Measuring Student Growth:

A Practical Guide to Educator Evaluation

WAYNE RESA
Leading... Learning for All

SERVICE
LEADERSHIP
COLLABORATION
EXCELLENCE

SECTION ❷:

Developing and Selecting Assessments of Student Growth for Use in Teacher Evaluation Systems

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

New legislation requires districts to establish clear methods whereby the measurement of student growth will be utilized as an important aspect of determining an educator's effectiveness through a rigorous evaluation process. The assessments utilized for the purpose of measuring student growth must be nationally normed or locally adopted assessments aligned to state standards, or based on achievement of individualized education program goals. In this section of the guidance document, considerations are provided to help districts develop and/or select assessments for purposes of supporting the evaluation process.

CRITERIA FOR SELECTING ASSESSMENTS

When selecting assessments for the purpose of measuring student growth, educators should consider the components of the assessment as well as the validity of its intended use. Analyzing the components of the assessment should include two major considerations:

1. **The assessment must align to content standards and learning objectives that will be taught during the interval of instruction that will be used** for assigning an effectiveness label to the educator. When examining assessments for alignment, content teams should consider the following:
    - Items on the test should reflect the essential content standards for the course/grade.
    - No items on the assessment should reflect standards that were not part of the instructional plan for the course/grade.
    - The number of items dedicated to specific standards should mirror the instructional emphasis intended for the course/grade. For example, if the curriculum in a reading course focuses 50% of the instructional time on reading comprehension, then 90% of the assessment items should not be measuring knowledge of vocabulary.
    - The items on the assessment should represent the full range of cognitive thinking required in the course/grade. For example, if the curriculum in a third grade math class requires students to solve word problems and explain reasoning, then the assessment should require students to demonstrate this same level of reasoning.
2. **The assessment must provide enough item flexibility so that both high- and low-achieving students may adequately demonstrate their understanding of the content.** In order to accurately determine student growth, the assessment must provide enough range to demonstrate learning on both ends of the ability continuum. When selecting assessments that include adequate range, educators should consider the following:
    - All students should be able to demonstrate growth on the assessment.
    - The items should include basic knowledge and skills, as well as items that will challenge the highest performing students.

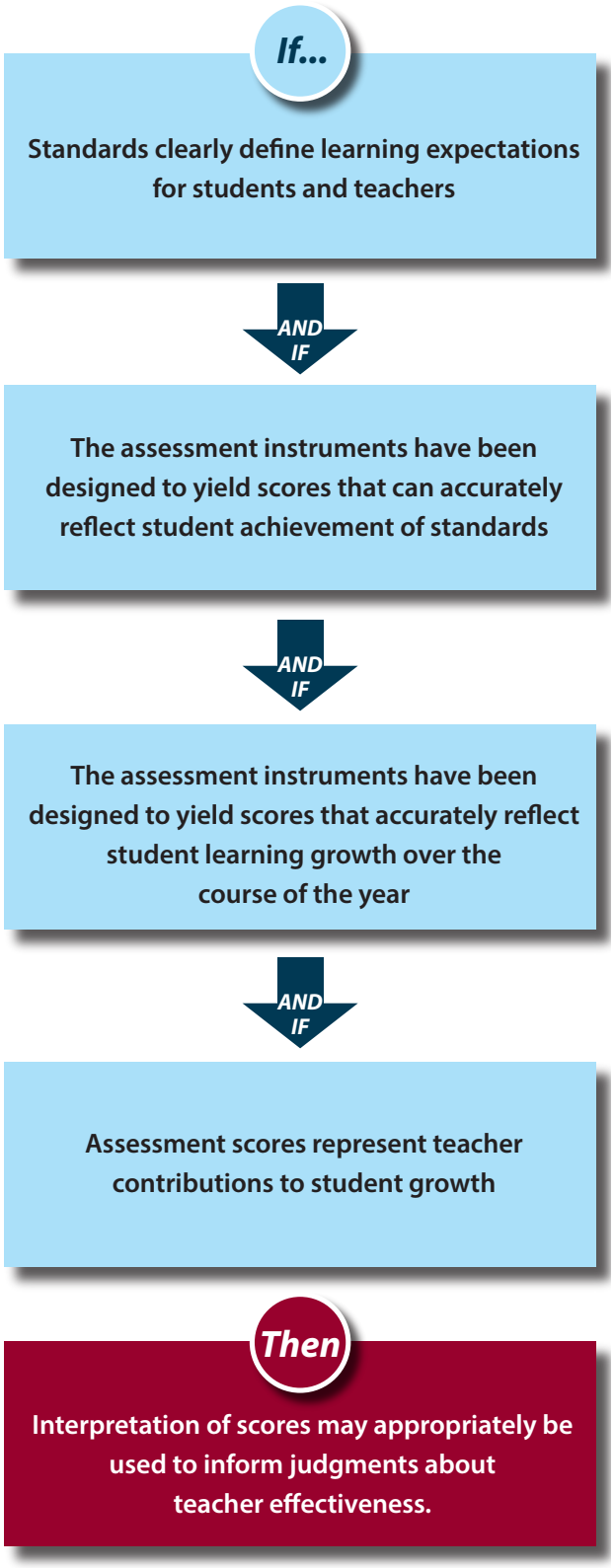
JUSTIFYING THE USE OF ASSESSMENTS IN TEACHER EVALUATION

The primary concept to be considered when selecting assessments for teacher evaluation is validity. **Validity is the extent to which an assessment measures what it is intended to measure and provides sound evidence for use as a part of the evaluation process.** As districts evaluate the purpose of assessments, they must consider validity.

Validation involves justifying the interpretation of scores as a determination of student growth over time and ensuring assessments are designed to report teacher effectiveness. Specific conclusions and decisions will be made on the basis of assessment performance; therefore, the following assurances regarding assessments should be verified:

- The purpose of the assessment is to measure student learning and growth over time.
- The standards of learning are clearly defined for both students and teachers.
- The assessment is of high quality, as measured by specific criteria and assurances. (Table 2.2)
- Student assessment scores accurately and fairly measure what students have learned in a given time period.
- Students' growth scores (based on assessments) accurately and fairly appraises the contributions of individual teachers.

FIGURE 2.1 JUSTIFYING THE USE OF ASSESSMENTS IN TEACHER EVALUATION



Adapted from Herman J.L., Heritage, M., and Goldschmidt, P. (2011). *Developing and Selecting Assessments of Students' Growth for Use in Teacher Evaluation Systems.*

ASSESSMENT INVENTORY

The first step in selecting assessments for the purpose of measuring student growth is to **determine which assessments are currently utilized throughout the district.** The assessment inventory is a tool designed to assist districts with the inventory process.

It is critical that all stakeholders understand the following:

- why specific assessments have been selected,
- the purpose the assessment serves,
- and how the district will use the data to impact student growth.

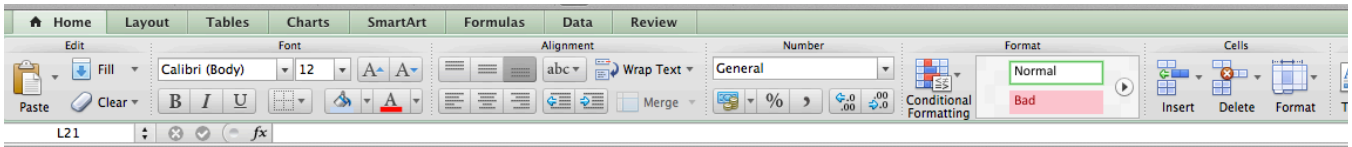
District and school leaders are encouraged to use and adapt this inventory to facilitate conversations about assessments with teachers, parents, and their local school board. This tool is also intended to identify gaps that exist across the district, in which an appropriate measure of student growth will need to be determined. Upon completion of this district-wide inventory, decisions can be made regarding which assessments are appropriate and should be used as a component of the evaluation system.

The assessment inventory is available in two electronic formats – Excel and Google Sheets. There are a variety of elements on the inventory that may be refined or adapted to provide the necessary data for collaborative district decisions. Elements from the assessment inventory include:

- Test Name
- Frequency of testing (i.e. monthly, quarterly, annually)
- Content standard alignment
- Intended purpose of the assessment
- Data collection methods and process for storing the data
- Identification of staff that will use the data
- Perceptions related to the appropriate use of the assessment
- Description of the processes in place to use the data
- Recommended reports to display student results
- Time between test administration and results provided to users of the data
- Intended decisions to be made from the data
- Information related to the assessment vendor/ developer
- Annual cost and funding sources

As district teams review these elements in collaborative teams, recommendations may be shared for the development and/or elimination of district assessments for the purpose of measuring student growth. If there are courses/subjects in which an assessment is not available or appropriate for use in a student growth model, educators should consider using criteria for the development of new assessments.

TABLE 2.1 ASSESSMENT INVENTORY



	A	B	C	D	E	F	G	H	I	J	K
1	Test Name	Entity Requiring Assessment	Which Students Are Eligible / Required to take this Assessment?	Grade/s Tested	Course / Subject Tested	Type of Assessment	Item Types	Type of Adminsitration	Accomodations	Number of Years Used by District	Expected Testing Time
2	M-Step	State	All	3-8, 11	Core	Summative	Varies	Online	yes	2	6+ hours
3											

CRITERIA FOR QUALITY ASSESSMENTS

Quality is a keystone of any assessment that will be used for the purpose of student growth. Districts must commit to **ensuring that all assessments associated with teacher evaluations are of high quality and have been verified as such through a defined process.** This process should be applied to vendor assessments, as well as those that have been locally developed. The following checklist may be used as a means to evaluate existing assessments for the purpose of educator evaluation.

TABLE 2.2 CRITERIA FOR QUALITY ASSESSMENTS

Criteria	Possible Evidence	Verified
Learning expectations are clear and reflect a progression (at minimum, the span of a grade level).	Sensitivity reviews by district team (e.g., curriculum director, department chair) Data analysis indicate consistent trends in achievement without significant fluctuations	
Specifications/blueprint for the assessment reflects the breadth and depth of learning expectations.		
Assessment tasks and items are designed to be accessible and fair for all students.		
Assessments are designed to accurately measure the growth of individual students.		
Assessments are designed to be instructionally sensitive to the standards being taught.		
Assessment items and tasks are consistent with the specifications/blueprint and are designed using item quality guidelines.	Item Quality Checklists *(included)*	
Assessment design, administration, and scoring procedures are likely to produce reliable results	Review of administration and scoring procedures (e.g., building principal, district team)	
Scores derived from the assessment are sufficiently precise and reliable.	Psychometric Analysis *(see Section 3: Measuring Student Growth)*	
Cut scores for student progress are justified.	Standard Setting Process *(see Section 4: Standard Setting for Student Growth)*	

STEPS FOR CREATING QUALITY LOCAL ASSESSMENTS

The criteria for quality assessments are recommended for justifying the decisions made in which student growth measures are attributed to teacher effectiveness. This process is essential for both locally developed and commercial assessments. Districts that elect to create new local assessments should employ processes and protocols to ensure consistent and quality results.

#	Step	Description	Review
1	**Assemble a Team**	Choosing a team of teachers that have knowledge in the content area will help verify that the assessment has content validity.	
2	**Establish the Purpose of the Assessment**	Clearly define the purpose of the assessment as a student growth measure.	*Michigan Assessment Consortium (MAC) Module 3:* **Determining the Outcome of the Common Assessment**
3	**Unpack/Deconstruct the Standards to Create Clear Learning Targets**	• **Important** – the standard or target is important to the discipline and/or important to the students' future. • **Leverage** – the standard or target has leverage, it represents enabling knowledge and/or skills. • **Appropriate** – the standard or target has been selected because it reasonably represents what the student can do at a given point in time (relative to the purpose of the assessment (e.g. pre-assessment, final end-of-course assessment). • **Clear** – the standard has been deconstructed so that it can be turned into a learning target and the standard is in language the student understands. • **Measureable** – the standard has been turned into a learning target that contains a verb that accurately represents the type of thinking, performance or disposition desired. • **Balance** – the collection of targets selected for one common assessment, or for a collection of common assessments for one grade and one academic year represent a variety of target types. For example, we are not unwittingly selecting a majority of targets that represent knowledge only.	*Michigan Assessment Consortium (MAC) Modules 4 and 5:* ***Identifying Learning Targets for the Common Assessment*** ***Matching the Assessment Methods to the Learning Target***

#	Step	Description	Review
4	**Ensure Access to the Content for ALL Students**	Provide appropriate accommodations that guarantee access to the assessment for students with special needs.	*Michigan Assessment Consortium (MAC) Module 6:* **Assessing Students with Special Needs**
5	**Create a Test Blueprint**	Design the assessment using the content and rigor of the learning targets. Choose the most appropriate assessment methods and quantity of items that will best reflect the standards taught for the instructional period. Consider the standard alignment, depth of knowledge (DOK), assessment method, item source, item sampling, and testing time.	*Michigan Assessment Consortium (MAC) Module 7:* **Writing the Test Blueprint**
6	**Select Items or Develop New Items**	Review content from item banks and make appropriate selections, based on the blueprint criteria. If there are no existing items to choose from, use quality guidelines for item creation.	*Michigan Assessment Consortium (MAC) Modules for Item Development 8-15:* **Module 8: Writing Selected Response Items – Part 1** **Module 8: Writing Selected Response Items – Part 2** **Module 9: Writing Constructed Response Items** **Module 10: Writing Performance Assessments** **Module 11: Using Portfolios to Assess Students** **Module 12: Developing and Using Scoring Guides and Rubrics** **Module 13: Editing the Draft Assessment Items** **Module 14: Detecting and Eliminating Bias and Distortion** **Module 15: Assembling the Assessment Instrument**

#	Step	Description	Review
7	Field Test Items	Evaluate the effectiveness of assessment items by analyzing real student data for item difficulty, item discrimination and bias. Analyze the reliability and validity of the assessment through psychometric principles and content expert reviews.	*Michigan Assessment Consortium (MAC) Modules for Field Testing 16-19:* **Module 16: Field Testing** **Module 17: Looking at Field Test Data** **Module 18: Reliability** **Module 19: Validity**
8	Determine Standards for Adequate Growth	Adhere to sound principles for standard setting and apply rigorous procedures to determine intervals of growth.	See section on Standard Setting for Student Growth
9	Assessment Review	Review the assessment periodically and revise to maintain alignment with standards and/or curriculum.	School/District Teams
10	Administer and Score the Assessment	Develop standardized procedures for the administration and scoring of the assessment. Establish procedures to ensure the validity of score interpretation as a measure of teacher effectiveness.	School/District Teams

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 TWO: Developing and Selecting Assessments of Student Growth for Use in Teacher Evaluation Systems**

Achieve. Student Assessment Inventory for School Districts. (2016) Retrieved March 23, 2016: **http://www.achieve.org/files/AchieveStudentAssessmentInventory.pdf**

Herman, J. L., Heritage, M., & Goldschmidt, P. (2011). *Developing and selecting assessments of student growth for use in teacher evaluation systems* (extended version). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Michigan Assessment Consortium. (2016). *Common Assessment Modules.* Retrieved on May 25, 2016: **http://www.michiganassessmentconsortium.org**

Ohio Department of Education. (2015). *Student Learning Objectives: A Guide to Using SLOs as a Locally-Determined Measure of Student Growth*. Retrieved on March 23, 2016: **https://education.ohio.gov/getattachment/Topics/Teaching/Educator-Evaluation-System/Ohio-s-Teacher-Evaluation-System/Student-Growth-Measures/Student-Learning-Objective-Examples/SLO_Guidebook.pdf.aspx**

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. Achieve/Aspen/Center of Assessment Policy Brief. Retrieved from: **http://www.achieve.org/files/TheRoleofInterimAssessments.pdf**

- **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

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