Growth curve modeling tutorial: What you need to know

Raúl Rojas
University of Texas at Dallas
Callier Center for Communication Disorders

Aquiles Iglesias
Temple University

November 19, 2011
San Diego, CA

Outline

- Development
- Traditional statistical methods
- Growth curve modeling (GCM)
- GCM advantages and disadvantages
- GCM requirements
- GCM in SPSS
- Step-by-step GCM

Development

- Development implies...***change***
- Change implies…***time***
- Goal: capture developmental changes over time
  - Describe
  - Predict

Patterns of change over time = Growth

Traditional statistical methods

- Analysis of variance (AN[C]OVA; MAN[C]OVA, etc.)
- Regression (multiple linear; hierarchical, etc.)
- Detection of significant differences between groups
  - Focus = **Inter**-individual variation
- Planned or Post-Hoc additional contrasts

- Univariate repeated measures
  - Individual differences in intercept *or* slope
  - Underestimates variability of distinct slopes
  - Participants with missing data excluded
  - Cannot use time-varying predictors
- Multivariate repeated measures
  - Participants with missing data excluded
  - Difficult to use time-varying predictors
  - Ordinary least squares estimation method

Intra-individual variation

- Do participants begin in the same place/level?
  - **Intercept** (initial status) of each participant
- Do participants grow at the same rate and follow the same trajectory?
  - **Slope** of each participant
- How much does each participant's *initial status* vary with the prototypical growth for the group?
- How can we measure this without additional testing?

- Consider both **intra**- *and* **inter**-individual variation in longitudinal analyses
- Multiple contextual/environmental differences
- How can we account for this?
- Multiple, nested levels of analyses
  - Level 1: Intra-individual variation
  - Level 2: Inter-individual variation
  - Level 3: Effects of context/environment

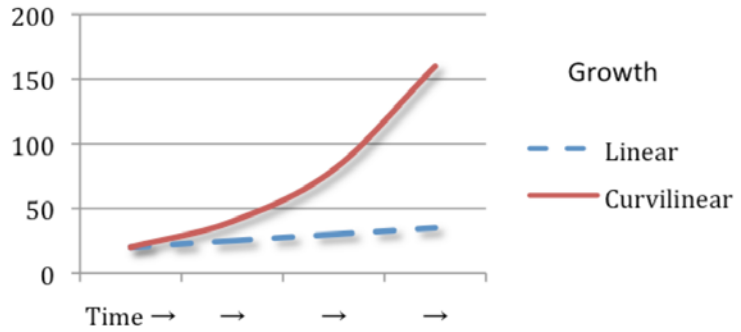
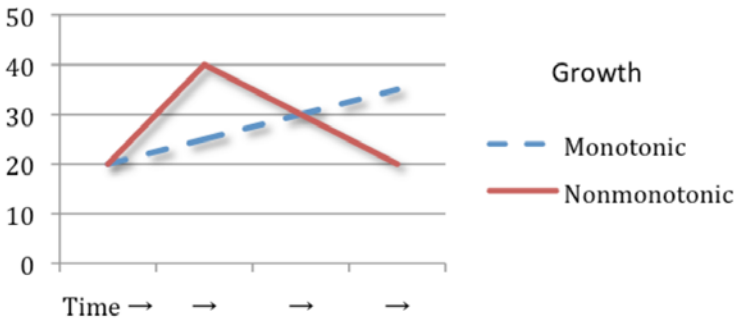
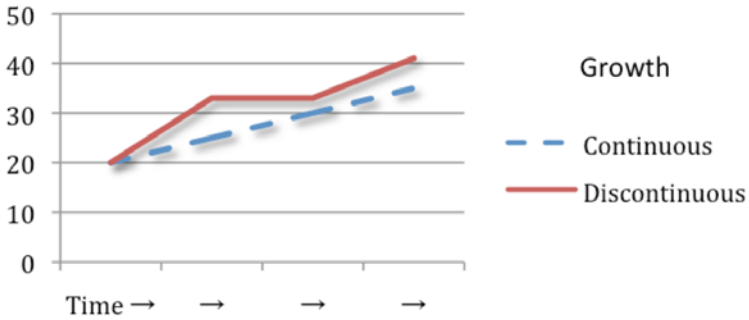
Growth curve modeling (GCM)

- Hierarchical linear modeling (HLM); Multilevel modeling (MLM)
- GCM a subset of HLM, is specifically designed for longitudinal analyses
- Although commonplace in other fields (e.g., medicine), GCM is relatively novel in communication sciences and disorders (CSD)

- GCM analyzes:
  - Starting point of growth (intercept; initial status)
  - Shape (functional form) of growth over time
  - Rate of growth (slope) over time
- Captures patterns of change (growth) at the intra-individual level
- Identifies growth predictors at the inter-individual level

- GCM permits us to ask different and novel research questions (Singer & Willett, 2003)
- For instance:

- [
- [
- I
- [
- I
- I

Dimension	Range	Example
Linearity	Linear Vs. Curvilinear	
Direction	Monotonic Vs. Non-monotonic	
Continuity	Continuous Vs. Discontinuous	

- Is growth monotonic or nonmonotonic?

(Iglesias & Rojas, 2012; Rojas, 2011)

GCM Advantages

- Measure growth throughout (not just at the end)
- Model shape of growth (linear, curvilinear, etc.)
- Calculate rate of growth (steady, acceleration, deceleration)
- Increased statistical power, relative to traditional approaches

- State-of-the-art approaches to handle missing data
  - Use all participants, even those with just one wave
  - Maximum likelihood estimation (full; restricted)
  - Multiple imputation
- Can use time-structured and unstructured data
  - Time-structured: evenly spaced observations (e.g., every 2 weeks)
  - Time-unstructured: non-evenly spaced observations (e.g., 1 week → 4 days → 2 months)

- Accepts multiple conditions, covariates
  - Time-invariant *and* time-varying covariates
- Able to conduct covariate *x* time interactions
  - Does a predictor's effect vary over time?
- Variance partitioned at the intra- and inter-individual levels
  - Onset of growth (intercept)
  - Rates of growth (slopes)
  - Intercept-slope covariance

GCM Disadvantages

- Depending on complexity of the growth curve model(s) (e.g., covariates; levels), some GCMs can be computationally intensive
- Lack of familiarity from journal reviewers
  - Following slide is from Singer & Willett (2011)

Part of the problem may be reviewers' ignorance

Comments received this year from two reviewers of a paper that fit individual growth models to 3 waves of data on vocabulary size among young children*:*

Reviewer A:

"I do not understand the statistics used in this study deeply enough to evaluate their appropriateness. I imagine this is also true of 99% of the readers of *Developmental Psychology*. … Previous studies in this area have used simple correlation or regression which provide easily interpretable values for the relationships among variables. … In all, while the authors are to be applauded for a detailed longitudinal study, … the statistics are difficult. … I thus think *Developmental Psychology* is not really the place for this paper."

Reviewer B:

"The analyses fail to live up to the promise…of the clear and cogent introduction. I will note as a caveat that I entered the field before the advent of sophisticated growth-modeling techniques, and they have always aroused my suspicion to some extent. I have tried to keep up and to maintain an open mind, but parts of my review may be naïve, if not inaccurate."

Source: http://www.ats.ucla.edu/stat/seminars/alda/default.htm

GCM Requirements

- Ok, you've convinced me, but what do I need?

1. Longitudinal (not cross-sectional) data
2. 3 waves of data or more
3. Outcome that is measured the same way at each time point, and that systematically changes over time (e.g., NDW; IRT)
4. A metric for tracking time (e.g., semesters)

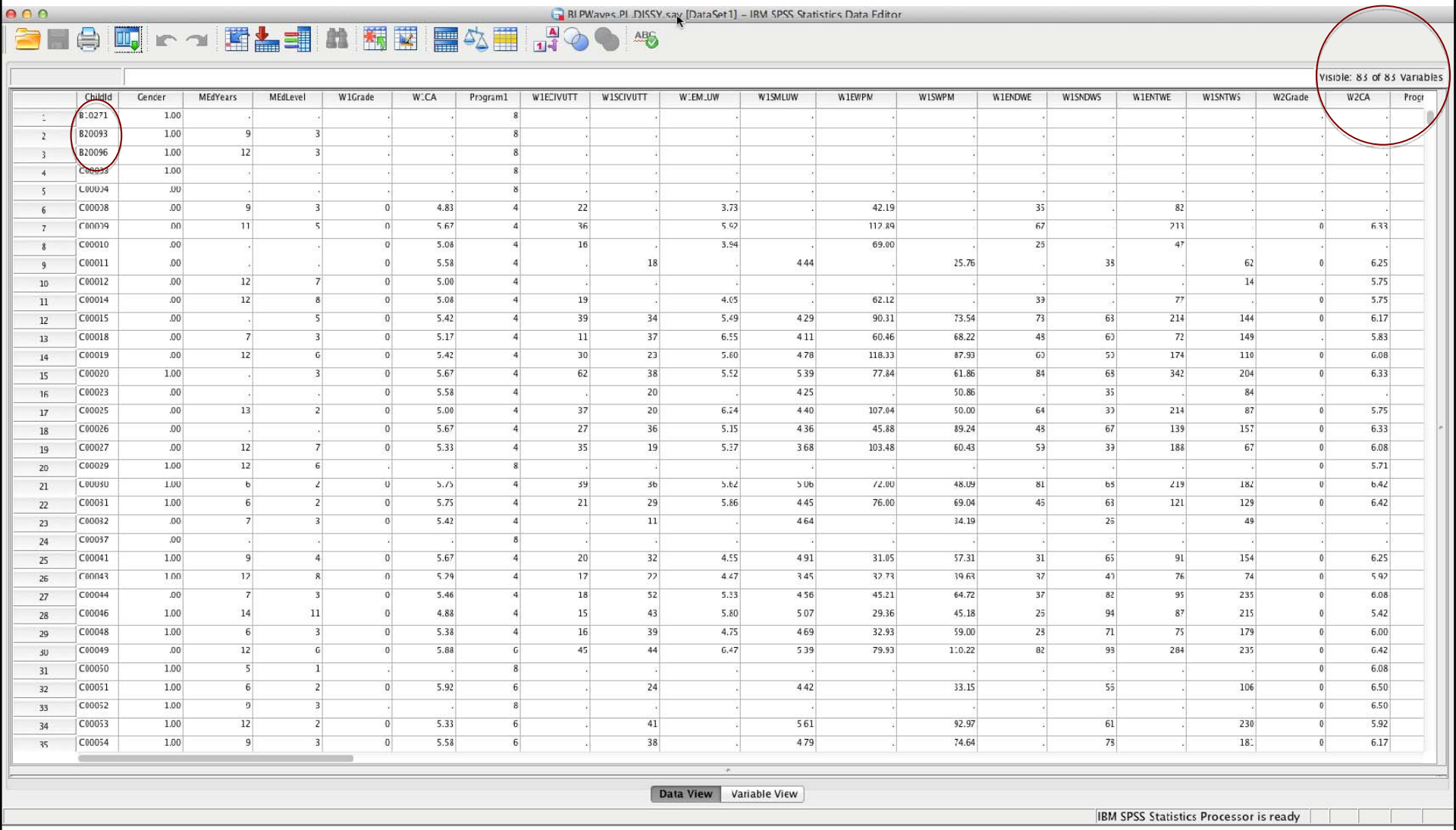
Why *three waves* of data or more?

- Having 2 or less, limits longitudinal analyses
  - Cannot identify shape of individual growth trajectories
  - Cannot distinguish true growth from measurement error
- Linear growth = 3 waves minimum
- Quadratic growth = 4 waves minimum
- Cubic growth = 5 waves minimum

GCM in SPSS

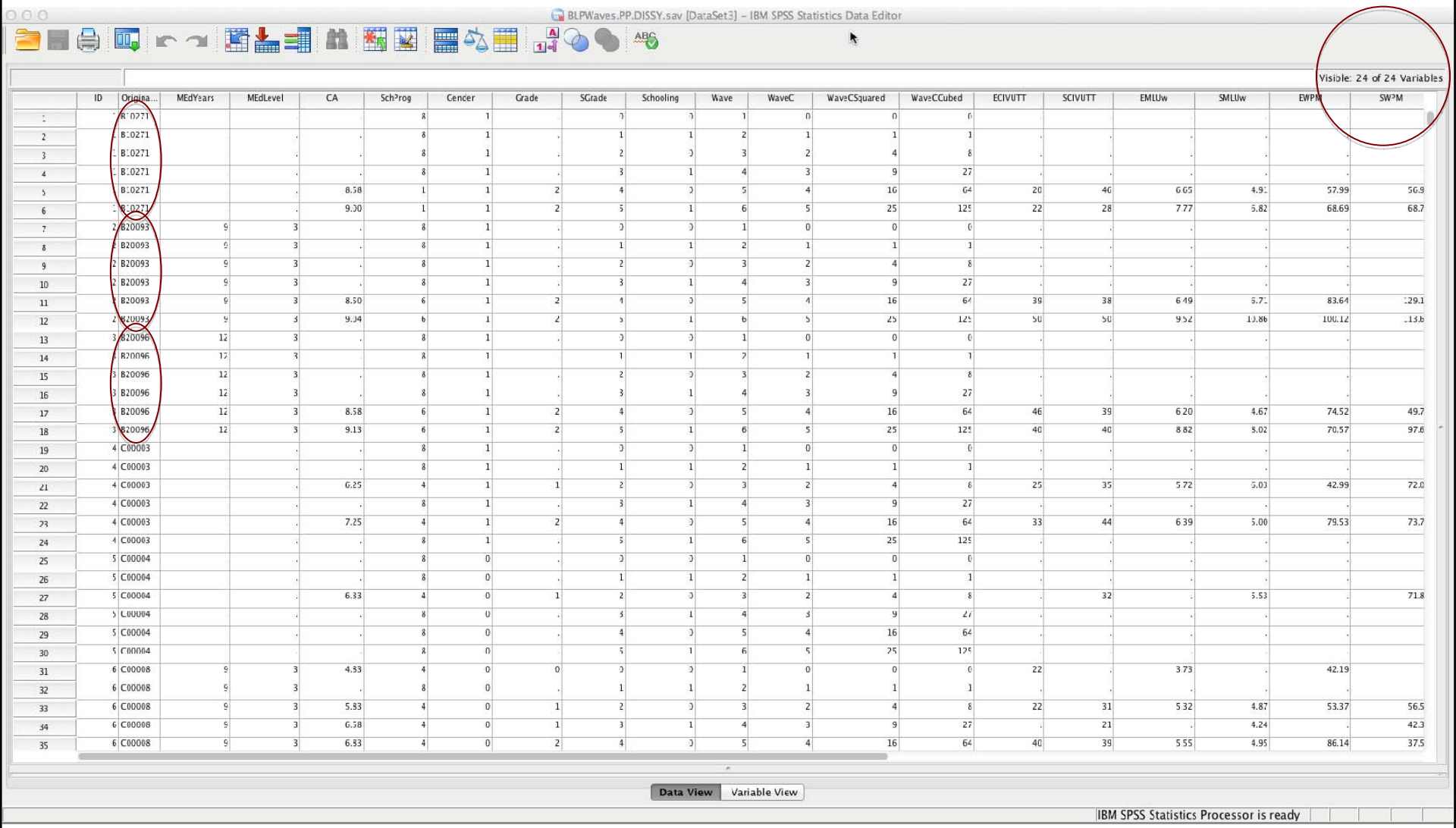
- A range of statistical packages are GCM-capable such as HLM, SAS, Stata, R, M*plus*, SPSS…
- Many packages require a person-*period* data set (rather than a person-*level* data set)
  - One row of data per wave of measurement
  - More rows, but less columns

Your traditional data set: person-level



	ChildId	Gencer	MEdYears	MEdLevel	W1Grade	W1CA	Program1	W1ECIVUTT	W1SCIVUTT	W1EMLUW	W1SMLUW	W1EWPM	W1SWPM	W1ENDWE	W1SNDWS	W1ENTWE	W1SNTWS	W2Grade	W2CA	Progr
1	B20271	1.00	.	.	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	
2	B20093	1.00	9	3	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	
3	B20096	1.00	12	3	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	
4	C00033	1.00	.	.	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	
5	C00034	.00	.	.	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	
6	C00038	.00	9	3	0	4.83	4	22	.	3.73	.	42.19	.	35	.	82	.	.	.	
7	C00039	.00	11	5	0	5.67	4	36		5.92		112.89		67		213		0	6.33	
8	C00010	.00	.	.	0	5.08	4	16	.	3.94	.	69.00	.	25	.	47	.	.	.	
9	C00011	.00	.	.	0	5.58	4	.	18	.	4.44	.	25.76	.	38	.	62	0	6.25	
10	C00012	.00	12	7	0	5.00	4	.	.	.	.	.	.	.	.	.	14	.	5.75	
11	C00014	.00	12	8	0	5.08	4	19	.	4.05	.	62.12	.	33	.	77	.	0	5.75	
12	C00015	.00	.	5	0	5.42	4	39	34	5.49	4.29	90.31	73.54	73	63	214	144	0	6.17	
13	C00018	.00	7	3	0	5.17	4	11	37	6.55	4.11	60.46	68.22	48	60	72	149	.	5.83	
14	C00019	.00	12	6	0	5.42	4	30	23	5.60	4.78	118.33	87.93	60	53	174	110	0	6.08	
15	C00020	1.00	.	3	0	5.67	4	62	38	5.52	5.39	77.84	61.86	84	68	342	204	0	6.33	
16	C00023	.00	.	.	0	5.58	4	.	20	.	4.25	.	50.86	.	35	.	84	.	.	
17	C00025	.00	13	2	0	5.00	4	37	20	6.24	4.40	107.04	50.00	64	33	214	87	0	5.75	
18	C00026	.00	.	.	0	5.67	4	27	36	5.15	4.36	45.88	89.24	48	67	139	157	0	6.33	
19	C00027	.00	12	7	0	5.33	4	35	19	5.37	3.68	103.48	60.43	53	33	188	67	0	6.08	
20	C00029	1.00	12	6	.	.	8	.	.	.	.	.	.	.	.	.	.	0	5.71	
21	C00030	1.00	6	2	0	5.75	4	39	36	5.62	5.06	72.00	48.09	81	65	219	182	0	6.42	
22	C00031	1.00	6	2	0	5.75	4	21	29	5.86	4.45	76.00	69.04	45	63	121	129	0	6.42	
23	C00032	.00	7	3	0	5.42	4	.	11	.	4.64	.	34.19	.	25	.	49	.	.	
24	C00037	.00	.	.	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	
25	C00041	1.00	9	4	0	5.67	4	20	32	4.55	4.91	31.05	57.31	31	65	91	154	0	6.25	
26	C00043	1.00	12	8	0	5.29	4	17	22	4.47	3.45	32.73	39.63	37	47	76	74	0	5.92	
27	C00044	.00	7	3	0	5.46	4	18	52	5.33	4.56	45.21	64.72	37	82	95	235	0	6.08	
28	C00046	1.00	14	11	0	4.88	4	15	43	5.80	5.07	29.36	45.18	25	94	87	215	0	5.42	
29	C00048	1.00	6	3	0	5.38	4	16	39	4.75	4.69	32.93	59.00	28	71	75	179	0	6.00	
30	C00049	.00	12	6	0	5.88	6	45	44	6.47	5.39	79.93	110.22	82	93	284	235	0	6.42	
31	C00050	1.00	5	1	.	.	8	.	.	.	.	.	.	.	.	.	.	0	6.08	
32	C00051	1.00	6	2	0	5.92	6	.	24	.	4.42	.	33.15	.	55	.	106	0	6.50	
33	C00052	1.00	9	3	.	.	8	.	.	.	.	.	.	.	.	.	.	0	6.50	
34	C00053	1.00	12	2	0	5.33	6	.	41	.	5.61	.	92.97	.	61	.	230	0	5.92	
35	C00054	1.00	9	3	0	5.58	6	.	38	.	4.79	.	74.64	.	73	.	182	0	6.17	

Your dataset on GCM: person-period



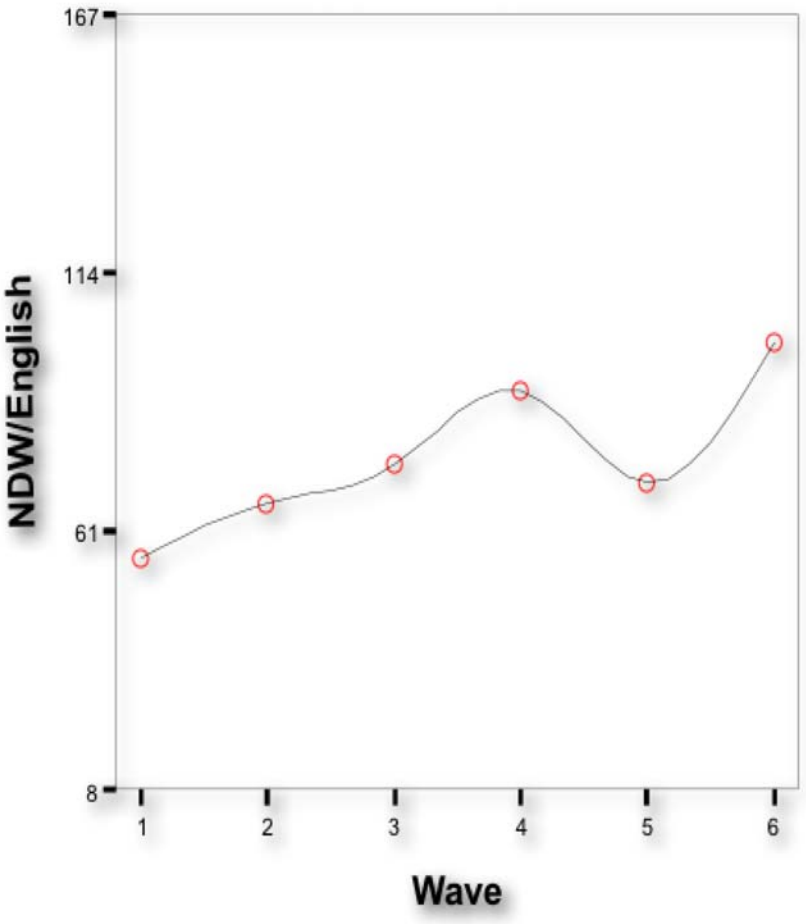
	ID	Origina...	MEdYears	MEdLevel	CA	SchProg	Cencer	Grade	SGrade	Schooling	Wave	WaveC	WaveCSquared	WaveCCubed	ECIVUTT	SCIVUTT	EMLUw	SMLUw	EWPM	SWPM
1	1	B10271				8	1		0	0	1	0	0	0						
2	1	B10271				8	1		1	1	2	1	1	1						
3	1	B10271				8	1		2	0	3	2	4	8						
4	1	B10271				8	1		3	1	4	3	9	27						
5	1	B10271			8.58	1	1	2	4	0	5	4	16	64	20	40	6.65	4.91	57.99	56.9
6	1	B10271			9.00	1	1	2	5	1	6	5	25	125	22	28	7.77	5.82	68.69	68.7
7	2	B20093	9	3		8	1		0	0	1	0	0	0						
8	2	B20093	9	3		8	1		1	1	2	1	1	1						
9	2	B20093	9	3		8	1		2	0	3	2	4	8						
10	2	B20093	9	3		8	1		3	1	4	3	9	27						
11	2	B20093	9	3	8.50	6	1	2	4	0	5	4	16	64	39	38	6.49	5.71	83.64	129.1
12	2	B20093	9	3	9.04	6	1	2	5	1	6	5	25	125	50	50	9.52	10.86	100.12	113.6
13	3	B20096	12	3		8	1		0	0	1	0	0	0						
14	3	B20096	12	3		8	1		1	1	2	1	1	1						
15	3	B20096	12	3		8	1		2	0	3	2	4	8						
16	3	B20096	12	3		8	1		3	1	4	3	9	27						
17	3	B20096	12	3	8.58	6	1	2	4	0	5	4	16	64	46	39	6.20	4.67	74.52	49.7
18	3	B20096	12	3	9.13	6	1	2	5	1	6	5	25	125	40	40	8.82	8.02	70.57	97.6
19	4	C00003				8	1		0	0	1	0	0	0						
20	4	C00003				8	1		1	1	2	1	1	1						
21	4	C00003			6.25	4	1	1	2	0	3	2	4	8	25	35	5.72	5.03	42.99	72.0
22	4	C00003				8	1		3	1	4	3	9	27						
23	4	C00003			7.25	4	1	2	4	0	5	4	16	64	33	44	6.39	5.00	79.53	73.7
24	4	C00003				8	1		5	1	6	5	25	125						
25	5	C00004				8	0		0	0	1	0	0	0						
26	5	C00004				8	0		1	1	2	1	1	1						
27	5	C00004			6.33	4	0	1	2	0	3	2	4	8		32		5.53		71.8
28	5	C00004				8	0		3	1	4	3	9	27						
29	5	C00004				8	0		4	0	5	4	16	64						
30	5	C00004				8	0		5	1	6	5	25	125						
31	6	C00008	9	3	4.83	4	0	0	0	0	1	0	0	0	22		3.73		42.19	
32	6	C00008	9	3		8	0		1	1	2	1	1	1						
33	6	C00008	9	3	5.83	4	0	1	2	0	3	2	4	8	22	31	5.32	4.87	53.37	56.5
34	6	C00008	9	3	6.58	4	0	1	3	1	4	3	9	27		21		4.24		42.3
35	6	C00008	9	3	6.83	4	0	2	4	0	5	4	16	64	40	39	5.55	4.95	86.14	37.5

Step-by-step GCM

- Exploratory analyses
- Growth curve model testing
  - Pseudo-R^2
  - Goodness-of-fit indices
  - ^{2} distribution
- Final growth curve model
  - Fixed effects and variance components
  - Prototypical growth curve trajectory

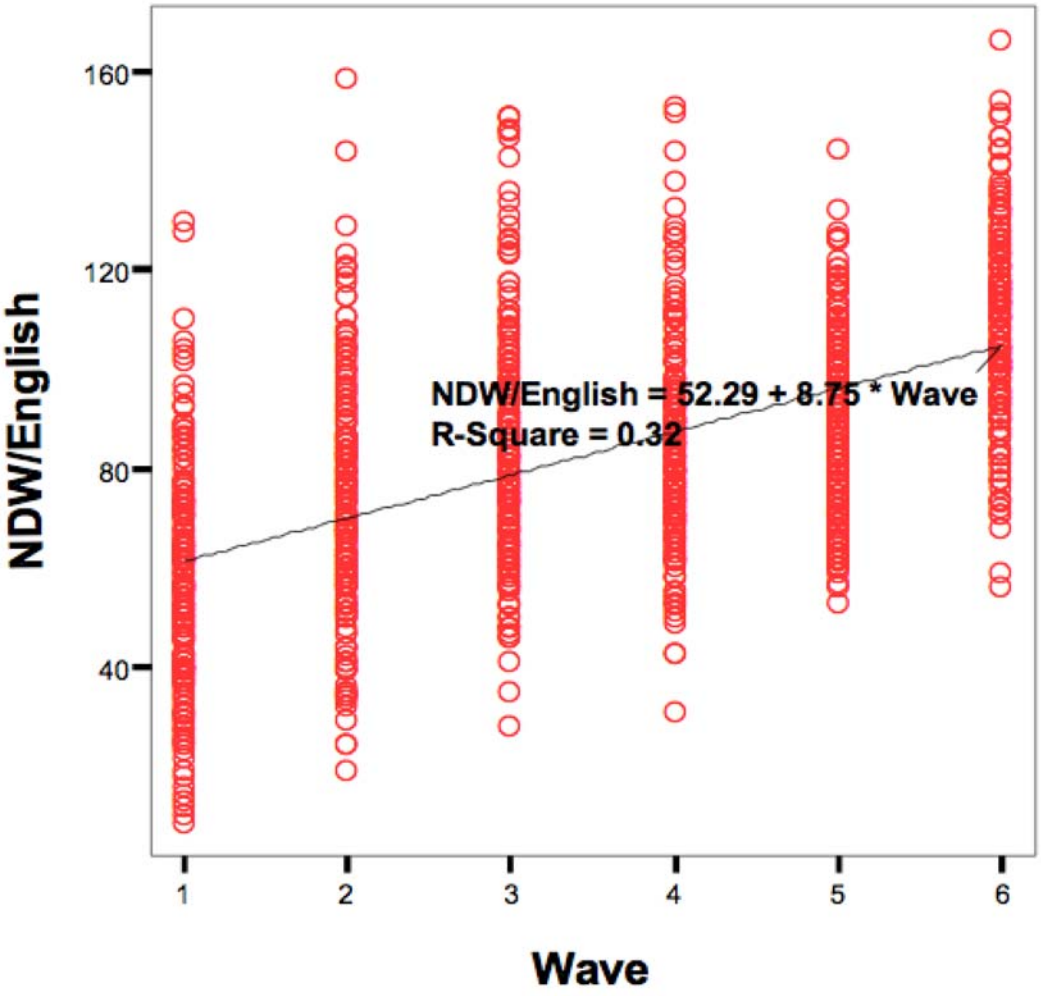
Exploratory Analyses

- Discerning patterns in raw data to inform modeling procedure
- Empirical growth plots demonstrate growth from individual participants (intra-individual variation)

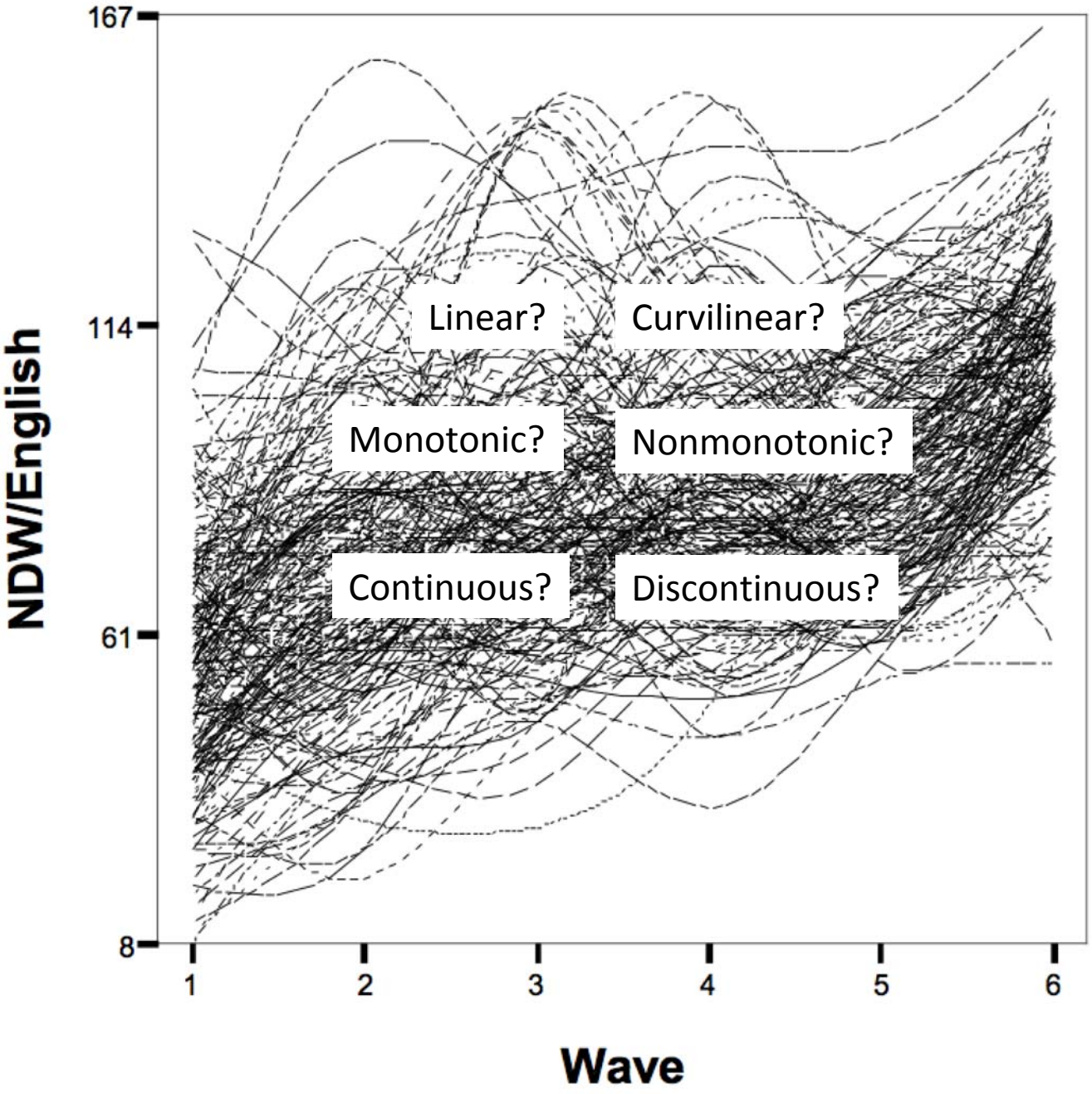


(Rojas, 2011)

- Fitted regression to gauge degree of inter-individual variation of participants in your sample



(Rojas, 2011)

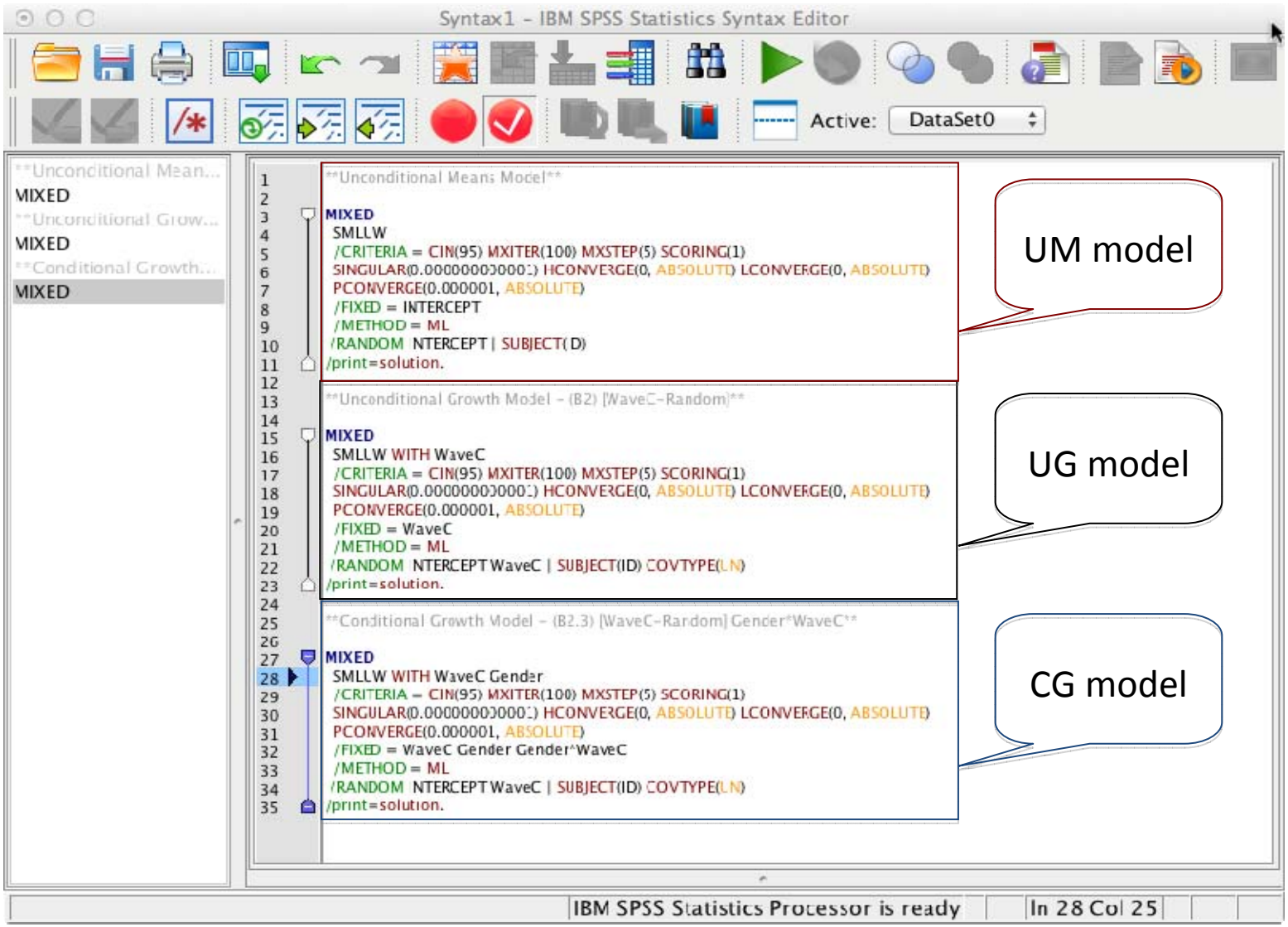

167
114
61
8
NDW/English
1
2
3
4
5
6
Wave
Linear?
Curvilinear?
Monotonic?
Nonmonotonic?
Continuous?
Discontinuous?

(Rojas, 2011)

Growth curve model testing

1. Unconditional means (UM) model
   - Intercept-only model (no time)
   - Baseline for unconditional growth model
2. Unconditional growth (UG) model(s)
   - Growth curve model with effect of time* (no additional covariates)
   - Slopes can be set as randomly varying* or fixed
   - Baseline for conditional growth model
3. Conditional growth (CG) model(s)
   - Add time-invariant and/or time-varying covariates

Sample SPSS-syntax for GCM



- Pseudo-R^2
  - Proportional variance reduction (intercept; slope)
  - Higher is better
- Goodness-of-fit indices
  - Negative 2 log likelihood (-2LL) deviance statistic*
  - Akaike's Information Criterion (AIC)
  - Schwarz's Bayesian Information Criterion (BIC)*
  - Lower is better
- ^{2} distribution
  - Confirms the -2LL differences between models

Final growth curve model

- Highest overall proportional variance reduction (Pseudo-R^2s)
- Best *fitting* model
  - Lowest -2LL across nested models
  - Lowest BIC across non-nested models

	Model A UM	Model B UG-Ln	Model C UG-Qd	Model D UG-Cb	Model E CG-Cb+G	Model F CG-Dc	Model G CG-Dc+G
Proportional variance reduction							
L1: Within-person variance R_ε^2		21%	22%	31%	**31%**	22%	22%
L2: B/w-person intercept R_0^2					**1%**		0%
L2: B/w-person linear slope R_1^2					**2%**		2%
L2: B/w-person quadratic slope R_2^2					**2%**		
L2: B/w-person cubic slope R_3^2					**2%**		
Goodness-of-fit							
-2LL	57888.5	56936.6*	56871.8*	56595.5*	**56577.6***	56902.2	56881.5*
BIC	57914.9	56989.3	56933.3	56727.2	**56744.5**	56963.7	56969.3

~$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

[1]*Note.* **Bold** indicates final model; Shading indicates non-applicable Pseudo-R^2

(Rojas, 2011)

- Fixed effects: directly interpretable*
  - Intercept [γ_{00}]
  - Slope(s) [$\gamma_{10,}$ $\gamma_{20,\ldots}$]
  - Covariate(s) [$\gamma_{01,}$ $\gamma_{02,\ldots}$]
  - Covariate *x* time interactions [$\gamma_{11,}$ $\gamma_{21,\ldots}$]

		Model A UM	Model B UG-Ln	Model C UG-Qd	Model D UG-Cb	Model E CG-Cb+G	Model F CG-Dc	Model G CG-Dc+G
Fixed effects								
Intercept	γ_{00}	68.86***	55.01***	57.59***	54.63***	**53.54*** **	54.25***	52.70***
Linear Slope	γ_{10}		4.24***	0.80~	13.67***	**11.31*** **	8.04***	8.95***
Quadratic Slope	γ_{20}			0.68***	-6.31***	**-4.96*** **	6.63*** Schooling	5.79*** Schooling
Cubic Slope	γ_{30}				-0.93***	**0.76*** **		
Gender (G)	γ_{01}					**2.13~**		3.01***
G x Linear Slope	γ_{11}					**4.62***		-1.75**
G x Quadratic Slope	γ_{21}					**-2.64** **		1.63* G x Schooling
G x Cubic Slope	γ_{31}					**0.34** **		

(Rojas, 2011)

- Variance components: individual differences
  - Intra-individual differences across each wave [σ_{ε}^2]
  - Inter-individual differences at initial status (intercept) [σ_0^2]
  - Inter-individual differences in rate of growth (slope) [σ_1^2,

	Model A UM	Model B UG-Ln	Model C UG-Qd	Model D UG-Cb	Model E CG-Cb+G	Model F CG-Dc	Model G CG-Dc+G
Variance components							
L1: Within-person variance σ_ε^2	328.24*	257.72*	255.09*	226.33*	**226.15***	256.28*	255.86*
L2: B/w-person intercept σ_0^2	152.78*	187.76*	181.75*	168.75*	**167.85***	186.50*	183.39*
L2: B/w-person linear slope σ_1^2		4.02*	3.94*	152.55*	**150.22***	16.60*	16.12*
L2: B/w-person quadratic slope σ_2^2				39.26*	**38.51***		
L2: B/w-person cubic slope σ_3^2				0.65*	**0.64***		

(Rojas, 2011)

- Intercept-slope covariance: impact of initial status on growth (strength and direction) [σ_{01}, σ_{02},…]
  - *Positive and significant*: ↑ initial status, on average, leads to ↑ growth
  - *Negative and significant*: ↑ initial status, on average, leads to ↓ growth
  - *Nonsignificant*: Lack of systematic relationship bw initial status and growth

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Covariance (σ_0^2, σ_1^2)	σ_{01}		-7.49*	-6.40	22.07	**19.40**	-13.57*	-12.27
Covariance (σ_0^2, σ_2^2)	σ_{02}				-18.10*	**-16.64**		
Covariance (σ_0^2, σ_3^2)	σ_{03}				2.41	**2.23**		
Covariance (σ_1^2, σ_2^2)	σ_{12}				-74.54*	**-73.22***		
Covariance (σ_1^2, σ_3^2)	σ_{13}				9.38*	**9.22***		
Covariance (σ_2^2, σ_3^2)	σ_{23}				-5.03*	**-4.94**		

(Rojas, 2011)

Table 18. Comparisons of Growth Curve Model Parameter Estimates for Words per Minute-Spanish[1,2]

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Fixed effects								
Intercept	γ_{00}	68.86***	55.01***	57.59***	54.63***	**53.54*** **	54.25***	52.70***
Linear Slope	γ_{10}		4.24***	0.80~	13.67***	**11.31*** **	8.04***	8.95***
Quadratic Slope	γ_{20}			0.68***	-6.31***	**-4.96*** **	6.63*** Schooling	5.79*** Schooling
Cubic Slope	γ_{30}				-0.93***	**0.76*** **		
Gender (G)	γ_{01}					**2.13~**		3.01***
G x Linear Slope	γ_{11}					**4.62***		-1.75**
G x Quadratic Slope	γ_{21}					**-2.64****		1.63* G x Schooling
G x Cubic Slope	γ_{31}					**0.34****		
Variance components								
L1: Within-person variance	σ_ε^2	328.24*	257.72*	255.09*	226.33*	**226.15***	256.28*	255.86*
L2: B/w-person intercept	σ_0^2	152.78*	187.76*	181.75*	168.75*	**167.85***	186.50*	183.39*
L2: B/w-person linear slope	σ_1^2		4.02*	3.94*	152.55*	**150.22***	16.60*	16.12*
L2: B/w-person quadratic slope	σ_2^2				39.26*	**38.51***		

(Rojas, 2011)

Table 18. (continued)

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
L2: B/w-person cubic slope	σ_3^2				0.65*	**0.64***		
Covariance (σ_0^2, σ_1^2)	σ_{01}		-7.49*	-6.40	22.07	**19.40**	-13.57*	-12.27
Covariance (σ_0^2, σ_2^2)	σ_{02}				-18.10*	**-16.64**		
Covariance (σ_0^2, σ_3^2)	σ_{03}				2.41	**2.23**		
Covariance (σ_1^2, σ_2^2)	σ_{12}				-74.54*	**-73.22***		
Covariance (σ_1^2, σ_3^2)	σ_{13}				9.38*	**9.22***		
Covariance (σ_2^2, σ_3^2)	σ_{23}				-5.03*	**-4.94**		
Proportional variance reduction								
L1: Within-person variance	R_ε^2		21%	22%	31%	**31%**	22%	22%
L2: B/w-person intercept	R_0^2					**1%**		0%
L2: B/w-person linear slope	R_1^2					**2%**		2%
L2: B/w-person quadratic slope	R_2^2					**2%**		
L2: B/w-person cubic slope	R_3^2					**2%**		

(Rojas, 2011)

Table 18. (continued)

	Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Goodness-of-fit							
-2LL	57888.5	56936.6*	56871.8*	56595.5*	**56577.6***	56902.2	56881.5*
BIC	57914.9	56989.3	56933.3	56727.2	**56744.5**	56963.7	56969.3

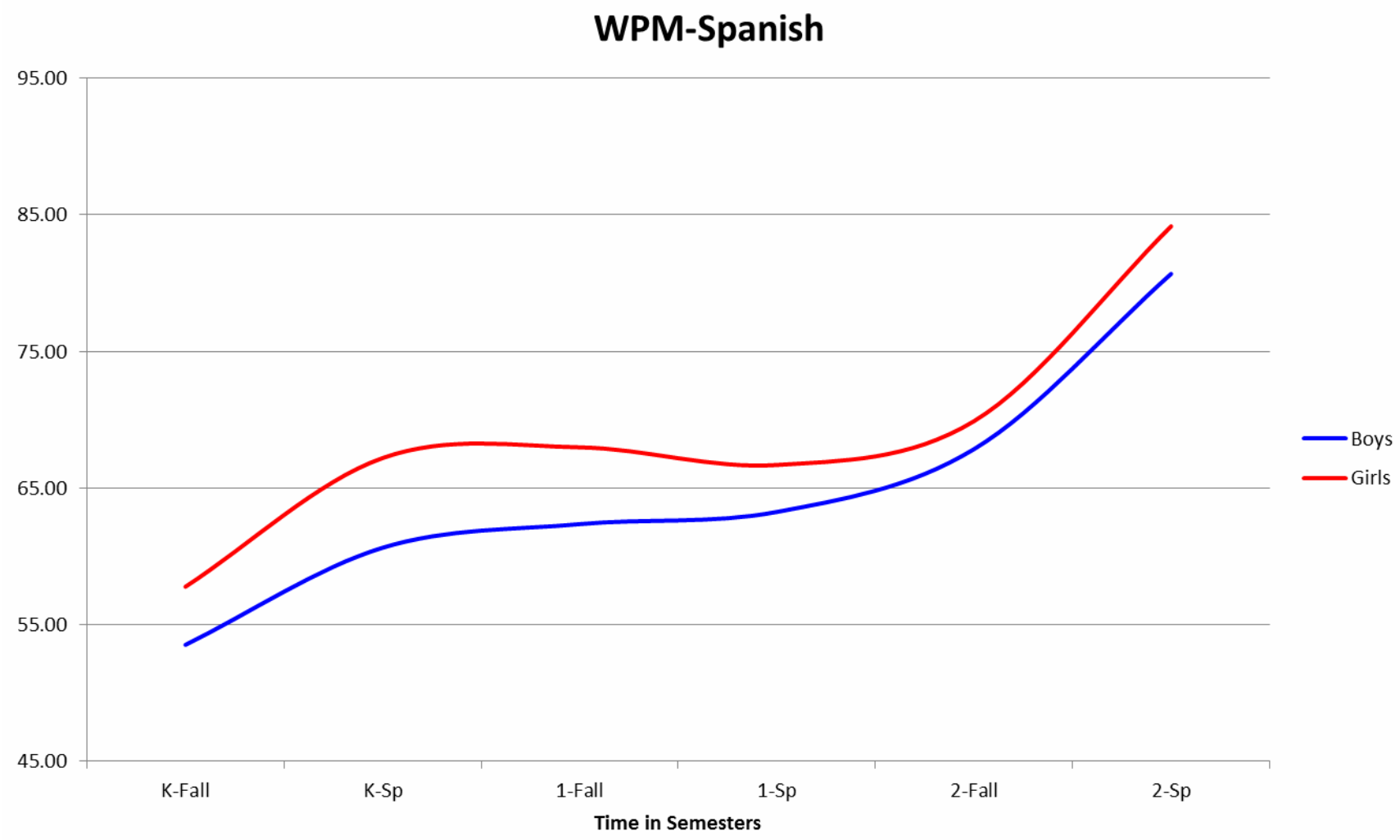
~$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

[1]*Note.* **Bold** indicates final model; Shading indicates non-applicable Pseudo-R^2

[2]*Key.* UM: Unconditional means model; UG-Ln: Unconditional linear growth model; UG-Qd: Unconditional quadratic growth model; UG-Cb: Unconditional cubic growth model; CG-Cb+G: Conditional cubic growth model with gender; CG-Dc: Conditional discontinuous growth model; CG-Dc+G: Conditional discontinuous growth model with gender; L1: Level-1 submodel; L2: Level-2 submodel

(Rojas, 2011)

Prototypical growth curve trajectory



What to look for in GCM-based work

- ✓ Longitudinal data with ≥3 waves of data collection
- ✓ Outcome measured the same way at each time point, and that systematically changes over time
- ✓ Defined metric of time
- ✓ Model testing procedure described*
- ✓ Table(s) with GCM data (at minimum: fixed effects + variance components + fit indices)*
- ✓ Prototypical growth curve trajectory

More information

- GCM “bible” from Singer and Willett (2003)
  - http://gseacademic.harvard.edu/alda/
- M*plus* website for structural equation modeling
  - http://www.statmodel.com/

THANK YOU!

raul.rojas@utdallas.edu