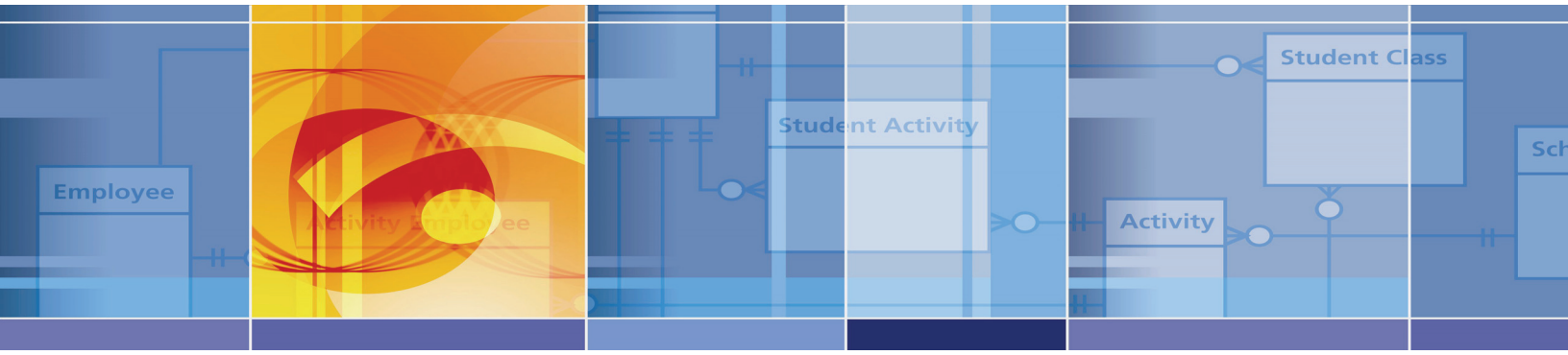

The Optimal Reference Guide:

Comparison of Growth and Value-Add Models

Growth Model Series – Part II

Extraordinary insight™ into today's education information topics

By Glynn D. Ligon

ESP Solutions Group

Table of Contents

Foreword

There are only a few really different approaches to growth models, but many different formulas for calculating them. If we understand which question each model answers, then making a selection among them will be easier.

This paper will examine, at a high level, the characteristics of different growth and value-add models. If you have not at least skimmed Part 1, then you should go back and do so. This paper is called Part 2 for a reason. This paper starts beyond the basics of growth measurement and assumes the caveats, limitations, and admonitions about them are understood.

We should define the term "model" as it is used in this paper. Model here means a high-level category of statistical techniques that answer a specific question for specific purposes. This paper does not give formulas or recommend specific statistical techniques (too often called models by their authors or proponents). The growth and value-add solutions being marketed can be used for one or more of the models described in this paper.

Background

Value-add models are simply a case within each of the growth model types. Interestingly, some people lump all value-add models into one type. They don't discuss the process of turning a simple growth model into a value-add model by merely including a comparison group or expected score.

Some of the confusion in this arena comes from the creative names people have given to their models. Even more arises from the idea that there is a different model if the calculation uses a different metric, for example vertical scale score vs. percentile rank.

Growth vs. Value-Add

1. The only practical difference between a "Growth" model and a "Value-Add" model is that to be value-add, the model must "control for" the influence of selected factors (e.g., demographics, prior performance, etc.) or the impact of an intervention (e.g., school, program, teacher, etc.) on student performance.

2. Controlling for these factors allows the interpreter of the results to say, "The growth beyond what was controlled for represents the value added by the school, for example."

3. When value is added by a school, the school is said to be effective— regardless of the status of students' performance.

4. Value-add models are used for several purposes beyond description of student performance.

    a. *Pay-for-performance:* When teachers or whole schools are rewarded for growth above and beyond what would have been predicted by the factors in the model.

    b. *Evaluation:* When a program is deemed effective because its students outperformed other similar students.

5. For this review, growth models and value-add models are analyzed together because growth models are merely a simple, unadjusted case of the same models that are labeled value-add.

The Model Comparison Chart

Describing the models and their characteristics in text became convoluted, so a chart was created to display all the information together. See the Model Comparison Chart at the end of this paper.

1. **Our Questions**

   The chart begins by stating the question asked about student performance, school effectiveness, or teacher effectiveness. Each question is composed for growth and for value-add.

2. **Related Evaluation Questions**

   Each question is restated to be measurable for groups or individuals.

3. **Model Names**

   The model names come from a review of the literature on growth and value-add models. From all the references, several were selected because they describe similar models to the one being described in the chart. The researcher or practitioner to whom the model name is attributed along with the model name they use are listed to help consolidate all the various terms being used for the same model.

4. **Use**

   Some models are better suited for different purposes. The preferred uses are shown for each.

5. **Pre-Measure**

   Acceptable metrics usable as pre-measures are shown as assessment scores by year and grade level. For group measures, the sequence of the cohort is indicated. For individual measures, the student is shown.

6. **Post-Measure**

   Acceptable metrics usable as post-measures are shown as assessment scores by year and grade level. For group measures, the sequence of the cohort is indicated. For individual measures, the student is shown.

7. **Calculation**

   The appropriate statistical calculation(s) is indicated using the pre- and post-measures.

Statistical significance tests are not included because they must be matched to the assumptions of the data used within each model. For example, interval data may use means and parametric tests; whereas, categorical data require non-parametric analyses, possibly employing the standard error of measurement.

For individual measures, the standard error of measurement is the metric of choice for determining an error band or confidence interval.

"School," as used in this comparison, may be a program, intervention, resource, or other factor thought to impact student learning outcomes.

Factors (covariables or predictors) for value-add formulas are listed on the chart for students, teachers, and schools.

The Real Growth and Value-Add Question

The questions require one additional modification. The stated questions apply only to those students with assessment scores. Unlike a status measure that under NCLB represents at least 95% of the eligible students, a longitudinal measure includes only those students with multiple measurements.

Each question could (should) be preceded by "For those students with whom the school had the opportunity to instruct for all the years included in this analysis…" Obviously, no growth model has data to measure the impact of the school on students that come and go within the timeframe of the model.

Is that really true? Keep in mind that some instances of these models (e.g., SAS) impute missing values. This means that they move forward with gaps in data for some students by plugging in an estimated value. Not a dreadful strategy, but users of that model must understand that for those students, the predicted or estimated performance has more error in it than for other students. (i.e., a combination of measurement error and sampling error).

With the emergence of universal statewide assessment programs, tracking student learning has improved. Thank you, NCLB. Today, a mobile student may bring along a pre-measure from another district in the same state—maybe even multiple years of pre-measures. That bonus adds to the number of students that can be included in an analysis.

How Many Years of Data Are Required for Each Model?

One might think that all models require more than one year of data for every student. In fact, the quasi-longitudinal model is a comparison of two status measures for different cohorts of students. Only one year of data is needed for each student because there is no expectation that the same students are in each year's cohort. This is the only model that includes all students tested in a year. (See Models That did not Make the Chart for possible exceptions to this statement.)

The *longitudinal model* requires at least two years of data for each student included.

The *trend comparison model* requires three or more years of data for each students.

The *trend-to-target model* can work with only one year of data, but the more years in the calculation, the better the prediction is assumed to be. Having more years of data in the calculation is a statistical plus for the model. However, changes in the assessments or academic standards being taught over the years are more likely with more years included.

Models That did not Make the Chart

Maybe there is another value-add model. Let's call it the quasi-value-add-model. Some evaluators and researchers have used a shortcut to growth and value-add measurement that requires only a single assessment score. Using a regression approach, the single score can be predicted, post-hoc, from demographic, context, and assessment data other than from the outcome measure. This is really a prediction of status, but it certainly meets the basic criteria for a value-add model.

In fact, several states have reported their annual status averages for schools in comparison groupings of similar schools based upon wealth, percent minority, percent economically disadvantage, etc. This quasi-value-add-status model is also represented by Just for the Kids, which reports a school's performance in comparison to the top performing schools with similar demographics. Any educator who runs a query on an assessment database to compare assessment performance across schools or districts of similar characteristics is running a quasi-value-add-status model.

These are certainly not growth models, but growth above or below some comparison is implied in the result.

I can support the use of these models for accountability, research, and evaluation. The reason—the extra information a decision maker receives from a growth or value-add model far out weights those limitations.

We must however, be ever vigilant that those decision makers understand the limitations—especially the real questions being answered. Even as a decision maker might be misled by only knowing the status of a school, that decision maker might also over interpret a positive growth result—especially a positive value-add result.

Over confidence in a value-add result is easy to understand. After all, value-add formulas are complex and they take into account so many factors that make comparison of status measures unfair. True, but they also impose different expectations for learning on low- and high-achieving students.

The Metric Metaphor Misconception—Metaphorically Vertical Scale

Many statisticians and psychometricians insist that an assessment must have a true vertical scale in order to properly measure growth. I happen to side with the others who are comfortable with the idea that the statistical models do not require either the predictor variables or the predicted variables to be on the same scale. After all, the non-assessment factors used in value-add models are all on different scales (e.g., family income, race/ethnicity, gender, age, etc.).

I have developed a liking for several metaphorically vertical metrics that should perform even better than others. The premier option is the true vertical scale. When there is none, an estimated vertical scale can work. These include:

- Standardized score within each grade level
- Percentile within each grade level
- Normal Curve Equivalent within each grade level

Selecting a Model

This chart will help in determining:

a. What your question is
b. What use you are pursuing
c. Whether or not you have the data to calculate the gains or value-add

Next, you call a statistician. Fortunately, with the issues detailed in Part 1 and the characteristics in Part 2's chart, you have a reasonable opportunity to understand whether or not the statistical approach recommended by the statistician will meet your real needs and wishes.

Now, you are ready to go back, re-read, and even argue with Part 1.

Model Comparison Chart

ESP Insight

There are only a few really different approaches to growth models, but many different formulas for calculating them. If we understand which question each model answers, then making a selection among them will be easier.

Our Questions	Restated Evaluation Questions		Model Names (Attribution) [Number of Measurements]	Use
	Groups	Individuals		
Growth: Did this year's students score higher than last year's students? **Value Add:** Is this school more effective than it was last year? Is this teacher more effective than he/she was last year?	**Growth:** How does this year's cohort compare to the prior year's cohort in the same grade level? **Value Add:** How does this year's cohort compare to the prior year's cohort in the same grade level after adjusting for selected factors or in comparison to a similar group?	Not Applicable	Quasi-Longitudinal (Ligon) Improvement (Hull) Successive Groups (Carlson) Growth Relative to Others (Gong) [One measurement per student; only model that includes all students tested]	**Growth:** School Improvement AYP Safe Harbor Accreditation **Value Add:** Pay for Performance
Growth: Did this year's students score higher than they did last year? **Value Add:** How effective was this school? How effective was this teacher?	**Growth:** How much did the cohort's performance change? **Value Add:** How much did the cohort change after adjusting for selected factors or in comparison to a similar group?	**Growth:** How much did the student's performance change? **Value Add:** How much did the student change after adjusting for selected factors or in comparison to similar students?	Longitudinal (Ligon) Simple Growth (Hull) Longitudinal (Carlson) Growth Relative to Others (Gong) Eductional Value-Added Assessment System (Sanders) Dallas Value-Added Accountability System (Webster) [Two measurements per student]	**Growth:** School Improvement Evaluation Accreditation **Value Add:** Pay for Performance Evaluation Research
Growth: Did the students learn at a faster pace this year than they did in the past? **Value Add:** Was this school effective in increasing the students' rate of growth? Was this teacher effective in increasing the students' rate of growth?	**Growth:** How did the cohort's trend in growth change? **Value Add:** How did the cohort's trend in growth change after adjusting for selected factors or incomparison to a similar group?	**Growth:** How did the student's trend in growth change? **Value Add:** How did the student's trend in growth change after adjusting for selected factors or in comparison to similar students?	Trend (Ligon) Change in Rate (Carlson) [Three or more measurements per student]	**Growth:** School Improvement **Value Add:** Pay for Performance Evaluation Research
Growth: Will these students be proficient by the time they leave this school or by 2014? **Value Add:** Is this school effective in increasing or keeping all students' learning on pace to be proficient by the target date? Is this teacher effective in increasing or keeping all students' learning on pace to be proficient by the target date?	**Growth:** Is the cohort on track to reach proficiency by the target date? **Value Add:** Is the cohort on track to reach proficiency by the target date after adjusting for selected factors or incomparison to a similar group?	**Growth:** Is the student on track to reach proficiency by the target date? **Value Add:** Is the student on track to reach proficiency by the target date or in comparison to similar students?	Target Date (Ligon) Growth to Proficiency (Hull) Growth Relative to a Standard (Gong) Eductional Value-Added Assessment System (Sanders) Dallas Value-Added Accountability System (Webster) Rate of Expected Academic Growth (Doran) [One or more measurements per student]	**Growth:** School Improvement AYP Growth Accreditation **Value Add:** Pay for Performance Evaluation Research

NOTES:

1. "School" can also be program, intervention, resource(s), or other factor thought to impact student learning.
2. Sources for model names are:
   a. Ligon, Glynn, Typical research references
   b. Hull, Jim, Center for Public Education, National School Boards Association
   c. Carlson, Dale, Assessment Consultant
   d. Gong, Brian, Center for Assessment
   e. Sanders, William, SAS
   f. Doran, Harold
3. Acceptable metrics may be referred to by different commercial or localized names that are equivalent to the textbook names used here.
4. Questions imply that they are for "students who have been enrolled for the time period being measured."
5. SEM (standard error of measurement) is the estimated standard deviation of the error in that method.
6. NCE is the old Chapter I normal curve equivalent.
7. Parametric refers to statistical techniques used on equal interval data—means, standard deviations, analysis of variance, etc.
8. Nonparametric statistics are medians, etc.

FACTORS (COVARIABLES, PREDICTORS) FOR VALUE-ADD FORMULAS:

Students: Prior scores for the same subject area, combined prior scores for multiple subject areas, teacher(s) of record, teacher(s) delivering instruction in a subject area, individual demographics (e.g., age, gender, race/ethnicity, socioeconomic status/economic disadvantaged status, census track, parent education level, etc.), program participation (e.g., Title I, ESL, Special Education, etc.), learning abilities/disabilities (e.g., handicapping conditions, 504, gifted/talented status, etc.), prior academic success (e.g., grades, promotion/retention, honors, etc.), discipline incidents, attendance, school of enrollment, district of enrollment, state of enrollment, school feeder pattern, instructional intervention/pedagogy, benchmark assessments, teacher observations, and others

Teachers: Degrees(s), experience, in-service training, degree-granting institution(s), individual demographics (e.g., age, gender, race/ethnicity, etc.), attendance, observation ratings, evaluation ratings, student performance (e.g., grades, promotion/retention, attendance, discipline, assessment scores, etc.), student factors (see above), school factors (see below), and others

Schools (Includes staff other than teachers, districts, programs, and other groupings of students): Expenditures, resources, building factors (e.g., age, size, condition, features, etc.), staffing, staff characteristics (same as teachers above), student factors (see above), teacher factors (see above)

ESP Solutions Group

Group Measures				Individual Measures			
Pre-Measure [Acceptable Metrics]	**Post-Measure [Acceptable Metrics]**	**Calculation [Statistical Model(s) for V-A]**	**Statistical Significance Test**	**Pre-Measure [Acceptable Metrics]**	**Post-Measure [Acceptable Metrics]**	**Calculation [Statistical Model(s) for V-A]**	**Statistical Significance Test**
$Score_1$, $Year_1$, $Grade_1$, Cohort 1	$Score_1$, $Year_2$, $Grade_1$, Cohort 2	**Growth:** Post-Measure Minus Pre-Measure **Value Add:** (Post-Measure Minus Pre-Measure) Minus (Post-Measure for Comparison Group)	**Growth:** (parametric, means; non-parametric, SEM) **Value Add:** (parametric, means; non-parametric, SEM)	Not Applicable	Not Applicable	Not Applicable	Not Applicable
[Cohort: Raw score, percent correct, scale score, percentile, grade equivalent, lexile, quantile, standard score, NCE, performance level, met standard] [Value-Add: Scale score, standard score, NCE]				Not Applicable			
$Score_1$, $Year_1$, Grade X_1, $Cohort_1$	$Score_2$, $Year_2$, Grade X_2, $Cohort_1$	**Growth:** Post-Measure Minus Pre-Measure **Value Add:** (Post-Measure Minus Pre-Measure) Minus (Post-Measure for Comparison Group) [Regression, Hierarchical Linear Model]	**Growth:** (parametric, means; non-parametric, SEM) **Value Add:** (parametric, means; non-parametric, SEM)	$Score_1$, $Year_1$, $Grade_1$, $Student_1$	$Score_2$, $Year_2$, $Grade_2$, $Student_1$	**Growth:** Post-Measure Minus Pre-Measure **Value Add:** (Post-Measure Minus Pre-Measure) Minus (Post-Measure for Comparison Students) [Regression, Hierarchical Linear Model]	**Growth:** (parametric, variance; non-parametric, SEM) **Value Add:** (parametric, variance; non-parametric, SEM)
[Cohort: Raw score, percent correct, scale score, percentile, percentile growth, grade equivalent, lexile, quantile, standard score, NCE, performance level, met standard] [Value-Add: Scale score, standard score, NCE]				[Cohort: Raw score, percent correct, scale score, percentile, percentile growth, grade equivalent, lexile, quantile, standard score, NCE, performance level, met standard] [Value-Add: Scale score, standard score, NCE]			
($Score_2$, $Year_2$, $Grade_2$, $Cohort_1$) minus ($Score_1$, $Year_1$, $Grade_1$, $Cohort_1$)	($Score_3$, $Year_3$, $Grade_3$, $Cohort_1$) minus ($Score_2$, $Year_2$, $Grade_2$, $Cohort_1$)	**Growth:** Post-Measure Minus Pre-Measure **Value Add:** (Post-Measure Minus Pre-Measure) Minus (Post-Measure Minus Pre-Measure for Comparison Group) [Regression, Hierarchical Linear Model]	**Growth:** (parametric, means; non-parametric, SEM) **Value Add:** (parametric, means; non-parametric, SEM)	($Score_2$, $Year_2$, $Grade_2$, $Student_1$) minus ($Score_1$, $Year_1$, $Grade_1$, $Student_1$)	($Score_3$, $Year_3$, $Grade_3$, $Student_1$) minus ($Score_2$, $Year_2$, $Grade_2$, $Student_1$)	**Growth:** Post-Measure Minus Pre-Measure **Value Add:** (Post-Measure Minus Pre-Measure) Minus (Post-Measure Minus Pre-Measure for Comparison Students) [Regression, Hierarchical Linear Model]	**Growth:** (parametric, variance; non-parametric, SEM) **Value Add:** (parametric, variance; non-parametric, SEM)
[Cohort: Raw score, percent correct, scale score, percentile, percentile growth, grade equivalent, lexile, quantile, standard score, NCE, performance level, met standard] [Value-Add: Scale score, standard score, NCE]				[Cohort: Raw score, percent correct, scale score, percentile, percentile growth, grade equivalent, lexile, quantile, standard score, NCE, performance level, met standard] [Value-Add: Scale score, standard score, NCE]			
$Scores_{1:n}$, $Years_{1:n}$, $Grade_{1:n}$, $Cohort_1$	Target Score minus Predicted Score	**Growth:** Post-Measure Minus Pre-Measure **Value Add:** (Post-Measure Minus Pre-Measure) Minus (Post-Measure Minus Pre-Measure for Comparison Group) [Regression, Hierarchical Linear Model]	**Growth:** (parametric, means; non-parametric, SEM) **Value Add:** (parametric, means; non-parametric, SEM)	$Scores_{1:n}$, $Years_{1:n}$, $Grade_{1:n}$, $Student_1$	Target Score minus Predicted Score	**Growth:** Post-Measure Minus Pre-Measure **Value Add:** (Post-Measure Minus Pre-Measure) Minus (Post-Measure Minus Pre-Measure for Comparison Students) [Regression, Hierarchical Linear Model]	**Growth:** (parametric, variance; non-parametric, SEM) **Value Add:** (parametric, variance; non-parametric, SEM)
[Cohort: Raw score, percent correct, scale score, percentile, percentile growth, grade equivalent, lexile, quantile, standard score, NCE, performance level, met standard] [Value-Add: Scale score, standard score, NCE]				[Cohort: Raw score, percent correct, scale score, percentile, percentile growth, grade equivalent, lexile, quantile, standard score, NCE, performance level, met standard] [Value-Add: Scale score, standard score, NCE]			

References

Betebenner, Damian W. (2008). *Norm- and Criterion-Referenced Student Growth.* Dover, NH: National Center for the Improvement of Educational Assessment.

Betebenner, D. & Doran, H. (2004). A Proposal for Modeling Student Growth as Outlined by HB 04-1433. School of Education, University of Colorado, Boulder, CO. Council of Chief State School Officers, Washington, DC.

Braun, Henri I. (2005). *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models.* Princeton, NJ: Educational Testing Service.

DePascale, Charles A. (2006). *Measuring Growth with the MCAS Tests: A consideration of vertical scales and standards.* National Center for the Improvement of Educational Assessment for the Massachusetts Department of Education.

Hull, James (2008). *Growth Models: A Guide for Informed Decision Making.* Center for Public Education, National School Boards Association.

Ligon, Glynn D. (2003). *Learning Growth Index for SARs.* Austin, TX: ESP Solutions Group.

Ligon, Glynn D. (2006). *Creating a Balanced Perspective on Growth.* Austin, TX: ESP Solutions Group.

Ligon, Glynn D. (2006). *Peer Review Guidance for the NCLB Growth Model Pilot Applications.* U.S. Department of Education.

Lissitz, Robert (2005). *Value Added Models in Education: Theory and Practice.* Maple Grove, MN: JAM Press.

Rand Corporation (2004). *The Promise and Peril of Using Value-Added Modeling to Measure Teacher Effectiveness.* Santa Monica, CA.

About ESP Solutions Group

ESP Solutions Group provides its clients with *Extraordinary Insight*™ into PK-12 education data systems and psychometrics. Our team is comprised of industry experts who pioneered the concept of "data-driven decision making" and now help optimize the management of our clients' state and local education agencies.

ESP personnel have advised school districts, all 52 state education agencies, and the U.S. Department of Education on the practice of K-12 school data management. We are regarded as leading experts in understanding the data and technology implications of the **No Child Left Behind Act (NCLB)**, **ED*Facts***, and the **Schools Interoperability Framework (SIF).**

Dozens of education agencies have hired ESP to design and build their student record collection systems, federal reporting systems, student identifier systems, data dictionaries, evaluation/assessment programs, and data management/analysis systems.

To learn how ESP can give your agency *Extraordinary Insight* into your PK-12 education data, email info@espsg.com.

This document is part of *The Optimal Reference Guide* Series, designed to help education data decision makers analyze, manage, and share data in the 21st Century.

Comparison of Growth and Value-Add Models, Growth Model Series – Part II.

ESP Solutions Group

(512) 879-5300
www.espsolutionsgroup.com