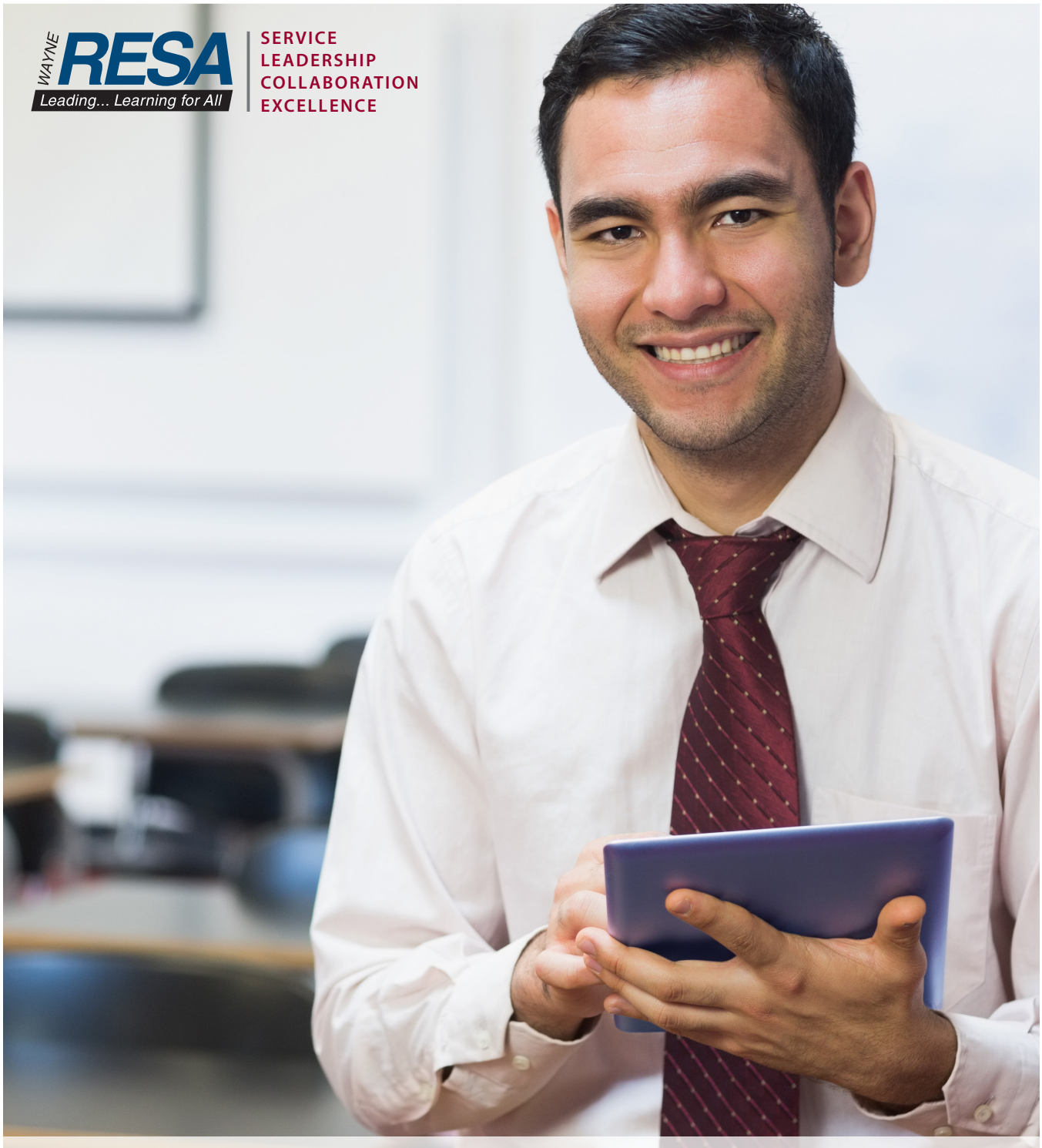
Measuring Student Growth:

A Practical Guide to Educator Evaluation

WAYNE RESA
Leading... Learning for All

SERVICE
LEADERSHIP
COLLABORATION
EXCELLENCE



SECTION 4:

Standard Setting for Student Growth

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.

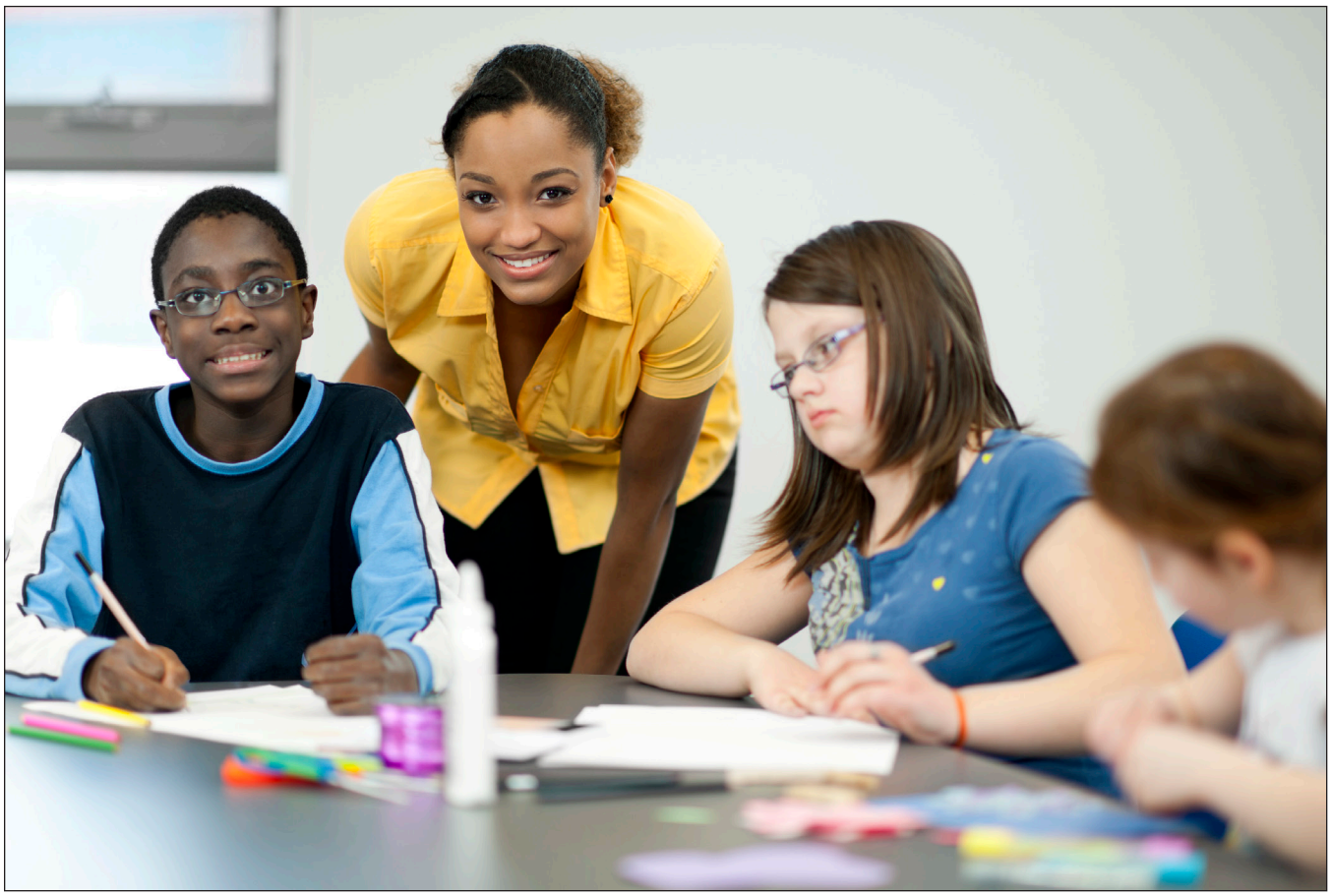
In order to do this work well, districts must commit to intentional implementation, which includes growing capacity and understanding. The Guidance Document in and of itself is not the final answer. Rather, the thoughtful reflection and implementation that occurs after the fact will be essential to any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



The standard setting process is research-based. It is a valid and reliable method for bringing stakeholders together with the intention of assigning meaning to achievement data.

Standard setting is a process for defining cut scores that is based on local school district achievement data. The standard setting process is used to find the cut scores for teacher evaluation. The purpose of the cut score is to indicate the threshold for each teacher evaluation category. The cut score establishes the difference between effective and minimally effective teacher performance based on student growth data.

The standard-setting steps include:

1. **Assemble the Team**
2. **Document the Process**
3. **Establish Data Sources for Student Growth**
4. **Clarify the selected Model of Growth Assumptions**
5. **Write Performance-Level Descriptions**
6. **Train Committee Members**
7. **Implement the Selected Standard-Setting Procedure**
8. **Evaluate the Process**
9. **Share the Recommendations**

STEP 1 *Assemble the Team*

The standard setting process involves the creation of **a committee of stakeholders whose task is to review the data and establish the cut scores for teacher effectiveness categories.** Committee members may include administrators, teachers, union representatives, and appropriate community members. Representatives from special populations should be included as well. The committee is formed at the district level to define cut scores that will be used across the school district. There may be reasons to apply this process to building data with building-based teams of educators and stakeholders.

- District leadership identifies the facilitator and the roles of committee representatives.
- Identify the facilitator who is trained in the data, the collaborative inquiry process, and the standard setting procedure. (See Appendix A)
- Assign committee representative to appropriate groupings, based on grade level, content area, and relevant considerations.

STEP 2 *Document the Process*

Establish the agenda, meeting norms, and schedule for the committee (See Appendix B).

Define the rules for the data to be included or excluded from the data sets. For example,

- student attendance rates,
- classroom enrollment,
- demographics,
- missing data,
- extenuating circumstances

Also consider how to handle incongruent data, such as how to round cut scores.

Clearly describe the selected standard setting procedure and the forms that will be used to collect data from the committee members.

Test the standard setting method using local data as a trial run for how the committee will proceed with this work.

STEP 3 *Establish Data Sources for Student Growth*

To implement the standard-setting process, the district will need to intentionally identify the data sources that will be used for the student growth component of the evaluation. The number and type of data sources utilized should consider current requirements of the Michigan legislation, but ultimately remains a district decision. Recommended practice would be to **include multiple measures of student growth with procedures designed to assure that the measurement of growth is comparable and consistent across grade levels and content areas.**

Table 1 provides a framework for summarizing the data sources, including consideration of grade level, content or essential/power standards, type of gain or growth score, and the instructional interval.

The instructional interval is an important consideration for such instructional time factors as semester classes or the timeframe of the evaluation benchmark periods. It is important to consider time to teach and opportunity to learn when comparing a teacher's growth data to criteria or norms that are developed based on an annual time frame. If, for example, a teacher has only had one semester to teach an advanced placement course that is used for student growth and the advanced placement test was developed based on norms for students who had one year of instruction, this would place the teacher at a disadvantage in showing comparable achievement.

TABLE 4.1 SUMMARIZING DATA SOURCES FOR GROWTH MEASUREMENT

Data Sources	Grade Level/s	Content	Type of Gain or Growth Score	Instructional Level
State Assessment Data				
District Purchased Assessment				
District Developed Assessment				
Classroom Assessment				
Collection of Evidence ***(e.g., portfolios, IEP goals)***				

STEP 4 *Clarify the Selected Model of Growth Assumptions*

According to Michigan legislation, growth is defined as the difference in student achievement measured at two points in time. There is no discussion of basing student growth on trajectories toward proficiency. **When implementing a standard setting process, the district should clarify assumptions about student growth and the explicit assessments to be used in this evaluation process.** By defining assumptions about growth in the standard setting process, there is the opportunity to differentiate the measurement of growth to address individual student needs. Clarification of assumptions about growth are integral to student learning objectives in teacher evaluation. Table 2 provides examples of assumptions about student growth that may need to be clarified with committee members.

TABLE 4.2 EXAMPLES OF ASSUMPTIONS ABOUT STUDENT GROWTH

Growth Models	Description
Catch Up Growth	Student is not at benchmark and needs to make catch-up growth to get to benchmark.
Keep Up Growth	Used to be known as the "bubble students". The achievement is near to benchmark.
Move Up Growth	Students are at or above benchmark and can be challenged to improve or move up to higher levels of achievement.

Adapted from Fielding, Lynn, Kerr, Nancy, and Rosier, Paul. *Annual Growth for All Students Catch-Up Growth for Those Who Are Behind.* The New Foundation Press, 2007.

STEP 5 *Write Performance Level Descriptions for Student Growth*

The performance level descriptions **define the parameters of student growth measurement that are then referenced by the team to determine data-based cuts for the categories of teacher effectiveness.** Table 4.3 (see page 69) provides examples of growth performance level descriptions that may vary based on the data type. The district team will need to consider their own data to develop meaningful performance level descriptors for your context. Remember, it is the task of the standard setting committee to assign data points, based on scores, percentages, percentiles, etc., to these descriptors.

STEP 6 *Train Committee Members*

Train the committee members on the goal of the committee. **They are there to set the cut scores for teacher effectiveness ratings based on district identified data using the performance level descriptors as the criteria.** Work with committee members to set aside assumptions and to focus on the purpose, tasks and outcomes. Train the committee members on the data reports they will be reviewing, building assessment literacy. Train the committee members on the standard setting protocol. Support the committee members with good facilitation using collaborative norms (See Appendix B).

TABLE 4.3 EXEMPLAR PERFORMANCE LEVEL DESCRIPTIONS BASED ON DIFFERENT DATA SOURCES

Data Source	Ineffective	Minimally Effective	Effective	Highly Effective
State Assessment Data *(Example of performance level description for Student Growth Percentile data)*	Student growth in the ineffective category is defined by MDE as...	Student growth in the minimally effective category is defined by MDE as...	Student growth in the effective category is defined by the MDE as….	Student growth in the highly effective category is defined by MDE as…...
District Purchased Assessment *(Example of Norm-Referenced Growth Data or Gain Scores)*	Student growth in the ineffective category is defined by parameters of an instructional interval (e.g., one year) and **decline, no, or minimal test growth** that is below district identified norms or standards. *(*Considering confidence intervals or standard error)*	Student growth in the minimally effective category is defined by parameters of an instructional interval (e.g., one year) and **marginal** test growth that is below district identified norms or standards. *(*Considering confidence intervals or standard error)*	Student growth in the effective category is defined by parameters of an instructional interval (e.g., one year) and **meets** test growth that is consistent with district identified norms or standards. *(*Considering confidence intervals or standard error)*	Student growth in the highly effective category is defined by parameters of an instructional interval (e.g., one year) and test growth **exceeding** district identified norms or standards. *(*Considering confidence intervals or standard error)*
District Developed Assessment *(Example of Common Assessment or Mastery of Standards)*	Student growth in the minimally effective category is defined by **mastery of few** grade level standards at grade expectancy as defined by a local benchmark assessment.	Student growth in the minimally effective category is defined by **mastery of some** grade level standards at grade expectancy as defined by a local benchmark assessment.	Student growth in the effective category is defined by **mastery of several** grade level standards at grade expectancy as defined by a local benchmark assessment.	Student growth in the highly effective category is defined by **mastery of most** grade level standards at grade expectancy as defined by a local benchmark assessment.
Classroom Assessments *(Example of Student Learning Targets/Objectives/ I Can Statements)*	Student growth in the ineffective category is defined by **mastery of few** learning objectives.	Student growth in the minimally effective category is defined by **mastery of some** learning objectives.	Student growth in the effective category is defined by **mastery of several** learning objectives.	Student growth in the highly effective category is defined by **mastery of most** learning objectives.
Collections of Evidence *(Examples may include Portfolios, Capstone Projects, IEP Goals)*	The **evidence is not present to specific IEP content area goals** or IEP goals do not include content area objectives. Students are unresponsive. Appropriate supports are not provided to the students.	**Evidence of progress is minimal in relation to IEP content area goals**, or IEP goals are not related to content area objectives. Students show little or no evidence of performance of IEP related goals; students perform tasks in a limited range of contexts, tasks are not meaningful or are not age-appropriate, failure to use appropriate supports.	**Evidence of progress demonstrates mastery of several IEP objectives**; IEP goals are relevant to content area objectives. The students show some evidence of performance of goals, in a limited variety of settings with opportunity for some interactions with peers, uses age-appropriate materials to perform some meaningful tasks with real-world applications, uses some appropriate supports.	**There is ample evidence of mastery of IEP goals**, which are clearly related to content area objectives. The student shows considerable evidence of performance related to goals, performs tasks in a variety of settings, engages in social interaction with a diverse range of age-appropriate peers, uses age-appropriate materials to perform meaningful tasks in a real-world context, and consistently uses appropriate supports.

STEP 7 *Implement the Standard Setting Procedure: The Vertical Scaling Method*

One recommended method of standard setting is based on the concept of vertical scaling. **Vertical scaling provides the opportunity to establish a system of cut scores across grade levels within the content area.** In this method the standard setting panel is comprised of representatives from adjacent grades (e.g., K - 1 - 2) who are knowledgeable in the content area and method of assessment for measuring growth. Ideally the vertical scaling teams would have grade level representatives in three year clusters, such as Grades K - **1** - 2; 1 - **2** - 3; 2 - **3** - 4, etc. The bolded grade is the primary focus of the standard setting.

Benefits to this method include the **opportunity for teachers to vertically align and identify non-negotiable standards and expectations for each grade level.** Consider beginning with a limited and specific amount of standards (e.g., 5 - 10 non-negotiable standards within the grade level for that content area) that can be refined in future iterations. For contexts of special education or personal curriculum, adjustments can be made to the amount of standards, which may be 1 - 5 standards).

Even with a rigorous standard setting process, conflicts and inconsistencies may occur for a variety of reasons that make this challenging work. These challenges have been recognized in the research on standard setting with growth measures.

SEVEN STEPS TO IMPLEMENT THE VERTICAL SCALING METHOD

1. **Prepare the data by grade level and content area.** Collect historical and trend data for up to 3 years, if available. Organize the data by year, grade and content area.

2. **Assign participants to grade level and content groupings, including adjacent grades.**
   (Grades K - **1** - 2; 1 - **2** - 3; 2 - **3** - 4, etc.).

3. **Develop the Performance Level Descriptors.**
   Establish the performance level descriptors, considering the historical data. Where evidence are not available, rely on generalization from other measures and discussions with content experts and stakeholders to shape the performance model.

4. **Train the participants on the data they will be reviewing, norms, and the standard setting process.**
   Review the performance level descriptors in detail to assure all participants are clear on the criteria for setting cut scores.

5. **Conduct the Standard-Setting Sessions with Cross Grade Participants.**
   Present the historical data in multiple rounds of the standard-setting procedure, cross grade or cross subject panels or in meta-panels.
   If cut scores are to be articulated across grades, it seems reasonable that the cut scores for a given grade be considered by individuals with strong interests in the performances of students in the adjacent grades. Where possible, all grade review should be included for at least one round, at or near the final round of review.

6. **Individual Setting of Cut Scores.**
   The task is for the participant to answer the question, *"Is it likely that the minimally qualified teachers at this category (e.g., Effective) would have a median score in this range?"* Participants are given data sets with scores vertically scaled from lowest grade to highest grade and from lowest score to highest score. Each participant independently reviews the data and marks the cut points, applying the performance level descriptors to the data.

TABLE 4.4 EXAMPLE OF PARTICIPANT RATING CHART FOR SETTING CUT SCORE USING GROWTH PERCENTILES

	Growth Percentile	Ineffective	Minimally Effective	Effective	Highly Effective
Grade 4	90 - 100				
	80 - 89				●
	70 - 79				
	60 - 69				
	50 - 59			●	
	40 - 49				
	30 - 39		●		
	20 - 29				
	10 - 19	●			
	0 - 9				

The chart is an example of how to organize the data ranges so that the participant would mark the score point or cut score for each category. For example, with fourth grade data, this participant set a cut score of 50 Growth Percentile as Effective.

7. **Summarize each participant's ratings and provide them with the feedback.**
   The purpose of this feedback is for the participants to see their agreement, to consider the impact, to clarify understanding of the data or performance level descriptors and to reach consensus. If discrepancies continue to exist, consider repeating the process. Another option would be to submit the recommendations for independent review as described below.

TABLE 4.5 EXAMPLE OF SUMMARY CUT SCORES FROM INDIVIDUAL PARTICIPANTS

Participant	Ineffective	Minimally Effective	Effective	Highly Effective
1	20 - 29	40 - 49	60 - 69	90 - 100
2	10 - 19	30 - 39	50 - 59	80 - 90
3	10 - 19	40 - 49	50 - 59	90 -100
4	60 - 69	70 - 79	80 - 89	90 - 100
5	50 - 59	60 - 69	70 - 79	90 - 100

As part of providing the participants with feedback, consider providing data that demonstrates the impact, based on recommended cut points. For example, provide them with the number or percent of teachers who would fall within each category based upon their cut scores. This additional review of the impact data will inform them as whether their recommendations will lead to desired outcomes when put into practice.

STEP 8 *Evaluate the Process*

Ask the committee members to provide reflection and feedback on the process and outcome of the standard setting. The information is essential to check for validity and to demonstrate the degree of support for the cut scores.

Ask them to answer the questions:

1. **Was it reasonable?**
2. **Can it be replicated?**

STEP 9 *Share the Recommendations*

Prepare a summary report to be shared with district constituents that includes the following components:

- Brief overview of the standard setting process
- Committee members involved in the standard setting
- Data and impact
- Cut scores for effectiveness categories
- Revisions per state law
- Include recommendations for future review committees

When the process is completed, the standard setting process may yield a rubric that includes the data points to be used for categorizing student growth data. An example of what this rubric might look like is demonstrated in Table 4.6:

TABLE 4.6 EXAMPLE OF STUDENT GROWTH RUBRIC FOR TEACHER EFFECTIVENESS CATEGORIES

Ineffective	Minimally Effective	Effective	Highly Effective
Less than X% of students demonstrated adequate growth on the ABC assessment. The standard setting team, based upon trend data at the district/grade or building level, will determine growth on an annual basis using three years of the most recent data (if available).	**Between X% and X% of students demonstrated adequate growth** on the ABC assessment. The standard setting team, based upon trend data at the district/grade or building level, will determine growth on an annual basis using three years of the most recent data (if available).	**Between X% and X% of students demonstrated adequate growth** on the ABC assessment. The standard setting team, based upon trend data at the district/grade or building level, will determine growth on an annual basis using three years of the most recent data (if available).	**At least X% of students demonstrated adequate growth** on the ABC assessment. The standard setting team, based upon trend data at the district/grade or building level, will determine growth on an annual basis using three years of the most recent data (if available).

FOLLOW UP WITH REVIEW AND ADJUSTMENTS

The cut scores may be reviewed and revised by officials or in situations in which the standard setting was disparate, a third party must make the final determinations. Furthermore, as there are changes with legislation, assessments, or observed data trends, it would be appropriate to revisit the standard setting process to make the needed adjustments to cut scores.

STANDARD SETTING PROTOCOL FOR EDUCATIONAL SETTINGS THAT ARE BASED ON COLLECTION OF EVIDENCE

Examples of settings in which achievement measures are based on collection of evidence may include: Alternative Education Schools, Career Technical Schools, Special Education Center Programs, Co-Op Programs, Special Disciplines, e.g., the Arts.

Examples of collection of evidence may include Portfolio Assessment Systems, IEP Goals, or other Performance data. The preferred method of standard setting with collections of evidence is the Generalized Holistic Method. In this method, the collections of evidence are scored by established criteria along essential dimensions that align to designated standards. Each collection of evidence is then independently rated as meeting performance level descriptions for the teacher effectiveness categories. Cut scores are based on mid-points between adjacent categories.

General steps to take with collection of evidence are described:

1. **Identify the artifacts or evidence** to be included.
2. **Review alignment of the artifacts/evidence to the state content standards.**
3. **Establish a scoring criteria for the collection of evidence**, appropriate to the skills and instructional objectives/benchmarks for the population or discipline.
4. **Create a rubric** for the scoring of the collection of evidence.
5. **Prepare the Performance Level Description** based on your data sources and scores.
6. **Follow a Generalized Holistic Method of Standard Setting.**
    a. **Organize the collection of evidence** by content area/grade range.
    b. **Establish rules for data to include/exclude and how to handle score incongruities**, e.g., rounding of scores.
    c. **Each collection of evidence is reviewed by 3 raters.**
    d. **Each participant reviews up to 8 collections of evidence,** making independent ratings. Using the rubric scores of the collections of evidence to establish cut points for effective, minimally effective, highly effective, and ineffective categories. (See Appendix C)
    e. **Provide data analysis and feedback to the participants,** including impact data.
    f. **Calculate the cut scores** based on mid-points between adjacent categories.
    g. **Facilitate discussion on the results,** emphasize variability in ratings.
    h. **Arrive at consensus or defer for independent review.**
7. **Assess participant ratings.**
8. **Summarize and report out.**

References

SECTION FOUR: Standard Setting for Student Growth

Cizek, Gregory J. and Bunch, Michael B. *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Sage Publications, 2007.

Fielding, Lynn, Kerr, Nancy, and Rosier, Paul. *Annual Growth for All Students Catch-Up Growth for Those Who Are Behind*. The New Foundation Press, 2007.

Garmstom, Robert J. and Wellman, Bruce M. *The Adaptive School: A Sourcebook for Developing Collaborative Groups*. Rowland & Littlefield Pub., 2013.

How to Design and Agenda for an Effective Meeting. Harvard Business Review. Taken 4/4/2016 from: **https://hbr.org/2015/03/how-to-design-an-agenda-for-an-effective-meeting** .

Protocols for Question Types in Group Facilitation. School Reform Initiative. Taken 4/4/2016 from: **http://www.nsrfharmony.org/free-resources/protocols/a-z** .

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/** .

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.

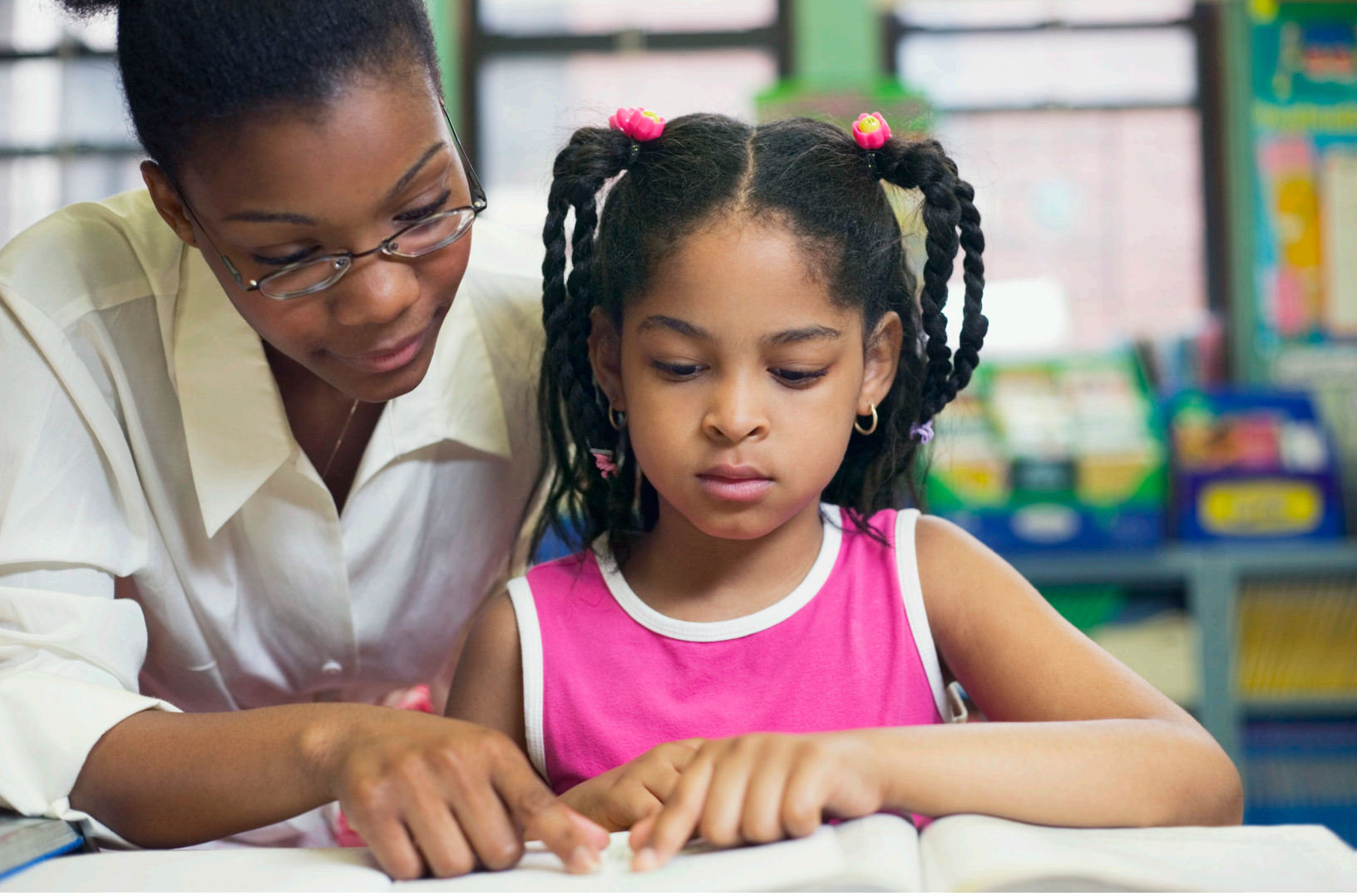
Terms for Statistics and Measurements	Definition	Answers the Question	Pros in Growth Measurement	Cons in Growth Measurement
Mean	Represents the arithmetic average of scores. It is a measure of central tendency.	What is the average gain for the data on hand?	• Easy to calculate. • Can be used when identifying 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.



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