Day 1, Lecture 2: Likelihood Based Modeling with Incomplete Data

John J. McArdle
Longitudinal Research Institute

CALDAR, August 2010

Overview

1. Likelihood Based Structural Modeling of Incomplete Data
2. Alternative Models of Two-Occasion Data
3. Longitudinal Change SEMs, Including Incomplete Data
4. Longitudinal Modeling of Complete Longitudinal WISC Data
5. Masking Longitudinal WISC Data

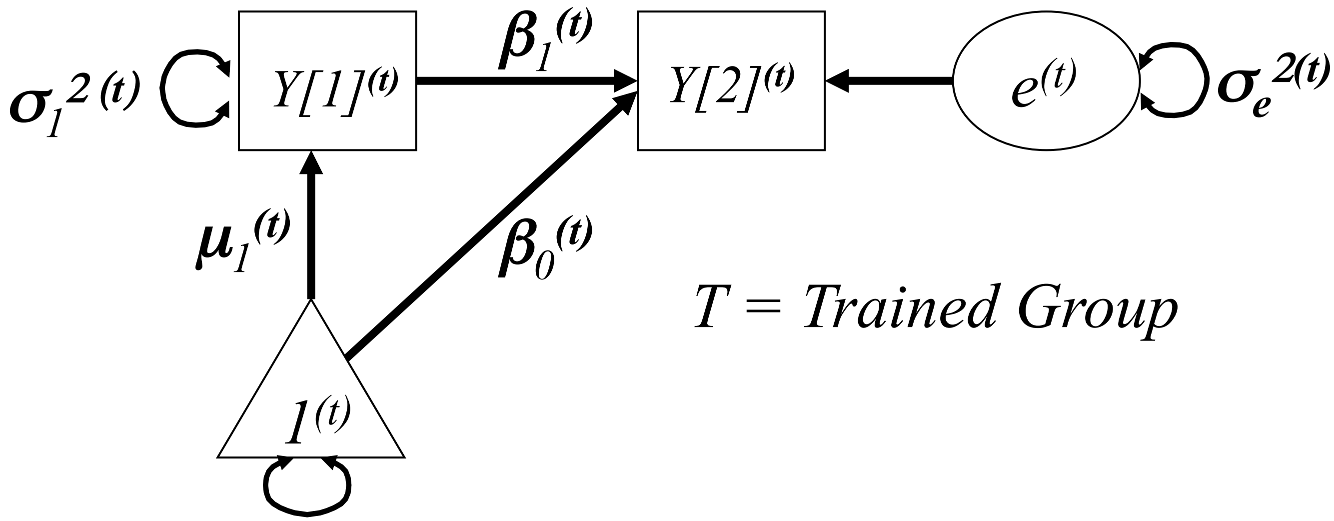
MG-SEM with incomplete data

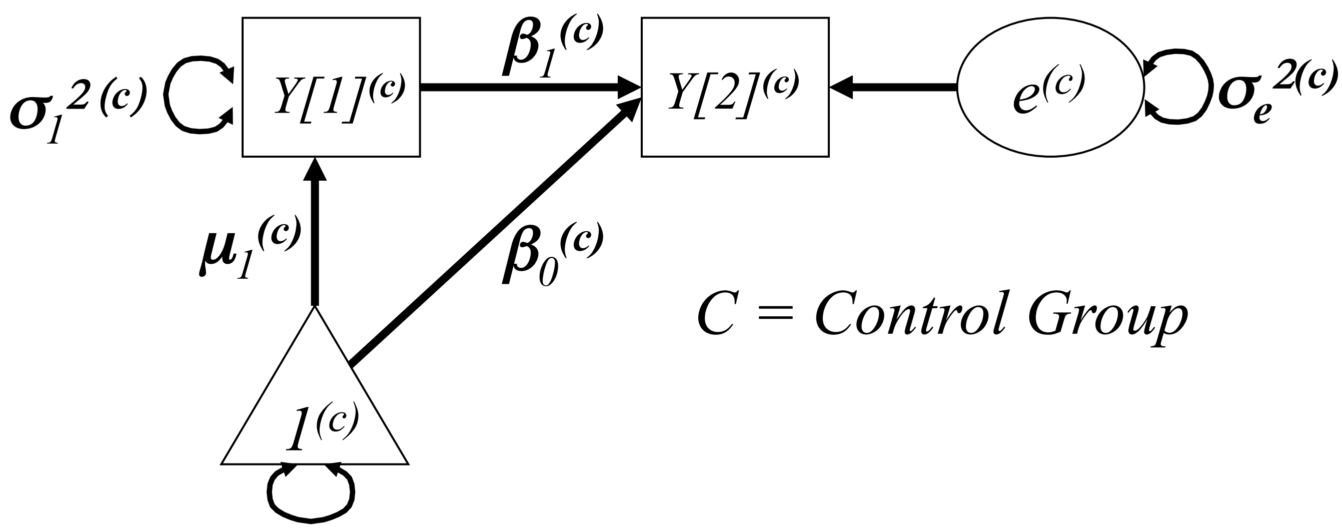
- MG-SEM offers the possibility of dealing directly with incomplete or missing data. In this approach the available data for each person is put in a group as *defined by the pattern of available data*.

- A model is used so we *assuming but do not testing* a model with *metric invariance* fitted to all groups – this is equivalent to the classic *Missing-At-Random* (MAR) assumptions (Rubin, 1987; Schaffer, 1999).

- If the MAR assumption is adequate, the MG-SEM can be used for a wide variety of incomplete/missing data patterns, including planned incomplete data.

A general multiple group auto-regression model with group differences





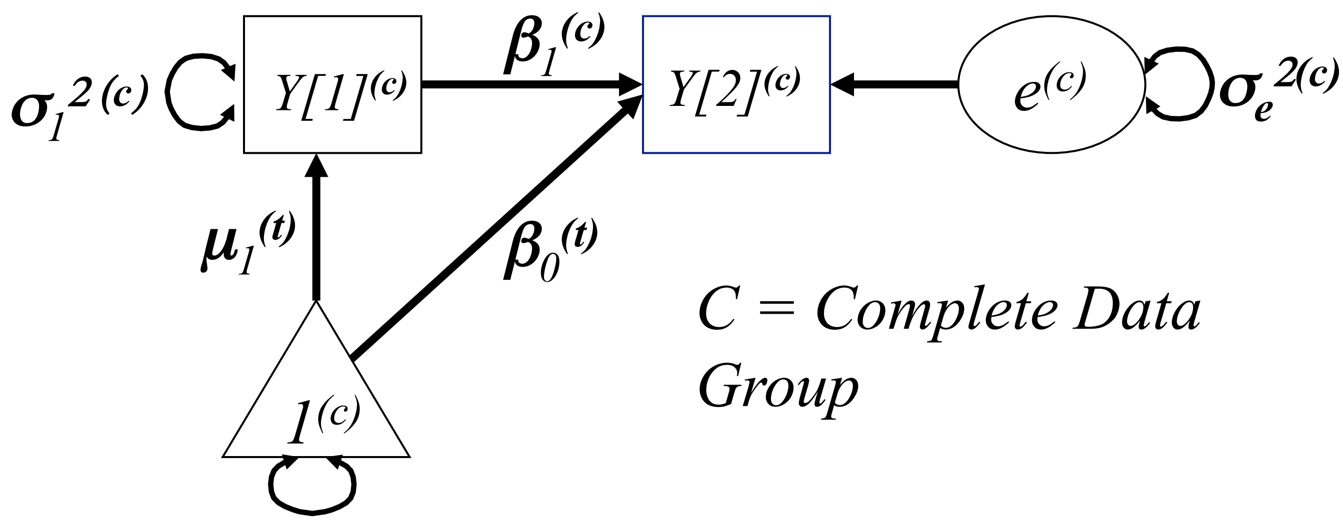
Incomplete regression SEM

- Assume a regression model based on moments -- scores for *Y*[1] and *Y*[2]

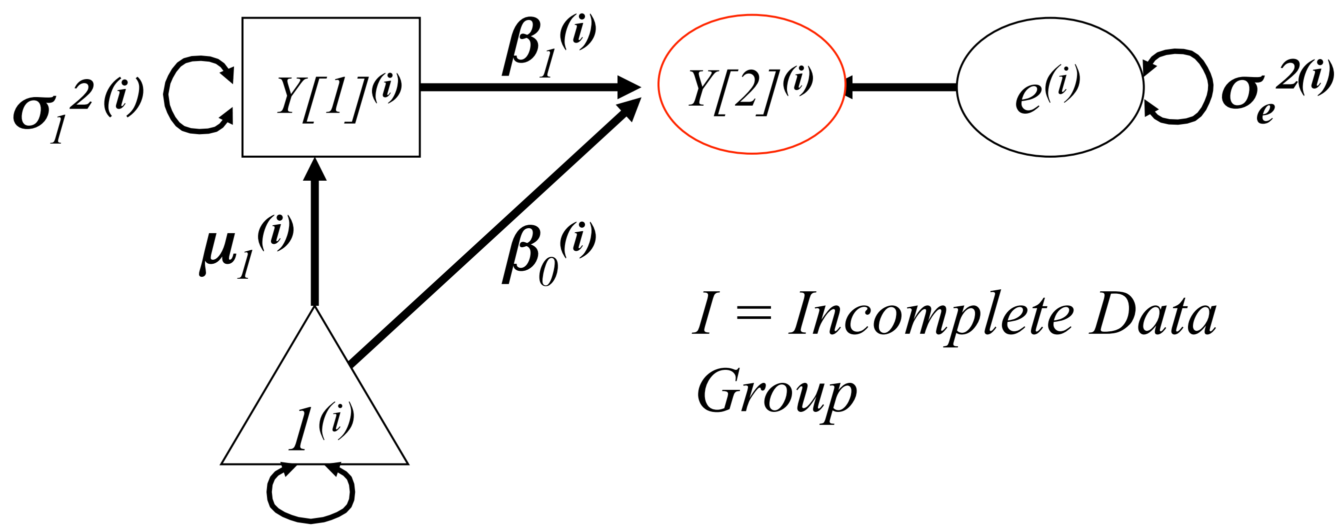
- Consider missing data in one group *Y*[1]
- Consider missing data in one group *Y*[2]

- Define overall model assumptions in regression matrix **A** and covariance matrix **S** but use **F** to define the *pattern of observed variables in each group*

A general multiple group auto-regression model with incomplete data differences



C = Complete Data Group



I = Incomplete Data Group

Standard "likelihood-based" calculations

- Now consider 3 groups with (a) all data, (b) data on Y[1] only, and (c) data on Y[2] only.
- In this case we write the joint likelihood as

  $$\boldsymbol{L^2 = L(\mathrm{a})^2 + L(\mathrm{b})^2 + L(\mathrm{c})^2} \qquad \text{where}$$

  $$\boldsymbol{L(\mathrm{a})^2 = N(\mathrm{a})^* f\{\ \mathrm{m}(1,2)-\mu(1,2) + \mathrm{C}(1,2)-\Sigma(1,2)\},}$$

  $$\boldsymbol{L(\mathrm{b})^2 = N(\mathrm{b})^* f\{\ \mathrm{m}(1)-\mu(1) + \mathrm{C}(1)-\Sigma(1)\}, \text{ and}}$$

  $$\boldsymbol{L(\mathrm{c})^2 = N(\mathrm{c})^* f\{\ \mathrm{m}(2)-\mu(2) + \mathrm{C}(2)-\Sigma(2)\}.}$$

- Assume that any model has a log likelihood L^2 based on the means (m) and covariances (C) for all N individuals, and we write

  $$\boldsymbol{L^2 = N * f\{\ \mathrm{m}(1,2)-\mu(1,2) + \mathrm{C}(1,2)-\Sigma(1,2)\}}$$

2. SEM for incomplete CESD data

M+3.0 input for auto-regression

```
TITLE: Simple CESD Auto-regression -- Complete Cases Only
DATA:          FILE = cesd0910.dat;
VARIABLE:      NAMES = cesd_09 cesd_10;
               USEVAR= cesd_09 cesd_10;
               MISSING= . ;

ANALYSIS:      TYPE=MEANSTRUCTURE;
MODEL:         cesd_10 ON cesd_09;
OUTPUT:                SAMPSTAT;
```

M+3.0 input for auto-regression

```
TITLE: Simple CESD Auto-regression -- Complete Cases Only
DATA:          FILE = cesd0910.dat;
VARIABLE:      NAMES = cesd_09 cesd_10;
               USEVAR= cesd_09 cesd_10;
               MISSING= . ;
ANALYSIS:      TYPE=MEANSTRUCTURE;
MODEL:         cesd_10 ON cesd_09;
OUTPUT:        SAMPSTAT RESIDUAL STANDARDIZED;
```

```
TITLE: Simple CESD Auto-regression model -- ALL data
DATA:          FILE = cesd0910.dat;
VARIABLE:      NAMES = cesd_09 cesd_10;
               USEVAR= cesd_09 cesd_10;
               MISSING=.;
ANALYSIS:      TYPE=MEANSTRUCTURE MISSING;
MODEL:         cesd_10 ON cesd_09;
OUTPUT:        SAMPSTAT RESIDUAL STANDARDIZED PATTERNS ;
```

Complete case auto-regression output

```
SUMMARY OF ANALYSIS
Number of observations                          560
Number of y-variables                             1
Number of x-variables                             1
              Means
             CESD_10      CESD_09
1            14.280        15.559
Covariances
              CESD_10      CESD_09
CESD_10        110.943
CESD_09         58.890       113.979
Correlations
              CESD_10      CESD_09
              ________
 CESD_10        1.000
 CESD_09        0.524
THE MODEL ESTIMATION TERMIN
TESTS OF MODEL FIT
Chi-Square Test of Model Fi
          Value
          Degrees of Freedo
          P-Value
Loglikelihood          H0 V
;
```

```
MODEL RESULTS

                       Estimates     S.E.  Est./S.

 CESD)10   ON
    CESD_09               0.517    0.036     14.547

 Residual Variances
    CESD10               80.372    4.803     16.733

 Intercepts
    CESD10                6.242    0.670      9.316
```

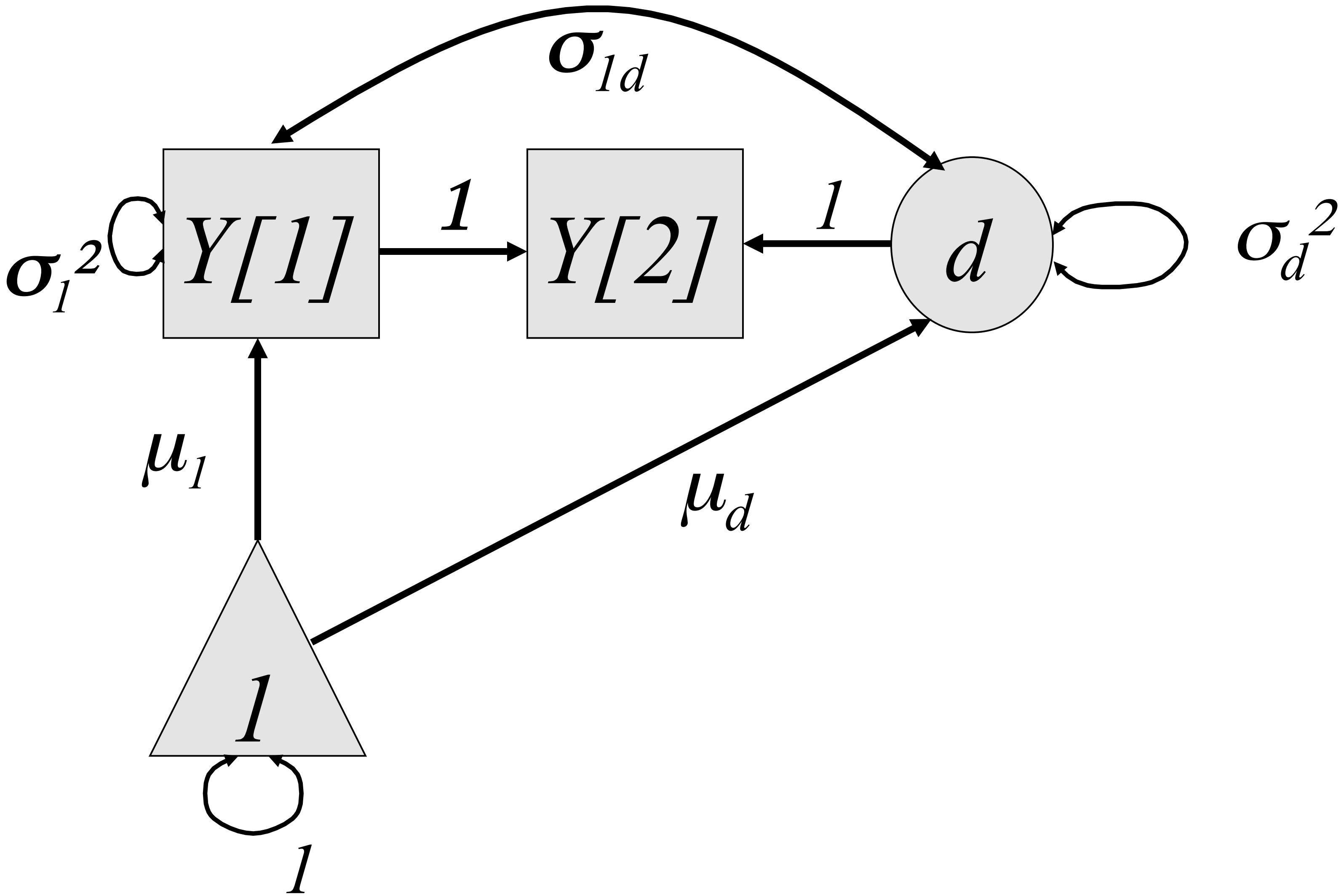
ALL data auto-regression output

```
Simple CESD auto-regression model with Incomplete data
SUMMARY OF ANALYSIS
Number of observations                             1104
Number of y-variables                                 1
Number of x-variables                                 1
Observed variables in the analysis  CESD9        CESD10
Estimator                                           ML
SUMMARY OF DATA
     Number of patterns           3
Simple CESD autoregression model with Incomplete data

SUMMARY OF MISSING DATA PATTERNS
     MISSING DATA PATTERNS
            1  2  3
 CESD_10    x  x
 CESD_09    x     x
     MISSING DATA PATTERN FREQUENCIES
    Pattern   Frequency     Pattern   Frequency   Pattern   Frequency
          1         560           2         112         3         432

COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100
     PROPORTION OF DATA PRESENT
           Covariance Coverage
                 CESD_10      CESD_09
                ________     ________
 CESD_10         0.609
 CESD_09         0.507        0.899
```

ALL data auto-regression output

```
SAMPLE STATISTICS
     ESTIMATED SAMPLE STATISTICS
           Means
            CESD_10       CESD_09
    1        14.570        16.135
           Covariances
            CESD_10       CESD_09
CESD_10      109.601
CESD_09       60.141        118.450
           Correlations
              CESD_10     CESD_09
CESD10         1.000
CESD_09        0.528          1.000
MAXIMUM LOG-LIKELIHOOD VALUE FOR THE
UNRESTRICTED (H1) MODEL IS   -6215.936
THE MODEL ESTIMATION TERMINATED NORMALLY
TESTS OF MODEL FIT
Chi-Square Test of Model Fit
         Value                               0.000
          Degrees of Freedom                     0
          P-Value                           0.0000
MODEL RESULTS
                   Estimates     S.E.  Est./S.E.
CESD_10   ON
    CESD_09            0.508    0.033     15.235
Residual Variances
    CESD_10           79.058    4.512     17.521
Intercepts
    CESD_10            6.377    0.642      9.928
```

Model D: A path diagram including a latent change score



M+ ALL case change score

```
TITLE:   CESD difference score model with ALL data
DATA:          FILE = cesd0910.dat;
VARIABLE:      NAMES = cesd_09 cesd_10;
               USEVAR= cesd_09 cesd_10;
               MISSING=.;

ANALYSIS:      TYPE=MEANSTRUCTURE MISSING;

MODEL:         y1 BY cesd_09 @1;
              y2 BY cesd_10 @1;
              y0 BY y1 @1;
               change BY y2 @1;
              y2 ON y1 @1 change @1;
              cesd_09 @0.0 cesd_10 @2.0 y1 @0 y2 @0
              [cesd_099 @0.0 cesd_10 @0.0 y1 @0 y2 @0];
              [y0*20 change*15];

OUTPUT:       PATTERNS SAMPSTAT;
```

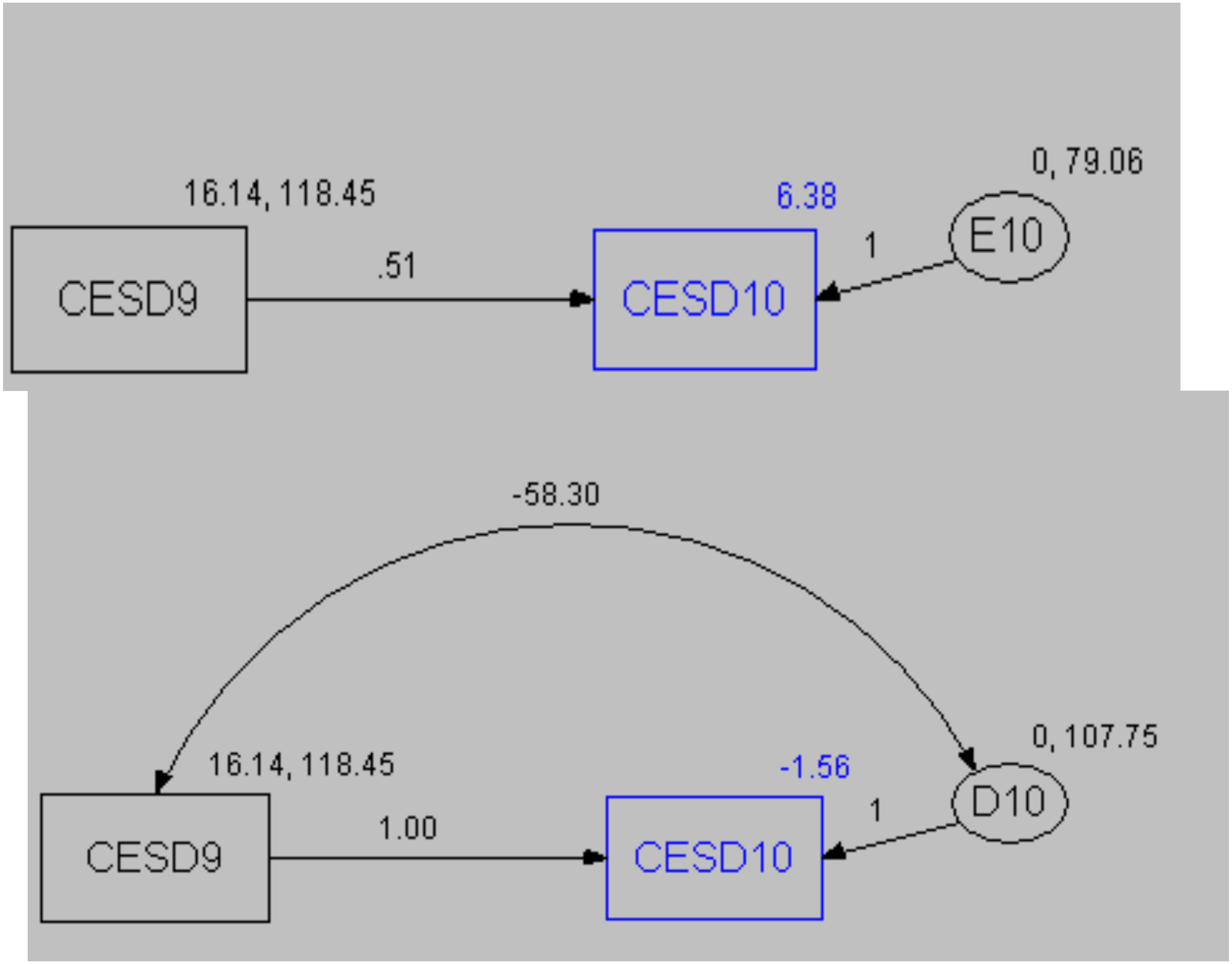
Complete Cases change score output

```
MODEL RESULTS (Complete Cases N=560)
                   Estimates     S.E.  Est./S.E.
 CHANGE   BY
    CESD_09            0.000    0.000      0.000
 CESD_10   ON
    CHANGE             1.000    0.000      0.000
 CESD_10   ON
    CESD_09            1.000    0.000      0.000
 CESD_09    WITH
    CHANGE           -55.056    5.212    -10.564

 Residual Variances
    CESD_09          113.776    6.799     16.733
    CESD_10            0.000    0.000      0.000

 Variances
    CHANGE           107.047    6.397     16.733

 Means
    CHANGE            -1.275    0.437     -2.916

 Intercepts
    CESD_09           15.557    0.451     34.514
    CESD_10            0.000    0.000      0.000
```

ALL data change score output

```
MODEL RESULTS (All data N=1104)
                    Estimates     S.E.  Est./S.E.
 CHANGE     BY
    CESD_09             0.000    0.000      0.000
 CESD_10    ON
    CHANGE              1.000    0.000      0.000
 CESD_10    ON
    CESD_09             1.000    0.000      0.000
 CESD_09     WITH
    CHANGE            -58.134    4.768    -12.193

Residual Variances
   CESD_09            117.656    5.264     22.352
   CESD_10              0.000    0.000      0.000
 Variances
    CHANGE            107.791    6.324     17.045

 Means
    CHANGE             -1.571    0.399     -3.936

 Intercepts
    CESD_09            16.134    0.340     47.479
    CESD_10             0.000    0.000      0.000
```

AMOS Input/Output of CESD problem

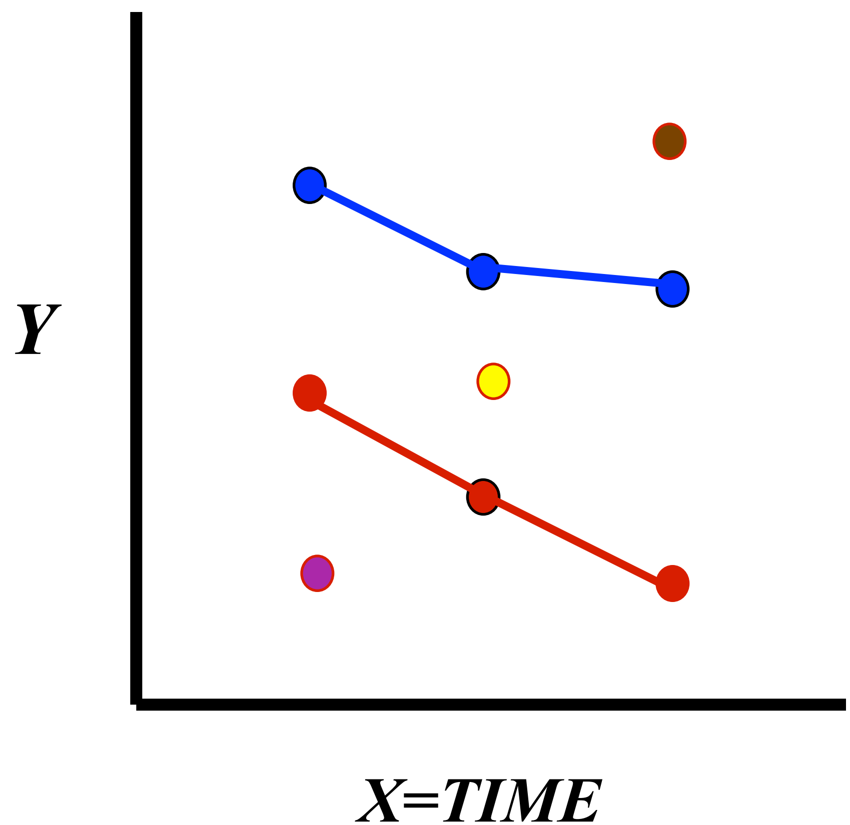


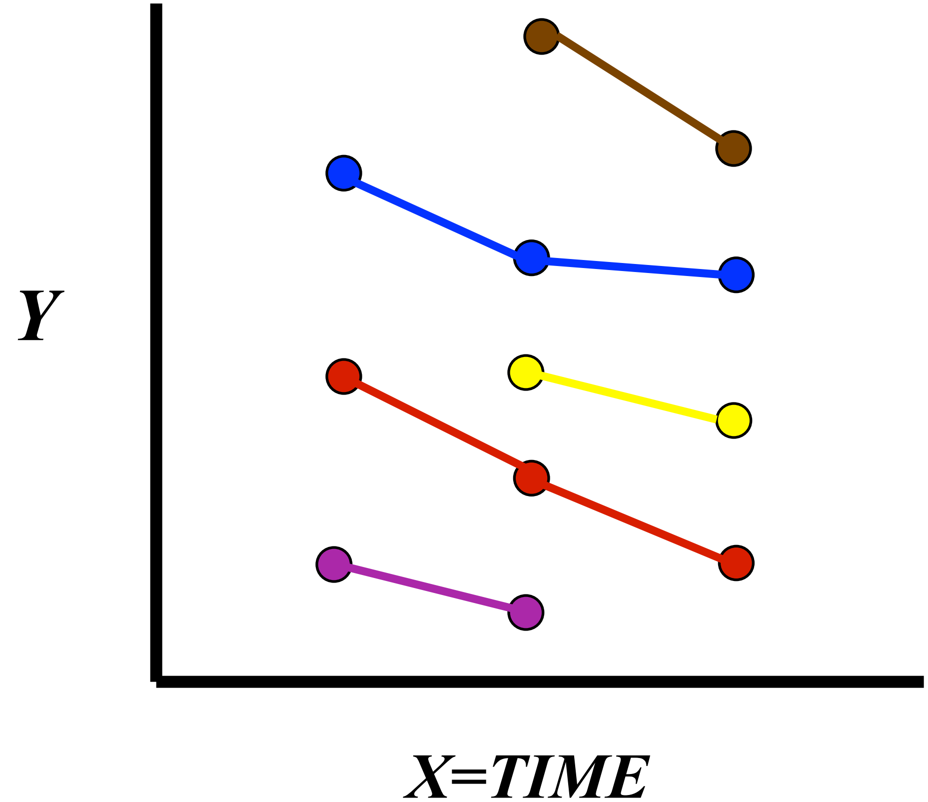
3. General SEMs with Incomplete Longitudinal Data

Dealing with incomplete data as groups with comparable patterns

- Practical problems occur in any longitudinal study, including initial subject self-selection and subject dropout (attrition) or designed lack of measurement
- There are many new and innovative analyses for these general problems --- including SEM, Multiple Imputation, Gibbs Sampling (Little & Rubin, 1987)
- The SEM presented here rely on a *likelihood-based* approach to incomplete patterns (e.g., multiple groups), and this approach is identical to most "unbalanced" or "mixed" or "multilevel" models
- In these SEM, no data are "imputed" – the parameters are estimated using "all available data" (not "complete cases").

Dealing with incomplete longitudinal data





Adding Incomplete Data

Adding Incomplete Pairs

A “typical” longitudinal study

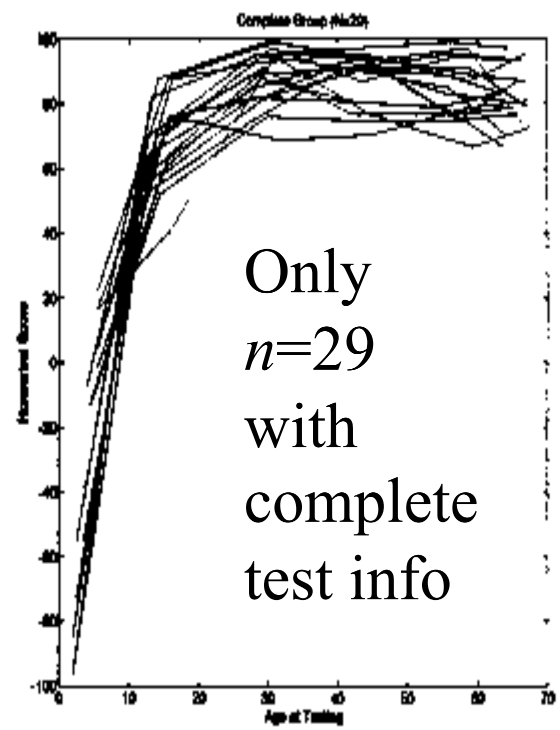


Figure [1]: Complete Bradway Longitudinal Non-Verbal data (N=29)

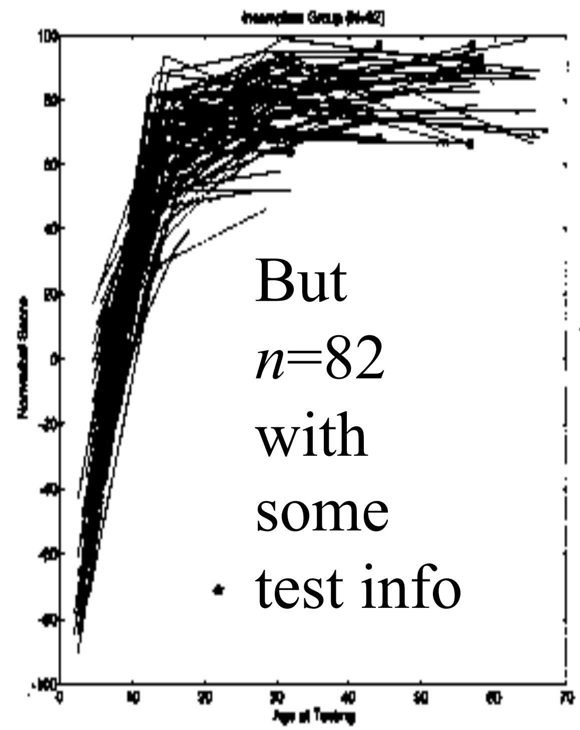
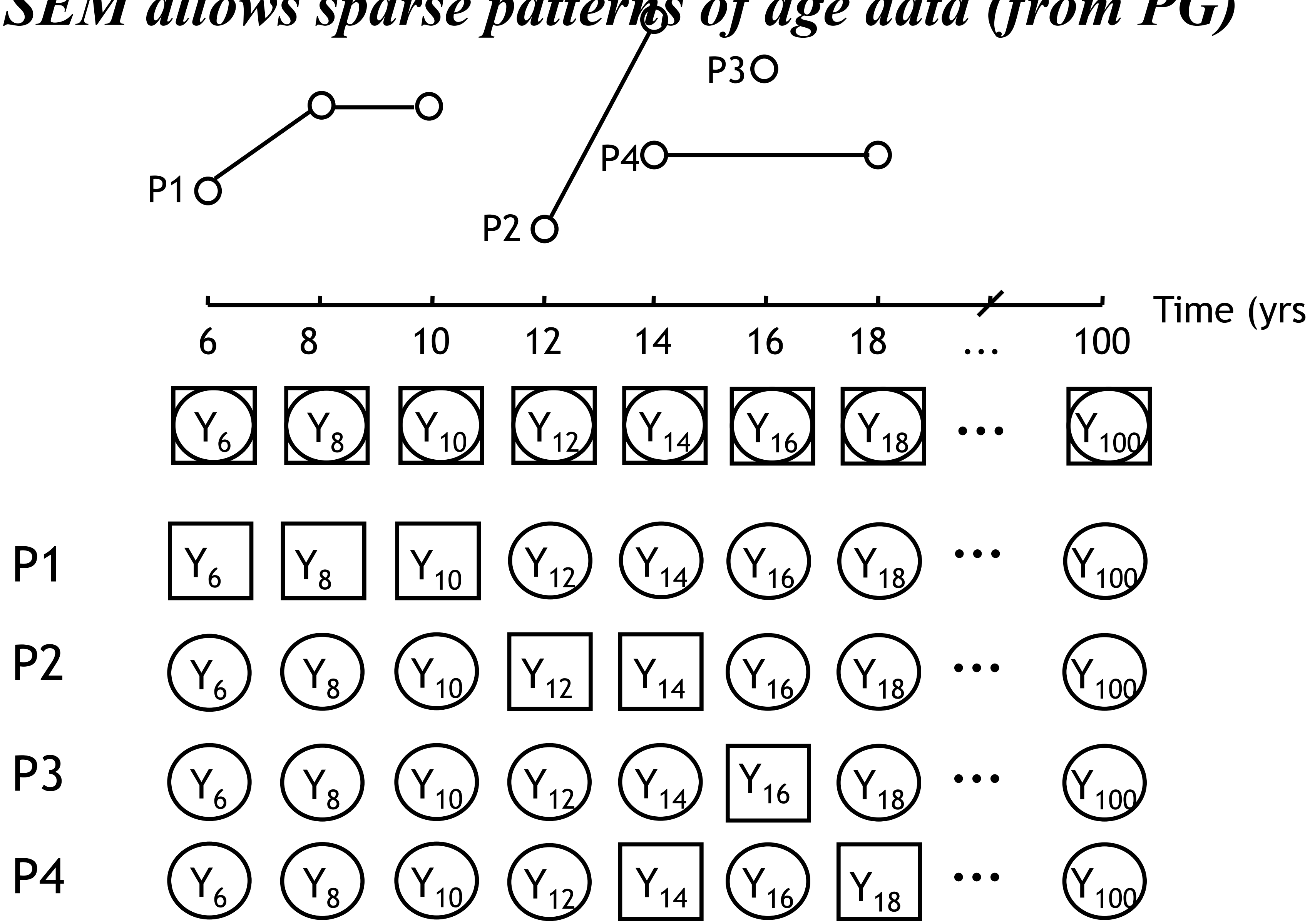


Figure [2]: Incomplete Bradway Longitudinal Non-Verbal data (N=82)

Note: These N=111 individual records have D=498 data points and were used in latent growth model fitting

Given assumptions of growth invariance, SEM allows sparse patterns of age data (from PG)



All SEM based on MLE modeling of incomplete data

- New statistical methods (e.g., Little & Rubin, 1987) offers ways to deal directly with incomplete or missing data
- Model parameters are likelihood-based *calculations* instead of score *imputations*
- Statistical estimation can assess MAR assumptions and correct the resulting MLE
- Leads to a variety of implications for planned incomplete or missing data
- But we need to start with an overall model for the data

SEM basis of Latent Curve models

1. We start with a "first level" model of random effects

$$\boldsymbol{Y[\mathrm{t}]_{\mathrm{n}} = i_{\mathrm{n}} + A[\mathrm{t}]\ \ s_{\mathrm{n}} + e[\mathrm{t}]_{\mathrm{n}}}$$

where ***i*** latent scores representing an individual's initial level or intercept; ***A*[t]** are group "basis" parameters represent some form of timing; ***s*** are latent slopes for the individual change over time, and the ***e*[t]** are errors of measurements.

2. There is also a second level of "fixed effects"

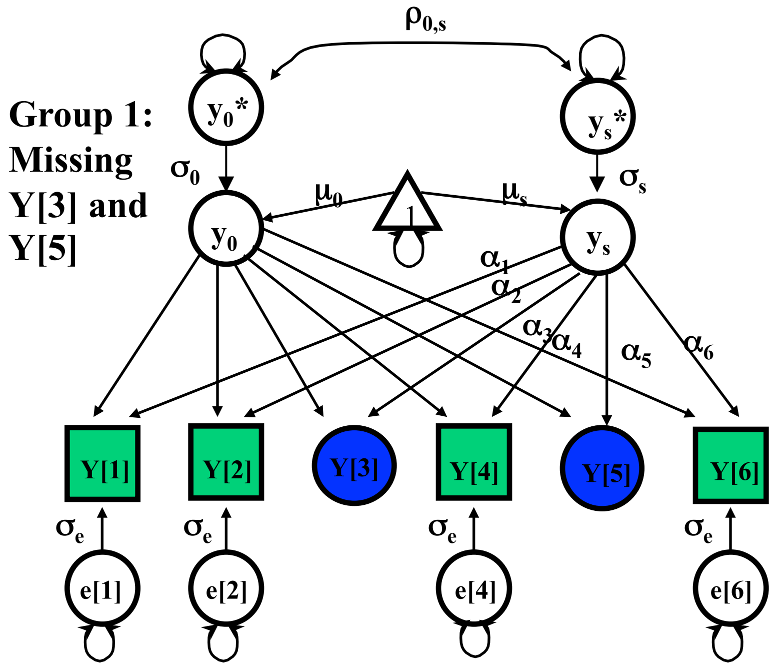
$$\boldsymbol{i_n = \mu_i + d_{in}}$$
$$\boldsymbol{s_n = \mu_s + d_{sn}}$$

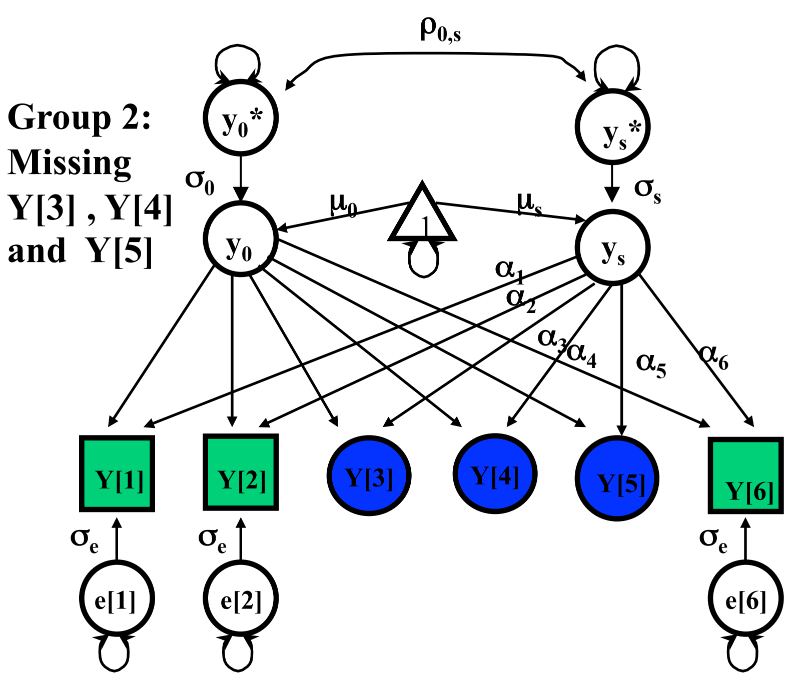
so the levels and slope scores have means ($\boldsymbol{\mu_j}$) and residuals ($\boldsymbol{d_j}$), and the residuals have variance components ($\boldsymbol{\sigma_j^2}$).

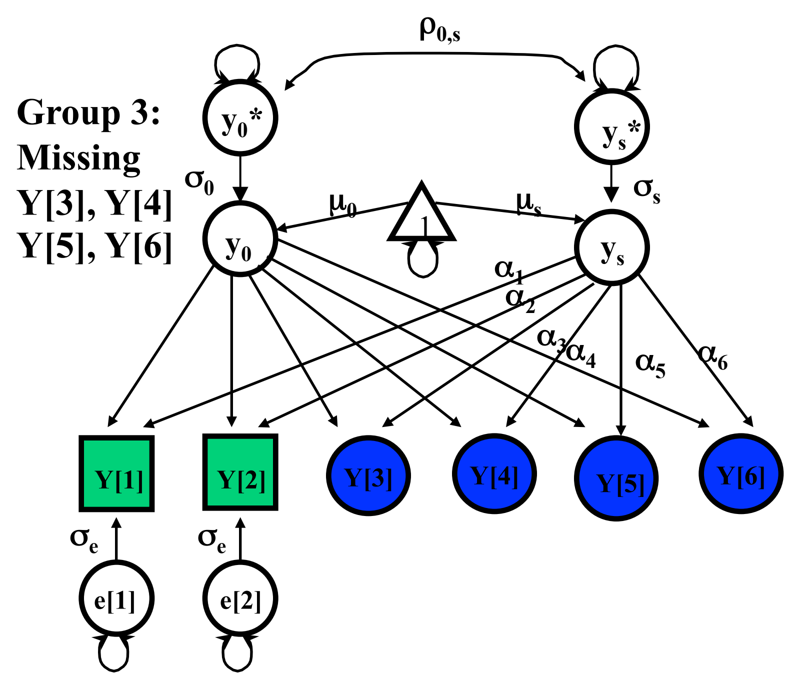
3. Fitted directly to overcome problems of "gain scores"

Three groups with different patterns of incomplete data.

Note: The key "untested" assumption is the groups all come from the same population, selected only by differences in the factor scores -- i.e., that factorial invariance holds.







Incomplete data SEM studies

- Initial simulation examples based on “node” (or “phantom”) variables with Horn (‘80)
- Multiple growth curve analysis in aging with Anderson (‘90) and Aber (‘91)
- Various studies with Hamagami (‘91-03)
- Statistical vector fields with Boker (’95)
- Behavioral genetics with Prescott (‘96-98)
- Test-retest designs with Woodcock (‘97)
- McArdle & Bell (2000) and others
- Many post-2000 simulation studies

Contemporary raw data SEM programs

- Assume an overall model based on independent groups of subjects
- Calculate the model parameters from the *model likelihood for each individual* (IML)
- Calculate the individual misfit from the *data likelihood for each individual* (IDL)
- Need to fit model with a special computer program such as Mx, M+, AMOS, etc.
- Some models more easily represented as Mixed models (SAS PROC MIXED, NLMIXED, etc.)

4. Complete Longitudinal WISC Data

An example of longitudinal WISC data

- Classic data from Osborne & Suddick (1972) as reported by McArdle & Epstein (1987)

- Repeated measures of WISC collected on *N*=204 children at ages 6, 7, 9 and 11 (or Grades 1, 2, 4 and 6)

- A variety of path and factor analytic SEM were used to evaluate models of development for the same data.

- Also used by McArdle (1988), McArdle & Aber (1990), McArdle & Nesselroade (1994), McArdle (1990, 2001)

Summary statistics for the WISC-Verbal

5 Variables: Verb_1 Verb_2 Verb_4 Verb_6 MothEduc

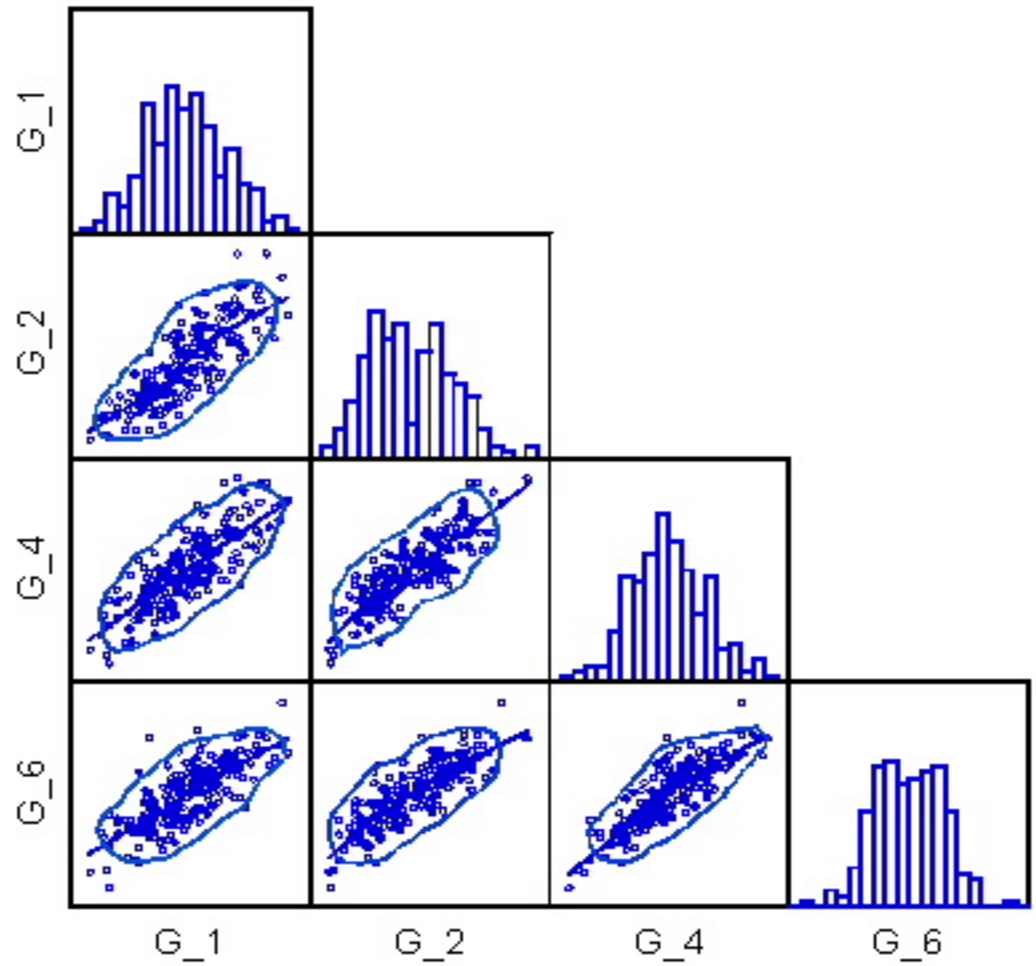
Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Verb_1	204	19.58501	5.80777	3995	3.33333	35.14881
Verb_2	204	25.41506	6.10647	5185	5.95238	39.85119
Verb_4	204	32.60745	7.31971	6652	12.60417	52.84226
Verb_6	204	43.74985	10.66495	8925	17.35119	72.58929
MothEduc	204	0.85294	0.76109	174.00000	0	2.00000

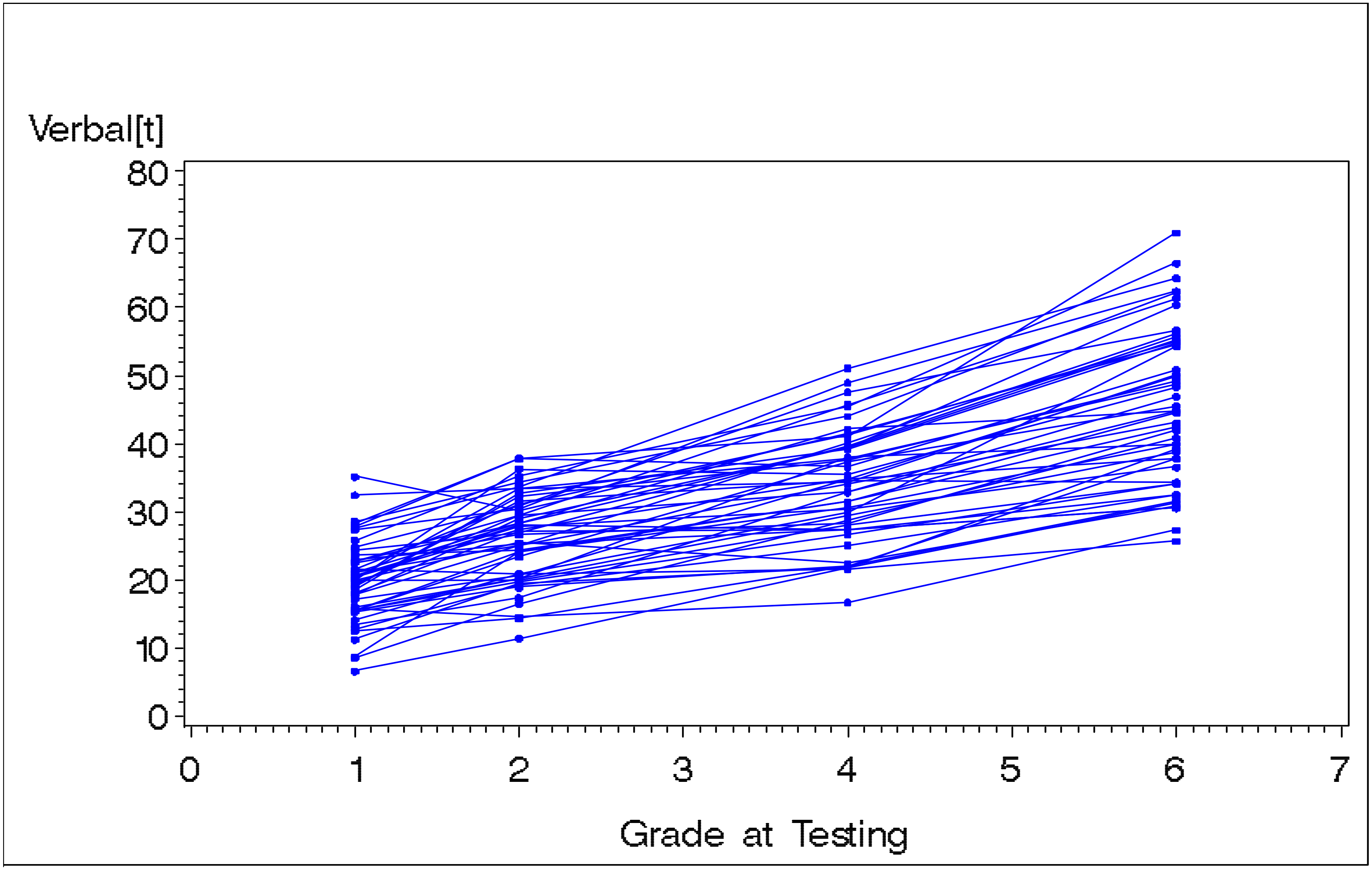
Pearson Correlation Coefficients, N = 204

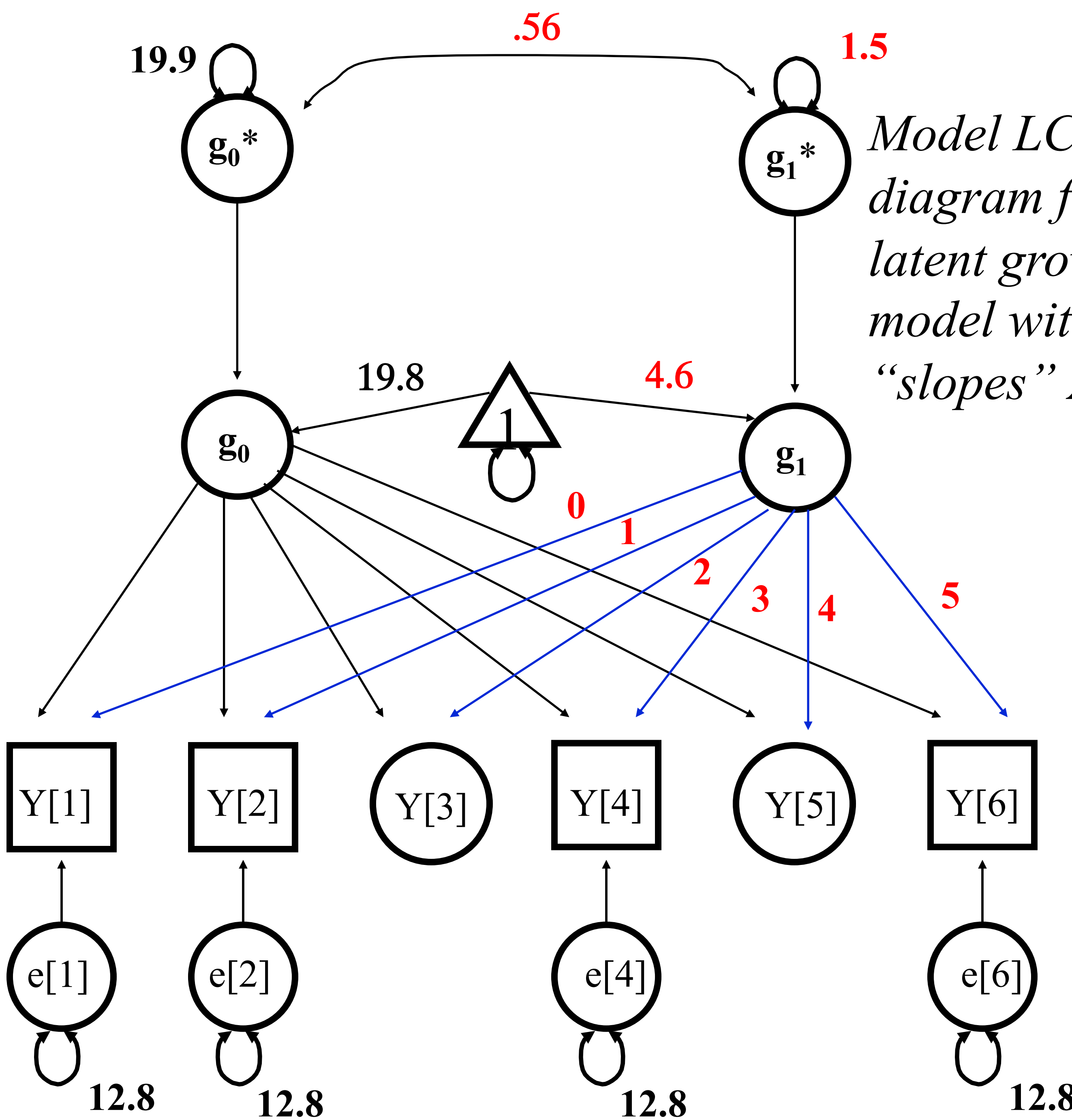
	Verb_1	Verb_2	Verb_4	Verb_6	MothEduc
Verb_1	1.00000	0.71803	0.72662	0.65415	0.45879
Verb_2	0.71803	1.00000	0.75665	0.72791	0.46060
Verb_4	0.72662	0.75665	1.00000	0.79746	0.47897
Verb_6	0.65415	0.72791	0.79746	1.00000	0.47778
MothEduc	0.45879	0.46060	0.47897	0.47778	1.00000

A scatterplot matrix for four occasions



Individual trajectories on the WISC-Verbal scales (N=50)





Model LC2: A path diagram for the latent growth model with linear "slopes" ***A***[*t*]=*t-1*

M+ programming for Linear LC2

```
TITLE:      Wisc_LC2.m+ Linear Latent Curve Model;
DATA:        FILE = Wisc4vpe.dat;
VARIABLE:    NAMES = V1 V2 V4 V6 P1 P2 P4 P6 Moeducat;
            USEVAR = V1 V2 V4 V6;

ANALYSIS:   TYPE = MEANSTRUCTURE;

MODEL:
              i BY V1-V6 @1;

             s BY  V1  @  0
                   V2  @  1
                    V4  @  3
                    V6  @  5;

             [i s];  ! Latent variables have means

            V1-V6 * 10 (Ve); ! Residual Variance equal
           [V1-V6 @0]; ! No additional intercepts

OUTPUT:     SAMPSTAT RESIDUAL STANDARDIZED;
```

```
                           Estimates     S.E.  Est./S.E.
 I        BY
    V1                        1.000    0.000      0.000
    V2                        1.000    0.000      0.000
    V4                        1.000    0.000      0.000
    V6                        1.000    0.000      0.000
 S        BY
    V1                        0.000    0.000      0.000
    V2                        1.000    0.000      0.000
    V4                        3.000    0.000      0.000
    V6                        5.000    0.000      0.000
 S         WITH
    I                         3.093    0.590      5.244
 Means
    I                        19.824    0.367     54.030
    S                         4.673    0.108     43.101
 Intercepts
    V1                        0.000    0.000      0.000
    V2                        0.000    0.000      0.000
    V4                        0.000    0.000      0.000
    V6                        0.000    0.000      0.000
 Variances
    I                        19.854    2.771      7.165
    S                         1.529    0.245      6.236
 Residual Variances
    V1                       12.828    0.898     14.283
    V2                       12.828    0.898     14.283
    V4                       12.828    0.898     14.283
    V6                       12.828    0.898     14.283
```

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	79.185
Degrees of Freedom	8
P-Value	0.0000

Loglikelihood

H0 Value	-2519.408
H1 Value	-2479.816

ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED)

Model Estimated Means/Intercepts/Thresholds

	V1	V2	V4	V6
1	19.824	24.498	33.844	43.191

Residuals for Means/Intercepts/Thresholds

	V1	V2	V4	V6
1	-0.239	0.917	-1.237	0.559

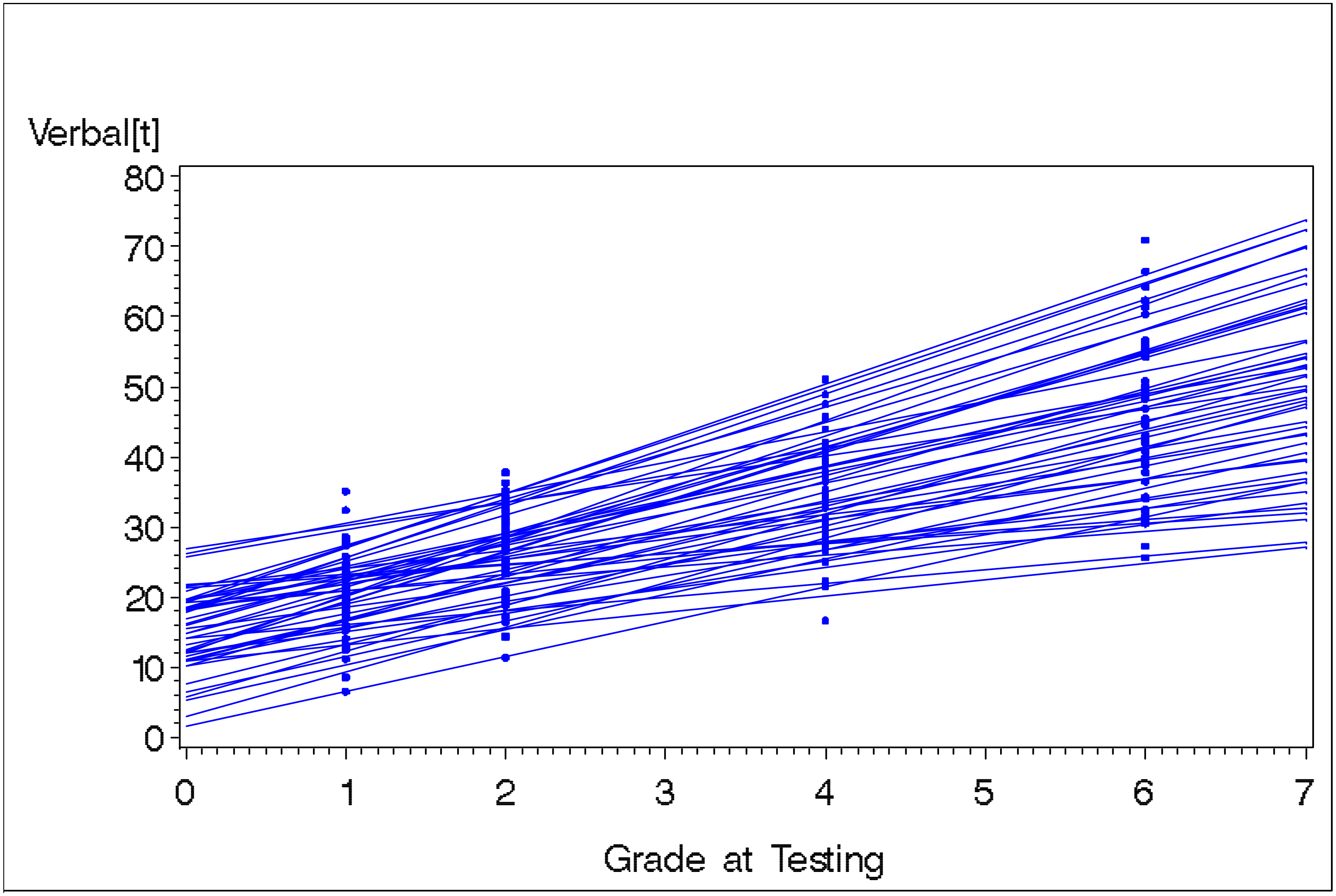
Model Estimated Covariances/Correlations/Residual Correlations

	V1	V2	V4	V6
V1	32.682			
V2	22.947	40.397		
V4	29.134	36.814	65.000	
V6	35.321	46.058	67.532	101.833

Residuals for Covariances/Correlations/Residual Correlations

	V1	V2	V4	V6
V1	0.883			
V2	2.393	-3.291		
V4	1.604	-3.159	-11.685	
V6	4.998	1.115	-5.584	11.350

Linear regression estimates of individual growth curves (N=50)



M+ Random Slopes programming

```
TITLE:      Wisc_RS2.m+ Equal Residual Linear Growth ;

DATA:       FILE= c:\ATI_2003\Data\Wisc4vpe.dat;

VARIABLE:   NAMES = V1 V2 V4 V6 P1 P2 P4 P6 Moeducat;
            USEVAR=V1 V2 V4 V6;

ANALYSIS:   TYPE=MEANSTRUCTURE;

MODEL:       i s | V1@0 V2@1 V4@3 V6@5;
             V1-V6 (Ve); ! Adding equal error variance

OUTPUT:          SAMPSTAT;
```

Chi-Square Test of Model Fit

Value	79.185
Degrees of Freedom	8
P-Value	0.0000

MODEL RESULTS

	Estimates	S.E.	Est./S.E.
I |			
V1	1.000	0.000	0.000
V2	1.000	0.000	0.000
V4	1.000	0.000	0.000
V6	1.000	0.000	0.000
S |			
V1	0.000	0.000	0.000
V2	1.000	0.000	0.000
V4	3.000	0.000	0.000
V6	5.000	0.000	0.000
S WITH			
I	3.093	0.590	5.244
Means			
I	19.824	0.367	54.030
S	4.673	0.108	43.101
Variances			
I	19.854	2.771	7.165
S	1.529	0.245	6.236
Residual Variances			
V1	12.828	0.898	14.283
V2	12.828	0.898	14.283
V4	12.828	0.898	14.283
V6	12.828	0.898	14.283

A Mixed-Effects Latent Curve model

- We start with a "first level" model of random effects

$$\boldsymbol{Y[\mathrm{t}]_n = i_n + A[\mathrm{t}]\ s_n + e[\mathrm{t}]_n}$$

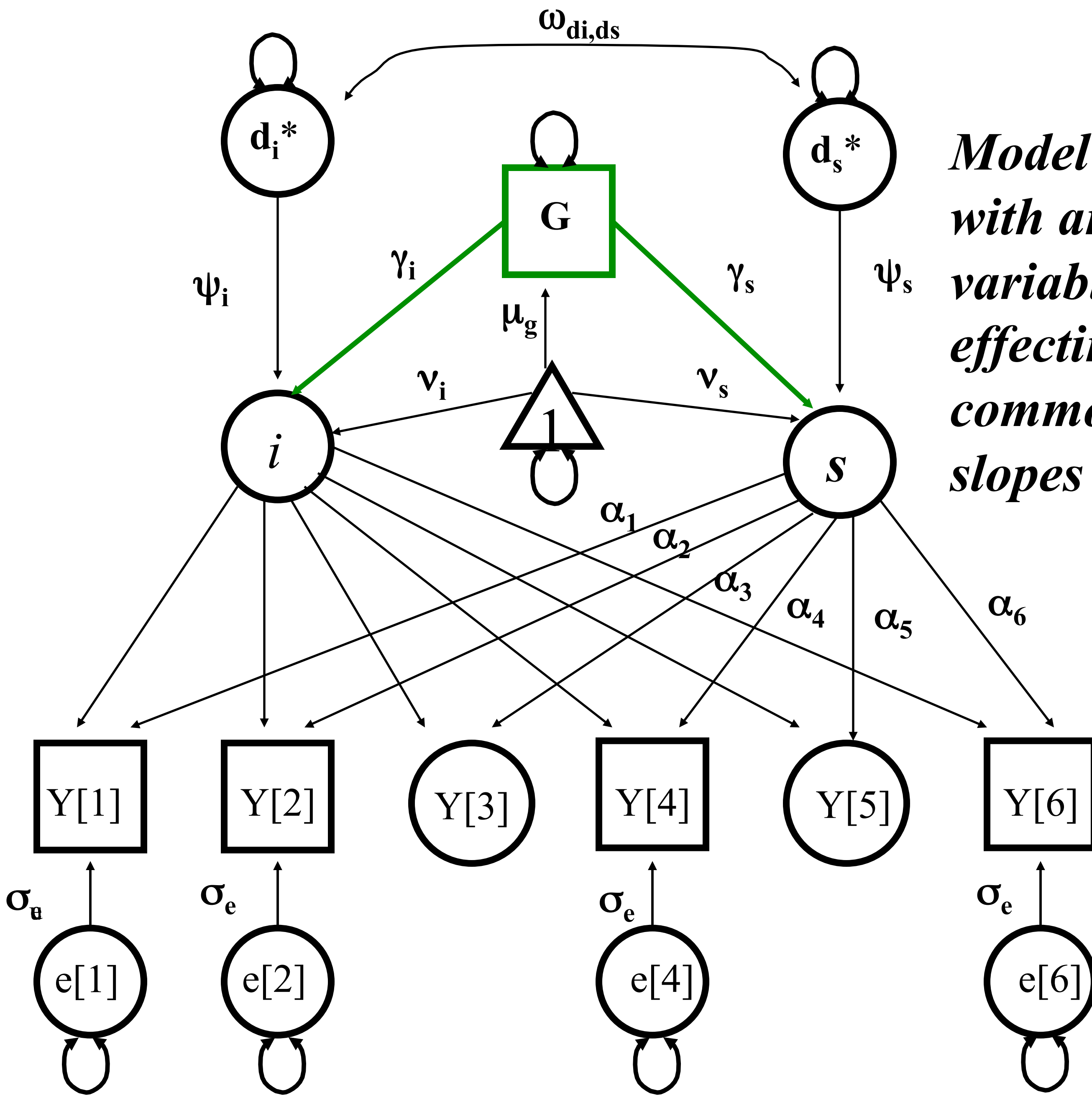
- We simultaneously fit a "second level" model of effects

$$\boldsymbol{i_n = \nu_i + \delta_i\ G_n + d_{in}}$$

$$\boldsymbol{s_n = \nu_s + \delta_s\ G_n + d_{sn}}$$

where independent groups in a variable $\boldsymbol{G}$ (dummy coded, effect coded, multiple codes, etc.) and the parameters include two 2nd level regression intercepts ($\boldsymbol{\nu_i}$, $\boldsymbol{\nu_s}$) and two 2nd level slopes ($\boldsymbol{\delta_i}$, $\boldsymbol{\delta_s}$) for the two latent variables ($\boldsymbol{g_i}$, $\boldsymbol{g_s}$).

- Now the terms "random coefficient" and "multi-level" model are useful descriptors of the overall model.
- Interpretation depends on scaling (centering) and coding.



Model [2A]: LGM with an external G variable effecting both the common levels and slopes

M+ Mixed-Effects LCM for WISC

```
TITLE:    Wisc-LCM3.m+ Latent Curve Model with Predictor;
DATA:     FILE= Wisc4vpe.dat;
VARIABLE:  NAMES = V1 V2 V4 V6 P1 P2 P4 P6 Moeducat;
           USEVAR=V1 V2 V4 V6 Moeducat;
ANALYSIS:  TYPE=MEANSTRUCTURE;
MODEL:      i BY V1-V6 @1;
            s BY V1   @   0
                  V2   @   1
                  V4   @   3
                  V6   @   5;
            V1-V6 * 10 (1);
           [V1-V6 @  0];
           [i*50 s*5];
            i s ON Moeducat;
  * Moeducat treated as a continuous [0,1,2];
OUTPUT:    SAMPSTAT STANDARDIZED RESIDUALS;
```

M+ results for Latent Curve Groups

Means

	V1	V2	V4	V6	MOEDUCAT
1	19.585	25.415	32.607	43.750	0.853

Covariances

	V1	V2	V4	V6	MOEDUCAT
V1	33.730				
V2	25.465	37.289			
V4	30.890	33.821	53.578		
V6	40.518	47.406	62.253	113.741	
MOEDUCAT	2.028	2.141	2.668	3.878	0.579

Correlations

	V1	V2	V4	V6	MOEDUCAT
V1	1.000				
V2	0.718	1.000			
V4	0.727	0.757	1.000		
V6	0.654	0.728	0.797	1.000	
MOEDUCAT	0.459	0.461	0.479	0.478	1.000

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	79.330
Degrees of Freedom	10
P-Value	0.0000

Chi-Square Test of Model Fit for the Baseline Model

Value	652.020
Degrees of Freedom	10
P-Value	0.0000

M+ results for Latent Curve

	Estimates	S.E.	Est./S.E.	Std	StdYX
I BY					
V1-V4	1.000	0.000	0.000	4.456	0.779
S BY					
V1	0.000	0.000	0.000	0.000	0.000
V2	1.000	0.000	0.000	1.236	0.195
V4	3.000	0.000	0.000	3.709	0.460
V6	5.000	0.000	0.000	6.182	0.613
I ON					
MOEDUCAT	3.195	0.428	7.460	0.717	0.544
S ON					
MOEDUCAT	0.635	0.136	4.681	0.514	0.390
S WITH					
I	1.923	0.498	3.863	0.349	0.349
Intercepts					
V1-V4	0.000	0.000	0.000	0.000	0.000
I	17.099	0.489	34.956	3.838	3.838
S	4.132	0.155	26.659	3.341	3.341
Residual Variances					
V1	12.827	0.898	14.283	12.827	0.393
V2	12.827	0.898	14.283	12.827	0.318
V4	12.827	0.898	14.283	12.827	0.197
V6	12.827	0.898	14.283	12.827	0.126
I	13.967	2.202	6.343	0.704	0.704
S	1.296	0.223	5.815	0.848	0.848

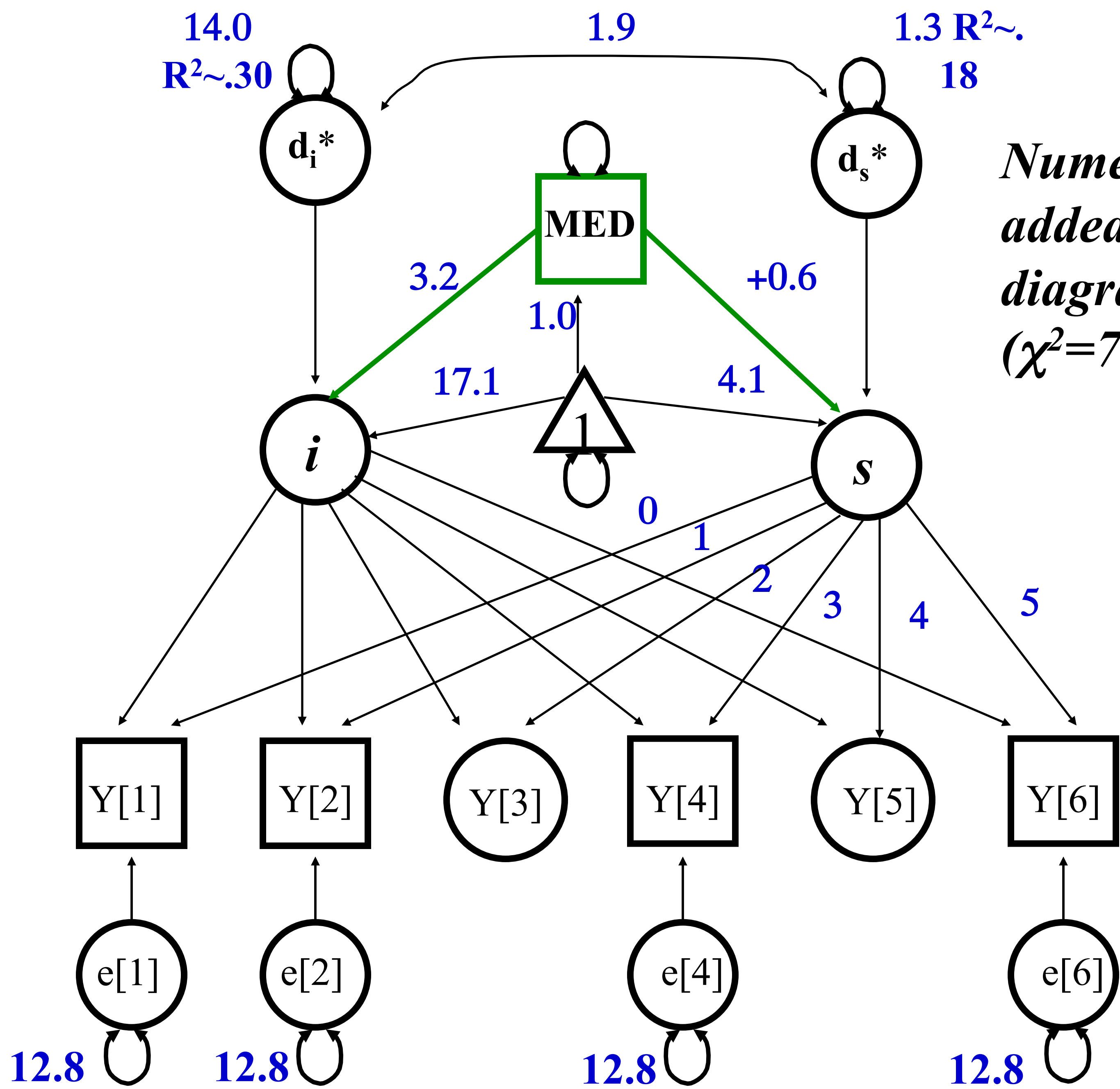
R-SQUARE

Observed

Variable	R-Square
V1	0.607
V2	0.682
V4	0.803
V6	0.874

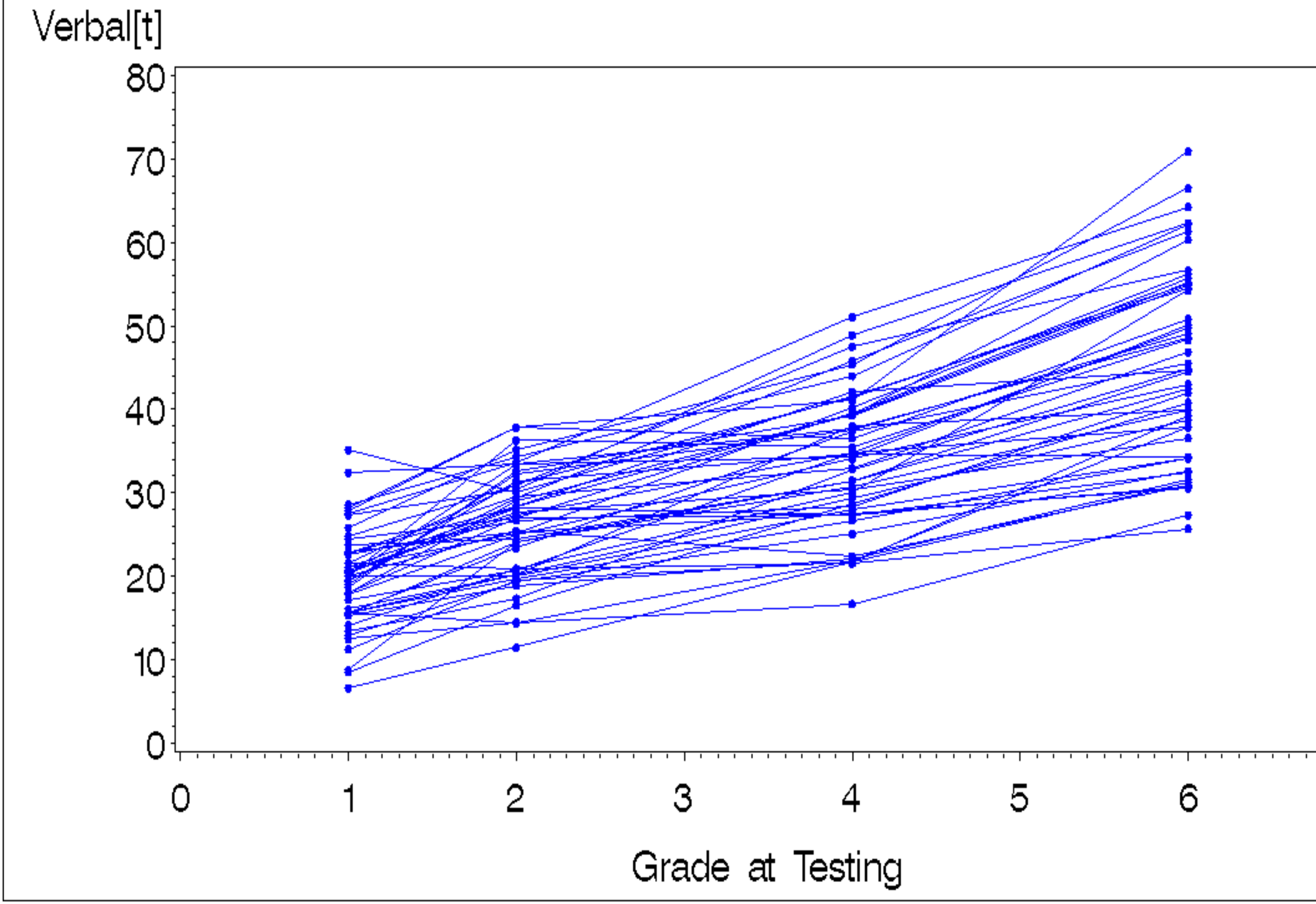
Latent

Variable	R-Square
I	0.296
S	0.152



Numerical results added to path diagram (χ^2=79 on df=10)

```
TITLE1 'Repeated Measures Analyses - WISC data from RTO Children';
TITLE2 'Growth Curve Analyses by Jack McArdle -- 2003-05-24 and 2010-08-03';

TITLE3 'Reading in raw data';
DATA wisc4vpe;
        INFILE 'Z:\Bank\WAIS\WISC\WISC_2003\wisc4vpe.dat';
    INPUT Verb_1 Verb_2 Verb_4 Verb_6 Perf_1 Perf_2 Perf_4 Perf_6 MomEduc;

TITLE2 'Creating Relational Input File for Standard ANOVAs';
DATA temp;
    SET wisc4vpe;
        FILE outfile LRECL=200 LINESIZE=200;
        grade1=1; grade2=2; grade4=4; grade6=6;
        PUT
        #1 _N_ MomEduc grade1 Verb_1 Perf_1
        #2 _N_ MomEduc grade2 Verb_2 Perf_2
        #3 _N_ MomEduc grade4 Verb_4 Perf_4
        #4 _N_ MomEduc grade6 Verb_6 Perf_6;
DATA wisc4vec;
        INFILE outfile;
        INPUT Subject MomEduc Grade Verbal Perform;
TITLE2 'Longitudinal WISC-Verbal Data';
SYMBOL1 I=JOIN COLOR=BLUE VALUE=DOT HEIGHT=1.25 WIDTH=1.25 LINE=1 REPEAT=5000;
DATA=wisc4vec (WHERE=(subject<50));
        TITLE1 ' '; PLOT Verbal*Grade=Subject / FRAME NOLEGEND;
        LABEL Verbal='Verbal[t]' Grade='Grade at Testing';

TITLE1 ' Mixed Effects Linear models -- No Centering, all data ';
PROC MIXED DATA=wisc4vec  NOCLPRINT METHOD=ML COVTEST IC; CLASS Subject;
        MODEL Verbal =  Grade MomEduc Grade*MomEduc / SOLUTION DDFM=BW CHISQ;
        RANDOM INTERCEPT Grade / SUBJECT=subject TYPE=UNR; RUN;
```



Verbal[t]
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
Grade at Testing

Covariance Parameters	4
Columns in X	4
Columns in Z Per Subject	2
Subjects	204
Max Obs Per Subject	4

Number of Observations Read	816
Number of Observations Used	816
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Log Like	Criterion
0	1	5453.82693148	
1	3	4972.97138292	0.00006935
2	1	4972.84762701	0.00000020
3	1	4972.84727024	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Var(1)	Subject	11.4178	2.5122	4.54	<.0001
Var(2)	Subject	1.2962	0.2229	5.81	<.0001
Corr(2,1)	Subject	0.1631	0.1625	1.00	0.3156
Residual		12.8276	0.8981	14.28	<.0001

The Mixed Procedure

Fit Statistics

-2 Log Likelihood	4972.8
AIC (smaller is better)	4988.8
AICC (smaller is better)	4989.0
BIC (smaller is better)	5015.4

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	480.98	<.0001

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
4972.8	8	4988.8	4989.0	4999.6	5015.4	5023.4

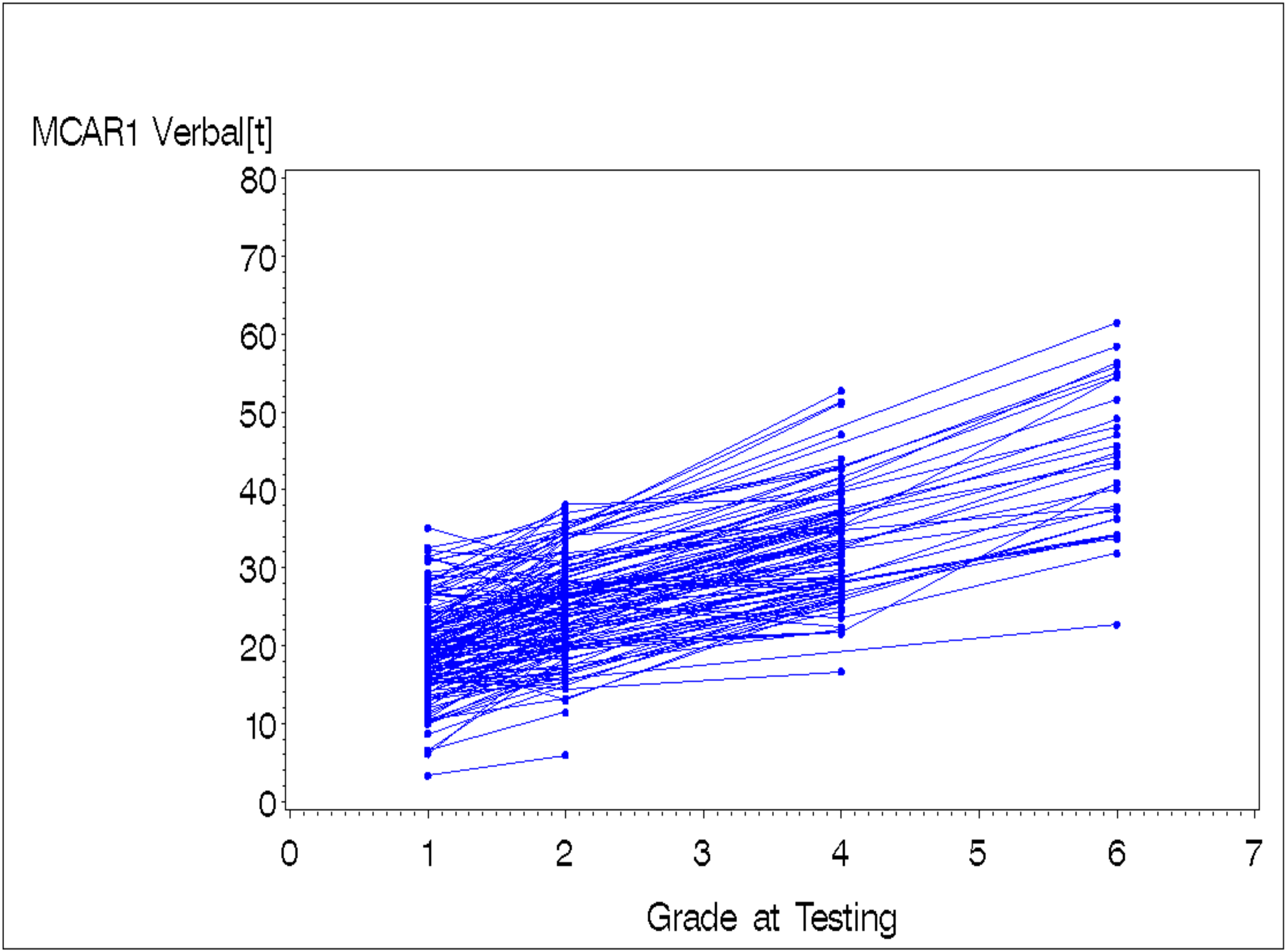
Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.9673	0.5138	202	25.24	<.0001
Grade	4.1316	0.1550	610	26.66	<.0001
MomEduc	2.5600	0.4500	202	5.69	<.0001
Grade*MomEduc	0.6352	0.1357	610	4.68	<.0001

5. Masking Incomplete Longitudinal WISC Data

Masking some of the Available WISC Data

```
TITLE2 'Scaling the Basis in different ways';
DATA wisc4vec;
        INFILE outfile;
        INPUT Subject MomEduc Grade Verbal Perform;
        Grade1 = Grade - 1;
        Grade4 = Grade - 4;
        GradeU = (Grade -1)/5;

TITLE2 'Creating random selections of dropout conditions ';
        Seed=20100803;
        Rand6 = INT(RANUNI(Seed)*7); /* random selection for dropout*/

        MCAR1 =Verbal; IF (Rand6 LT Grade) THEN MCAR1 = .;
        MCAR2 =Verbal; IF (Rand6 LT Grade+1) THEN MCAR2 = .;

        MAR1  =Verbal; IF (Rand6 LT Grade + MomEduc) THEN MAR1=.;
        MAR2  =Verbal; IF (Rand6 LT Grade+1+MomEduc) THEN MAR2=.;

        NMAR1 =Verbal; IF (Verbal GT 44) THEN NMAR1=.;
        NMAR2 =Verbal; IF (Verbal GT 33) THEN NMAR2=.;

        RUN;
```


MCAR1 Verbal[t]
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
Grade at Testing

Results from MCAR1 Selection

Number of Observations Read	816
Number of Observations Used	427
Number of Observations Not Used	389

Convergence criteria met.

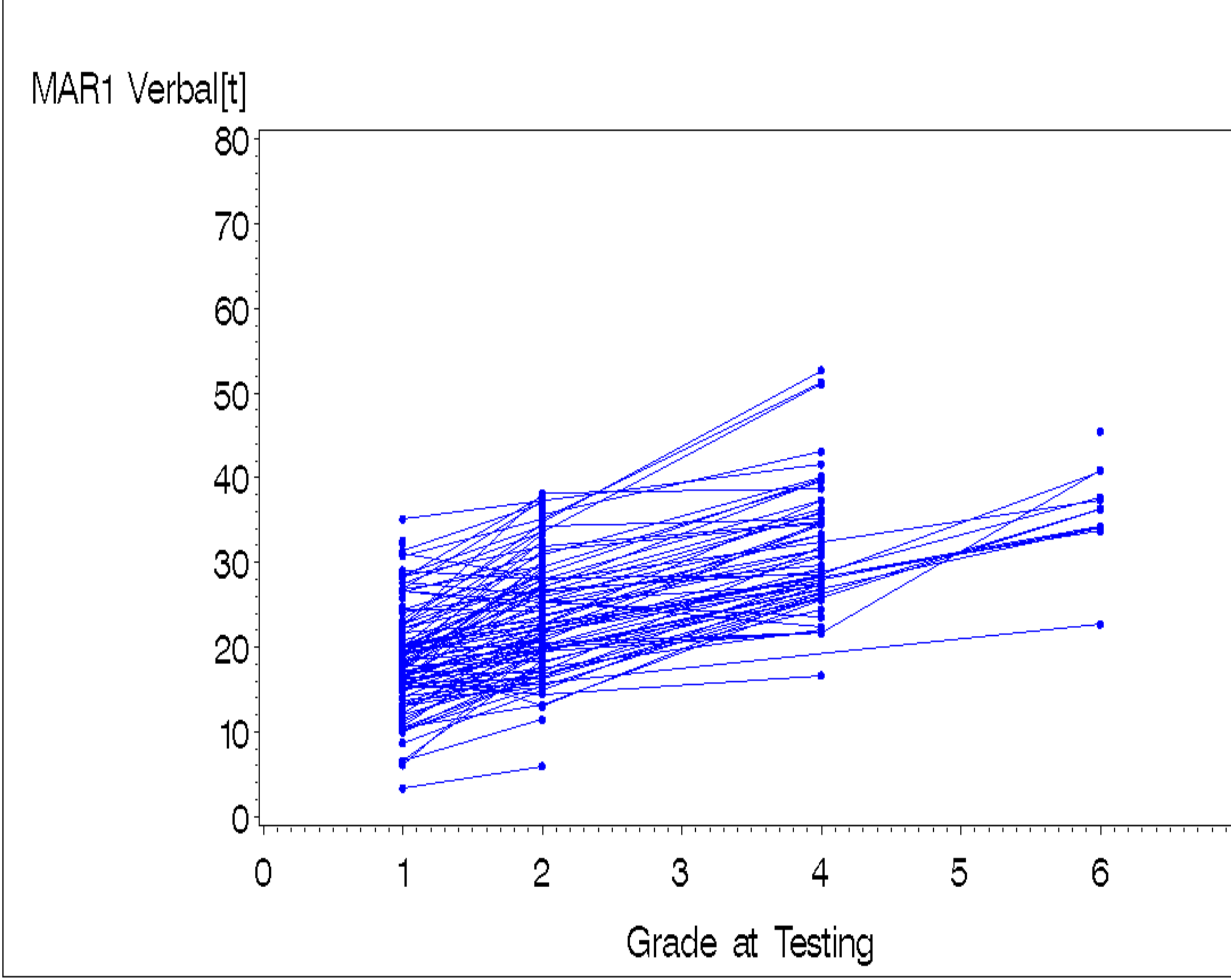
Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Var(1)	Subject	11.6689	3.4372	3.39	0.0003
Var(2)	Subject	0.5058	0.3266	1.55	0.0607
Corr(2,1)	Subject	0.6139	0.5295	1.16	0.2463
Residual		11.5611	1.3431	8.61	<.0001

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
2574.0	8	2590.0	2590.4	2600.7	2616.6	2624.6

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	13.0308	0.6171	199	21.11	<.0001
Grade	4.0740	0.2223	224	18.33	<.0001
MomEduc	2.8575	0.5413	199	5.28	<.0001
Grade*MomEduc	0.4616	0.1889	224	2.44	0.0153


MAR1 Verbal[t]
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
Grade at Testing

Results from MAR1 Selection

Number of Observations Read	816
Number of Observations Used	334
Number of Observations Not Used	482

45 1 2025.92124861 0.00000000 BUT Convergence criteria met.

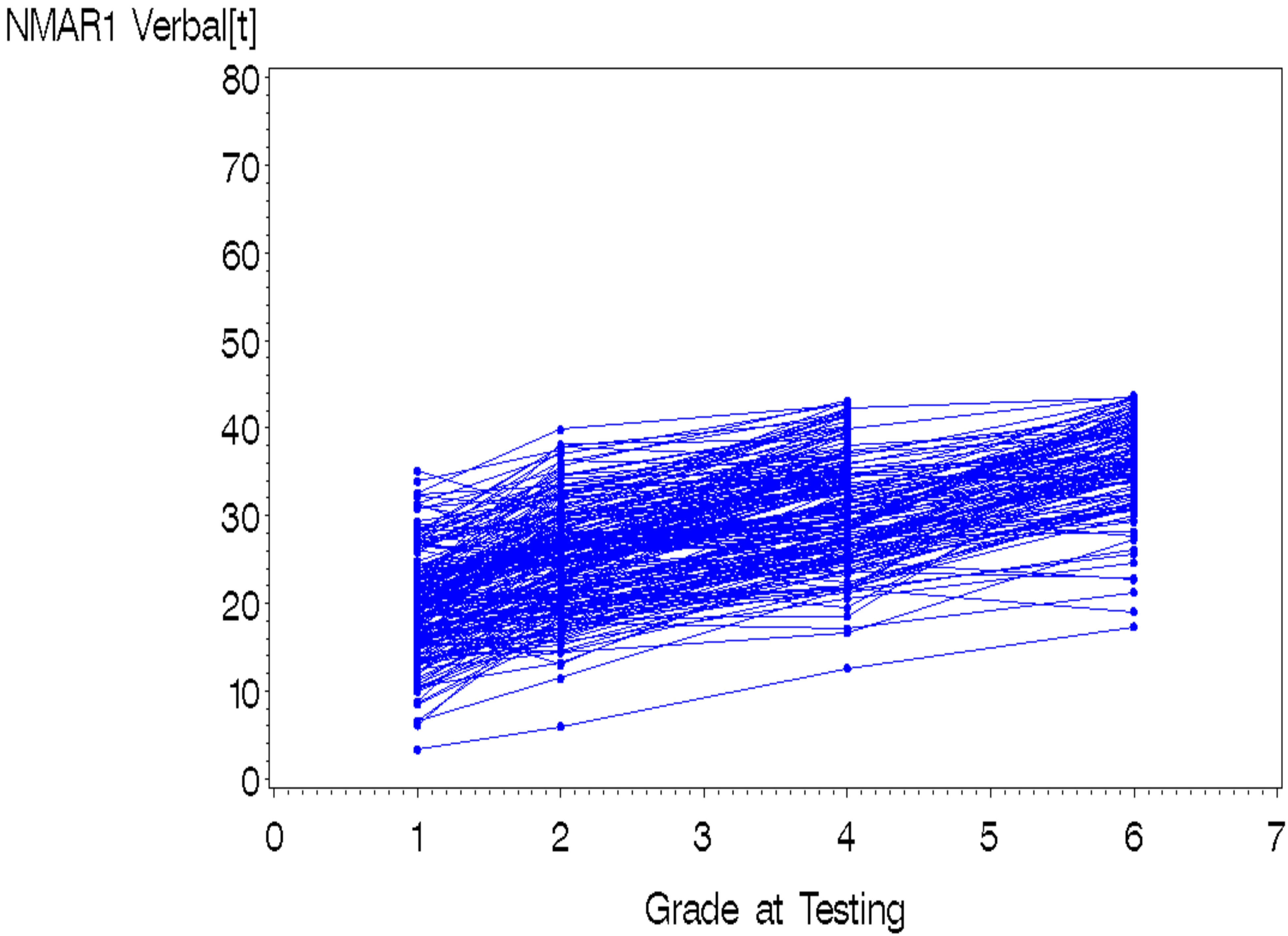
Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Var(1)	Subject	17.4276	4.5775	3.81	<.0001
Var(2)	Subject	0.6265	0.4866	1.29	0.0990
Corr(2,1)	Subject	0.06891	0.3878	0.18	0.8590
Residual		10.8350	1.5556	6.97	<.0001

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
2025.9	8	2041.9	2042.4	2052.7	2068.5	2076.5

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	13.0693	0.6797	182	19.23	<.0001
Grade	3.9882	0.2373	148	16.81	<.0001
MomEduc	2.9352	0.6968	182	4.21	<.0001
Grade*MomEduc	0.5977	0.2911	148	2.05	0.0418


NMAR1 Verbal[t]
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
Grade at Testing

Results from NMAR1 Selection

Number of Observations Read	816
Number of Observations Used	709
Number of Observations Not Used	107

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Var(1)	Subject	16.7713	2.9898	5.61	<.0001
Var(2)	Subject	0.3013	0.1499	2.01	0.0222
Corr(2,1)	Subject	-0.00574	0.2441	-0.02	0.9812
Residual		10.8227	0.8367	12.93	<.0001

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
4141.8	8	4157.8	4158.0	4168.5	4184.4	4192.4

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	13.6243	0.5557	202	24.52	<.0001
Grade	3.7558	0.1161	503	32.34	<.0001
MomEduc	3.4520	0.4970	202	6.95	<.0001
Grade*MomEduc	0.1387	0.1157	503	1.20	0.2309

References

- Ghisletta, P. & McArdle, J.J. (2001). Structural modeling of children's height. *Structural equation modeling*, 10 (1), 19-39.

- McArdle, J.J. & Hamagami, E. (1992). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research, 18* (3), 145-166.

- McArdle, J.J. & Bell, R.Q. (2000). Recent trends in modeling longitudinal data by latent growth curve methods. In Little, T.D., Schnabel, K.U. & Baumert, J. (Eds.). *Modeling longitudinal and multiple-group data: practical issues, applied approaches, and scientific examples*. Mahwah, NJ: Erlbaum (69-108).

- Singer, J.D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24* (4), 323-355.

- Verbeke, G. & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.

Recent References on Latent Growth

- Ferrer, E., Hamagami, F., & McArdle, J.J. (2003). Modeling latent growth curves with incomplete data using different types of structural equation modeling and multilevel software. *Structural Equation Modeling*.
- McArdle, J.J. & Aber, M.S. (1990). Patterns of change within latent variable structural equation modeling. In A. von Eye (Eds.). *New Statistical Methods in Developmental Research,* New York: Academic Press, 151-224.
- McArdle, J.J. & Epstein, D.B. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58* (1), 110-133.
- McArdle, J.J. & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. Marcoulides & R. Schumacker (Eds.), *Advanced Structural Equation Modeling Techniques* (89-124). Hillsdale, N.J.: Erlbaum.
- McArdle, J.J. & Nesselroade, J.R. (2003). Growth curve analyses in contemporary psychological research. In J. Schinka & W. Velicer (Editors), *Comprehensive Handbook of Psychology, Volume Two* (pp. 447-480). New York: Pergamon Press.