A Practitioner's Guide to Growth Models

Katherine E. Castellano
University of California, Berkeley

Andrew D. Ho
Harvard Graduate School of Education

February 2013

A Practitioner's Guide to Growth Models

Authored By:

Katherine E. Castellano, University of California, Berkeley

Andrew D. Ho, Harvard Graduate School of Education

A paper commissioned by the
Technical Issues in Large-Scale Assessment (TILSA)
and
Accountability Systems & Reporting (ASR)
State Collaboratives on Assessment and Student Standards
Council of Chief State School Officers

5 - Growth Models of Interest

The main chapters of this guide review seven individual growth models in turn. The ordering of the chapters is primarily pedagogical, beginning with more simple models and proceeding to more complex models. We attempted to select the most widely used growth models and label them by their most common names. However, some models (i.e., the residual gain model) are less commonly used but serve as a conceptual "missing link" between contrasting statistical foundations. A list of equivalent or closely related models is provided in each chapter. There is also an appendix relating these models to those associated with Council of Chief State School Officers (CCSSO) publications about growth models. The seven growth models of interest in this report follow:

- Gain Score
- Trajectory
- Categorical
- Residual Gain
- Projection
- Student Growth Percentile
- Multivariate

6.1 Question 1: What *Primary Interpretation* does the Growth Model Best Support?

One of the central tenets of modern validity theory is that the target of validation is not a model but a use or interpretation of model results. A model suited for one interpretation may not be well suited to support an alternative interpretation. Thus, a natural starting point for growth model classification is the identification of the interpretations that particular growth models best support.

Growth models summarize — typically by quantifying — student performance over two or more time points. They result in metrics that describe individuals and/or groups. This guide identifies three fundamental interpretations that growth metrics can support:

1. **Growth Description**: How much growth? A growth metric may support inferences about the absolute or relative magnitude of growth for an individual or group.
2. **Growth Prediction**: Growth to where? A growth metric may support inferences about the future status of a student or group given current and past achievement.
3. **Value-added**: What caused growth? A growth metric may support inferences about the causes of growth by associating growth with particular educators (e.g., teachers or principals) and schools.

Table 1.5

Classification Scheme for Growth Models

	Primary Interpretation		
Statistical Foundation	**Growth Description**	**Growth Prediction**	**Value-Added**
Gain-Based Model Chapters 1-3: Based on score gains and trajectories on a vertical scale over time	• **Gain-Score** Chapter 1: Gains, average gains, slopes • **Categorical** Chapter 3: Changes and transitions between categories	• **Trajectory** Chapter 2: Extrapolation of gains into the future • **Categorical (a.k.a. Transition, Value Table)** Chapter 3: Implicit momentum toward higher categories in the future	• **Gains/Slopes as Outcomes** Chapter 1.4: Establishes links between average gains and classroom/school membership
Conditional Status Model Chapters 4-6: Expresses scores in terms of expectations based on past scores	• **Residual Gain** Chapter 4: Simple difference between status and expected status given past scores • **Student Growth Percentile (a.k.a the Colorado Model)** Chapter 6: Percentile rank of status given past scores	• **Projection (a.k.a. Prediction, Regression)** Chapter 5: Empirically predicted future score given past scores • **Student Growth Percentile (a.k.a. the Colorado Model)** Chapter 6: Continuation of current percentile rank into the future	• **Covariate-Adjustment** Chapter 4.4: Establishes links between average conditional status and classroom/school membership
Multivariate Model Chapter 7: Uses entire student score histories as an outcome to associate higher-than-expected scores with particular educators	• Generally not used for this purpose	• Generally not used for this purpose	• **Multivariate (a.k.a. EVAAS, Cross-Classified, Persistence Models)** Chapter 7

6.2 Question 2: What is the *Statistical Foundation* Underlying the Growth Model?

This guide also classifies growth models by their underlying statistical foundation. Although statistical methods can be intimidating and model descriptions can be opaque, we find that models can be classified into one of three categories: gain-based models, conditional status models, and multivariate models. These three categories make up the rows of Table 1.5, which cross-classifies growth models by Questions 1 and 2. This table represents a central conceptual framework for this guide. The following subsections briefly describe each statistical foundation in more detail and reference some of their corresponding models.

6.2.1 Gain-based models

The first type of statistical foundation underlies models that are based on gains, average gains, or score trajectories over time. We call these *gain-based models*. A gain or gain score is the simple difference between two scores at different points in time. The gain score can be extrapolated over future time points to support predictions. When there are more than two data points for an individual, the gain can be generalized over multiple time points by averaging and expressing progress as an average change per unit of time.

A common feature to all gain-based models is an implicit or explicit recognition of a *vertical scale*, a common scale that allows scores to be compared across different grade-level tests. Vertical scales support interpretable score differences over the time and grade range of interest. A gain-based statistical foundation is consistent with an intuitive definition of growth: the difference between where one was and where one is. However, vertical scales are difficult to design and maintain, and many useful questions about performance over time do not require vertical scales. This motivates a contrasting statistical foundation underlying a second class of growth models.

6.3 Question 3: What are the *Required Data Features* for this Growth Model?

The selection of a growth model can be motivated by both the advantages it offers and the constraints it satisfies. The selection of a desired model may necessitate alternative or additional data structures. In some cases, the cost of meeting data requirements may outweigh the benefits of the desired model.

In general, all growth models rely on the usual expectations for test reliability and validation. These are not trivial requirements, but this section focuses on requirements for growth, above and beyond the requirements for test score interpretations at a single time point. If low reliability threatens interpretations of test scores at a single time point, the problems will only compound as these scores are reconfigured to support growth inferences. Similarly, all the growth models in this guide require student data that is linked longitudinally over at least two time points.

6.3.1 Vertical scales

Some assessments are scaled across grades with what is known as a "vertical scale." A vertical scale links the reporting test score scale across several grade levels so that a test score from one grade can be meaningfully compared to a test score in a subsequent or previous grade

Vertical scales are necessary for gain-based models and are implicit in intuitive notions of growth. If a test has a defensible vertical scale, a user can take a simple difference of individual scores over time and interpret this as a gain regardless of the starting point on the continuum. In some cases, vertical scales are not formally supported but are implicit and loosely operationalized. An example of this is the categorical model where no vertical scale is claimed, but transitions across performance category boundaries are treated as gains, an interpretation that requires meaningful linkages in cut scores defining the performance categories across grades.

6.3.2 Proficiency cut scores articulated across grades

Some growth models afford growth predictions, often with inferences about trajectories toward some future standard such as "Proficiency" or "College and Career Readiness." These models proliferated under the Growth Model Pilot Program of 2005 (U.S. Department of Education, 2005) that required students to be "on track" to proficiency. Most growth models do not require a proficiency cut score to make a prediction, but the prediction is ultimately referenced to the cut score. In these cases, model predictions require *articulated* cut scores across grades, in other words, proficiency cut scores that maintain some consistent relative stringency or pattern of stringency across grades.

Such cut scores are determined through standard setting procedures in which a committee first defines what proficient students should know and be able to do and then sets cuts by taking into account characteristics of the test scale, item content and difficulty levels, and the qualitative description of proficiency. For many growth models, this process requires consideration of the definitions of proficiency in all other grade levels. Without articulated cut scores, nonsensical conclusions can arise, including a student who is on track to some future standard in one year and three years, but not in two years (Ho, Lewis, & Farris, 2009). Lack of articulation leads to unpredictable relationships between stringency of standards and the grade of entry, the time horizon to proficiency, and target year by which standards must be reached, respectively.

6.3.3 Multiple cut scores articulated across grades

Many accountability and evaluation policies focus primarily on students reaching a single achievement level, usually designated as "Proficiency." Some policies also operationalize performance levels that support finer grain distinctions at higher and lower score points.

Performance level descriptors may include Below Basic, Basic, Proficient, and Advanced, and some states include an even finer resolution of categories below proficiency. Standard-setting processes help to set these cut scores and elaborate on the descriptions for each category.

Categorical models, sometimes known as transition matrix models or value tables, use such ordered performance level categories to determine whether students are making adequate gains toward a standard. Such models rely heavily on the assumption that the performance level categories have been articulated within and across grades. Moreover, the same performance level category in different grades should reflect the same *relative* degree of mastery. As an extension of the previous argument for proficiency cut scores, any growth model that uses multiple cut scores to document growth must have well-articulated standards across grades to avoid counterintuitive results.

Model / Characteristics	Gain Score	Trajectory	Categorical	Residual Gain	Projection	Student Growth Percentile	Multivariate
Brief Description	Describes growth with simple differences or average gains over time	Extends gains or average gains in a predictable, usually linear fashion into the future	Defines growth by transitions among status categories (e.g., Basic, Proficient, Advanced) over time	Describes growth as the difference between current status and expected status given past scores	Uses past scores to predict future scores through regression equations	Percentile rank of current status in a reference group of students with similar past scores	Uses entire student score histories, including other subjects and teachers, to detect higher than expected student scores associated with particular teachers
Aliases, Variants, Close Extensions	Growth Relative to Self, Raw Gain, Simple Gain, Slope, Average Gain, Gains/Slopes-as-Outcomes, Trajectory Model	Growth-to-Standards Model, Gain-Score Model	Transition Model, Transition Matrix Model, Value Table	Residual Difference Model, Covariate Adjustment Model, Regression Model, Percentile Rank of Residuals	Regression Model, Prediction Model	The Colorado Model, Percentile Growth Trajectories, Conditional Status Percentile Ranks	Sanders Model, EVAAS, TVAAS, Tennessee Model, Layered Model, Variable Persistence Model, Cross-Classified Model
Primary Question(s) Addressed	How much has a student learned on an absolute scale?	If this student continues on this trajectory, where is she likely to be in the future?	How has this student grown in terms of transitions through categories over time? In which category will she likely be in the future?	How much higher or lower has this student scored than expected given her past scores?	Given this student's past scores, and based on patterns of scores in the past, what is her predicted score in the future?	What is the percentile rank of a student compared to students with similar histories? What is the minimum SGP a student must maintain to reach a target future standard?	Is this teacher associated with higher scores for his or her students than expected given all available scores and other teacher effects?
Q1: Primary Interpretation	Growth Description	Growth Prediction	Growth Description and Growth Prediction	Growth Description	Growth Prediction	Growth Description and Growth Prediction	Value Added
Q2: Statistical Foundation	Gain-Based	Gain-Based	Gain-Based	Conditional Status	Conditional Status	Conditional Status	Multivariate
Q3: Required Data Features	Vertical scale	Vertical scale	Articulated cut scores across years and grades. Values for value tables. Implicit vertical scale.	An interpretable scale. Assumptions of linear regression must be met.	Interpretable future scale or future standard.	Large sample sizes for reliable estimation.	For high-stakes value-added uses, many years of student data required for stable teacher effects.
Q4: Group-Level Interpretations	Average gain	Average trajectory or percentage of on-track students	Average across value tables or percentage of on-track students	Average residual gain	Average future prediction or percentage of on-track students	Median or average SGP, percentage of on-track students	Only group-level interpretations: Teacher- and school-level "effects"
Q5: Setting Standards	Requires judgment about adequate gain or adequate average gain. Requires understanding of the scale or can be norm-referenced.	Set by defining a future standard and a time horizon to meet the standard.	Set by defining cut scores for categories and values in value table. Requires judgmental cut scores to define adequacy of both individual and aggregate values.	Requires judgment about adequate residual gain. Requires understanding of the scale or can be norm-referenced.	Set by defining a future standard and a time horizon to meet the standard.	Requires judgment about an adequate SGP or median/average SGP. Predictions require a future standard and a time horizon to meet the standard.	Standards required to support absolute or relative distinctions among teacher/school effects, e.g., awards/sanctions to top/bottom 5%.
Q6: Misinterpretations and Unintended Consequences	Intuitive but dependent on vertical scales that can impart undesired dependencies between growth and initial status or socioeconomic status. Can be inflated by dropping initial scores.	Less of an empirical prediction than an aspirational and descriptive prediction. Requires defensible vertical scale over many years. Can be inflated by dropping initial scores.	Loss of information due to categorization of scores. Requires careful articulation of cut scores across grades and years: assumes an implicit vertical scale. Can be inflated by dropping initial scores.	Not a "gain" but a difference from actual and expected status. Violations of linear regression assumption can lead to distortions. Can be inflated by dropping initial scores.	The "projection" metaphor can be confused with "trajectory" when it is in fact a prediction. Maximizing predictive accuracy can diminish incentives to address low-scoring students.	Sometimes misinterpreted as the percentile rank of gain scores. Sometimes overinterpreted as supporting value-added inferences. Can be inflated by dropping initial scores.	Naming fallacy: calling a metric "value-added" does not make it so. Can be unreliable. Detached from theories about improving teaching. Can be inflated by dropping initial scores.

GAIN SCORE MODEL

Aliases and Variants:
- Growth Relative to Self
- Raw Gain
- Simple Gain
- Gains/Slopes-as-Outcomes, Trajectory Model

Primary Interpretation:
Growth Description

Statistical Foundation:
Gain-based model

Metric/Scale:
Gain score – on the common test score scale

Data: Vertically-scaled tests and test scores from two time points

Group-Level Statistic:
Average Gain – describes average change in performance from Time 1 to Time 2

Set Growth Standards:
Determining a minimum gain score needed for "adequate growth"

Operational Examples:
- Pretest/Posttest experimental designs
- Quick growth summaries
- A basis for trajectory models

CATEGORICAL MODEL

Aliases and Variants:
- Transition Model
- Transition Matrix Model
- Value Table

Primary Interpretation:
Growth description and growth prediction

Statistical Foundation:
Gain-based model

Metric/Scale: Change in performance level categories (categorical scale)

Data: Performance levels articulated across years (implicit vertical scale), student status expressed by performance level, and values for transitions if value tables are used

Group-Level Statistic:
Percentage of students "on track" to proficiency or average value across value tables

Set Growth Standards:
Define cut scores for performance levels and values for value tables; specify rules for students being counted as "on track"; establish what average value is good enough

Operational Examples:
NCLB Growth Model (e.g., Delaware and Iowa)

Figure 1.1
Illustration of the Gain Score Model

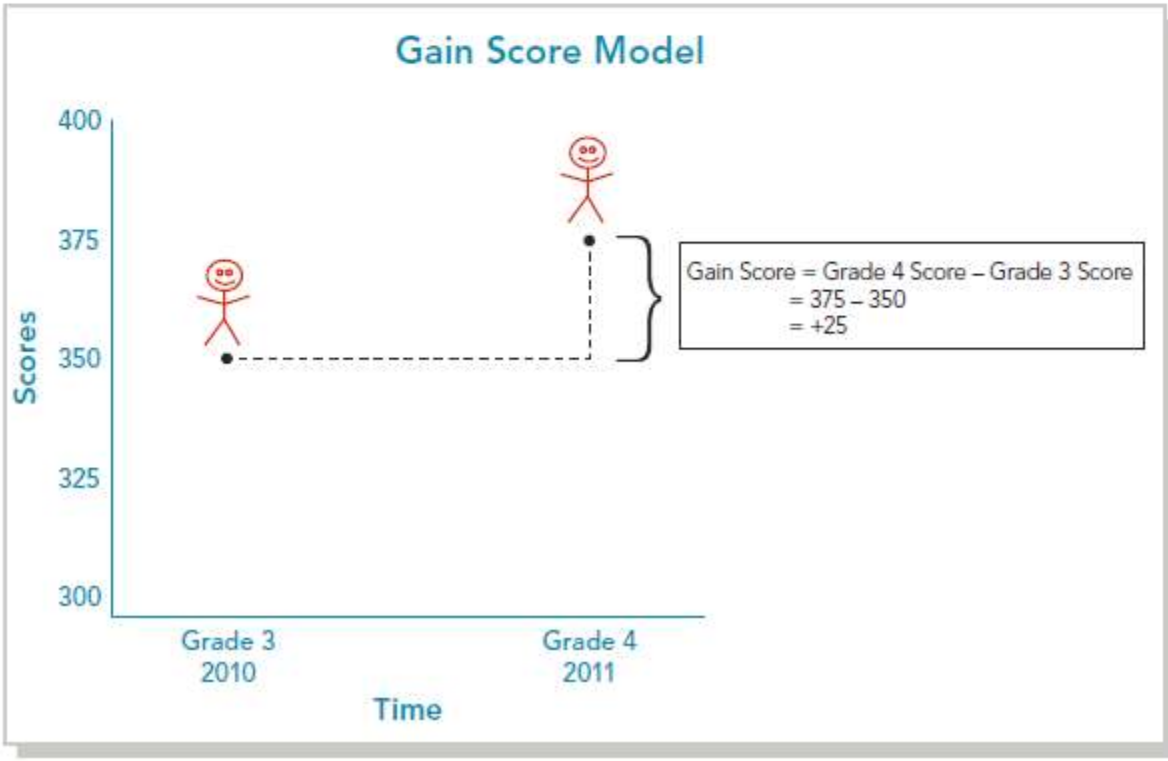
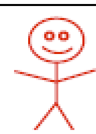


Table 3.1
Example of a Transition Matrix

	Performance Level in Grade 4			
Performance Level in Grade 3	Below Basic	Basic	Proficient	Advanced
Below Basic				
Basic				
Proficient				
Advanced				