Types of Growth Models

Joseph Stevens
Department of Educational Methodology, Policy, and Leadership
University of Oregon

Presented at the NCAASE Coordinating Session at SREE, Washington, DC, March, 2012

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

Contact Information:

Joseph Stevens, Ph.D.
College of Education
5267 University of Oregon
Eugene, OR 97403
(541) 346-2445
stevensj@uoregon.edu

Presentation available at: http://www.uoregon.edu/~stevensj/sree2012.pdf
And on NCAASE web site soon: http://www.ncaase.com/

This work was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R32C110004 awarded to the University of Oregon. The opinions expressed are those of the author and do not necessarily represent views of the Institute or the U.S. Department of Education.

Presentation Purpose

- Briefly describe some types of accountability models and my own definitions of some key terms and characteristics of growth models
- Some information about goals of NCAASE work:
  - Explore how different models work to describe growth at student and school levels (our work does not include classroom/teacher level)
  - Explicitly extend work on growth models to special education students
  - Study impact of common challenges like attrition and mobility
  - Study effect of design choices on results (e.g., including covariates or not; including process variables like OTL or not)
  - Study growth within and between formative, interim, and summative assessments
- NCAASE goals do not include "evaluating" existing systems or attempting to decide "best" model

http://www.uoregon.edu/~stevensj/sree2012.pdf

3

Brief Background

- In the pre-NCLB era, larger variety of accountability models in use, usually implemented at the district level
- NCLB standardized accountability systems and required state level systems and school level reporting
  - Swept away existing systems and assessments
  - NCLB depended wholly on single year status models with exception of the two year "safe harbor" calculation
- In 2005, flexibility allowed for Pilot Growth Models and now further expansion under Race To The Top
- More variety in models now in use across the country
- Limited research or evidence for reliability and validity of many models now in use

http://www.uoregon.edu/~stevensj/sree2012.pdf

4

How to Measure Educational Effectiveness?

- Estimating educational impact is a complex task; a problem in research or program evaluation design (see Stevens, 2005)
- Challenge of separating "intake" from "impact" of the school (Raudenbush & Willms, 1995)
- Intake includes confounding effects of pre-existing differences, out-of-school learning, family, poverty, community, etc.
- For valid inference, need to disentangle multiple influences on student learning; correctly identify effect of teacher, program, school
- Growth models have promise of providing greater control over individual differences between students; students serve as their own controls; some control over preexisting conditions

http://www.uoregon.edu/~stevensj/sree2012.pdf

NCAASE Issues to Study

Study issues important in implementing growth models:

- Number and timing of assessments
- Effect of score scales on estimates
- Methods for reporting growth
- Precision and accuracy of models and how affected by model variations
- Effect of covariates (e.g., OTL, prior achievement)
- Attrition and Missing Data
  - Cohort Stability
  - Missing data, drop-outs, mobility and transfer
- Functional Form of Growth
- Inclusion of and identification of Special Education Students

What is the Purpose/Question?

- What is achievement growth of a student over time?

Not one analysis or model that is "best" for all questions.

Different questions/purposes may require a different analysis or model.

standards in three years?

- Is this school an effective school?

What is "Growth"

- Federal definitions:
  - NCLB
  - RTTT, recent flexibility: " **'Student growth' is the change in student achievement for an individual student between two or more points in time** (U. S. Department of Education, 2011, p. 9)."

- Growth that is conditional on time is an absolute growth model
- Growth that is conditional on prior achievement is a relative growth model (Briggs & Betebenner, 2009)

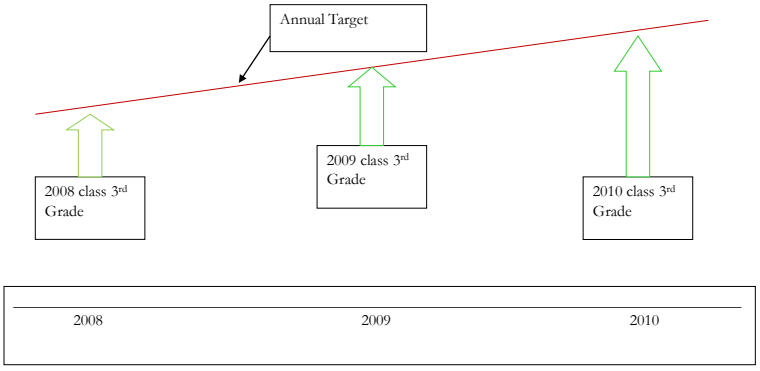
Types of "Growth" Models

- Status and AYP
- Different Groups Improvement
- Same Groups Improvement
- Transition Matrix
- Simple Growth
- Residual and Gain Score Models
- Value-Added Models
- Multilevel Growth Models
- Piecewise Growth Models

NCLB Status and Improvement Models

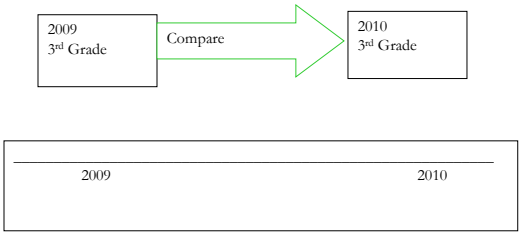
- Status models provide a snapshot of performance—a census
- Interpretation via comparison to a performance standard or benchmark
- Another variant in NCLB is the different groups improvement model, embodied in the NCLB "safe harbor" provision
  - Different cohorts of students compared from one year to another
  - Cohort stability, attrition and mobility important issues
- Status models considered growth models by some, however:
  - Individual student progress over time not tracked
  - Requires assumption that student population has remained stable from one year to the next

Types of Growth Models: Status Models—AYP



NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

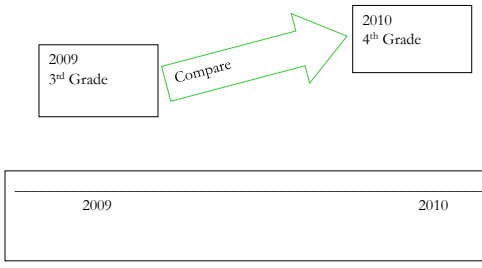
Different Groups Improvement Models (Safe Harbor*)



* If 10% or more of the 2010 group meet AYP than in the 2009 group, school is held "safe". Note that different cohorts of students are being compared.

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

Same Groups Improvement Models



Unmatched cohorts compared over time. Attrition and mobility are major challenges unless probability sampling is used (e.g., NAEP). Popular model before NCLB.

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

Transition Matrix

- Change tracked at the performance standard level (e.g., Below Basic, Basic, Proficient, Advanced)
- Table with previous performance levels as rows and current performance levels as columns
- Each cell indicates percent of students that moved from year to year
  - The diagonal cells indicate students at same level
  - Cells below the diagonal show the students with lower performance
  - Cells above the diagonal show the students with higher performance
  - Can be combined across classes and grades to show total performance for the school
  - Not really growth if students are not tracked individually
  - Allows scores from tests on different scales to be aggregated (e.g., special education students taking alternate assessments)

Example Transition Matrix

	2010 Percent Proficient			
2009 Percent Proficient	Below Basic	Basic	Proficient	Advanced
Below Basic	68%	30%	2%	0%
Basic	16%	61%	20%	3%
Proficient	7%	24%	58%	11%
Advanced	0%	6%	29%	65%

Simple Growth Models

- Simple growth models track individual student scores over two or more measurement occasions
  - Results are often displayed graphically
  - Analysis is usually quite simple (arithmetic rather than statistical estimation)
- Commonly used in classroom and school applications
- Can provide useful information for progress monitoring or instructional intervention

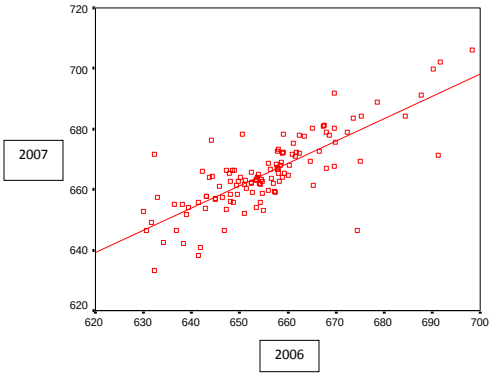
Simple Growth Models

Student	2009 6th grade score	2010 7th grade score	Change in scores
Student 1	610	645	+ 35
Student 2	573	585	+ 12
Student 3	559	604	+ 45
…	…	…	…
Student n	601	623	+ 22
School average	**581**	**618**	**+ 37**

Residual Models

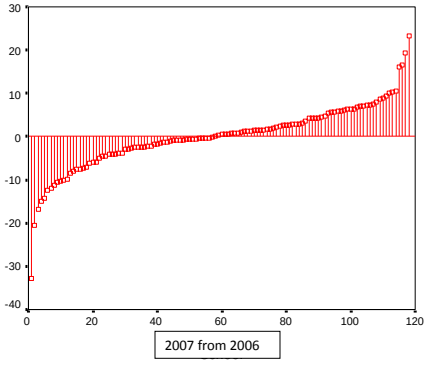
- Simplest residual model predicts current performance from prior achievement
- Difference between predicted and observed scores in the current year is the residual
- Residual gains near zero indicate growth consistent with prediction, positive scores indicate greater than predicted growth and negative scores indicate performance lower than predicted growth
- Residual gain scores can be averaged to obtain a group or school growth measure
- Problem of RTM

Residual Model



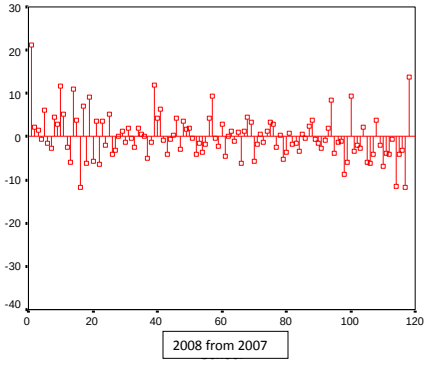
19

Residual Model



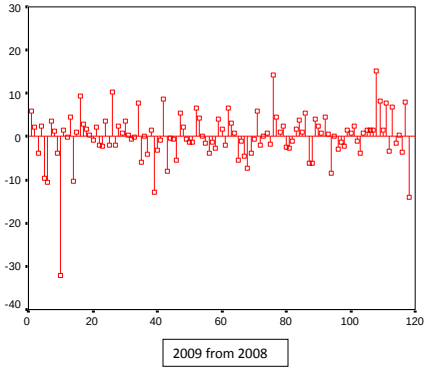
20

Residual Model Stability



21

Residual Model Stability



22

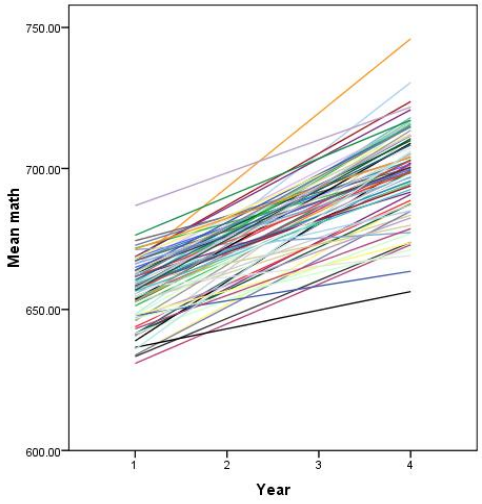
Value-Added Models (VAM)

- Term comes from economics literature; refers to the way in which the value of a product increases from input to output as a result of a production process
- Problem in educational applications is we don't measure input or, usually, the production process
- VAMs try instead to control inputs ad hoc to estimate production function
- Term used indiscriminately
- A number of varieties including:
  - Gain and residual models
  - Persistence Models
  - Some VAMs are mixed models (see next slide)

Hierarchical Linear Models, Mixed Models

- School performance measured more accurately if school nesting is taken into account (Raudenbush & Willms, 1995; Teddlie & Reynolds, 2000)
- Separates individual differences from school differences
- Protects against certain statistical problems (i.e., aggregation bias, misestimated standard errors; enables the investigation of randomly varying slopes)
- HLM models (Raudenbush & Bryk, 1989; Webster & Mendro, 1995)
- TVAAS (Sanders, Saxton, & Horn, 1997)
- EVAAS models (Sanders, 2005)
- Some evidence indicating multilevel models are more accurate than difference or residual gain score models (Koretz & Hamilton, 2006)

Example HLM School Average Growth



25

Piecewise models and seasonal growth

- Number and timing of assessments in growth models very important
- Tracking annual performance may lead to confounding of school effects with out-of-school "seasonal" effects
- Schools may differ in average summer learning
  - Different proportion of students attending summer school
  - Different composition of students; more advantaged students show summer gains or less "summer slide" than less advantaged students
- Piecewise models can be used to examine seasonal effects:

Level-1 (measurement occasions): $Y_{tij} = \pi_{0ij} + \pi_{1ij}(First\ Grade) + \pi_{2ij}(Summer\ Break) + \pi_{3ik}(Second\ Grade) + e_{tij}$

Level-2 (students): $\pi_{0ij} = \beta_{p0j} + \beta_{pqj}(a_{Pij}) + r_{0ij}$

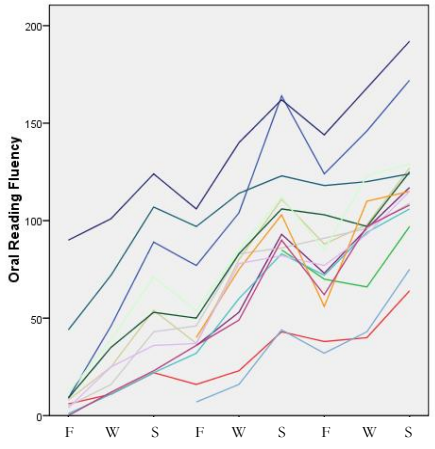
$\pi_{pij} = \beta_{p1j} + \beta_{pqj}(a_{Pij}) + r_{1ij}$, for each slope parameter

Level-3 (schools): $\beta_{p0j} = \gamma_{000} + \gamma_{pq1}(W_{sj}) + u_{00j}$

$\beta_{pkj} = \gamma_{pq1} + \gamma_{pq1}(W_{sj}) + u_{10j}$, for each slope parameter

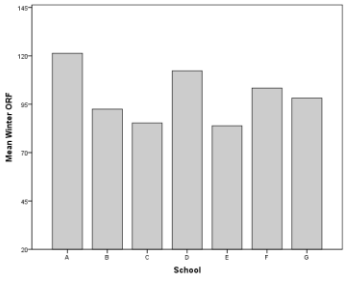
Individual Student Growth in Reading Fluency
(fall, winter, spring for grades 1-3)

Note annual summer drop in performance

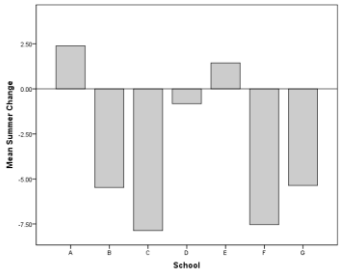


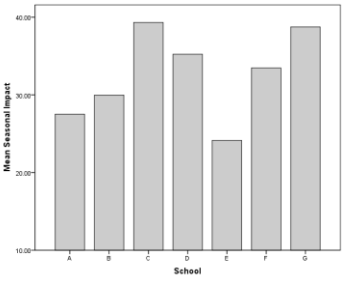
27

Status: 6 Elementary Schools



Summer Growth: 6 Elementary Schools





Seasonally adjusted change: 6 Elementary Schools

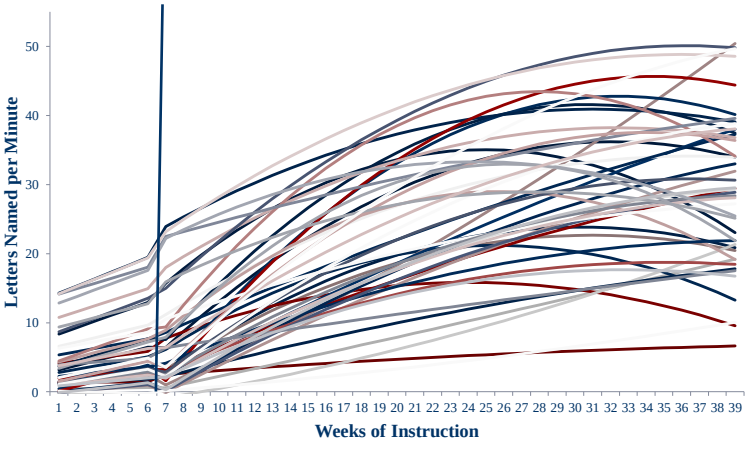
28

Growth Models

- Many variations and different statistical approaches
- Require tracking individual students over time, advantage in that:
  - Each student acts as his or her own control (see Stevens, 2005)
  - Focus is on the outcome of interest, student learning
- Vertical scale (debated by some); may not be necessary for relative growth models
- Functional form important especially for absolute growth models
- Models also differ in how growth is represented and interpreted
  - Projection models
  - Growth to standard
  - Individual trajectories
  - Normative interpretations

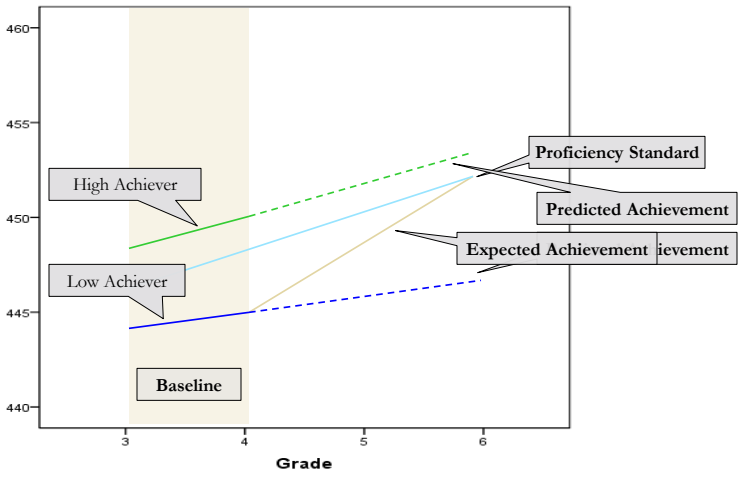
NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

Functional Form



30

Example: Growth to Proficiency Model

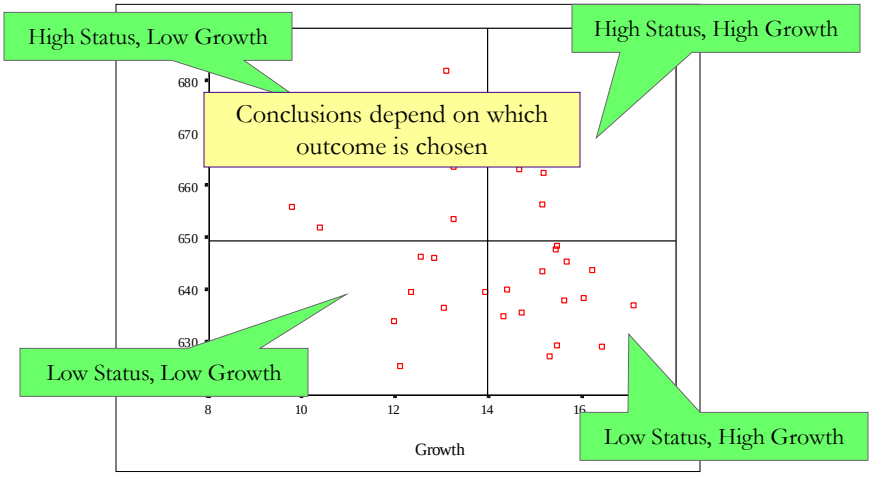


31

Other Important Features of Growth Models

- Number of Measurements/Timing of measurement
- Use of covariates/conditioning (NCLB; England)
- Stability, reliability
  - Number of assessments important
  - *N*-size an issue
  - Strength of prediction important in prior achievement models
- Regression to the mean
- Handling of missing data
- Estimation methods
- Cohort stability
- Causal interpretations only with caution

Model Choice Matters in Inferences About School Performance (an old slide but a good point)



Cohort Stability-Status?

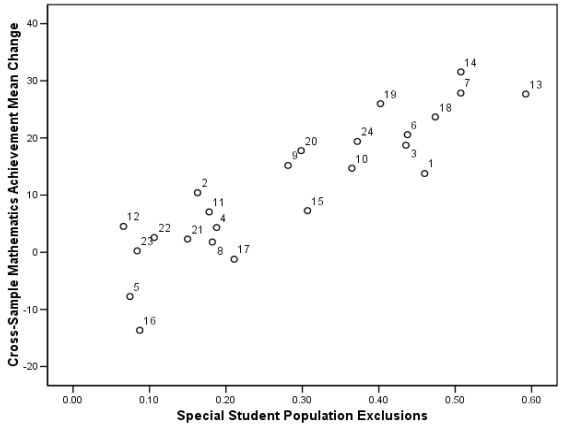


Figure 2. Cross-Sample School Growth Rate Change in Mathematics as a Function of the Proportion of Students from Special Populations Excluded from the Accountability Sample (Zvoch & Stevens, 2006)

Cohort Stability-Growth?

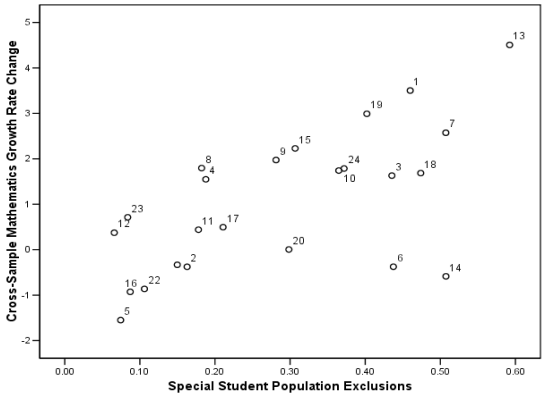


Figure 2. Cross-Sample School Growth Rate Change in Mathematics as a Function of the Proportion of Students from Special Populations Excluded from the Accountability Sample (Zvoch & Stevens, 2006)

Contact Information

Joseph Stevens, Ph.D.
College of Education
5267 University of Oregon
Eugene, OR 97403
(541) 346-2445
stevensj@uoregon.edu

Presentation available at:
http://www.uoregon.edu/~stevensj/hice10.pdf