



NATIONAL COMPREHENSIVE CENTER
FOR **TEACHER QUALITY**

Measuring Student Growth in Non-Tested Subjects, Grades, and with At-Risk Student Populations

Lynn Holdheide

**Vanderbilt University/National Comprehensive Center
for Teacher Quality**

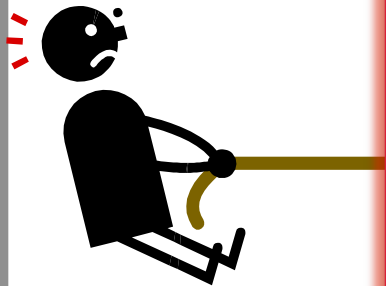
SEDL's Designing and Implementing Teacher Evaluation Systems Institute

July 27–28, 2011

LEARNING POINT Associates®
An Affiliate of American Institutes for Research®



What We Will Do Today



A Review and Discussion of

- Measuring Student Growth
- Pros and Cons of State and District Approaches
- Steps to Ensure Rigor



An expectation that the measures are

- Rigorous
- Across two points in time
- Comparable

U.S. DOE Priorities

- Increasing effective and highly effective teachers
 - Number and/or percentage
 - Retention and equitable distribution
- Method for determining and identifying effective and highly effective teachers
 - Must include multiple measures
 - Effectiveness evaluated, in significant part, on the basis of student growth
 - Supplemental measures may include, e.g., multiple observation-based instruments

Make an Argument

Student growth should not be a factor in teacher evaluation.

- General agreement that growth needs to be a component
- Significant challenges surface when determining **HOW** it can be done in a **FAIR** and **ACCURATE** manner and with limited **RESOURCES**

The Challenge: Measuring Contributions to Student Learning Growth for...

- Teachers of non-tested subjects (e.g., social studies, K-2, art, drama, band)
- Teachers of certain student populations and/or situations in which standardized test scores are not available or utilized
 - Teachers of students assessed on alternate assessments
 - Smaller teacher caseloads for some student groups (e.g., students with disabilities, English language learners).

Brainstorming Activity

How is student growth currently measured in non-tested subjects?

Race-to-the-Top Definition of Student Growth

- Student growth means the change in student achievement (as defined in this notice) for an individual student
 - **between two points in time**
- A state may also include other measures that are
 - **rigorous**
 - **comparable across classrooms (pg 11)**

Potential Definition

- **Rigorous**

- Rigorous measures may mean that *high expectations for student progress towards college- and career-readiness* are exhibited
 - Measure designed to measure students' mastery of grade-level standards for that subject.

Slide courtesy of Laura Goe, ETS

Potential Definition

- **Between two points in time**
 - May mean assessments that occur as close as possible to the beginning and end of a course, so that the maximum growth towards subject/grade standards can be shown.
 - Examples:
 - Pre- and post-test given in a particular course
 - Student portfolio demonstrating mastery of standards-based knowledge and skills (measured over time)

Slide courtesy of Laura Goe, ETS

Potential Definition

- **Comparable across classrooms** may have two interpretations:
 - The measures used to show students' growth for a particular subject are the same or very similar across classrooms within a district or within a state
 - The measures used in *non-tested* subjects and grades are as rigorous as those in the *tested* subjects and grades

Slide courtesy of Laura Goe, ETS

Brainstorming Activity

How is student growth currently measured in non-tested subjects?

- ☒ Rigorous
- ☒ Across two points in time
- ☒ Comparable

Range of State and District Approaches



- Existing measures
- Rigorous new measures
- Portfolios/products/performance/projects
- Student learning objectives
- School-wide or team-based value-added



Measures must be rigorous, between two points in time, and comparable across classrooms.

Measuring Teachers' Contributions to Student Learning Growth: A Summary of Current Models

Model	Description
Subject & grade alike team models	Teachers meet in grade-specific and/or subject-specific teams to consider and agree on appropriate measures that they will all use to determine their individual contributions to student learning growth
Pre- and post-tests model	Identify or create pre- and post-tests for every grade and subject
Student learning objectives	Teachers assess students at beginning of year and set objectives, then assess again at end of year; principal or designee works with teacher, determines success
School-wide value-added	Teachers in tested subjects & grades receive their own value-added score; <i>all other teachers get the school-wide average</i>

Existing Measures



Strengths of this Measure	Challenges for this Measure
<ul style="list-style-type: none">• Already exist• Teacher familiarity and use• Not creating additional assessments/work• Depending on type, could be formative in nature	<ul style="list-style-type: none">• Validity is a concern whenever a measure is used in a way that was not intended• Concern over content validity• Fidelity and/or standardization

Delaware/Tennessee, Rhode Island

- Assembled group of practitioners
- Tightly facilitated meetings
- Group recommended measures
- Expert panel approves measures

National RTI Center

- Progress Monitoring Tools
- Tiers I, II, and III
- <http://www.rti4success.org/chart/progressMonitoring/progressmonitoringtoolschart.htm>

New Measures



Strengths of this Measure

- Tests can be made to match specific grade/subject standards
- Assessments can be created to meet standards of validity and reliability
- Same assessment given across district/teachers

Challenges for this Measure

- More tests!
- Time and cost intensive approach
- Paper-and-pencil tests may not be appropriate as the sole measure, particularly in subjects requiring students to demonstrate knowledge and skills (art, music, etc.).
- Capacity to build valid and reliable assessments.

Hillsborough County, Florida

- Race to the Top Grantee
- Each course has a pre- and post-assessment
- Scores are averaged over a three-year period to determine teacher effectiveness

Use products/portfolio/performance/projects



Strengths of this Measure	Challenges for this Measure
<ul style="list-style-type: none">• Evidence of growth can be documented over time using performance rubrics• Portfolios and projects can reflect skills and knowledge that are not readily measured by paper-and-pencil tests	<ul style="list-style-type: none">• Training for inter-rater reliability• Logistical challenge for group raters• Ensuring rigor

Five New York districts participating in the AFT Innovation Project

- Like Delaware, teachers identify existing measures already used in classrooms
- Must develop pre-tests to establish knowledge and skills students need prior to project
- Panel of experts/practitioners will evaluate and approve measures

Student Learning Objectives



Strengths of this Measure	Challenges for this Measure
<ul style="list-style-type: none">Teachers benefit from direct involvement in assessing students' knowledge and skillsSLOs are applicable to all teachers, subjects and students (such as students with disabilities or English language learners)	<ul style="list-style-type: none">Comparability across classrooms will be problematicVery resource-intensive for principal or district personnelConcern over rigor

Austin, TX is piloting SLOs; Rhode Island is using SLOs in a new teacher evaluation system to establish teachers' contribution to student learning in both tested and non-tested subjects/grades

Evaluation System's Purpose/Goals

- Tendency to oversimplify this step
- Purpose should drive all decisions regarding
 - Measurement selection and weight
 - Evaluation format (e.g., frequency of observations, pre- and post-observation conferences)
 - Data collection needs
- **Higher stakes point to measures that are technically defensible (e.g., valid & reliable)**
- Improved teacher capacity points to measures that identify effective teaching practices

Goe, L., Holdheide, L. & Miller, T. (in press). *A Practical Guide to Designing Comprehensive Teacher Evaluation Systems*. Washington, DC: National Comprehensive Center for Teacher Quality.

Team Activity

What Is The Purpose?

As a team, determine **5** key purposes that best describe the goals of the teacher evaluation system.

Take Five!

Aligning Purpose(s) and Measures

Purpose of Evaluation of Teacher Effectiveness	Growth Models	Classroom Observation	Analysis of Artifacts	Portfolios	Teacher Self-reports	Student/parent ratings	Other
Determine whether a teacher's students are meeting achievement growth expectations	✓		✓				
Gather evidence for making contract renewal and tenure decisions	✓	✓					
Determine types of assistance and support a teacher may need	✓	✓	✓		✓	✓	
Determine whether a teacher's performance qualifies him or her for additional compensation or incentive pay (rewards)	✓	✓					
Gather information on a teacher's ability to work collaboratively with colleagues		✓		✓	✓		✓
Determine how students and parents perceive a teacher's instructional efforts				✓		✓	

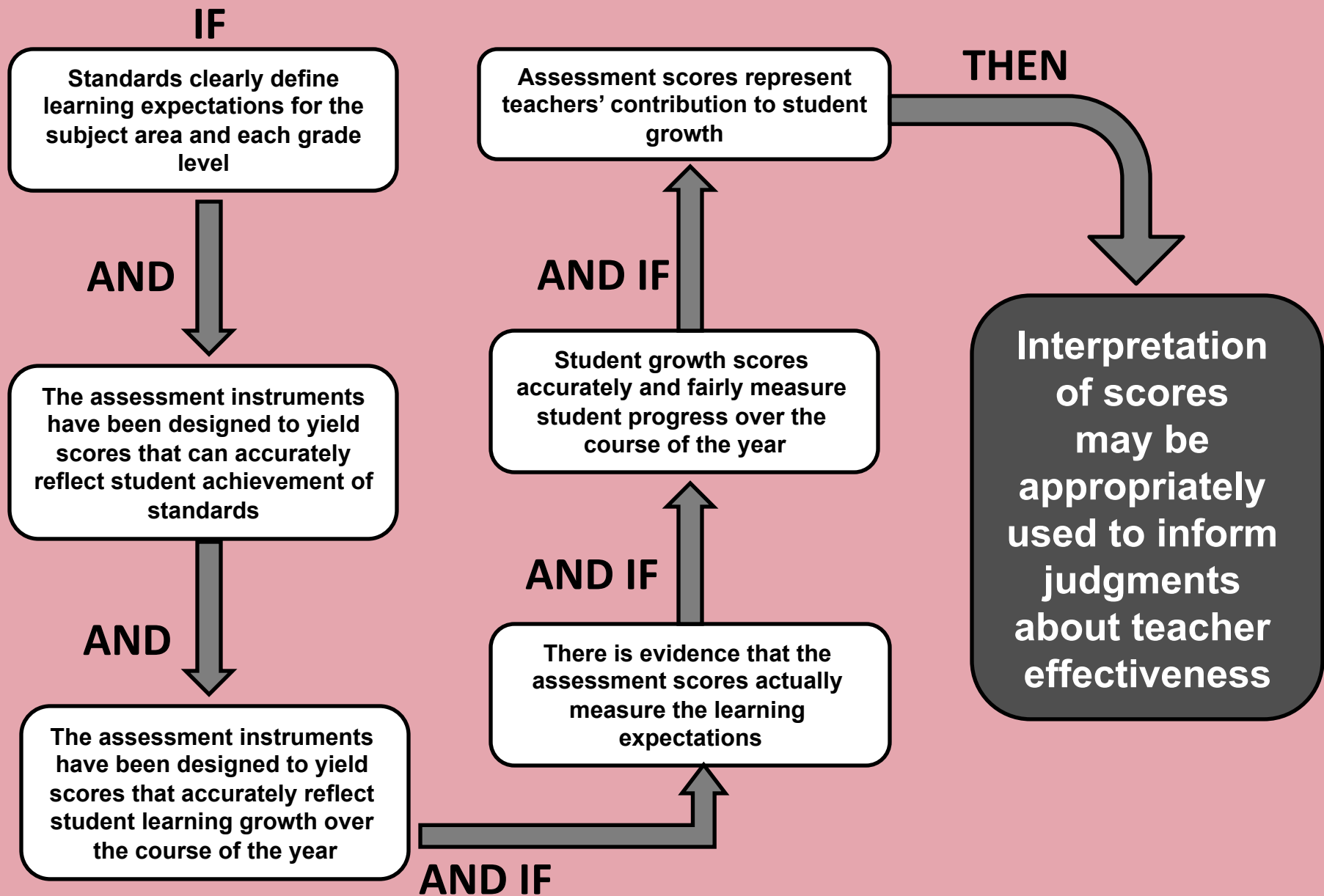
American Institutes for Research: Aligning Purposes and Criteria: Example Worksheet. From *A Practical Guide to Evaluating Teacher Effectiveness*, Little, Bell, and Goe. The National Comprehensive Center for Teacher Quality. April 2009. <http://www.tqsource.org/publications/practicalGuide.php> The guide is based on the TQ Center research synthesis *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis* by Goe, Bell, and Little (2008).

Evaluation System's Purpose/Goals

- Higher stakes point to measures that are technically defensible (e.g., valid & reliable)
- What are the implications on using measures of student growth?

The Process of Validity

- Currently little research
- Starts with content validity (defining the criteria and standards to be measured)
- Curricular and instructional strategies
- Measurement designed using psychometric properties meeting technical standards
- Validity verified by using multiple data points and comparing results to other measures
- Lengthy process!



Propositions that justify the use of these measures for evaluating teacher effectiveness. (Adaptation based on Bailey & Heritage, 2010, and Perie & Forte (in press)) (Herman, Heritage & Goldschmidt, 2011). Slide used courtesy of Margaret Heritage.

State Content Standards

Is there a consensus on the competencies students should achieve in this content area?



- Focus on proficiency: experts and practitioners define the knowledge, concepts, and skills students should acquire for each subject and grade level based on content **standards**
- Content standards form the basis on which measures can be either identified or developed

Factors to Consider

Can these measurements be applied to all grades and student populations?



- Vertically equated scales makes the achievement measure applicable to a broader range of grades and ability levels
- Students with disabilities and English language learners instructed on the general education curriculum assessed w/ same measures (with potential need for accommodations)
- Different measures for students on alternate standards
- IEPs as basis for student growth presents legal and other potentially contentious issues

Factors to Consider

What resource and human capacity limitations and strengths need to be factored into the decision on measurements?



- Does the district have the human capacity to implement these assessments with fidelity?
- What are the training needs?
- What type of resources are required to ensure implementation fidelity?

Brainstorming Activity

How is student growth currently measured in non-tested subjects?

☒ Reliability

☒ Validity

Identifying/Developing Assessments

- Set assessment goals and objectives
- Secure the assistance of assessment professionals
- Implement professional development
 - Training in how to
 - conduct assessments
 - interpret results
 - explain results to students/parents
 - use results to improve instruction
- Implement Assessments
 - Field test/pilot
- Evaluate results
 - Checks for reliability
 - Comparison/correlations to other measures

Model Highlight: Ensuring Rigor

Austin's Reach Program
includes a rubric for
determining the rigor of
teacher-created student
learning objectives
(SLOs)

Austin Independent School District

Student Learning Objectives:

- Teachers determine two SLOs for the semester/year
- One SLO must address all students, other may be targeted
- Use broad array of assessments
- Assess student needs more directly
- Align classroom, campus, and district expectations
- Aligned to state standards/campus improvement plans
- Based on multiple sources of student data
- Assessed with pre and post assessment
- Targets of student growth
- Peer collaboration

Slide courtesy of Laura Goe, ETS

Austin Reach Program: Rubric for Determining SLO Rigor (DRAFT)

Student Learning Objective Rigor Rubric

4 Exemplary	3 Proficient	2 Progressing	1 Does not meet standard
<u>Assessment</u> <ul style="list-style-type: none"> Variety of levels of questions (Beginning, Progressing, Proficient, Advanced) At least one very challenging question Sufficient number of items Grade level appropriate Extends and deepens knowledge Measures what is intended 	<u>Assessment</u> <ul style="list-style-type: none"> Variety of levels of questions (Beginning, Progressing, Proficient, Advanced) Sufficient number of items Grade level appropriate Measures what is intended 	<u>Assessment</u> <ul style="list-style-type: none"> Addresses 2 or 3 levels of questions Spread of questions is insufficient Grade level appropriate Mostly measures what is intended 	<u>Assessment</u> <ul style="list-style-type: none"> Addresses only 1 level of questions Insufficient number of questions Not grade level appropriate Does not measure what is intended
<u>Objective</u> <ul style="list-style-type: none"> Reflects a high need Yearlong objective Grade level appropriate Deepens and extends knowledge for all students 	<u>Objective</u> <ul style="list-style-type: none"> Reflects a significant need Yearlong objective Grade level appropriate 	<u>Objective</u> <ul style="list-style-type: none"> Addresses a need Yearlong objective Grade level appropriate 	<u>Objective</u> <ul style="list-style-type: none"> Does not address a need Not a yearlong objective Not grade level appropriate
<u>Growth Target</u> <ul style="list-style-type: none"> Addresses more than 75% of students Substantial growth expected (2 or more years) Students and teachers exceeding expectations 	<u>Growth Target</u> <ul style="list-style-type: none"> Addresses 75% of students (exceptions for sped, small classes, etc) Significant individual growth (at least one year) Pushes students and teachers to exceed typical expectations 	<u>Growth Target</u> <ul style="list-style-type: none"> Addresses fewer than 75% of students Moderate individual growth (less than one year) Students and teachers barely meet expectations 	<u>Growth Target</u> <ul style="list-style-type: none"> Does not address 75% of students Minor individual student growth (less than ½ year) Students and teachers do not meet expectations

Considerations

- Establishing and Maintaining Rigor
 - What assurances will the state employ to ensure rigor?
 - How will states ensure the goals are measureable, aligned to the state standards, and assessed using a quality measure (with fidelity!)?
- Valid and Reliable Measures
 - **Teacher A:** Uses a rigorous measure, develops high-level student goals, and assesses student achievement against the state standards
 - **Teacher B:** Develops a teacher-made assessment which may or may not be aligned to state standards and due to lack of standardization, outcomes are suspect

Considerations

- Training and Personnel
 - Teacher competencies in
 - Writing quality goals
 - Establishing baseline
 - Conducting assessments with fidelity
 - Establishing growth trajectory
 - Interpreting and using data
- Administrative Oversight
 - Time and subjectivity

Considerations

- Adequate Growth
 - Need to collect and analyze data over time
 - Teacher competencies
 - Students functioning at lower levels
- Comparability
 - Individualized goals
 - Determining quality

The Challenge: Determining How to Attribute Learning Gains to Teachers

- Should teachers be held to the same level of accountability if a student
 - Is only in classroom for a portion of the year?
 - Has a high rate of school absences?
 - Fails to complete assessments that will be used for determining teachers' contribution to student growth?
- Which teacher should be held accountable in a co-teaching situation?
 - Various co-teaching models make it difficult to evaluate teachers
- Involve teachers in the problem solving process!

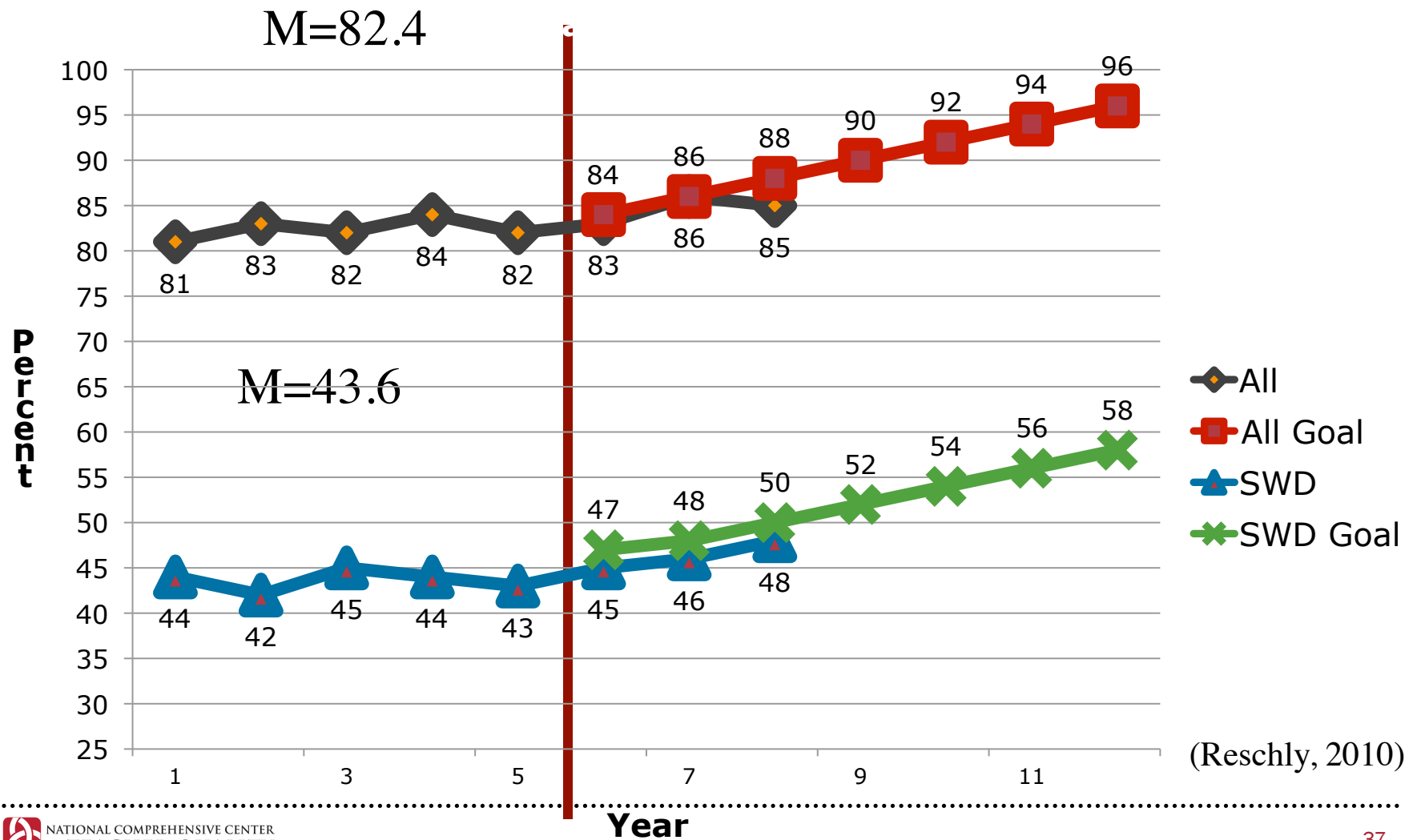
Factors to Consider

How will related personnel be factored into the system?



- “Caseload educators” (e.g., counselors, librarians, school psychologists) may be held accountable for all students, groups of students, and even educating other teachers
- Contribution of growth in educational successes and other types of outcomes (needs based)
 - Percent of students applying to college
 - Percent of students who attend school regularly
 - Percent of students who eat a healthy and nutritious lunch

Graduation Rates and Goals



School-Wide Value-Added



●Pros

- Prompt implementation
- Identical process
- Promotes collaboration and shared responsibility for all students

●Cons

- Questions concerning rigor and comparability when judgments are made on students whom teachers never taught
- Could devalue contributions those teachers make to student learning
- Provides no information about their effectiveness in teaching their subject matter

Considerations

- Partner with national and regional comprehensive centers.
- Engage stakeholders (teachers, administrators, parents, school board members, union representatives, business leaders, etc.) in decision-making processes early and often.
- If lacking grade-level and subject standards, adopt such standards.
- Conserve resources by encouraging districts to join forces with other districts or regional groups.

Considerations

- Consider whether human resources and capacity are sufficient to ensure fidelity of implementation.
- Develop a communication strategy to increase awareness and buy-in (FAQs on website, public meetings, news “blasts” to email subscribers).
- Establish a plan to evaluate measures to determine whether they can effectively differentiate among teacher performance.
- Examine correlations among measures.
- Evaluate processes and data each year and make needed adjustments.



Lynn Holdheide, Vanderbilt University

P: 615-477-1880

E-Mail: lynn.holdheide@vanderbilt.edu