3.998
HISP	0.094	0.705	0.133
BLACK	0.900	0.385	2.339
MOTHED	-0.003	0.106	-0.028
HOMERES	-0.060	0.069	0.864
EXPECT	-0.251	0.074	-3.406
DROPTHT7	1.616	0.451	3.583
EXPEL	0.698	0.337	2.068
ARREST	1.093	0.384	2.842

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

	Estimates	S.E.	Est./S.E.
C#2 ON			
FEMALE	-1.610	0.450	-3.577
HISP	1.144	0.466	2.458
BLACK	-0.961	0.656	-1.465
MOTHED	-0.234	0.139	-1.684
HOMERES	0.102	0.094	1.085
EXPECT	0.056	0.089	0.628
DROPTHT7	0.570	0.657	0.869
EXPEL	1.217	0.397	3.068
ARREST	1.133	0.580	1.951
Intercepts			
C#1	0.492	0.535	0.921
C#2	-0.533	0.627	-0.849

51

Output Excerpts A Multilevel Growth Mixture Model For LSAY Math Achievement (Continued)

	Estimates	S.E.	Est./S.E.
Between Level			
C#1 ON			
LUNCH	2.265	0.706	3.208
MSTRAT	-2.876	2.909	-0.988
C#2 ON			
LUNCH	-0.088	1.343	-0.065
MSTRAT	-0.608	2.324	-0.262

52

Between-Level Latent Classes

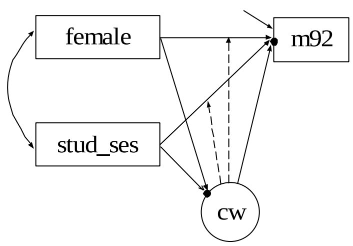
53

Regression Analysis

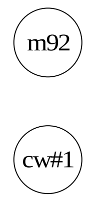
54

NELS Two-Level Regression With Latent Classes For Students

Within (Students)



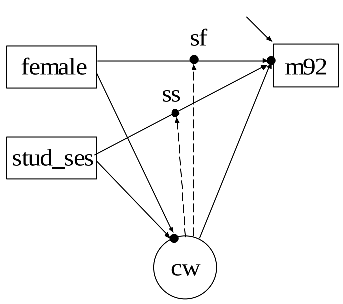
Between (Schools)



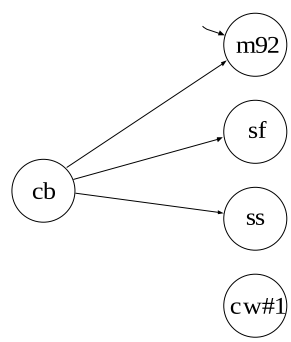
55

NELS Two-Level Regression With Latent Classes For Students And Schools

Within (Students)



Between (Schools)



56

Model Results For NELS Two-Level Regression Of Math Score Related To Gender And Student SES

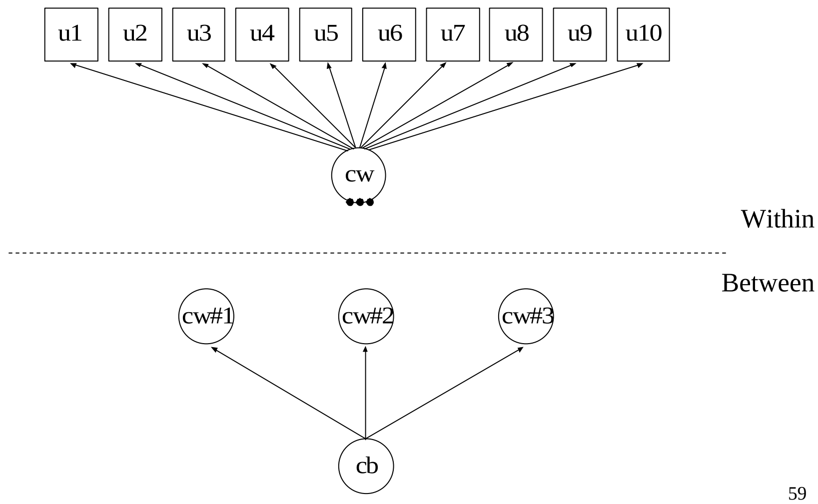
Model	Loglikelihood	# parameters	BIC
(1) Conventional 2-level regression with random intercepts and random slopes	-39,512	10	79,117
(2) Two-level regression mixture, 2 latent classes for students	-39,368	12	78,848
(3) Two-level regression mixture, 3 latent classes for students	**-39,280**	**19**	**78,736**
(4) Two-level regression mixture, 2 latent classes for schools, 2 latent classes for students	-39,348	19	78,873
(5) Two-level regression mixture, 2 latent classes for schools, 3 latent classes for students	-39,260	29	78,789

57

Latent Class Analysis

58

Example 4: Two-Level LCA With Categorical Latent Class Indicators And A Between-Level Categorical Latent Variable



Input For Two-Level Latent Class Analysis

```
TITLE:          this is an example of a two-level LCA with
                categorical latent class indicators and a between-
                level categorical latent variable
DATA:           FILE = ex4.dat;
VARIABLE:       NAMES ARE u1-u10 dumb dumw clus;
                USEVARIABLES = u1-u10;
                CATEGORICAL = u1-u10;
                CLASSES = cb(5) cw(4);
                WITHIN = u1-u10;
                BETWEEN = cb;
                CLUSTER = clus;
ANALYSIS:       TYPE = TWOLEVEL MIXTURE;
                PROCESSORS = 2;
                STARTS = 100 10;
MODEL:
                %WITHIN%
                %OVERALL%
                %BETWEEN%
                %OVERALL%
                cw#1-cw#3 ON cb#1-cb#4;
```

Input For Two-Level Latent Class Analysis (Continued)

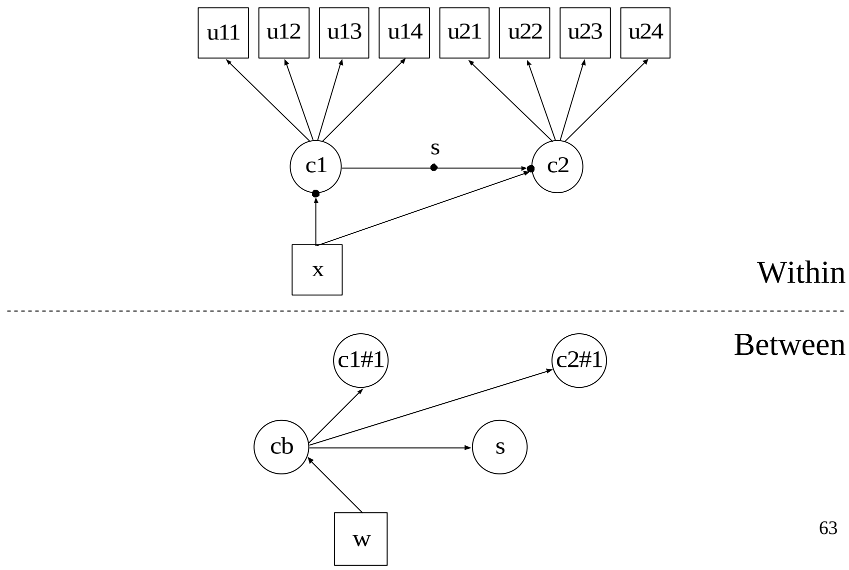
```
MODEL cw:
            %WITHIN%
            %cw#1%
            [u1$1-u10$1];
            [u1$2-u10$2];
            %cw#2%
            [u1$1-u10$1];
            [u1$2-u10$2];
            %cw#3%
            [u1$1-u10$1];
            [u1$2-u10$2];
            %cw#4%
            [u1$1-u10$1];
            [u1$2-u10$2];
OUTPUT:     TECH1 TECH8;
```

61

Latent Transition Analysis

62

Example 8: Two-Level LTA With A Covariate And A Between-Level Categorical Latent Variable



63

Input For Two-Level LTA

```
TITLE:      this is an example of a two-level LTA with a covariate
            and a between-level categorical latent variable
DATA:       FILE = ex8.dat;
VARIABLE:   NAMES ARE u11-u14 u21-u24 x w dumb dum1 dum2 clus;
            USEVARIABLES = u11-w;
            CATEGORICAL = u11-u14 u21-u24;
            CLASSES = cb(2) c1(2) c2(2);
            WITHIN = x;
            BETWEEN = cb w;
            CLUSTER = clus;
ANALYSIS:   TYPE = TWOLEVEL MIXTURE;
            PROCESSORS = 2;
MODEL:
            %WITHIN%
            %OVERALL%
            c2#1 ON c1#1 x;
            c1#1 ON x;
            %BETWEEN%
            %OVERALL%
            c1#1 ON cb#1;
```

64

Input For Two-Level LTA (Continued)

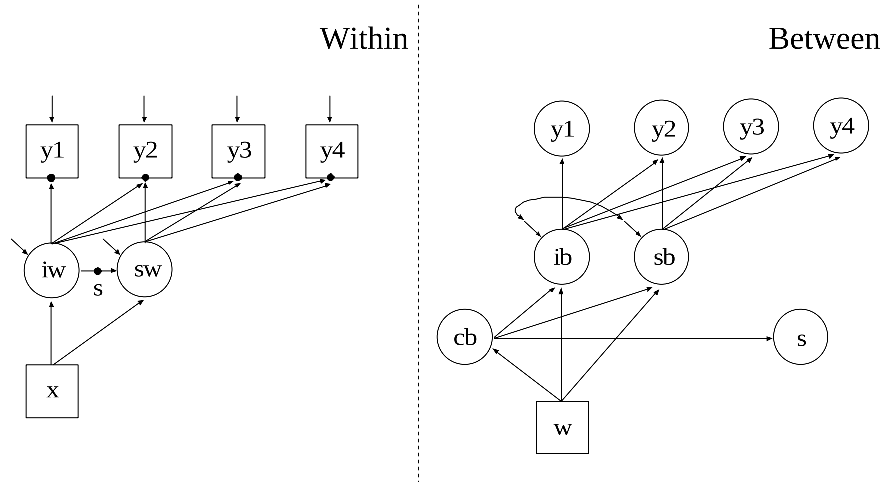
```
              c2#1 ON cb#1;
              cb#1 ON w;
MODEL cb:
              %WITHIN%
              %cb#1%
              c2#1 ON c1#1;
MODEL c1:
              %BETWEEN%
              %c1#1%
              [u11$1-u14$1] (1-4);
              %c1#2%
              [u11$1-u14$1] (5-8);
MODEL c2:
              %BETWEEN%
              %c2#1%
              [u21$1-u24$1] (1-4);
              %c2#2%
              [u21$1-u24$1] (5-8);
OUTPUT:       TECH1 TECH8;
```

65

Growth Modeling

66

Example 5: Two-Level Growth Model For A Continuous Outcome (Three-Level Analysis) With A Between-Level Categorical Latent Variable



67

Input For Two-Level Growth Model

```
TITLE:      this is an example of a two-level growth model for a
            continuous outcome (three-level analysis) with a
            between-level  categorical latent variable
DATA:       FILE = ex5.dat;
VARIABLE:   NAMES ARE y1-y4 x w dummy clus;
            USEVARIABLES = y1-w;
            CLASSES = cb(2);
            WITHIN = x;
            BETWEEN = cb w;
            CLUSTER = clus;
ANALYSIS:   TYPE = TWOLEVEL MIXTURE RANDOM;
            PROCESSORS = 2;
MODEL:
            %WITHIN%
            %OVERALL%
            iw sw | y1@0 y2@1 y3@2 y4@3;
            y1-y4 (1);
            iw sw ON x;
            s | sw ON iw;
```

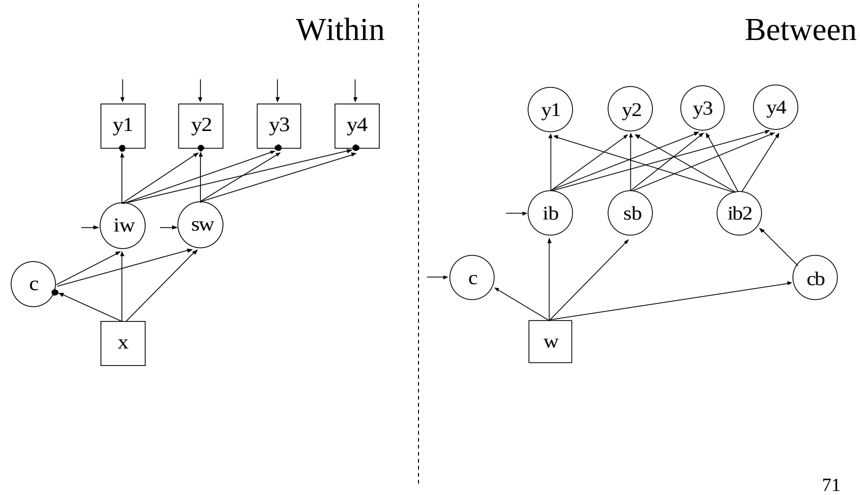
68

Input For Two-Level Growth Model (Continued)

```
              %BETWEEN%
              %OVERALL%
              ib sb | y1@0 y2@1 y3@2 y4@3;
              y1-y4@0;
              ib sb ON w;
              cb#1 ON w;
              s@0;
              %cb#1%
              [ib sb s];
              %cb#2%
              [ib sb s];
OUTPUT:       TECH1 TECH8;
```


Growth Mixture Modeling

Example 6: Two-Level GMM (Three-Level Analysis) For A Continuous Outcome With A Between-Level Categorical Latent Variable



71

Input For Two-Level GMM (Three-Level Analysis)

```
TITLE:      this is an example of a two-level GMM (three-level
            analysis) for a continuous outcome with a between-
            level categorical latent variable
DATA:       FILE = ex6.dat;
VARIABLE:   NAMES ARE y1-y4 x w dummyb dummy clus;
            USEVARIABLES = y1-w;
            CLASSES = cb(2) c(2);
            WITHIN = x;
            BETWEEN = cb w;
            CLUSTER = clus;
ANALYSIS:   TYPE = TWOLEVEL MIXTURE;
            PROCESSORS = 2;
MODEL:
            %WITHIN%
            %OVERALL%
            iw sw | y1@0 y2@1 y3@2 y4@3;
            iw sw ON x;
            c#1 ON x;
            %BETWEEN%
            %OVERALL%
```

72

Input For Two-Level GMM (Continued)

```
   ib sb | y1@0 y2@1 y3@2 y4@3;
   ib2 | y1-y4@1;
   y1-y4@0;
   ib sb ON w;
   c#1 ON w;
   sb@0; c#1;
   ib2@0;
   cb#1 ON w;
MODEL c:
   %BETWEEN%
   %c#1%
   [ib sb];
   %c#2%
   [ib sb];
MODEL cb:
   %BETWEEN%
   %cb#1%
   [ib2@0];
   %cb#2%
   [ib2];
OUTPUT: TECH1 TECH8;
```


Forthcoming in Mplus Version 5

New, simple WLS estimator for high-dimensional 2-level models with categorical outcomes

Example 1: Grilli & Rampichini (2007)

Five items on job satisfaction of 2,432 graduates (level 1) from 36 degree programs (level 2): "How satisfied are you with…"

5-point scale (1 = very much satisfied, …, 5 = very unsatisfied)

	Loadings				*Thresholds*			
	Within (W)		*Between (B)*					
Item					$\gamma_{1,h}$	$\gamma_{2,h}$	$\gamma_{3,h}$	$\gamma_{4,h}$
1. Earnings	--	1.09	0.75	--	-3.96	-1.39	1.24	3.29
2. Career	1[a]	1[a]	1[a]	--	-3.64	-0.92	1.46	3.37
3. Consistency	2.30	--	0.32	1[a]	-1.80	0.19	1.98	3.39
4. Professionalism	2.25	0.41	0.31	0.27	-1.79	1.07	3.36	5.27
5. Interests	2.85	--	0.09	0.47	-2.34	0.34	2.83	4.62
Factor Variance	0.75	3.09	0.71	0.45				

Estimation by ML using numerical integration in the Mplus program (Muthen & Muthen, 1998 – 2007)

Example 2: 2nd-Generation JHU PIRC Trial Aggression Items

Item Distributions for Cohort 3: Fall 1st Grade (n=362 males in 27 classrooms)

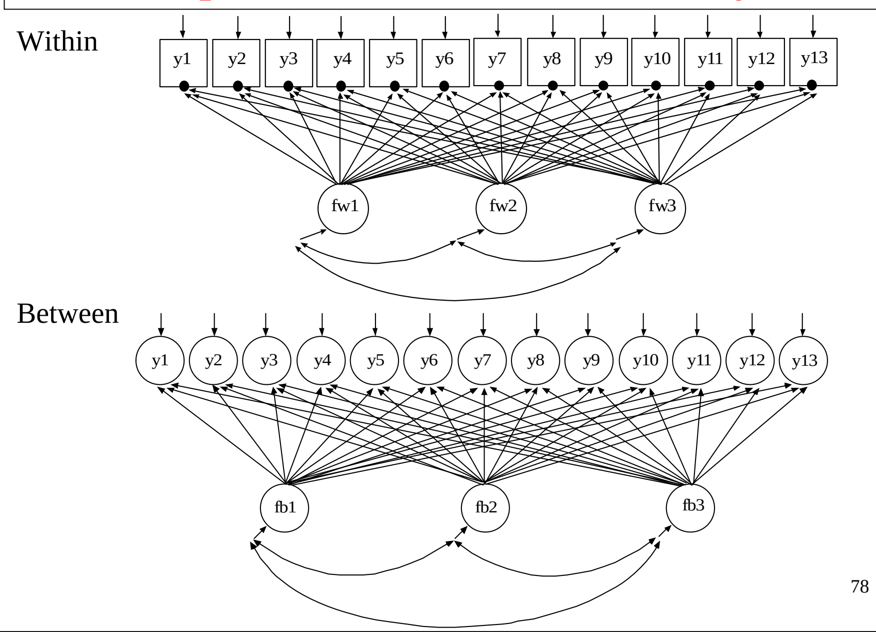
	Almost Never (scored as 1)	*Rarely (scored as 2)*	*Sometimes (scored as 3)*	*Often (scored as 4)*	*Very Often (scored as 5)*	*Almost Always (scored as 6)*
Stubborn	42.5	21.3	18.5	7.2	6.4	4.1
Breaks Rules	37.6	16.0	22.7	7.5	8.3	8.0
Harms Others	69.3	12.4	9.40	3.9	2.5	2.5
Breaks Things	79.8	6.60	5.20	3.9	3.6	0.8
Yells at Others	61.9	14.1	11.9	5.8	4.1	2.2
Takes Others' Property	72.9	9.70	10.8	2.5	2.2	1.9
Fights	60.5	13.8	13.5	5.5	3.0	3.6
Harms Property	74.9	9.90	9.10	2.8	2.8	0.6
Lies	72.4	12.4	8.00	2.8	3.3	1.1
Talks Back to Adults	79.6	9.70	7.80	1.4	0.8	1.4
Teases Classmates	55.0	14.4	17.7	7.2	4.4	1.4
Fights With Classmates	67.4	12.4	10.2	5.0	3.3	1.7
Loses Temper	61.6	15.5	13.8	4.7	3.0	1.4

Example 2: Hypothesized Aggressiveness Factors

- Verbal aggression
  - Yells at others
  - Talks back to adults
  - Loses temper
  - Stubborn
- Property aggression
  - Breaks things
  - Harms property
  - Takes others' property
  - Harms others
- Person aggression
  - Fights
  - Fights with classmates
  - Teases classmates

77

Example 2: Two-Level Factor Analysis



78

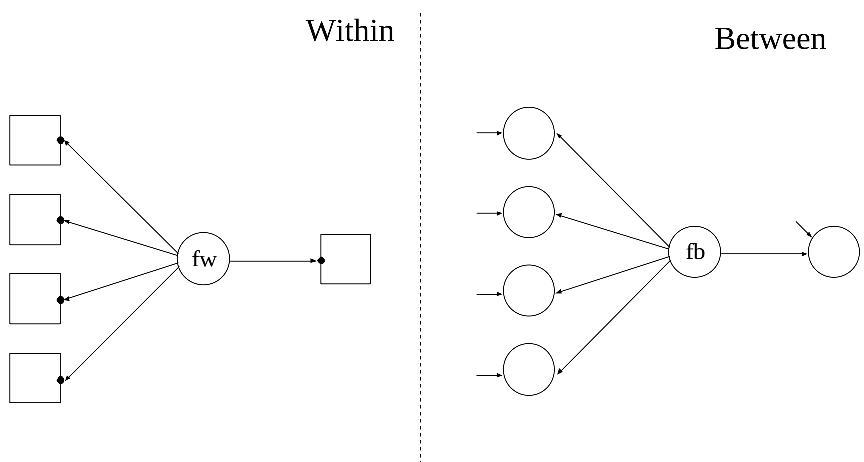
Reasons For Finding Dimensions

Different dimensions may have different

- predictors
- effects on later events
- growth curves
- treatment effects

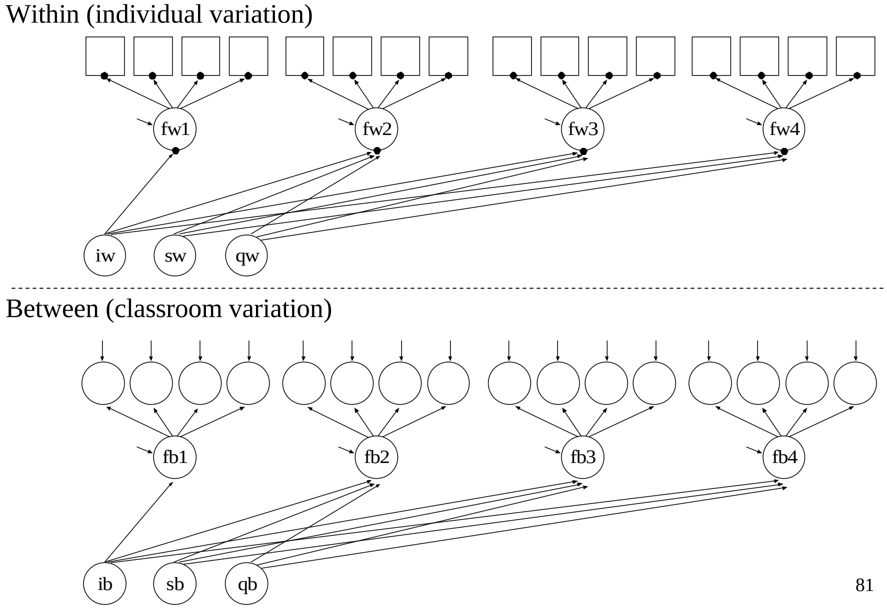
79

Example 3: Predicting Juvenile Delinquency From First Grade Aggressive Behavior. Two-Level Logistic Regression



80

Example 4: Aggression Development Grades 1-3



81

ML Versus Other Estimators

- ML uses numerical integration, which is heavy with many dimensions of integration (many factors and/or random effects).
  - Example 1 has 4 dimensions: 2 factor for level 1 and 2 factors for level 2 (Note: level 2 residuals are zero – otherwise 9 dimensions)
  - Example 2 has potentially 3 + 3 (+13) dimensions
  - Example 3 has 7 dimensions
  - Example 4 has potentially about 30 dimensions
  - Monte Carlo integration possible, but gives only approximate results
- MCMC
  - challenges include parameterizations to ensure good mixing, "label switching" for factors, difficulties in determining identification status, similar in computational demands to ML with Monte Carlo integration
  - Goldstein & Browne (2002, 2005)

82

ML Versus Other Estimators

Two choices

- Keep ML or MCMC estimation and simplify model (few random effects)
- Keep model and simplify estimator

ML versus Other Estimators

- New simple alternative: two-level, limited information WLS
  - computational demand virtually independent of number of factors/random effects
  - high-dimensional integration replaced by multiple instances of one- and two-dimensional integration
  - possible to explore many different models in a time-efficient manner
  - generalization of the Muthen (1984) single-level WLS
  - variables can be categorical, continuous, censored, combinations
  - residuals can be correlated (no conditional independence assumption)
  - model fit chi-square testing
  - can produce unrestricted level 1 and level 2 correlation matrices for EFA

Example 1: Using Two-Level WLS Estimation

Item	Loadings: Within (W)		Loadings: Between (B)	
1. Earnings	--	1.09	0.75	--
2. Career	1[a]	1[a]	1[a]	--
3. Consistency	2.30	--	0.32	1[a]
4. Professionalism	2.25	0.41	0.31	0.27
5. Interests	2.85	--	0.09	0.47

- The Grilli & Rampichini (2007) model has 2 level 1 factors and 2 level 2 factors (35 par.'s). This model is rejected by WLS: Chi-square (10) = 140.6, p = 0.0
- Model exploration using WLS estimated level 1 and level 2 correlation matrices suggests an alternative model with an unrestricted 2-factor model on level 1, a single factor on level 2, AND non-zero level 2 residual variances (39 parameters): Chi-square (6) = 9.8, p = 0.13
- The alternative model would give 8-dimensional integration with ML

85

Example 2: Aggression EFA On Within-Level Correlation Matrix Estimated With Two-Level WLS

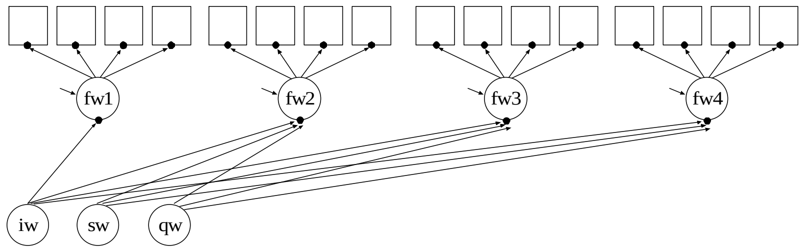
Within-Level Loadings

	1	2	3
Stubborn	0.07	**0.70**	0.05
Breaks Rules	0.25	0.31	0.37
Harms Others	**0.52**	0.16	0.27
Breaks Things	**0.84**	0.16	-0.01
Yells at Others	0.15	**0.64**	0.13
Takes Others' Property	**0.57**	0.00	0.37
Fights	0.20	0.21	**0.63**
Harms Property	**0.73**	0.21	0.10
Lies	0.48	0.28	0.24
Talks Back to Adults	0.29	**0.71**	0.23
Teases Classmates	0.11	0.19	**0.62**
Fights With Classmates	0.10	0.31	**0.63**
Loses Temper	0.12	**0.75**	0.04

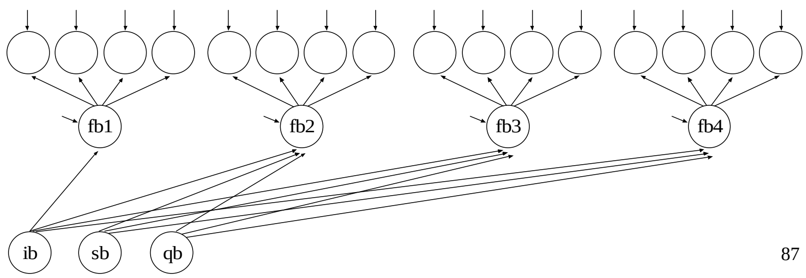
86

Example 4: Aggression Development Grades 1-3

Within (individual variation)

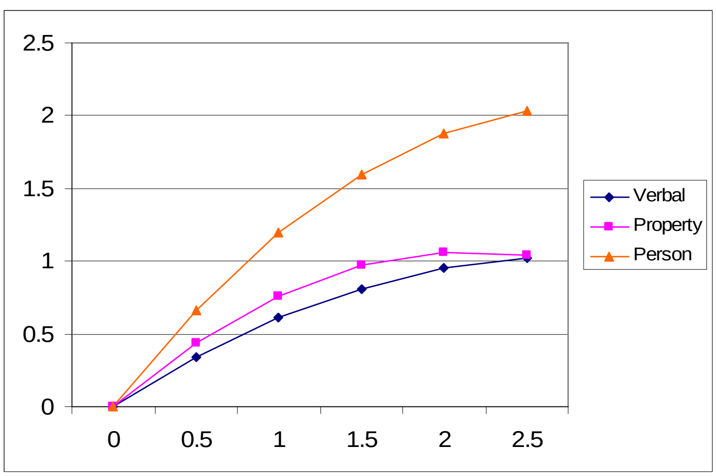


Between (classroom variation)



87

Example 3: Aggression Development Grades 1-3



88

References

(To request a Muthén paper, please email bmuthen@ucla.edu.)

Cross-sectional Data

Asparouhov, T. (2005). Sampling weights in latent variable modeling. Structural Equation Modeling, 12, 411-434.

Chambers, R.L. & Skinner, C.J. (2003). Analysis of survey data. Chichester: John Wiley & Sons.

Goldstein, H. & Browne, W.J. (2005). Multilevel factor analysis model for continuous and discrete data. In A. Maydeu-Olivares & J.J. McArdle (eds.), Contemporary psychometrics: A festschrift for Roderick P. McDonald (pp.453-475). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Grilli, L. & Rampichini, C. (2007). Multilevel factor models for ordinal variables. Structural Equation Modeling, 14, 1-25.

Harnqvist, K., Gustafsson, J.E., Muthén, B. & Nelson, G. (1994). Hierarchical models of ability at class and individual levels. Intelligence, 18, 165-187. (#53)

Heck, R.H. (2001). Multilevel modeling with SEM. In G.A. Marcoulides & R.E. Schumacker (eds.), New developments and techniques in structural equation modeling (pp. 89-127). Lawrence Erlbaum Associates.

References (Continued)

Hox, J. (2002). Multilevel analysis. Techniques and applications. Mahwah, NJ: Lawrence Erlbaum.

Kaplan, D. & Elliott, P.R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. Structural Equation Modeling: A Multidisciplinary Journal, 4, 1-24.

Kaplan, D. & Ferguson, A.J (1999). On the utilization of sample weights in latent variable models. Structural Equation Modeling, 6, 305-321.

Kaplan, D. & Kresiman, M.B. (2000). On the validation of indicators of mathematics education using TIMSS: An application of multilevel covariance structure modeling. International Journal of Educational Policy, Research, and Practice, 1, 217-242.

Korn, E.L. & Graubard, B.I (1999). Analysis of health surveys. New York: John Wiley & Sons.

Kreft, I. & de Leeuw, J. (1998). Introducing multilevel modeling. Thousand Oakes, CA: Sage Publications.

Larsen & Merlo (2005). Appropriate assessment of neighborhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression. American Journal of Epidemiology, 161, 81-88.

Longford, N.T., & Muthén, B. (1992). Factor analysis for clustered observations. Psychometrika, 57, 581-597. (#41)

References (Continued)

Ludtke Marsh, Robitzsch, Trautwein, Asparouhov, Muthen (2007). Analysis of group level effects using multilevel modeling: Probing a latent covariate approach. Submitted for publication.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585. (#24)

Muthén, B. (1990). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society meeting in Princeton, N.J., June 1990. UCLA Statistics Series 62. (#32)

Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. Journal of Educational Measurement, 28, 338-354. (#37)

Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox & I. Kreft (eds.), Multilevel Modeling, a special issue of Sociological Methods & Research, 22, 376-398. (#55)

Muthén, B., Khoo, S.T. & Gustafsson, J.E. (1997). Multilevel latent variable modeling in multiple populations. (#74)

Muthén, B. & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (ed.), Sociological Methodology 1995, 216-316. (#59)

Neale, M.C. & Cardon, L.R. (1992). Methodology for genetic studies of twins and families. Dordrecth, The Netherlands: Kluwer.

91

References (Continued)

Patterson, B.H., Dayton, C.M. & Graubard, B.I. (2002). Latent class analysis of complex sample survey data: application to dietary data. Journal of the American Statistical Association, 97, 721-741.

Prescott, C.A. (2004). Using the Mplus computer program to estimate models for continuous and categorical data from twins. Behavior Genetics, 34, 17- 40.

Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods. Second edition. Newbury Park, CA: Sage Publications.

Skinner, C.J., Holt, D. & Smith, T.M.F. (1989). Analysis of complex surveys. West Sussex, England, Wiley.

Snijders, T. & Bosker, R. (1999). Multilevel analysis. An introduction to basic and advanced multilevel modeling. Thousand Oakes, CA: Sage Publications.

Stapleton, L. (2002). The incorporation of sample weights into multilevel structural equation models. Structural Equation Modeling, 9, 475-502.

Vermunt, J.K. (2003). Multilevel latent class models. In Stolzenberg, R.M. (Ed.), Sociological Methodology (pp. 213-239). New York: American Sociological Association.

92

References (Continued)

Longitudinal Data

Choi, K.C. (2002). Latent variable regression in a three-level hierarchical modeling framework: A fully Bayesian approach. Doctoral dissertation, University of California, Los Angeles.

Khoo, S.T. & Muthén, B. (2000). Longitudinal data on families: Growth modeling alternatives. Multivariate applications in substance use research, J. Rose, L. Chassin, C. Presson & J. Sherman (eds.), Hillsdale, N.J.: Erlbaum, pp. 43-78. (#79)

Masyn, K. E. (2003). Discrete-time survival mixture analysis for single and recurrent events using latent variables. Doctoral dissertation, University of California, Los Angeles.

Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (ed.) Sociological Methodology (pp. 453-480). Boston: Blackwell Publishers.

Muthén, B. (1997). Latent variable growth modeling with multilevel data. In M. Berkane (ed.), Latent variable modeling with application to causality (149-161), New York: Springer Verlag.

Muthén, B. & Masyn, K. (in press). Discrete-time survival mixture analysis. Journal of Educational and Behavioral Statistics, Spring 2005.

References (Continued)

Raudenbush, S.W. & Bryk, A.S. (2002). Hierarchical linear models: Applications and data analysis methods. Second edition. Newbury Park, CA: Sage Publications. Snijders, T. & Bosker, R. (1999). Multilevel analysis. An introduction to basic and advanced multilevel modeling. Thousand Oakes, CA: Sage Publications.

Seltzer, M., Choi, K., Thum, Y.M. (2002). Examining relationships between where students start and how rapidly they progress: Implications for conducting analyses that help illuminate the distribution of achievement within schools. CSE Technical Report 560. CRESST, University of California, Los Angeles.

General

Mplus Analysis

Asparouhov, T. & Muthén, B. (2003a). Full-information maximum-likelihood estimation of general two-level latent variable models. In preparation.

Asparouhov, T. & Muthén, B. (2003b). Maximum-likelihood estimation in general latent variable modeling. In preparation.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. Behaviormetrika, 29, 81-117.

References (Continued)

Muthén, B. & Asparouhov, T. (2003b). Advances in latent variable modeling, part II: Integrating continuous and categorical latent variable modeling using Mplus. In preparation.

Numerical Integration

Aitkin, M. A general maximum likelihood analysis of variance components in generalized linear models. Biometrics, 1999, 55, 117-128.

Bock, R.D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46, 443-459.

95