# New Developments in Mplus Version 7: Part 3

Bengt Muthén & Tihomir Asparouhov

Mplus
www.statmodel.com

Presentation at Utrecht University
August 2012

# Table of Contents I

# Table of Contents II

- Cross-Classified / Multiple Membership Applications

Advanced topics:

- 2-mode path analysis
- Cross-classified SEM
- Random item IRT

## 2-Mode Path Analysis: Random Contexts in Gonzalez et al.

Gonzalez, de Boeck, Tuerlinckx (2008). A double-structure structural equation model for three-mode data. Psychological Methods, 13, 337-353.

- A population of situations that might elicit negative emotional responses
- 11 situations (e.g. blamed for someone else's failure after a sports match, a fellow student fails to return your notes the day before an exam, you hear that a friend is spreading gossip about you) viewed as randomly drawn from a population of situations
- 4 binary responses: Frustration, antagonistic action, irritation, anger
- n=679 high school students
- Level 2 cluster variables are situations and students
- 1 observation for each pair of clustering units

Research questions: Which of the relationships below are significant?
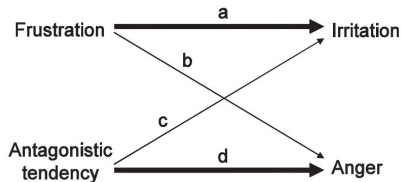Are the relationships the same on the situation level as on the subject
level?



*Figure 3.* Graphical representation of the research questions. *a*,
*b*, *c*, and *d* are effect parameters.

# 2-Mode Path Analysis Input

| | |
|---|---|
| VARIABLE: | NAMES = frust antag irrit anger student situation; |
| | CLUSTER = situation student; |
| | CATEGORICAL = frust antag irrit anger; |
| DATA: | FILE = gonzalez.dat; |
| ANALYSIS: | TYPE = CROSSCLASSIFIED; |
| | ESTIMATOR = BAYES; |
| | BITERATIONS = (10000); |
| MODEL: | %WITHIN% |
| | irrit anger ON frust antag; |
| | irrit WITH anger; |
| | frust WITH antag; |
| | %BETWEEN student% |
| | irrit ON frust (1); |
| | anger ON frust (2); |
| | irrit ON antag (3); |
| | anger ON antag (4); |
| | irrit; anger; irrit WITH anger; |
| | frust; antag; frust WITH antag; |

```
                %BETWEEN situation%
                irrit ON frust (1);
                anger ON frust (2);
                irrit ON antag (3);
                anger ON antag (4);
                irrit; anger; irrit WITH anger;
                frust; antag; frust WITH antag;
OUTPUT:         TECH8 TECH9 STDY;
PLOT:           TYPE = PLOT2;
```

# 2-Mode Path Analysis: Monte Carlo Simulation Using the Gonzalez Model

M is the number of cluster units for both between levels, $\beta$ is the common slope, $\psi$ is the within-level correlation, $\tau$ is the binary outcome threshold. Table gives bias (coverage).

| Para | M=10 | M=20 | M=30 | M=50 | M=100 |
|------|------|------|------|------|-------|
| $\beta_1$ | 0.13(0.92) | 0.05(0.89) | 0.00(0.97) | 0.01(0.92) | 0.01(0.94) |
| $\psi_{2,11}$ | 0.11(1.00) | 0.06(0.96) | 0.01(0.98) | 0.00(0.89) | 0.02(0.95) |
| $\psi_{2,12}$ | 0.15(0.97) | 0.06(0.92) | 0.05(0.97) | 0.03(0.87) | 0.01(0.96) |
| $\tau_1$ | 0.12(0.93) | 0.01(0.93) | 0.00(0.90) | 0.03(0.86) | 0.00(0.91) |

Small biases for $M = 10$. Due to parameter equalities information is combined from both clustering levels. Adding unconstrained level 1 model: tetrachoric correlation matrix.

# Cross-Classified SEM

- General SEM model: 2-way ANOVA. $Y_{pijk}$ is the $p-$th variable for individual $i$ in cluster $j$ and cross cluster $k$

$$Y_{pijk} = Y_{1pijk} + Y_{2pj} + Y_{3pk}$$

- 3 sets of structural equations - one on each level

$$Y_{1ijk} = \nu + \Lambda_1 \eta_{ijk} + \varepsilon_{ijk}$$
$$\eta_{ijk} = \alpha + B_1 \eta_{ijk} + \Gamma_1 x_{ijk} + \xi_{ijk}$$

$$Y_{2j} = \Lambda_2 \eta_j + \varepsilon_j$$
$$\eta_j = B_2 \eta_j + \Gamma_2 x_j + \xi_j$$

$$Y_{3k} = \Lambda_3 \eta_k + \varepsilon_k$$
$$\eta_k = B_3 \eta_k + \Gamma_3 x_k + \xi_k$$

# Cross - Classified SEM

- The regression coefficients on level 1 can be a random effects from each of the two clustering levels: combines cross-classified SEM and cross classified HLM
- Bayesian MCMC estimation: used as a frequentist estimator.
- Easily extends to categorical variables.
- ML estimation possible only when one of the two level of clustering has small number of units.

# Monte Carlo Simulation of Cross-Classified SEM

- 1 factor at the individual level and 1 factor at each of the clustering levels, 5 indicator variables on the individual level

$$y_{pijk} = \mu_p + \lambda_{1,p}f_{1,ijk} + \lambda_{2,p}f_{2,j} + \lambda_{3,p}f_{3,k} + \varepsilon_{2,pj} + \varepsilon_{3,pk} + \varepsilon_{1,pijk}$$

- M level 2 clusters. M level 3 clusters. 1 unit within each cluster intersection. More than 1 unit is possible. Zero units possible: sparse tables

- Monte Carlo simulation: Estimation takes less than 1 min per replication

Table: Absolute bias and coverage for cross-classified factor analysis model

| Param | M=10 | M=20 | M=30 | M=50 | M=100 |
|-------|------|------|------|------|-------|
| $\lambda_{1,1}$ | 0.07(0.92) | 0.03(0.89) | 0.01(0.95) | 0.00(0.97) | 0.00(0.91) |
| $\theta_{1,1}$ | 0.05(0.96) | 0.00(0.97) | 0.00(0.95) | 0.00(0.99) | 0.00(0.94) |
| $\lambda_{2,p}$ | 0.21(0.97) | 0.11(0.94) | 0.10(0.93) | 0.06(0.94) | 0.00(0.92) |
| $\theta_{2,p}$ | 0.24(0.99) | 0.10(0.95) | 0.04(0.92) | 0.05(0.94) | 0.02(0.96) |
| $\lambda_{3,p}$ | 0.45(0.99) | 0.10(0.97) | 0.03(0.99) | 0.01(0.92) | 0.03(0.97) |
| $\theta_{3,p}$ | 0.75(1.00) | 0.25(0.98) | 0.15(0.97) | 0.12(0.98) | 0.05(0.92) |
| $\mu_p$ | 0.01(0.99) | 0.04(0.98) | 0.01(0.97) | 0.05(0.99) | 0.00(0.97) |

Perfect coverage. Level 1 parameters estimated very well. Biases when the number of clusters is small $M = 10$. Weakly informative priors can reduce the bias for small number of clusters.

# Cross-Classified Models: Types Of Random Effects

- Type 1: Random slope.
  %WITHIN%
  s | y ON x;
  *s* has variance on both crossed levels. Dependent variable can be within-level factor. Covariate x should be on the WITHIN = list.

- Type 2: Random loading.
  %WITHIN%
  s | f BY y;
  *s* has variance on both crossed levels. *f* is a within-level factor. The dependent variable can be a within-level factor.

- Type 3: Crossed random loading.
  %BETWEEN level2a%
  s | f BY y;
  *s* has variance on crossed level 2b and is defined on crossed level 2a. *f* is a level 2a factor, *s* is a level 2b factor. This is a way to use the interaction term $s \cdot f$.

# Cross-Classified Interaction Model:
## Random Items, Generalizability Theory

- Items are random samples from a population of items.
- The same or different items may be administered to individuals.
- Suited for computer generated items and adaptive testing.
- 2-parameter IRT model

$$P(Y_{ij} = 1) = \Phi(a_j\theta_i + b_j)$$

- $a_j \sim N(a, \sigma_a)$, $b_j \sim N(b, \sigma_b)$: random discrimination and difficulty parameters
- The ability parameter is $\theta_i \sim N(0,1)$
- Cross-classified model. Nested within items and individuals. 1 or 0 observation in each cross-classified cell.
- Interaction of two latent variables: $a_j$ and $\theta_i$: Type 3 crossed random loading
- The model has only 4 parameters - much more parsimonious than regular IRT models.

# Random Item 2-Parameter IRT Model Setup

VARIABLE:

    NAMES = u item individual;
    CLUSTER = item individual;
    CATEGORICAL = u;

ANALYSIS:

    TYPE = CROSS RANDOM;
    ESTIMATOR = BAYES;

MODEL:

    %WITHIN%

    %BETWEEN individual%
    s | f BY u;
    f@1 u@0;
    %BETWEEN item%
    u s;

- Fox (2010) Bayesian Item Response Theory. Section 4.3.3. Dutch Six Graders Math Achievement. Trends in International Mathematics and Science Study: TIMMS 2007
- 8 test items, 478 students

Table: Random 2-parameter IRT

| parameter | estimate | SE |
|---|---|---|
| average discrimination $a$ | 0.752 | 0.094 |
| average difficulty $b$ | 0.118 | 0.376 |
| variation of discrimination $a$ | 0.050 | 0.046 |
| variation of difficulty $b$ | 1.030 | 0.760 |

- 8 items means that there are only 8 clusters on the %between item% level and therefore the variance estimates at that level are affected by their priors. If the number of clusters is less than 10 or 20 there is prior dependence in the variance parameters.

- Using factor scores estimation we can estimate item specific parameter and SE using posterior mean and posterior standard deviation.

Table: Random 2-parameter IRT item specific parameters

| item | discrimination | SE | difficulty | SE |
|------|----------------|-------|------------|-------|
| Item 1 | 0.797 | 0.11 | -1.018 | 0.103 |
| Item 2 | 0.613 | 0.106 | -0.468 | 0.074 |
| Item 3 | 0.905 | 0.148 | -1.012 | 0.097 |
| Item 4 | 0.798 | 0.118 | -1.312 | 0.106 |
| Item 5 | 0.538 | 0.099 | 0.644 | 0.064 |
| Item 6 | 0.808 | 0.135 | 0.023 | 0.077 |
| Item 7 | 0.915 | 0.157 | 0.929 | 0.09 |
| Item 8 | 0.689 | 0.105 | 1.381 | 0.108 |

# Random Item 2-Parameter IRT: TIMMS Example, Comparison With ML

Table: Random 2-parameter IRT item specific parameters

| item | Bayes random discrimination | Bayes random SE | ML fixed discrimination | ML fixed SE |
|---|---|---|---|---|
| Item 1 | 0.797 | 0.110 | 0.850 | 0.155 |
| Item 2 | 0.613 | 0.106 | 0.579 | 0.102 |
| Item 3 | 0.905 | 0.148 | 0.959 | 0.170 |
| Item 4 | 0.798 | 0.118 | 0.858 | 0.172 |
| Item 5 | 0.538 | 0.099 | 0.487 | 0.096 |
| Item 6 | 0.808 | 0.135 | 0.749 | 0.119 |
| Item 7 | 0.915 | 0.157 | 0.929 | 0.159 |
| Item 8 | 0.689 | 0.105 | 0.662 | 0.134 |

- Bayes random estimates are shrunk towards the mean and have smaller standard errors: shrinkage estimate

- One can add a predictor for a person's ability. For example adding gender as a predictor yields an estimate of 0.283 (0.120), saying that males have a significantly higher math mean.
- Predictors for discrimination and difficulty random effects, for example, geometry indicator.
- More parsimonious model can yield more accurate ability estimates.

# Random Item Rasch IRT Example

- De Boeck (2008) Random item IRT models
- 24 verbal aggression items, 316 persons

$$P(Y_{ij} = 1) = \Phi(\theta_i + b_j)$$
$$b_j \sim N(b, \sigma)$$
$$\theta_i \sim N(0, \tau)$$

Table: Random Rasch IRT - variance decomposition

| parameter | person $\tau$ | item $\sigma$ | error |
|---|---|---|---|
| estimates(SE) | 1.89(0.19) | 1.46(0.53) | 2.892 |
| variance explained | 30% | 23% | 46% |

# Random Item Rasch IRT Example:
## Simple Model Specification

MODEL:

    %WITHIN%

    %BETWEEN person%
    y;

    %BETWEEN item%
    y;

- An old dilemma
- Two new solutions
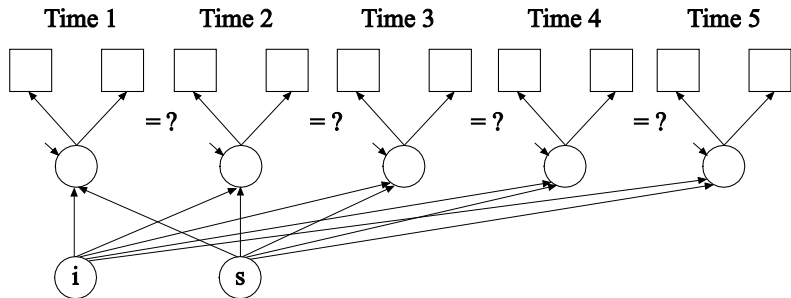
Single-level analysis with $p \times T = 2 \times 5 = 10$ variables, $T = 5$ factors.

- ML hard and impossible as T increases (numerical integration)
- WLSMV possible but hard when $p \times T$ increases and biased unless attrition is MCAR or multiple imputation is done first
- Bayes possible
- Searching for partial measurement invariance is cumbersome
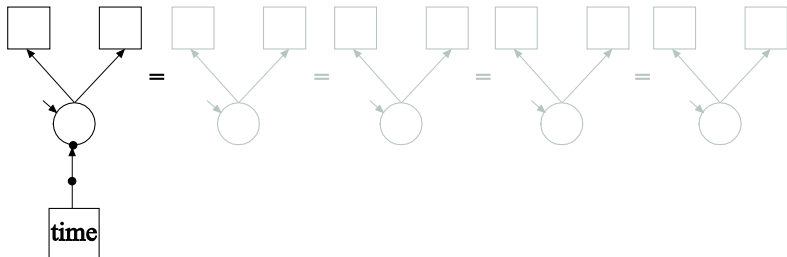
Two-level analysis with $p = 2$ variables, 1 within-factor, 2-between factors, **assuming full measurement invariance across time**.

- ML feasible
- WLSMV feasible (2-level WLSMV)
- Bayes feasible

# Measurement Invariance Across Time

- Both old approaches have problems
  - Wide, single-level approach easily gets significant non-invariance and needs many modifications
  - Long, two-level approach has to assume invariance

- New solution no. 1, suitable for small to medium number of time points
  - A new wide, single-level approach where time is a fixed mode
- New solution no. 2, suitable for medium to large number of time points
  - A new long, two-level approach where time is a random mode
  - No limit on the number of time points
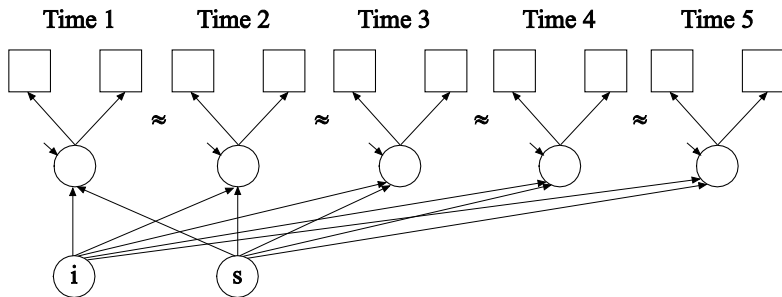
Single-level analysis with $p \times T = 2 \times 5 = 10$ variables, $T = 5$ factors.

- Bayes ("BSEM") using approximate measurement invariance, still identifying factor mean and variance differences across time

- New solution no. 2, time is a random mode
- A new long, two-level approach
  - Best of both worlds: Keeping the limited number of variables of the two-level approach without having to assume invariance
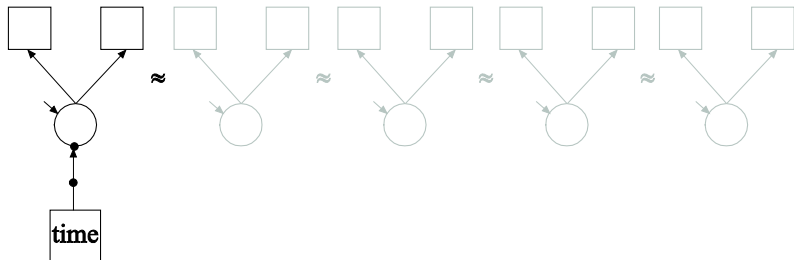
Two-level analysis with $p = 2$ variables.

- Bayes twolevel random approach with random measurement parameters and random factor means and variances using Type=Crossclassified: Clusters are time and person

Randomized field experiment in Baltimore public schools with a classroom-based intervention aimed at reducing aggressive-disruptive behavior among elementary school students (Ialongo et al., 1999).

This analysis:

- Cohort 1
- 9 binary items at 8 time points, Grade 1 - Grade 7
- $n = 1174$

# Aggressive-Disruptive Behavior in the Classroom: ML vs BSEM

- Traditional ML analysis
  - 8 dimensions of integration
  - Computing time: 25:44 with Integration = Montecarlo(5000)
  - Increasing the number of time points makes ML impossible
- BSEM analysis
  - 156 parameters
  - Computing time: 4:01
  - Increasing the number of time points has relatively less impact

```
                USEVARIABLES = stub1f-tease7s;
                CATEGORICAL = stub1f-tease7s;
                MISSING = ALL (999);
DEFINE:         CUT stub1f-tease7s (1.5);
ANALYSIS:       ESTIMATOR = BAYES;
                PROCESSORS = 2;
MODEL:          f1f by stub1f-tease1f* (lam11-lam19);
                f1s by stub1s-tease1s* (lam21-lam29);
                f2s by stub2s-tease2s* (lam31-lam39);
                f3s by stub3s-tease3s* (lam41-lam49);
                f4s by stub4s-tease4s* (lam51-lam59);
                f5s by stub5s-tease5s* (lam61-lam69);
                f6s by stub6s-tease6s* (lam71-lam79);
                f7s by stub7s-tease7s* (lam81-lam89);
                f1f@1;
```

```
                [stub1f$1-tease1f$1] (tau11-tau19);
                [stub1s$1-tease1s$1] (tau21-tau29);
                [stub2s$1-tease2s$1] (tau31-tau39);
                [stub3s$1-tease3s$1] (tau41-tau49);
                [stub4s$1-tease4s$1] (tau51-tau59);
                [stub5s$1-tease5s$1] (tau61-tau69);
                [stub6s$1-tease6s$1] (tau71-tau79);
                [stub7s$1-tease7s$1] (tau81-tau89);
                [f1f-f7s@0];
                i s q | f1f@0 f1s@0.5 f2s@1.5 f3s@2.5 f4s@3.5
                f5s@4.5 f6s@5.5 f7s@6.5;
                q@0;
MODEL
PRIORS:         DO(1,9) DIFF(lam1#-lam8#) ~ N(0,.01);
                DO(1,9) DIFF(tau1#-tau8#) ~ N(0,.01);
OUTPUT:         TECH1 TECH8;
```

|  | Estimate | Posterior S.D. | One-Tailed P-Value | 95% C.I. Lower 2.5% | Upper 2.5% |  |
|---|---|---|---|---|---|---|
| **Means** |  |  |  |  |  |  |
| I | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |  |
| S | 0.238 | 0.068 | 0.000 | 0.108 | 0.366 | * |
| Q | -0.022 | 0.011 | 0.023 | -0.043 | 0.000 | * |
| **Variances** |  |  |  |  |  |  |
| I | 9.258 | 2.076 | 0.000 | 6.766 | 14.259 | * |
| S | 0.258 | 0.068 | 0.000 | 0.169 | 0.411 | * |
| Q | 0.001 | 0.000 | 0.000 | 0.001 | 0.001 |  |

|  | Estimate | Posterior S.D. | One-Tailed P-Value | 95% C.I. Lower 2.5% | Upper 2.5% | |
|---|---|---|---|---|---|---|
| **F1F      BY** | | | | | | |
| STUB1F | 0.428 | 0.048 | 0.000 | 0.338 | 0.522 | * |
| BKRULE1F | 0.587 | 0.068 | 0.000 | 0.463 | 0.716 | * |
| HARMO1F | 0.832 | 0.082 | 0.000 | 0.677 | 0.985 | * |
| BKTHIN1F | 0.671 | 0.067 | 0.000 | 0.546 | 0.795 | * |
| YELL1F | 0.508 | 0.055 | 0.000 | 0.405 | 0.609 | * |
| TAKEP1F | 0.717 | 0.072 | 0.000 | 0.570 | 0.839 | * |
| FIGHT1F | 0.480 | 0.052 | 0.000 | 0.385 | 0.579 | * |
| LIES1F | 0.488 | 0.054 | 0.000 | 0.386 | 0.589 | * |
| TEASE1F | 0.503 | 0.055 | 0.000 | 0.404 | 0.608 | * |
| **...** | | | | | | |
| **F7S      BY** | | | | | | |
| STUB7S | 0.360 | 0.049 | 0.000 | 0.273 | 0.458 | * |
| BKRULE7S | 0.512 | 0.068 | 0.000 | 0.392 | 0.654 | * |
| HARMO7S | 0.555 | 0.074 | 0.000 | 0.425 | 0.716 | * |
| BKTHIN7S | 0.459 | 0.063 | 0.000 | 0.344 | 0.581 | * |
| YELL7S | 0.525 | 0.062 | 0.000 | 0.409 | 0.643 | * |
| TAKEP7S | 0.500 | 0.069 | 0.000 | 0.372 | 0.634 | * |
| FIGHT7S | 0.515 | 0.067 | 0.000 | 0.404 | 0.652 | * |
| LIES7S | 0.520 | 0.070 | 0.000 | 0.392 | 0.653 | * |
| TEASE7S | 0.495 | 0.064 | 0.000 | 0.378 | 0.626 | * |

# Displaying Non-Invariant Items: Time Points With Significant Differences Compared to the Mean (V = 0.01)

| Item   | Loading          | Threshold     |
|--------|------------------|---------------|
| stub   | 3                | 1, 2, 3, 6, 8 |
| bkrule | -                | 5, 8          |
| harmo  | 1, 8             | 2, 8          |
| bkthin | 1, 2, 3, 7, 8    | 2, 8          |
| yell   | 2, 3, 6          | -             |
| takep  | 1, 2, 5          | 1, 2, 5       |
| fight  | 1, 5             | 1, 4          |
| lies   | -                | -             |
| tease  | -                | 1, 4, 8       |

- Observations nested within time and subject
- A large number of time points can be handled via Bayesian analysis
- A relatively small number of subjects is needed

# Intensive Longitudinal Data

- Time intensive data: More longitudinal data are collected where very frequent observations are made using new tools for data collection. Walls & Schafer (2006)
- Typically multivariate models are developed but if the number of time points is large these models will fail due to too many variables and parameters involved
- Factor analysis models will be unstable over time. Is it lack of measurement invariance or insufficient model?
- Random loading and intercept models can take care of measurement and intercept invariance. A problem becomes an advantage.
- Random loading and intercept models produce more accurate estimates for the loadings and factors by borrowing information over time
- Random loading and intercept models produce more parsimonious model

TITLE:          this is an example of longitudinal modeling using a
                cross-classified data approach where observations are
                nested within the cross-classification of time and subjects

MONTECARLO:
                NAMES = y1-y3;
                NOBSERVATIONS = 7500;
                NREPS = 1;
                CSIZES = 75[100(1)];! 75 subjects, 100 time points
                NCSIZE = 1[1];
                WITHIN = (level2a) y1-y3;
                SAVE = ex9.27.dat;

ANALYSIS:
                TYPE = CROSS RANDOM;
                ESTIMATOR = BAYES;
                PROCESSORS = 2;

MODEL
POPULATION:

```
%WITHIN%
s1-s3 | f by y1-y3;
f@1;
y1-y3*1.2; [y1-y3@0];
%BETWEEN level2a% ! across time variation
s1-s3*0.1;
[s1-s3*1.3];
y1-y3*.5;
[y1-y3@0];
%BETWEEN level2b% ! across subjects variation
f*1; [f*.5];
s1-s3@0; [s1-s3@0];
```

```
TITLE:      this is an example of a multiple indicator
            growth model with random intercepts and
            factor loadings using cross-classified
            data
DATA:       FILE = ex9.27.dat;
VARIABLE:   NAMES = y1-y3 time subject;
            USEVARIABLES = y1-y3 timescor;
            CLUSTER = subject time;
            WITHIN = timescor (time) y1-y3;
DEFINE:     timescor = (time-1)/100;
ANALYSIS:   TYPE = CROSSCLASSIFIED RANDOM;
            ESTIMATOR = BAYES;
            PROCESSORS = 2;
            BITERATIONS = (1000);
MODEL:      %WITHIN%
            s1-s3 | f BY y1-y3;
            f@1;
            s | f ON timescor; !slope growth factor s
            y1-y3; [y1-y3@0];
            %BETWEEN time% ! time variation
            s1-s3; [s1-s3]; ! random loadings
            y1-y3; [y1-y3@0]; ! random intercepts
            s@0; [s@0];
            %BETWEEN subject% ! subject variation
            f; [f]; ! intercept growth factor f
            s1-s3@0; [s1-s3@0];
            s; [s]; ! slope growth factor s
OUTPUT:     TECH1 TECH8;
```

# Cross-Classified Analysis
## of Aggressive-Disruptive Behavior in the Classroom

- Teacher-rated measurement instrument capturing aggressive-disruptive behavior among a sample of U.S. students in Baltimore public schools (Ialongo et al., 1999).
- The instrument consists of 9 items scored as 0 (almost never) through 6 (almost always)
- A total of 1174 students are observed in 41 classrooms from Fall of Grade 1 through Grade 6 for a total of 8 time points
- The multilevel (classroom) nature of the data is ignored in the current analyses
- The item distribution is very skewed with a high percentage in the Almost Never category. The items are therefore dichotomized into Almost Never versus the other categories combined
- We analyze the data on the original scale as continuous variables and also the dichotomized scale as categorical

- For each student a 1-factor analysis model is estimated with the 9 items at each time point
- Let $Y_{pit}$ be the $p-$th item for individual $i$ at time $t$
- We use cross-classified SEM. Observations are nested within individual and time.
- Although this example uses only 8 time points the models can be used with any number of time points.

- Model 1: Two-level factor model with intercept non-invariance across time

$$Y_{pit} = \mu_p + \zeta_{pt} + \xi_{pi} + \lambda_p \eta_{it} + \varepsilon_{pit}$$

- $\mu_p$, $\lambda_p$ are model parameters, $\varepsilon_{pit} \sim N(0, \theta_{w,p})$ is the residual
- $\zeta_{pt} \sim N(0, \sigma_p)$ is a random effect to accommodate intercept non-invariance across time
- To correlate the factors $\eta_{it}$ within individual $i$

$$\eta_{it} = \eta_{b,i} + \eta_{w,it}$$

- $\eta_{b,i} \sim N(0, \psi)$ and $\eta_{w,it} \sim N(0, 1)$. The variance is fixed to 1 to identify the scale in the model
- $\xi_{pi} \sim N(0, \theta_{b,p})$ is a between level residual in the between level factor model
- Without the random effect $\zeta_{pt}$ this is just a standard two-level factor model

MODEL:

```
        %WITHIN%
        f BY y1-y9*1 (11-19);
        f@1;

        %BETWEEN t1%
        y1-y9;

        %BETWEEN id%
        y1-y9;
        f BY y1-y9*1 (11-19);
```

- Model 2: Adding latent growth model for the factor

$$\eta_{it} = \alpha_i + \beta_i \cdot t + \eta_{w,it}$$

- $\alpha_i \sim N(0, v_\alpha)$ is the intercept and $\beta_i \sim N(\beta, v_\beta)$ is the slope. For identification purposes again $\eta_{w,it} \sim N(0,1)$
- The model looks for developmental trajectory across time for the aggressive-disruptive behavior factor

```
MODEL:      ! s = beta, fb = alpha
            %WITHIN%
            f BY y1-y9*1 (11-19);
            f@1;
            s | f ON time;

            %BETWEEN t1%
            y1-y9;
            s@0; [s@0];

            %BETWEEN id%
            y1-y9;
            f BY y1-y9*1 (11-19);
            s*1; [s*0];
```

- Model 3: Adding measurement non-invariance
- Replace the fixed loadings $\lambda_p$ with random loadings $\lambda_{pt} \sim N(\lambda_p, w_p)$
- The random loadings accommodate measurement non-invariance across time
- All models can be estimated for continuous and categorical scale data

# Aggressive-Disruptive Behavior Example Continued: Model 3 Setup

MODEL:
$\qquad$ %WITHIN%
$\qquad$ s1-s9 | f BY y1-y9;
$\qquad$ f@1;
$\qquad$ s | f ON time;
$\qquad$ %BETWEEN t1%
$\qquad$ y1-y9;
$\qquad$ f@0; [f@0];
$\qquad$ s@0; [s@0];
$\qquad$ s1-s9*1; [s1-s9*1];
$\qquad$ %BETWEEN id%
$\qquad$ y1-y9;
$\qquad$ f*1; [f@0];
$\qquad$ s*1; [s*0];
$\qquad$ s1-s9@0; [s1-s9@0];

```
Within Level
 Residual Variances
    Y1              1.073       0.022       0.000       1.029       1.119
    Y9              0.630       0.014       0.000       0.604       0.658
    F               1.000       0.000       0.000       1.000       1.000

Between ID Level
 Variances
    Y1              0.146       0.016       0.000       0.118       0.180
    Y9              0.052       0.009       0.000       0.035       0.068
    F               1.316       0.080       0.000       1.172       1.486
    S               0.026       0.003       0.000       0.020       0.032

Between T1 Level
 Means
    Y1              1.632       0.120       0.000       1.377       1.885
    Y9              1.232       0.096       0.000       1.044       1.420
    S1              0.679       0.023       0.000       0.640       0.732
    S9              0.705       0.043       0.000       0.628       0.797
 Variances
    Y1              0.080       0.138       0.000       0.025       0.372
    Y9              0.047       0.109       0.000       0.017       0.266
    S1              0.002       0.004       0.000       0.000       0.013
    S9              0.010       0.079       0.000       0.003       0.052
```

- Model 4: Adding measurement non-invariance also across individuals
- Replace the loadings $\lambda_{pt}$ with random loadings

$$\lambda_{pit} = \lambda_{pi} + \lambda_{pt}$$

where $\lambda_{pt} \sim N(\lambda_p, w_p)$ and $\lambda_{pi} \sim N(0, w_i)$

- The random loadings accommodate measurement non-invariance across time and individual
- Model 4: Adding factor variance non-invariance across time. Can be done either by adding (a) introducing a factor model for the random loadings or (b) introducing a random loadings for the residual of the factor.
- We choose (b). $Var(f) = 0.51 + (0.7 + \sigma_t)^2$ where $\sigma_t$ is a mean zero random effect

```
model:
%within%
s1-s9 | f by y1-y9;
s | f on time; f@0.51;
ss | e by f; e@1;

%between T1%
y1-y9;
s@0; [s@0];
s1-s9*1; [s1-s9*1];
ss*0.6; [ss@0.7];

%between ID%
y1-y9;
f*1; [f@0];
s*1; [s*0];
s1-s9*1; [s1-s9@0];
ss@0; [ss@0];
```

# Aggressive-Disruptive Behavior Example Continued: Results For Categorical Analysis

```
Between ID Level
 Variances
    Y1              0.113      0.039      0.000      0.045      0.195
    Y9              0.194      0.049      0.000      0.108      0.299
    S1              0.082      0.029      0.000      0.041      0.152
    S9              0.030      0.025      0.000      0.001      0.089
    F               1.789      0.241      0.000      1.353      2.300

Between T1 Level
 Means
    S1              0.821      0.094      0.000      0.647      1.018
    S9              1.049      0.125      0.000      0.809      1.299
 Thresholds
    Y1$1           -0.864      0.156      0.000     -1.183     -0.558
    Y2$1           -0.800      0.120      0.000     -1.041     -0.568
 Variances
    Y1              0.120      0.231      0.000      0.035      0.666
    Y9              0.073      0.158      0.000      0.019      0.416
    S1              0.023      0.055      0.000      0.005      0.137
    S9              0.050      0.110      0.000      0.011      0.302
```

- Other extensions of the above model are possible, for example the growth trend can have time specific random effects: $f$ and $s$ can be free over time

- The more clusters there are on a particular level the more elaborate the model can be on that level. However, the more elaborate the model on a particular level is, the slower the convergence

- The main factor $f$ can have a random effect on each of the levels, however the residuals $Y_i$ should be uncorrelated on that level. If they are correlated through another factor model such as, $fb$   $by$   $y_1 - y_9$, then $f$ would be confounded with that factor $fb$ and the model will be poorly identified

- On each level the most general model would be (if there are no random slopes) the unconstrained variance covariance for the dependent variables $Y_i$. Any model that is a restriction of that model is in principle identified

# Aggressive-Disruptive Behavior Example Conclusions, Continued

- Unlike ML and WLS multivariate modeling, for the time intensive Bayes cross-classified SEM, the more time points there are the more stable and easy to estimate the model is
- Bayesian methods solve problems not feasible with ML or WLS
- Time intensive data naturally fits in the cross-classified modeling framework
- Asparouhov and Muthén (2012). General Random Effect Latent Variable Modeling: Random Subjects, Items, Contexts, and Parameters

- Jeon & Rabe-Hesketh (2012). Profile-Likelihood Approach for Estimating Generalized Linear Mixed Models With Factor Structures. JEBS
- Longitudinal growth model for student self-esteem
- Each student has 4 observations: 2 in middle school in wave 1 and 2, and 2 in high school in wave 3 and 4
- Students have multiple membership: Membership in middle school and in high school with a random effect from both
- $Y_{tsmh}$ is observation at time $t$ for student $s$ in middle school $m$ and high school $h$
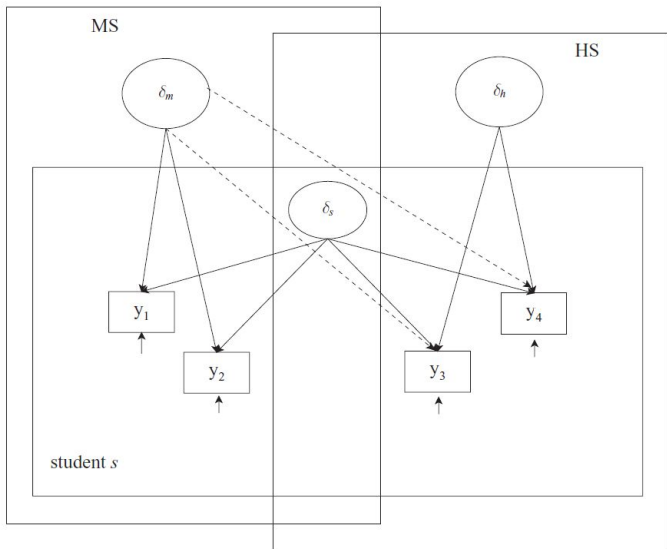
- The model is

$$Y_{tsmh} = \beta_1 + \beta_2 T2 + \beta_3 T3 + \beta_4 T4 + \delta_s + \delta_m \mu_t + \delta_h \lambda_t + \varepsilon_{tsmh}$$

  where T2, T3, T4 are dummy variables for wave 2, 3, 4

- $\delta_s$, $\delta_m$ and $\delta_h$ are zero mean random effect contributions from student, middle school and high school
- $\mu_t = (1, \mu_2, \mu_3, \mu_4)$
- $\lambda_t = (0, 0, 1, \lambda_4)$, i.e., no contribution from the high school in wave 1 and 2 because the student is still in middle school
- $\varepsilon_{tsmh}$ is the residual
- Very simple to setup in Mplus

MODEL:

        %WITHIN%
        fs BY y1-y4@1;
        [y1-y4];

        %BETWEEN mschool%
        fm BY y1@1 y2-y4;
        y1-y4@0; [y1-y4@0];

        %BETWEEN hschool%
        fh BY y1@0 y2@0 y3@1 y4;
        y1-y4@0; [y1-y4@0];