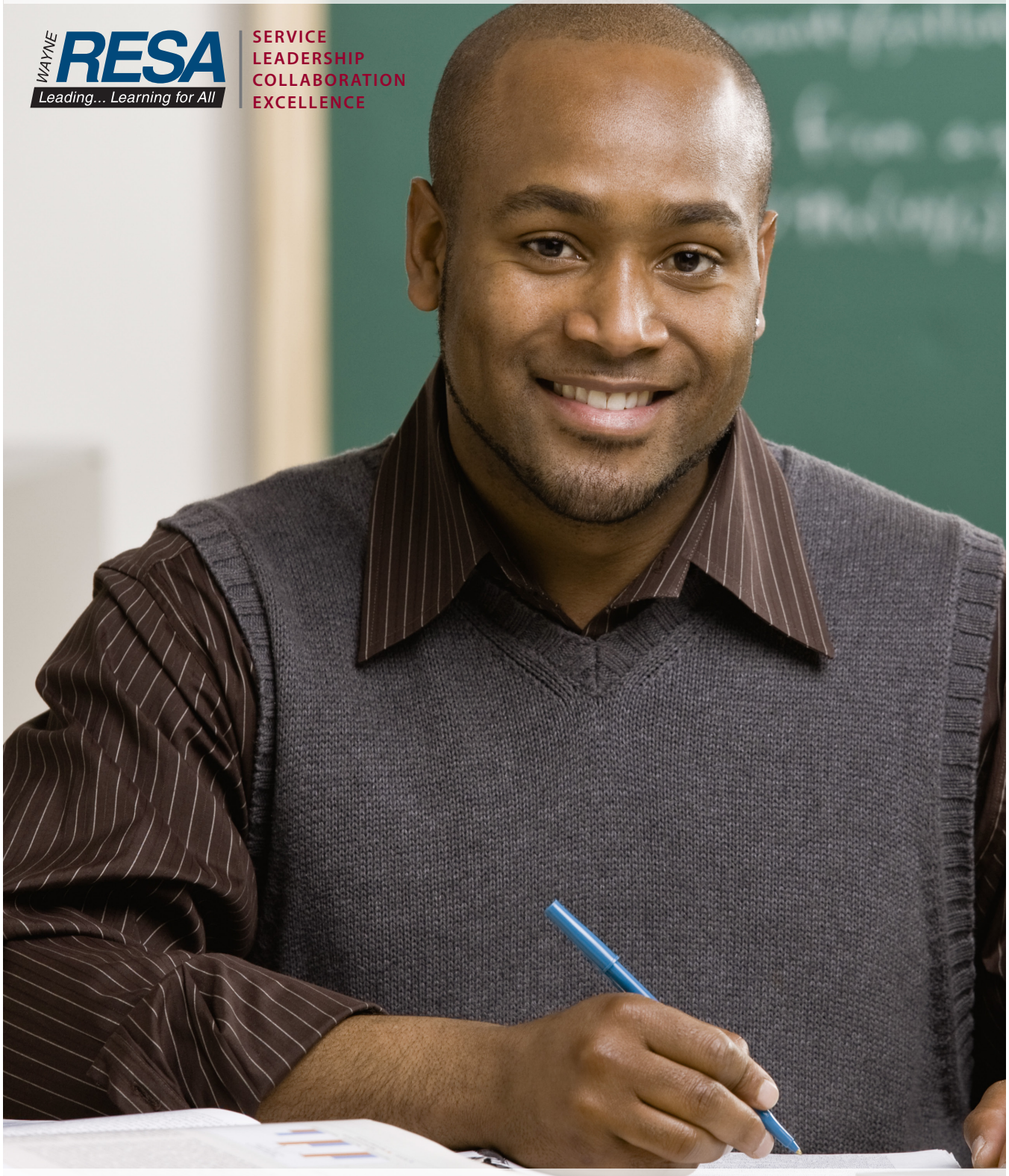
Measuring Student Growth:

A Practical Guide to Educator Evaluation

WAYNE RESA
Leading... Learning for All

SERVICE
LEADERSHIP
COLLABORATION
EXCELLENCE



SECTION 1: Growth Models

Table of Contents

Participants

The development of this ***Student Growth Guidance Document*** has been a collaborative effort involving many educators from across Wayne County, Michigan. These educators have been dedicated to identifying fair, transparent and appropriate methods for measuring student growth throughout the educator evaluation process. Teachers, administrators, central office leaders and ISD staff worked together to understand the research related to student growth models and the best ways with which to implement those models in today's educational environment.

The guidance suggested in this document is based upon a year and a half of study, analysis, debate and thoughtful reflection. This guidance document was not designed with the intention of being read cover to cover. Rather, each section could be read as a stand-alone to further your understanding of student growth. Targeted professional learning will be an important component as you implement this process. The intent of this guidance is to provide several methods whereby a district may be able to measure student growth for purposes of conducting evaluations. The list of participants below reflects the dedicated educators that contributed to this work:

NAME	TITLE	DISTRICT/ORGANIZATION
Brian Amorose	Teacher	Lincoln Park Public Schools
Blaine Armstrong	Principal	Flat Rock Public Schools
Mary Ruth Bird	District Data Coordinator	South Redford Public Schools
Dr. Sandra Brock	Director of Instructional Programs and Services	Northville Public Schools
Chris Buehner	Teacher	South Redford Public Schools
Dr. Patricia Drake	Special Education Data Consultant	Wayne RESA
Mike Flannery	Teacher	Henry Ford Public School Academy
Tom Martin	Principal	Woodhaven Brownstown Public Schools
Angelyn Maxon	Principal	Riverview Public Schools
Dr. Joseph Musial	Assessment Consultant	Wayne RESA
Linda Lazar	Title I Resource Teacher	Dearborn Public Schools
Lena Nemeth	Assessment Consultant	Wayne RESA
Brittany O'Brien	Teacher	Riverview Public Schools
Joe Orban	Director of Data Analysis and Assessment	Wayne Westland Community Schools
Stacy Peterson	Curriculum Director	Woodhaven Brownstown Public Schools
Dr. Lori Roy	Teacher	Livonia Public Schools
Dr. Paul Salah	Associate Superintendent, Educational Services	Wayne RESA
Dr. Sybil St. Clair	Executive Director of Research, Evaluation, Assessment and Accountability	Detroit Public Schools
Stephen Taylor	Coordinator of Student Services	Livonia Public Schools
Cindy Taraskiewicz	Assessment Consultant	Wayne RESA
Mari Treece	Manager/Educational Services	Wayne RESA
Dr. Ellen Vorenkamp	Assessment Consultant	Wayne RESA
Michelle Wagner	Teacher (Instructional Coach)	Van Buren Public Schools
Amy West	Teacher	Allen Park Public Schools

Graphic Design/Document Layout: Kate de Fuccio, Wayne RESA

WAYNE RESA
Leading... Learning for All

SERVICE
LEADERSHIP
COLLABORATION
EXCELLENCE

EDUCATION SERVICES

33500 Van Born Road
Wayne, Michigan 48184
wwww.resa.net

Paul Salah, Ed.D.
Associate Superintendent
(734) 334 1587
(734) 334 1729 FAX

July, 2016

Dear Educator:

Measuring student growth for purposes of educator evaluation is, in my summation, the most challenging aspect of assigning effectiveness labels to educators. Our country has grappled with the following question for several years: How does student growth align with an educator's effectiveness?

Wayne County educators decided that continuing to wait for an answer to this question was fruitless and potentially damaging to the education profession. Yes, damaging is a strong word, and I feel appropriate given the current climate of the education community. The focus of using student growth should be upon the improvement of teaching and learning and thus, logical, fair measures must be implemented. Selecting random cuts based upon proficiency or guesswork is not only inappropriate but also harmful. Harmful because until we solve the student growth quandary, people from many walks of life will not be focused upon teaching and learning, which is the single most important consideration for helping children achieve at high levels. Thus, as a Wayne County, we decided to be proactive and create an approach that determines effectiveness in a fair, thoughtful and transparent way.

This project began during the Winter of 2015 with a small group of dedicated educators grappling with the research, orchestrating a plan, and making a commitment to developing solutions rather than waiting for answers.

We read…

As an internal Wayne RESA team, a group of seven people began by delving into the research. We studied works by Stiggins, Popham and Darling Hammond. We studied the recommendations of Michigan Council for Educator Effectiveness along with works like the Widget Effect and Standard Setting by Cizek and Bunch. We explored the work of other states related to Student Learning Objectives, Formative Assessment, Assessment Choice and overall systems of high quality student growth.

We developed a team…

After some internal study amongst the Wayne RESA group, we invited fourteen school districts and Public School Academies from across Wayne County to come together around a common purpose—developing guidance regarding student growth. Our goal was to challenge the paradigms of the research, continue the learning and foster the voices of teachers, principals and central office administrators toward a common end—fair, transparent methods for measuring student growth. We also met with a subcommittee of Superintendents in order to help facilitate the thinking and development of this process.

After learning…

The team divided into sub-groups with a focus upon key areas related to student growth. As a result of continued debate, thinking and dialogue, a comprehensive Guidance Document designed to provide districts with choice was created. The Guidance Document that follows is designed to give districts options related to Student Growth.

In order to do this work well, districts must commit to intentional implementation, which includes growing capacity and understanding. The Guidance Document in and of itself is not the final answer. Rather, the thoughtful reflection and implementation that occurs after the fact will be essential to any district's success.

I want to thank each and every person that participated in this work. I truly valued the journey we embarked upon and am hopeful that the education community will benefit.

Sincerely,

Dr. Paul Salah
Associate Superintendent, Educational Services
Wayne RESA

List of Tables

List of Figures

What are the different types of growth models?

Growth models measure the amount of academic progress students make between two points in time. There are numerous types of growth models but most tend to fall into six general categories:

1. **Improvement Model**
2. **Performance Index Model**
3. **Simple Growth Model**
4. **Growth to Proficiency Model**
5. **Student Learning Objective Model**
6. **Student Growth Percentile Model**

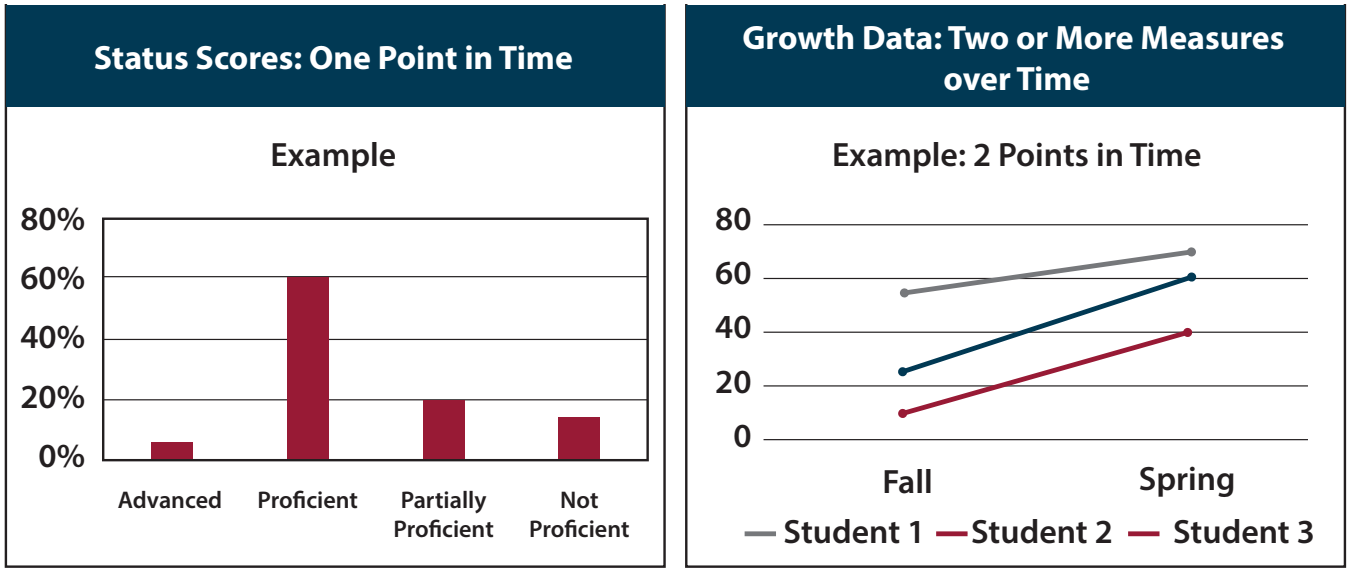
Each of these categories encompasses several variations depending on the model's purpose and available data. Because growth models are relatively new in education, and different models continue to be developed, these six categories may not capture all models.

Growth Based on Cohorts of Students vs. Longitudinal Data

It is important to make a distinction between status scores and growth data ***(Figure 1.1).*** **A status score is a measure taken at one point in time.** Each group of students is their own cohort. There are growth models that use status scores, comparing the performances of cohorts of students.

Growth data that are collected for the same cohort or group of students, looking at their progress in a longitudinal manner typically **involves assessment of students at two or more points in time.** There are growth models based on these longitudinal measurements. Often, these models also involve the use of targets as a basis for making comparisons and judgments of sufficient progress.

FIGURE 1.1 WHAT IS THE DIFFERENCE BETWEEN STATUS SCORES AND GROWTH DATA?



MODEL 1 *The Improvement Model*

The Improvement Model compares the scores of one cohort, or class of students in a particular grade to the scores of a subsequent cohort of students in the same grade. This model is based on achievement status—for example, students scoring at proficient or above. The difference in scores over time however would be considered growth for the educator's performance.

For example, if fifty-five percent of last year's fourth graders scored at or above proficient and sixty percent of this year's fourth graders reached proficiency, then, using the Improvement Model, this educator showed five percentage points in growth when considering fourth grade scores ***(Table 1.1).***

It does not measure growth among individual students or even the same cohort of students. The model actually compares two totally distinct groups of students, or in this example, last year's fourth graders to this year's fourth graders. The benefit of the Improvement Model is that it is fairly easy to implement and understand. While it does not track how individual students progress, it provides some indication of whether more students in a particular grade level are achieving proficiency from year to year. However, the change in the percent of students reaching proficiency may be due to different characteristics of the students in each cohort rather than a change in educator effectiveness, which can be perceived as a significant limitation of the Improvement Model. For example, the difference between last year's fourth graders' performance and this year's fourth graders could have been due to an increase in class sizes due to closing a nearby school in the district or a significant increase in special populations because of factors outside of the school district's control.

TABLE 1.1 EXAMPLE OF SUMMARY CUT SCORES FROM INDIVIDUAL PARTICIPANTS

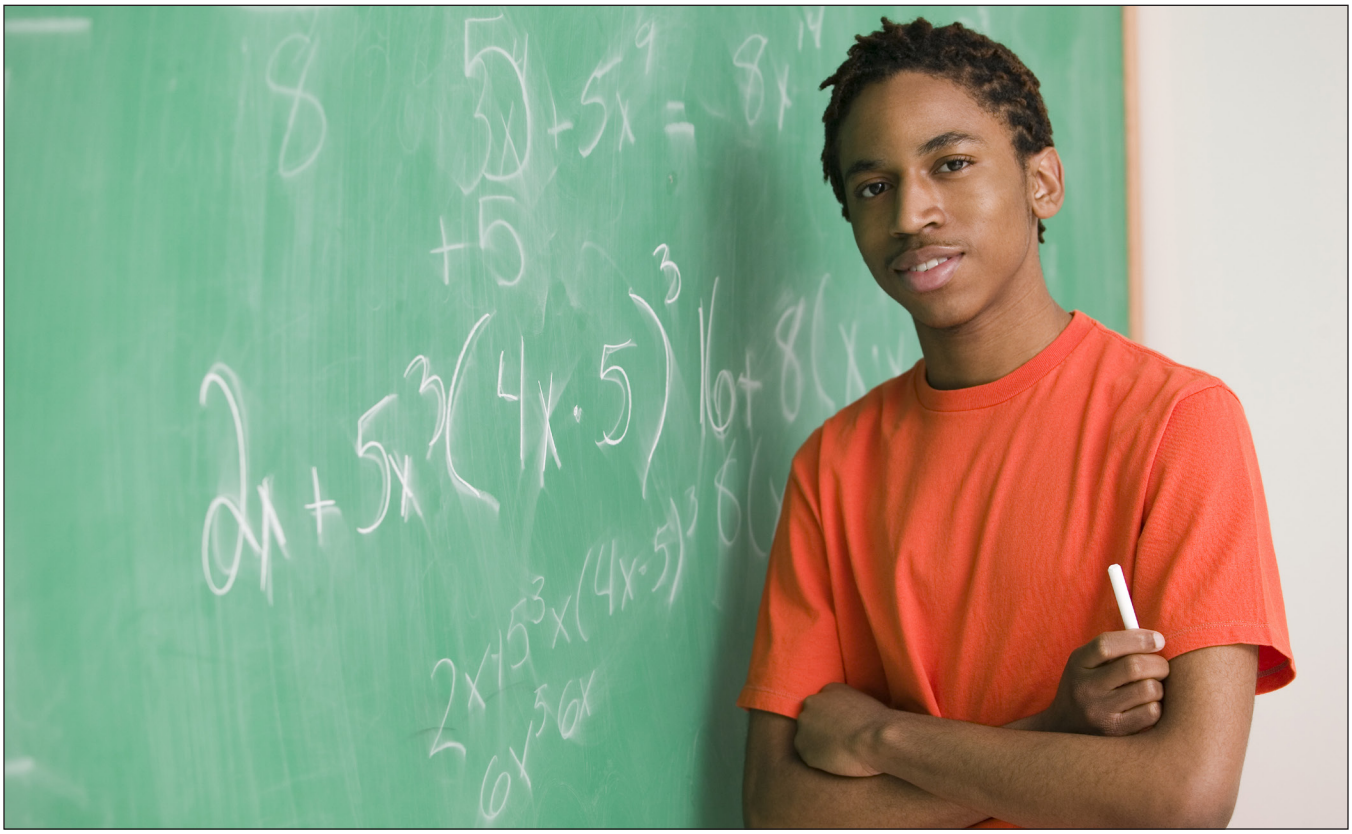
Achievement Level	Last Year's Grade 4	This Year's Grade 4	Change or "Growth"
Percent Proficient	55%	60%	+5%

MODEL 2 *The Performance Index Model*

Most Performance Index Models are status type models that give credit to educators for getting more students out of the lowest achievement levels even if they haven't reached proficiency. Just as with Status Models, Performance Index Models can be used as an Improvement Model. And just as with Improvement Models they do not necessarily measure the academic growth of individual students, but account for change in students' performance from one year to the next. There is, however, one important distinction: This model can be used to recognize change for students who are not proficient. As the example below shows, the educator received partial credit for the students scoring at the basic level but not below basic level.

In statistics, an index combines several indicators into one. Grade Point Average (GPA) is an index that we are all familiar with. It covers several indicators—grades students earn in various courses—and it is weighted in favor of the highest grades, an "A" is worth four points, a "B" is three points, a "C" is two points, and so on. To calculate the GPA, it is a matter of elementary math: Add up the grade points, divide by the number of courses, and the result is the GPA. The GPA shows how close students come to earning A's across their classes with straight A's earning a perfect 4.0 GPA.

Performance Index Models follow this same general principle. Think of it as the GPA for a school or educator where the goal is to determine how close the school or educator comes to getting all students to proficiency. It does so by **measuring student performance based on the percent of students scoring in each achievement level. More points are then awarded for students scoring at the highest levels**, just as students earn more points for higher grades. **The points are averaged and the result is the index.**



HOW A PERFORMANCE INDEX MODEL WORKS

The following example demonstrates the use of classroom data using the Index Model for educator evaluation. The index awards points for achievement as follows: ***(Table 1.2)***

TABLE 1.2 HOW A PERFORMANCE MODEL WORKS

Achievement Level	Points Possible
Students at Proficient and above	100 points
Students at Basic	50 points
Students at Below Basic	0 points

A perfect score of one hundred points means that all students reached proficiency. The educator would earn seventy-two points as shown in the table below ***(Table 1.3).*** Using an Improvement Model, this same educator would earn only fifty-five points for the percent of students who reached proficient.

TABLE 1.3 PERFORMANCE INDEX MODEL

Achievement Level	This Year's Grade	Computation	Points Awarded
Proficient +	55%	.55 x 100 points	55 points
Basic	25%	.25 x 50 points	13 points
Below Basic	20%	.20 x 0 points	4 points
Index Score for School or Teacher			**72 points**

When comparing performance from year to year, a Performance Index will include changes that occurred at the low end of student achievement and can also be designed to include changes among students scoring at proficient or better. For more information on how to determine the ratings to associate with points, refer to the Standard Setting section of this document.

Performance Index Models help schools and educators to concentrate on the growth of all students at all levels. There is credit applied for those students who are not yet proficient. Another advantage is that it can be used in situations where there is only one year of data. The index calculation helps to make the points awarded comparable across grade levels, content areas, and years.

Most Performance Index systems do not require sophisticated data systems. Keep in mind, however, that these models generally do not measure the growth of individual students from year to year. They also do not capture change within each achievement level. For example, if a team set a cut score of two hundred for "basic" and three hundred for "proficient," educators would not get credit for students whose scores improved from two hundred to two hundred ninety-eight. They would get credit for students who improved from two hundred ninety-nine to three hundred one. Establishing more achievement levels would help to capture these changes, making the model a more accurate measure of growth.

MODEL 3 *Simple Growth Model*

The Simple Growth Model is easy to calculate. **This model is most appropriate when working with assessments that provide scaled scores. A scaled score is the total number of correct questions (raw score) that have been converted into a consistent and standardized scale. The model is used to determine the difference in scaled scores from one point in time to the next.** But unlike the Improvement and most Performance Index Models, which compare successive cohorts at the same grade level (fourth graders in our hypothetical classroom), **Simple Growth Models actually document change in the scores of individual students as they progress from a baseline or pre-measure to the outcome or post-measure of learning.** The growth is calculated for each student who took both measures and the change or gain scores are then averaged for the educator's class or school.

HOW A SIMPLE GROWTH MODEL WORKS

This hypothetical educator has five fourth graders who took the fourth grade pre-assessment at the beginning of the year. The changes in scores are calculated in the table below for each student and a class average is reported in Table 1.4.

One drawback of this model is that only those students who took both assessments are included in the educator's growth calculation. Another limitation is that the points themselves provide no information. A fifty-point gain may or may not mean the student has met a set target or is on track to meet it in the future. For Simple Growth Models to be useful, experts, educators, and in many cases, policymakers must make informed judgments about how much growth is enough.

TABLE 1.4 SIMPLE GROWTH MODEL

Student	Pre Test Score	Post-Test Score	Change
Student A	350	400	+50
Student B	370	415	+45
Student C	380	415	+35
Student D	325	390	+65
Student E	316	370	+60
Class or School Average	347	398	+51

MODEL 4 *Growth to Proficiency Model*

While Simple Growth Models measure individual student growth, they do not indicate if students are developing the skills they need to meet state standards of proficiency. **Growth to Proficiency Models—also known as Growth to Standards or On-Track—are designed to show whether students are on track to meet standards for proficient and above.** Although there are several variations, the **key ingredient across all Growth to Proficiency Models is that educators get credit when a student's progress keeps them on pace to reach an established benchmark—usually proficient—at a set time in the future,** typically within three to four years or by the end of high school (Davidson and Davidson 2004).

The advantages to this model are

- that schools and educators are recognized for producing gains in student performance even if their students score below proficient and
- there is more incentive to focus on all students below the proficiency level, not just the "bubble kids."

However, without targets, the model itself cannot determine which students are on track for proficiency. No matter what model is chosen, Growth or Status, it is up to the district to determine the process for setting goals in order to determine how much students should know and when they should know it. Then the model can be designed to determine which students are meeting those targets.

HOW A GROWTH TO PROFICIENCY MODEL WORKS

Our hypothetical classroom has five students whose growth targets were established at the end of fourth grade based on meeting proficiency in seventh grade ***(Table 1.5).*** Growth targets are based on the yearly growth needed to reach the seventh grade proficient score.

TABLE 1.5 HOW A GROWTH TO PROFICIENCY MODEL WORKS

5th Grade Proficient Score	7th Grade Proficient Score
400	600

In the example below, three of five students met the proficient target and therefore do not have to meet a growth target. Two students did not meet the proficiency target: One met his growth target while the other student did not meet hers. This means four out of five students met criteria, or eighty percent, which exceeds the seventy-five percent goal for this year. Therefore, the growth data indicate that the teacher's performance was satisfactory.

TABLE 1.6 PROFICIENCY TO GROWTH MODEL

Student	This Year's Grade 4 Score	This Year's Grade 4 Score	Change	Was the student's score Proficient?	What is the student's growth target?	Did Non-Proficient students hit growth target?	Did the student meet criteria?
Student A	350	400	+50	**Yes**	—	—	**Yes**
Student B	370	415	+45	**Yes**	—	—	**Yes**
Student C	380	415	+35	**Yes**	—	—	**Yes**
Student D	325	390	+65	**NO**	59	**Yes**	**Yes**
Student E	310	370	+60	**NO**	64	**NO**	**NO**
Class Goal	Goal=75% of students will meet criteria. 80% of students met goal.					+51	**Yes**

MODEL 5 *The Student Learning Objectives Model*

A Student Learning Objective (SLO) is a measure of an educator's impact upon student learning within a given interval of instruction. A SLO is a measurable, long-term academic goal informed by available data that an educator or educator team sets at the beginning of the instructional interval for all students or subgroups of students. The teacher and students work toward the SLO growth targets throughout the instructional interval and use interim, benchmark, and formative assessment data to assess progress toward the goal. At the end of the interval of instruction, the teacher meets with a principal or building team to discuss attainment of the SLO and determine the teacher's impact upon student learning.

HOW THE STUDENT LEARNING OBJECTIVES MODEL WORKS

Educators must understand assessment data and identify student achievement trends to set rigorous, yet realistic student growth targets that align with state standards, district priorities, and course objectives. These growth targets should include specific indicators of growth that demonstrate learning between two points in time.

In this hypothetical classroom, the teacher has assessed her students with a pre-test that has 100 points possible. The teacher then reviews where the students started and develops a growth target for each student, considering what will be reasonable, appropriate, and rigorous. The students are given another assessment as a post-test then a calculation of the difference is conducted. The difference becomes the growth score that is compared to the growth target. The teacher then determines the percent of students who met their target and compares that to the class target.

TABLE 1.7 SIMPLE GROWTH MODEL

Student	Pre Test Score (Out of 100)	Post-Test Score (Out of 100)	Growth	Target Growth	Met Growth Target?
Student A	20	48	+28	+20	**Yes**
Student B	24	49	+25	+20	**Yes**
Student C	28	40	+12	+20	**NO**
Student D	45	55	+10	+15	**NO**
Student E	46	46	+0	+15	**NO**

The use of Student Learning Objectives has many benefits. The SLO process promotes learning, reflective teaching practices, the retention of teachers, and it aligns with many quality administrative and school improvement high impact initiatives.

Though the use of Student Learning Objectives is a promising practice, it is not without its challenges. SLOs can be a powerful solution if implemented with care and purpose, but they are not an easy solution. There can be a misconception that SLOs are the quick and easy fix to the challenge of assessing student growth, but in reality, much time and effort is required to execute the SLO process in a credible manner. Time and effort are necessary for planning, communicating, training, and monitoring SLO implementation to make the hoped for improvement in teacher effectiveness and student learning.

MODEL 6 *Student Growth Percentile Model*

The Student Growth Percentile (SGP) Model uses the calculation of student growth percentile scores to describe a student's growth compared to other students with similar test score histories (their academic peers). Although the calculations for SGPs are complex, percentiles are a familiar method of measuring students in comparison to their peers. Percentiles range from 0 to 99 and indicate how many scores in the comparison group are below that score. For example, if a student receives an SGP of 85, it means she demonstrated more growth than 85% of the students in the same grade and subject who had similar prior test scores.

HOW THE STUDENT GROWTH PERCENTILE WORKS

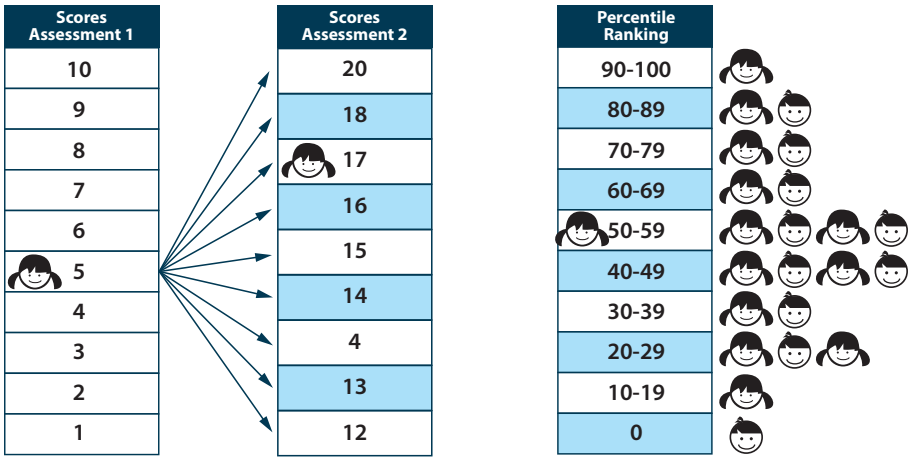
The student growth percentile score is typically calculated by the assessment developer. For example, assessments such as the NWEA and STAR provide a growth percentile calculation. The interpretation of the student growth percentile score is subject to the interpretation of the state or district. In the case of state assessment data, some states will set the cut point for establishing whether or not a teacher's performance is effective. With district purchased assessments, the district would be wise to engage in a process to review the data and define a context for establishing the percentiles that will be demarcated as effective versus ineffective. It can also demonstrate a student's growth and academic progress, even if she is not yet meeting the standard.

SGPs show how a student's achievement at the end of the year compares with that of other students who started the year at the same level. For example, if a student scored 5 on last year's test, their score at the end of this year would be compared with the scores of all the other students who scored 5 last year. The student's SGP would be the percentile rank (from 1 to 99) within this group of similar peers. If the student's SGP is 50, it means that his/her growth in test scores is right in the middle: half of the similar students who scored 5 last year scored higher than she did this year, and half of them scored lower.

The middle SGP score provides a simple indicator of how well the typical student in a class performed relative to similar students. This score is called the median growth percentile (MGP), and it is useful because, unlike a simple average, it doesn't change much if one or two students do unusually well or unusually poor relative to their peers. The MGP does not account for variations among students or classes, nor does it indicate what caused improvement.

Median growth percentile calculations do not try to adjust for differences in student characteristics. Neither SGPs nor value-added modeling indicates what might have caused improvements, nor do they reveal whether other students would make similar improvements if taught by that teacher.

FIGURE 1.2 STUDENT GROWTH PERCENTILE MODEL



What are the limitations of growth models?

Growth models hold great promise of evaluating schools on the amount of growth their students achieve. But growth models, especially value-added models, are relatively new in the education realm and their limitations are still being debated within the research community. Please note, however, that the research community is almost united in the opinion that growth models provide more accurate measures of schools than the current status models alone. Moreover current status models also suffer from many of the same limitations. While none of these issues should preclude states or districts from considering implementing a growth model, they do need to be acknowledged so the model developed will be the most effective tool for its purpose.

MEASURES OF ACHIEVEMENT CAN BE GOOD, BUT NONE IS PERFECT

This guide doesn't debate the pros and cons of standardized testing; there are plenty of publications that do. But it is necessary to discuss limitations and how they can affect the reliability of a growth measurement.

As discussed earlier, it's important to use tests that are appropriate for growth models. Growth can be measured without tests, but any tests used should have the following features:

- They cover the lower and upper ranges of student performance, rather than cluster test content around the knowledge and skills that constitute "proficient."
- They are vertically aligned and scaled to more accurately measure student achievement from year to year.
- They are aligned with state standards.

Unfortunately, while some tests are clearly better than others, there is no perfect measure of achievement, a statement to which even the most ardent supporter of standardized testing would agree.

One of the problems with tests used for growth models is that gain scores over time tend to be what statisticians call "noisier" than measures of achievement at a single point in time. By this, statisticians mean that gain scores tend to fluctuate even though a student's true ability typically does not change much from year to year. This happens because on any given test a student's performance is a result of his or her true ability and random influences, like distractions, during the test and the selection of items—effects that statisticians call measurement error. When scores from the two tests are subtracted from each other, as in Simple Growth models, the measurement error increases so the "true" performance becomes less clear.

There are statistical adjustments to minimize variance in the data, such as including scores from other subjects and previous years. Another way to minimize the effect of unreliable data is to create rolling averages by averaging growth over multiple years to provide a more stable picture of performance. However, such adjustments will add to the complexity of the growth model and may make it difficult to explain to educators why two schools (or teachers) with similar achievement gains received different ratings of effectiveness.

A SINGLE GROWTH MODEL DOES NOT SERVE ALL PURPOSES

Growth data can be helpful in many ways, and it is tempting to create one growth model and use it for multiple purposes. Policymakers and educators should resist this temptation. Although a single model could save a lot of time and money, many researchers strongly discourage using just one model, because trying to pull distinct pieces of information from one model would likely lead to false conclusions.

COMPARISON OF GROWTH MODELS

The following table provides a summary of the growth models described above that may assist the reader in making comparisons.

TABLE 1.8 GROWTH MODEL SUMMARY

Model	Description	Primary Question Asked	Benefits
Improvement Model	Compares the percent proficient from one year to the next with the same grade or teacher in that grade.	Are students in the teacher's class showing improvement from year to year?	• Easy to implement/ understand. • Indicates whether more students in a grade level are reaching proficiency from year to year.
Performance Index Model	Compares performance ranges by multiplying the percent proficient by an index score that converts the data to a "Performance Index" score. The performance index score is then compared to a standard or criteria for effectiveness.	How much gain are students showing, overall?	• The performance index can be used to credit gains made by proficient and non-proficient students. • The index creates comparability across teachers and content area.
Simple Growth Model	Describes growth with simple differences or average gains over time.	How much has the student learned on this scale?	• Simple to calculate. • Can be used with individual students or with groups of students.
Growth to Proficiency Model	Compares student gains to overall proficiency targets, including targets for non-proficient students based on projected trajectory to reach proficiency.	Are students proficient and making progress toward proficiency?	• Credits gains for students below proficiency.
Student Learning Objective Model	Establishes growth goals for students.	Are students meeting learning targets?	• Educators define learning targets. • Promotes learning, reflection, and retention of teachers. • Aligns with school improvement high impact initiatives.
Student Growth Percentile Model	Provides a ranking of a student's change in score when grouped with others who started at the same baseline score.	How does this student's growth rank in comparison to others of comparable prior ability?	• The model recognizes gains of all students, regardless of proficiency.

Primary Interpretation	Required Data Features	Setting Standards	Limitations
Growth description. Average group gains typically based on the percent proficient for different cohorts of students.	The use of the same assessment from year to year. Assumes the assessment has a vertical scale to be comparative from year to year.	Requires judgment about adequate gain or average gain. Requires understanding of the scale or can be norm-referenced.	• Does not measure growth for individual students. • Does not measure same cohort of students, change may be due to different student characteristics from year to year, not educator effectiveness.
Growth description. Provides a summary score that can be compared to a standard of effectiveness.	Can be used with raw scores, scaled scores, and percent proficient. Simple to calculate.	Requires setting of standards to determine performance bands, index weighting, and criteria for determining effectiveness of performance index scores.	• Does not measure growth for students from year to year. • Can mask change within score ranges.
Growth description. Simple difference between two points in time, typically compared to a norm or criterion to establish adequacy of the gains.	Vertically scaled data or scaled scores. Must control for instructional intervals between pre- and post-measurement.	Requires data review to establish adequacy of gains	• Requires business rules for missing data—students must take two measures. • Data can be manipulated to enhance gains. Requires control of instructional intervals when making comparisons between teachers.
A combination of growth description and growth prediction. Provides a description of overall proficiency including the accomplishment of projected gains for non-proficient students.	Vertically scaled data to be used to project proficiency in 3 to 5 years. Requires articulated cut scores across grades.	Standards are based on future scale or future standard. Requires setting targets for students based on projections of future performance. Requires setting of standard for effectiveness rating at teacher level.	• Does not attend to growth of consistently proficient students. • Problematic with student groups who will not achieve proficiency due to disability or disadvantage.
Growth prediction. Provides student level information that can be aggregated to reflect teacher performance.	Can be used with criterion referenced and/or standardized assessments at interim, benchmark, summative intervals.	Requires setting standards for future individual student learning objectives and for classroom targets based on a future standard of score.	• Time consuming. • Concerns with comparability of objectives and data sources. • Does not translate to year-to-year comparisons.
Growth description and growth prediction. Is used to interpret "on-track" students.	Requires large data sets to calculate the student growth percentile. Best with scaled score data.	Requires judgment about an adequate Student Growth Percentile or median/ average Student Growth Percentile. Predictions require a future standard and a time horizon to meet the standard.	• Sometimes misinterpreted as the percentile rank of gain scores. • Sometimes over interpreted as supporting value-added inferences. • Can be inflated by dropping initial scores.

Glossary

Terms for Statistics and Measurements	Definition	Answers the Question	Pros in Growth Measurement	Cons in Growth Measurement
Assessment Literacy	Refers to an educator's comprehensive understanding of assessment and its role in learning.	How well do I use assessment to improve the learning of my students?	• Is essential for teachers and administrators to understand the assessment data they have available and are using to define and analyze student growth.	• Requires professional development and opportunities to apply understandings of assessment in a meaningful context. • Requires time. • Requires motivation of educators to participate in, learn and apply assessment literacy to their work.
Confidence Interval	A range represented by a lower limit number and upper limit number.	How confident are you that the true mean falls between the two numbers? We say we are 95% confident.	• Provides a good visual for a measure of central tendency (true mean).	• It is not symmetric around the mean resulting in a possible low normal and a high normal.
Criterion Referenced Data	Tests and assessments are designed to measure student performance against a fixed set of predetermined criteria or learning standards.	What are students expected to know and be able to do at this point in their education?	• Criterion referenced assessments are preferable in comparing student performance to previous learning or rating performance aligned to a learning expectation.	• Criterion assessments can be time-consuming and complex, expensive to implement, and do not readily allow comparisons among students.
Interim Assessments	Assessments that are administered between annual assessments. For example, an interim assessment might occur in the fall, winter, and spring to be compared to annual spring assessments.	Is student learning on track toward annual performance expectations? Is sufficient curriculum being covered for students to meet annual assessment expectations?	• Interim assessments provide the ability to gather and compare data within a single year and over the course of multiple years. • The data provide longitudinal information for making comparisons over time. • Administrators often use the data to track student growth.	• There is concern with the amount of time that students spend taking tests with interim assessments. • Time for teachers to review the data and to understand how to use the data to adjust curriculum and instruction can be a problem. • The method assumes that growth is linear when that may not be the best trajectory for the student's developmental level or the skills being assessed.

Terms for Statistics and Measurements	Definition	Answers the Question	Pros in Growth Measurement	Cons in Growth Measurement
Mean	Represents the arithmetic average of scores. It is a measure of central tendency.	What is the average gain for the data on hand?	• Easy to calculate. • Can be used when identifying growth based on average number of students or averages of norm referenced data.	• Masks trends in the distribution of student gains from high to low. • Does not describe range of data. It is affected by extreme scores (outliers).
Median	Represents the mid-point in a distribution of scores. One-half of the scores fall below it and above it. It is a measure of central tendency.	What is the mid-point within the data set? Or what is the 50th percentile score?	• Requires the ranking of the data (or scores) from lowest to highest. It is a stable measure because it is not impacted by extreme scores (outliers). • It permits one to determine at which point a child is represented in terms of percentiles. • Can be more "fair" in representing data trends within the distribution of scores than a solitary mean score. • Most useful with student growth percentile data.	• Represents aggregate data. One should conduct quality assurance checks to ensure that the data entry was correct prior to calculating. • Should use a software with large data sets (Excel).
Mode	The mode is the value that appears most often in the data set.	What is the most common gain observed within the data set?	• Identifies the gain that is most commonly demonstrated across students.	• Time to organize the data for analysis and interpretation. • Does not represent the range of gains in student growth. It may take on a bi-modal shape or two modes. • Requires a context to be meaningful, e.g., a specific teacher's data set with additional explanation of factors.

Terms for Statistics and Measurements	Definition	Answers the Question	Pros in Growth Measurement	Cons in Growth Measurement
Norm Referenced Data	Norm-referenced data compares the individual's performance to that of others, usually of the same age or grade level.	How does this individual's performance compare to others?	• Data can be compared across individuals. • Data can be represented in equal interval units, such as standard scores or percentiles. • There is control for central tendency.	• Norm-referenced data may be too far removed from classroom instruction to be appropriate in teacher evaluation. • The representativeness of the sample may not match the local norms in performance or sampling. It makes no mention of content mastery, rather, it asks how a student did compared to her norm.
Percentile	A score that represents the ranking of scores from highest to lowest. For example, a score at the 75th percentile means that the score is greater than or equal to 75% of the persons taking the test.	How does this individual's score rank in comparison to others?	• The percentile provides a ranking or comparison that describes the relative standing of the individual in terms of the percent who performed equal and less well on the task. • Can be simple to calculate. It is misleading if examining scores from a highly gifted student population.	• Is often confused with a percentage score. • The percentile does not communicate the spread of scores from one another but the placement of the individual's score from high to low. • Calculation tools may vary in regard to central tendency in score dispersion.
Percentage	A ratio or number that expresses a fraction of 100.	What is the ratio of success on this task?	• The percent is simple to calculate. • The percent can be used to represent the ratio of students meeting certain criteria or levels of performance. Is often used by teachers when grading students. • Can be helpful to monitor growth and to summarize performance.	• Can be misused as a target for educator evaluation purposes, especially when used without a context of past performance, years of trend data, and analysis of what is reasonable within growth measurement timeframes.
Performance Level Descriptor	The performance level descriptor is the written criterion for the categories of a rubric.	What is the criterion that distinguishes each category?	• Is customized to the context of data, content, and categories. • Provides a standard against which raters classify data into categories.	• Requires clearly written descriptors.

Terms for Statistics and Measurements	Definition	Answers the Question	Pros in Growth Measurement	Cons in Growth Measurement
Predicted Score	A method of growth measurement in which past scores are used as a basis for projecting future scores.	Given the student's past scores or patterns of scores in the past, what is the predicted score for the future?	• Requires the setting of a future standard of performance and a time frame to meet the standard.	• Predicted scores can be confused with "trajectory". • Emphasis on predicted scores can diminish incentive to work with low achieving students.
Progress Monitoring	A method of assessing a student's academic performance, to quantify a student's rate of improvement or responsiveness to instruction, and to evaluate the effectiveness of instruction. Can be implemented with individual students or a class.	Is the student making progress with instruction and/or intervention?	• Repeated brief and targeted assessments are used that are aligned directly to the instruction of skill(s). • Can be easily represented in graphs. • Can be used with targets or goals.	• Identifying a method of progress monitoring that aligns with instruction. • The focus of the progress monitoring may be too narrow for educator evaluation purposes. • Requires training and monitoring of how the data are used to adjust instruction. • There is no gold standard in the number of observations needed to witness growth (e.g., 3 or 10 observations?)
Reliability	Reliability refers to the consistency of scores over time or the ability of a measure to be repeated with the same or similar results. It is inappropriate to say that a test is reliable because reliability is a function of data or scores on hand.	Are the data from this assessment consistent? If I did this again, would I get the same results?	• Relatively easy to calculate. • Strong reliability indicates that the method is stable.	• Requires some statistical calculation skill or access to calculation tools. • Tests or assessments that are highly reliable may not be sensitive to changes that are age/grade appropriate and meaningful to the individual. • Tests or assessments that have low reliability cannot be trusted to yield consistent information. It is a paradox when attempting to measure change. • High stakes testing requires reliability coefficients ≥ .90.

Terms for Statistics and Measurements	Definition	Answers the Question	Pros in Growth Measurement	Cons in Growth Measurement
Standard Deviation	A statistical method of analyzing the amount of variance around a score.	How much might the score vary due to factors other than ability?	• The standard deviation is an important statistic for describing the amount of error surrounding a score. • It is useful in understanding change in test scores between administrations. For example, if two scores are within the same standard deviation that would indicate that there was little change and the difference in scores may be due to normal fluctuations in the test scores/ data.	• The standard deviation is often not used, not available, or not referenced when analyzing test score data. • Requires some understanding of test scores and statistics to analyze and reference in the context of student growth measurement.
Standard Setting	Process for defining gains that requires judgment about adequate gain or adequate average gain. Requires understanding of the measurement scale or can be norm-referenced.	What are the cut points for differentiating teacher effectiveness categories using student growth data?	• A cut score is established based on performance level criteria. • Involves stakeholders. • Can be revised based on new information. • Provides a context for understanding data and making meaning of growth data categories.	• Can be a time-consuming process. • Requires training and understanding of data, measurement, and performance criteria. • Requires attention to business rules and clarity of terms.
Student Learning Objective (SLO)	A specific learning goal and a specific measure of student learning used to track progress toward the goal.	What is the expectation of learning and method of tracking progress toward that goal?	• The SLO in the context of educator evaluation reinforces best teaching practice, encourages collaboration, relies on teacher skill, and is considered to be helpful in connecting teacher practice to student skill.	• It can be difficult to identify and develop high quality assessments across all grades and subjects. • There are challenges to creating appropriate growth targets for classrooms in which students are starting at different achievement levels. • There are challenges to setting attainable yet rigorous targets with the proper "gain" size.

Terms for Statistics and Measurements	Definition	Answers the Question	Pros in Growth Measurement	Cons in Growth Measurement
Trajectory	A trajectory extends gains or average gains in a predictable, usually linear fashion into the future. Trajectories may be used when using growth-to-benchmark models or gain-score models.	If this student continues on this trajectory, where is she likely to be in the future?	• The trajectory is set by defining a future standard and a time horizon to meet the standard.	• The prediction is descriptive and aspirational. • Requires defensible vertical scaling over many years. • Can be inflated by dropping initial scores.
Validity	Validity is the extent to which a concept, conclusion or measurement is well-founded and corresponds accurately to the real world.	Does the assessment measure the skill, construct, or content it purports to measure?	• Validity is important to ensure the test is measuring the intended content.	• Sometimes persons mistake face validity as sufficient to determine the quality of the content.
Vertical Scaling	Vertical Scaling is the method based on Item Response Theory for assuring the items of a test are aligned to show growth.	Does the vertical scaling represent the developmental appropriateness of performance standards progression over grade levels?	• Vertical scaling provides consistent scores across grade levels and is advantageous for measuring growth.	• The procedure requires sophisticated statistical methods.

References

- **SECTION ONE: Growth Models**

  Hull, Jim. Center for Public Education. Taken from **http://www.centerforpubliceducation.org/Main-Menu/Policies/Measuring-student-growth-At-a-glance/Measuring-student-growth-A-guide-to-informed-decision-making.html#sthash.ObX8ECGd.dpuf**

- **GLOSSARY**

  American Institutes for Research. Progress Monitoring Taken on April 27, 2016 from **http://www.rti4success.org/essential-components-rti/progress-monitoring**

  American Institutes for Research. Student Learning Objectives: Benefits, Challenges, and Solutions. November 2012. Taken on April 27, 2016 from **https://westcompcenter.org/wp-content/uploads/2013/04/SLO-Benefits-and-Challenges-American-Institutes-for-Research.pdf** .

  Assessment Literacy: Making Sense of K-12 Assessment for Learning. Taken on April 27, 2016 from **http://assessmentliteracy.org/about/** .

  Castellano, Katherine E. and Ho, Andrew D. *A Practitioner's Guide to Growth Models*. Council of Chief State School Officers, 2013.

  Wood, John. The Importance of Comparability in Today's Assessment Environment. Taken April 27, 2016 from **http://assessmentliteracy.org/blog/importance-data-comparability-todays-assessment-environment/** .

Notes

WAYNE RESA
Leading... Learning for All

SERVICE
LEADERSHIP
COLLABORATION
EXCELLENCE

Wayne RESA
33500 Van Born Road • Wayne, MI 48184-2497
734.334.1300 • 734.334.1620 FAX • www.resa.net