

All kids learn



Northwest Evaluation Association

A Consumer's Guide to Using Tests for High Stakes Purposes

Nate Jensen, Ph.D.
Research Scientist, NWEA

October 2014

Two Questions

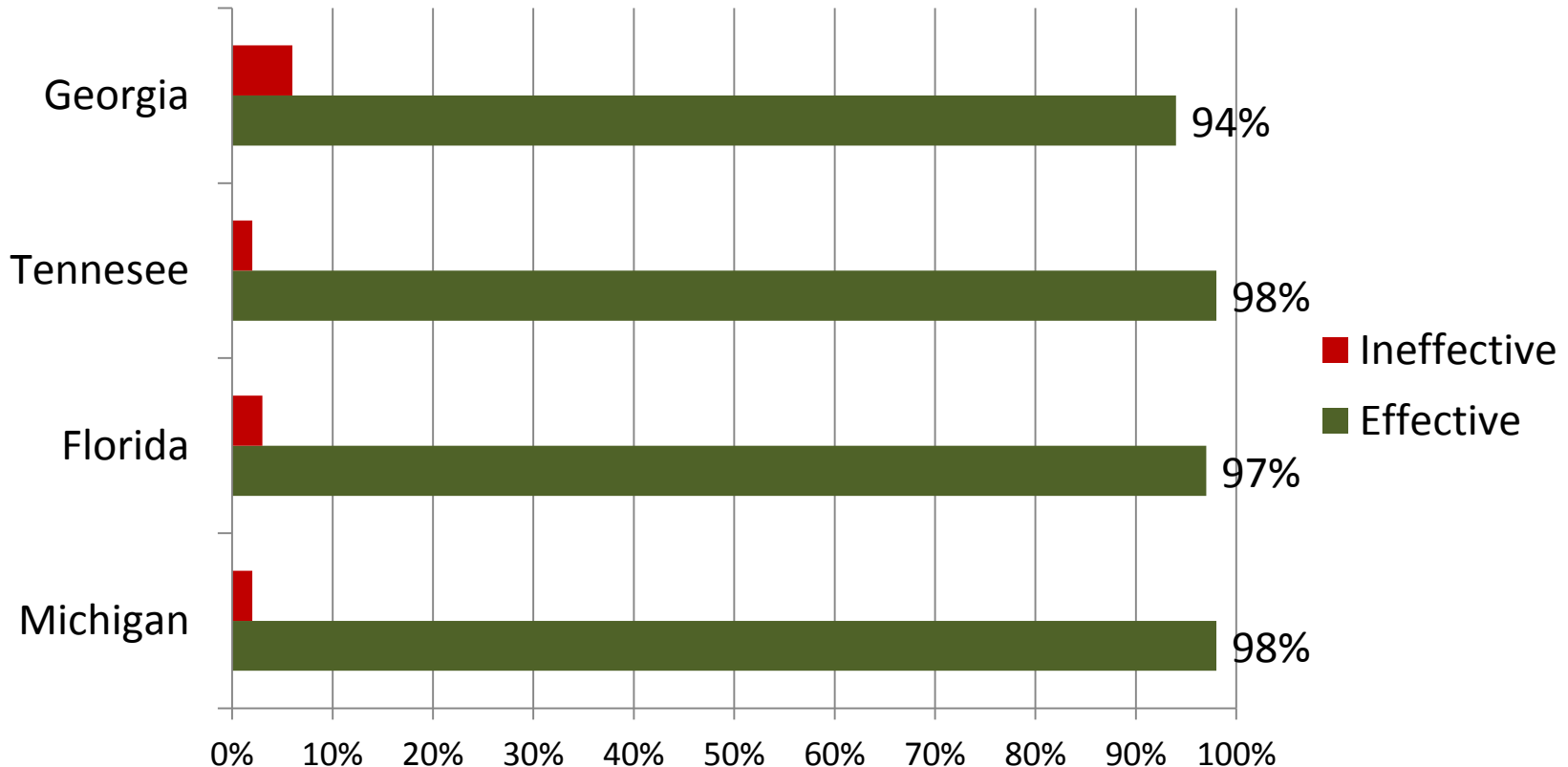
- How are your teachers responding to your school/district approach to evaluation?
- How easy/difficult have you found the Student Learning and Growth (SLG) goal setting process to be (and have you already written your SLG goals)?

Setting the Stage

- What is “high-stakes test use”?
 - Teacher/Principal evaluation
 - 3rd grade retention
 - Gifted and Talented Identification

Unfortunately, instructional use of student assessment results is becoming secondary – not perceived as the “stakes” on which we are most focused

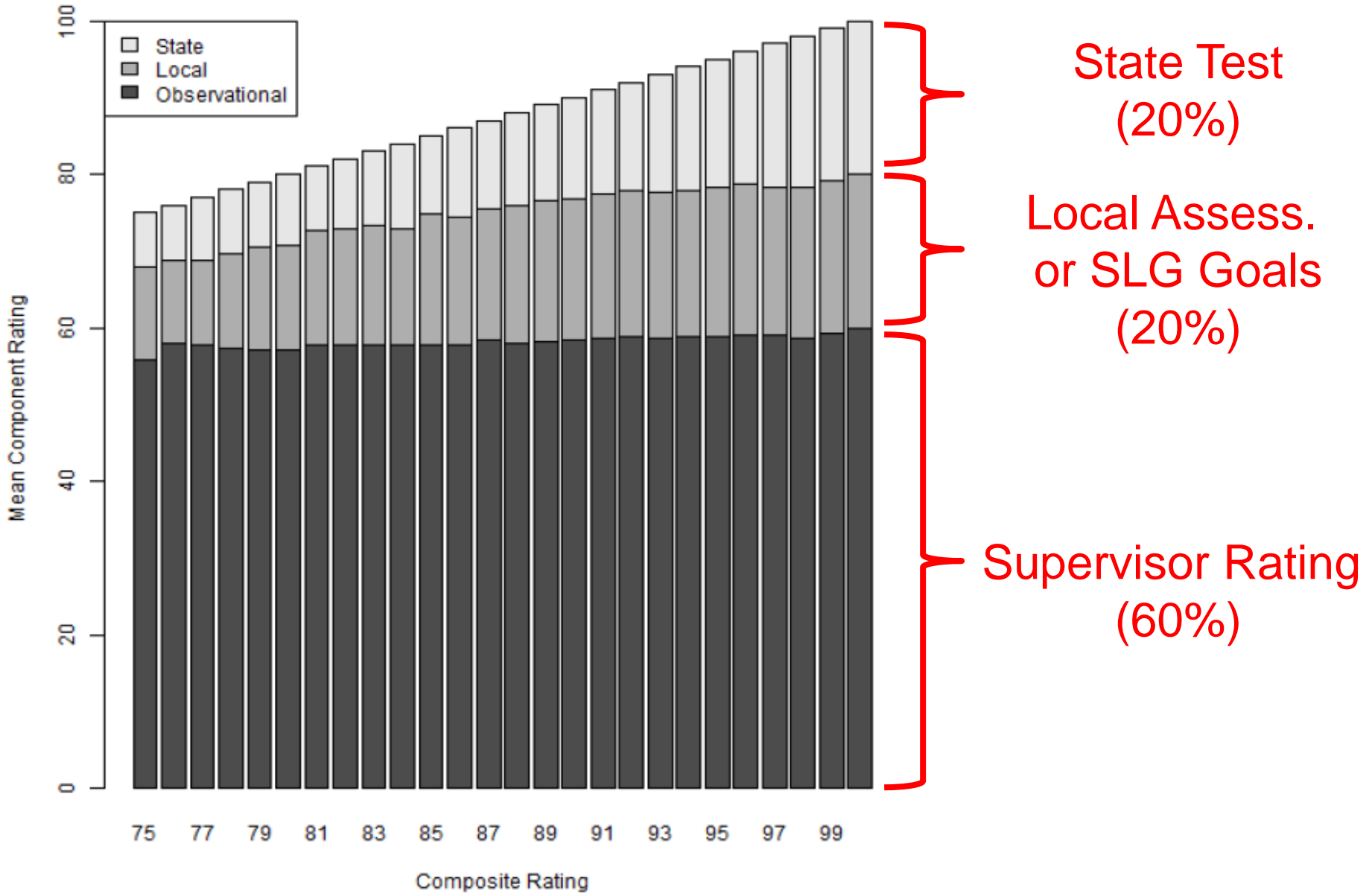
How Did We Get to This Point?



“...paint a picture of a K-12 system that remains hesitant to differentiate between the best and weakest performers—as well as among all those in the middle doing a solid job who still have room to improve.”

http://www.edweek.org/ew/articles/2013/02/06/20evaluate_ep.h32.html

Mean Component Ratings by Composite Rating

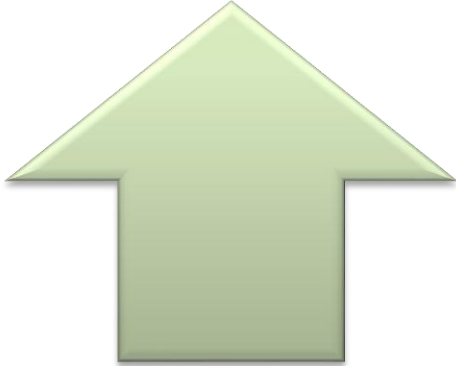
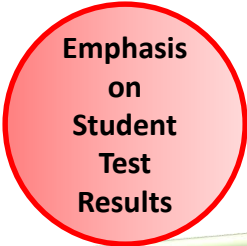



Contrasting Teacher Evaluation Approaches

Bottom-Up States (OR)

Much more flexibility and local control

“Promoting professional growth and continuous improvement...leading to improved student achievement”



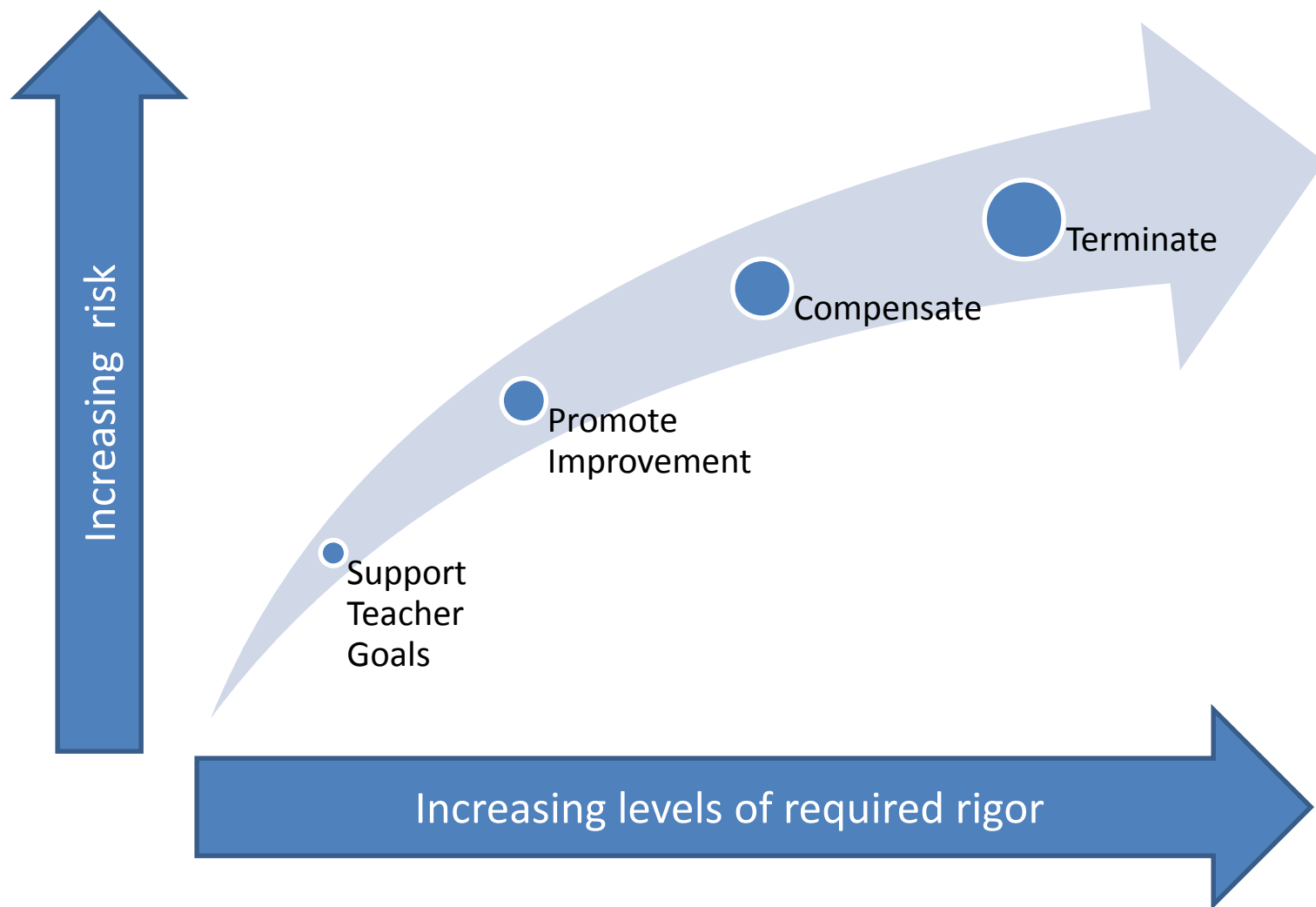
Emphasis
on
Student
Test
Results

Top-Down States (NY)

Lots of guidance and oversight from the State

Goal is to ensure “that there is an effective teacher in every classroom.”

Be Considerate of the Continuum of Stakes Involved



TEACHER SLG GOAL SETTING TEMPLATE

Teacher: _____ Contract Status: _____

School: _____

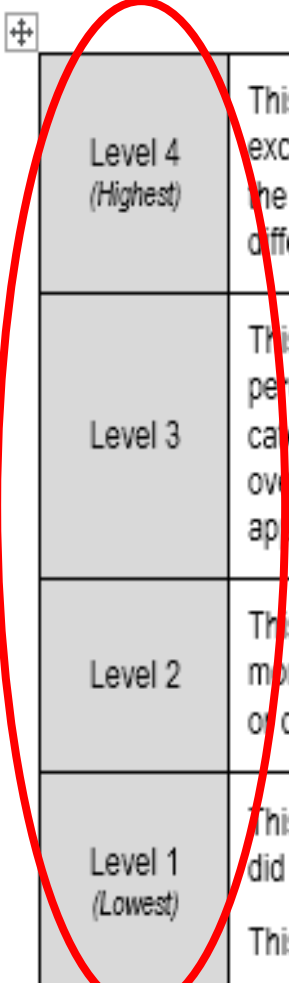
Administrator

Grade Level:

Goal Type:

SLG Scoring Rubric

This SLG scoring rubric is used for scoring individual SLG goals based on evidence submitted by the teacher and supervisor/evaluator. This rubric applies to both teacher and administrator evaluations.

Goal-Setting Conference	
Conti	
Asse:	
Conti	
Base	
Stud (Targ	
Ratio	
Strat	
Profe and s	
Level 4 (Highest)	This category applies when approximately 90% of students met their target(s) and approximately 25% of students exceeded their target(s). This category should only be selected when a substantial number of students surpassed the overall level of attainment established by the target(s). Goals are very rigorous yet attainable, and differentiated (as appropriate) for all students.
Level 3	This category applies when approximately 90% of students met their target(s). Results within a few points, a few percentage points, or a few students on either side of the target(s) should be considered "met". The bar for this category should be high and it should only be selected when it is clear that all or almost all students met the overall level of attainment established by the target(s). Goals are rigorous yet attainable and differentiated (as appropriate) for all students.
Level 2	This category applies when 70-89% of students met their target(s), but those that missed the target missed by more than a few points, a few percentage points or a few students. Goals are attainable but might not be rigorous or differentiated (as appropriate) for all students.
Level 1 (Lowest)	This category applies when less than 70% of students meet the target(s). If a substantial proportion of students did not meet their target(s), the SLG was not met. Goals are attainable, but not rigorous. This category also applies when results are missing or incomplete.

Four Primary Issues with Using Tests for “High Stakes”

1. Selection of an appropriate assessment
 - Used for the purpose for which it was designed (proficiency vs. growth)
 - Can accurately measure achievement for all students
2. Implications of narrow SLG goal focus
3. Alignment between content assessed and content taught
4. Adjust for and consider context



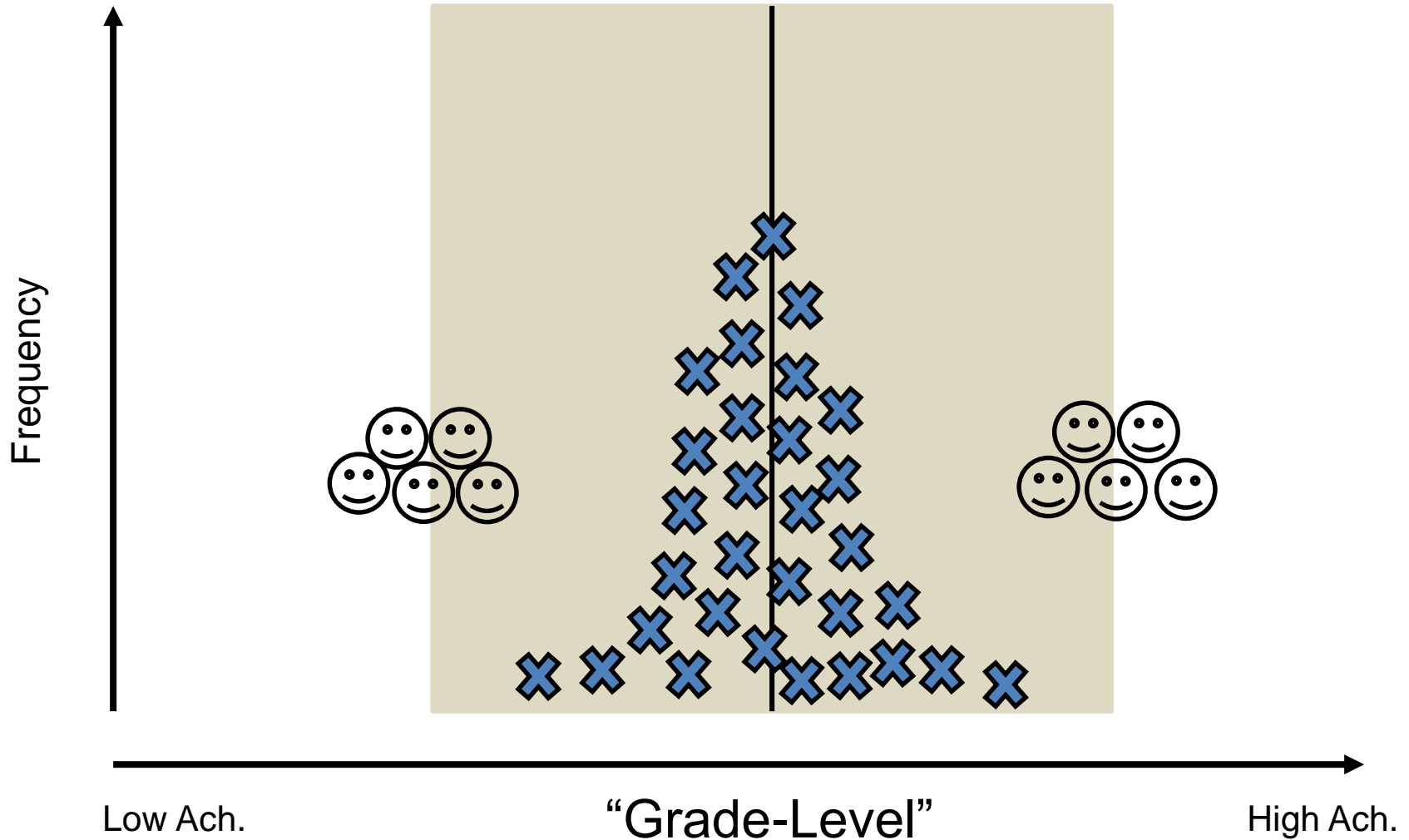
Test Selection for SLG Goals

The Purpose and Design of the Instrument is Significant

- Many assessments are not designed to measure growth
- Others do not measure growth equally well for all students

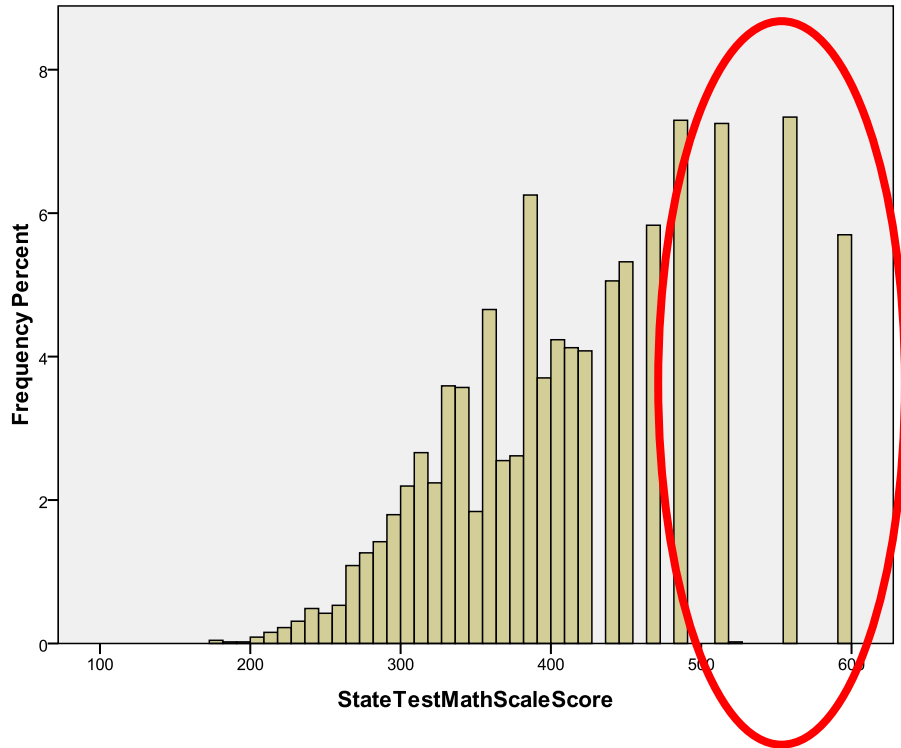


Simple Visual of a Grade-Level Test

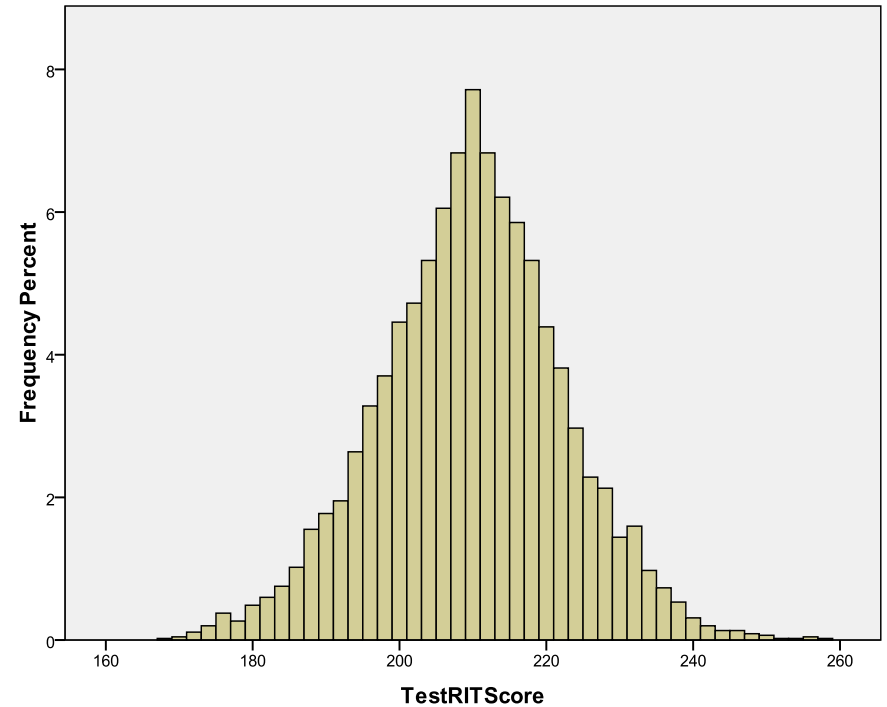


Tests Not Equally Accurate for All Students

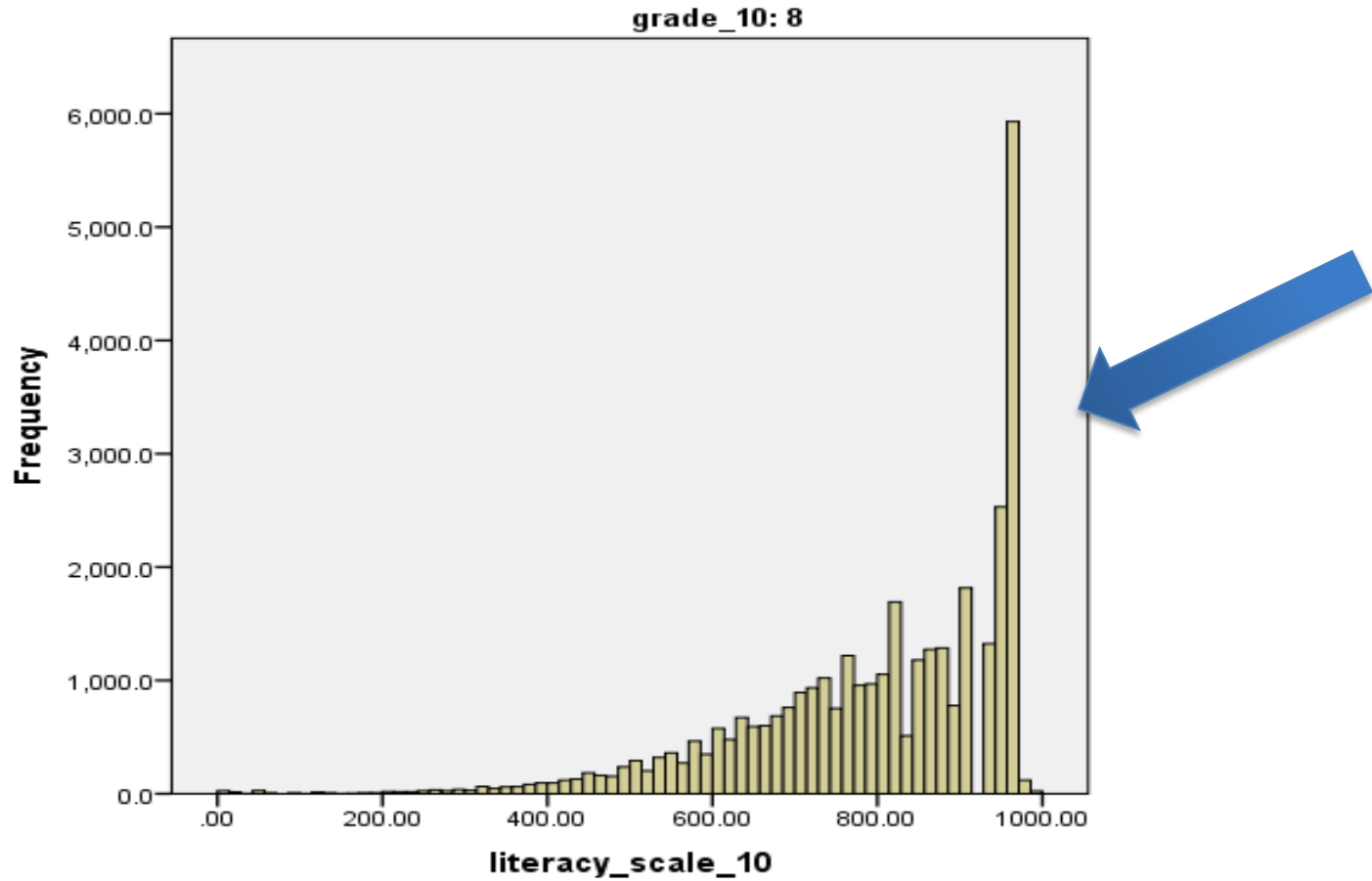
- Proficiency – CA STAR



Grade 5 Math

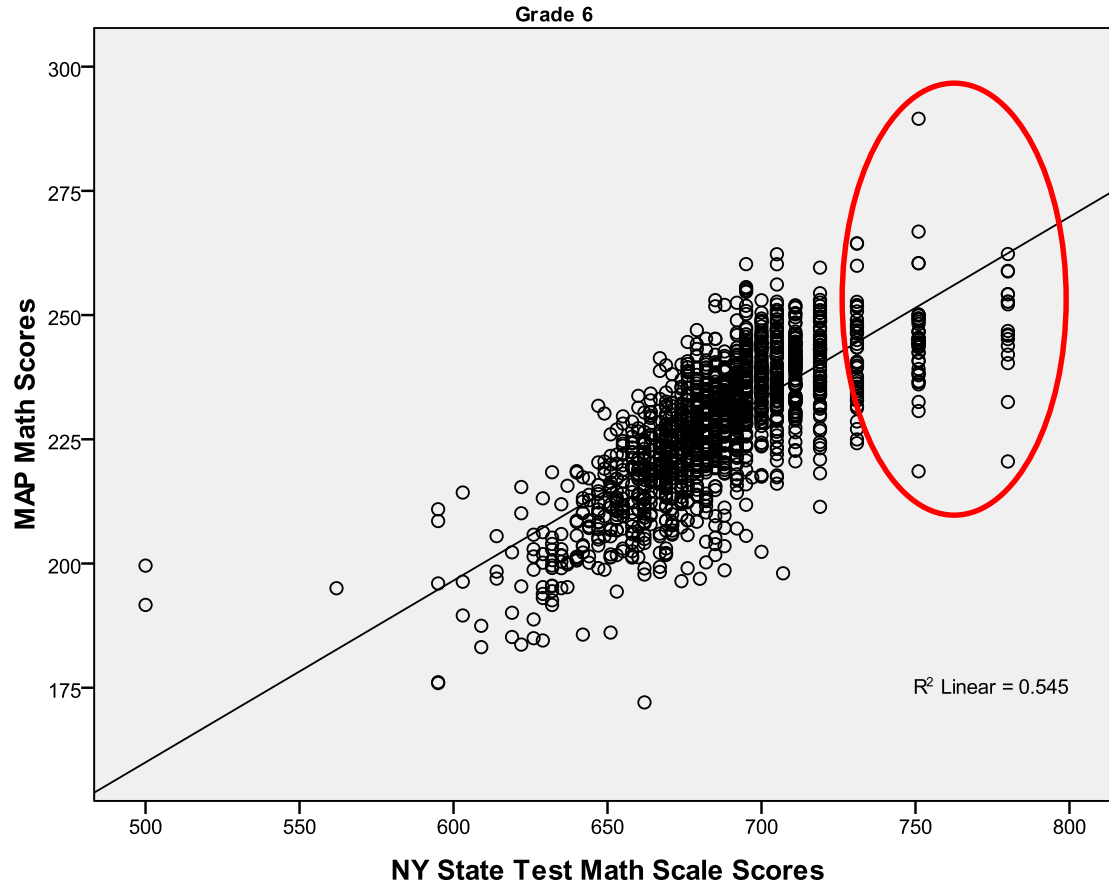


Some States Worse Than Others

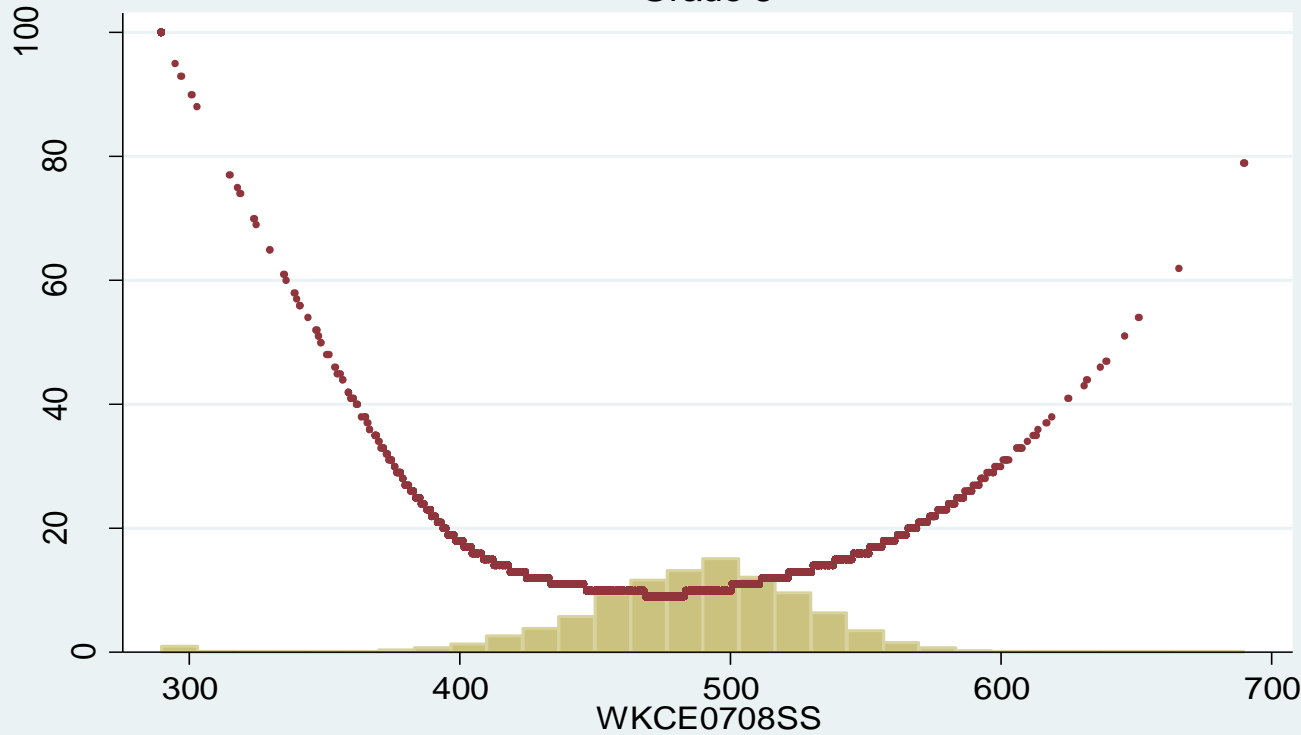


Tests Not Equally Accurate for All Students

- Grade 6 Math:
New York



WKCE 0708 Math Grade 5

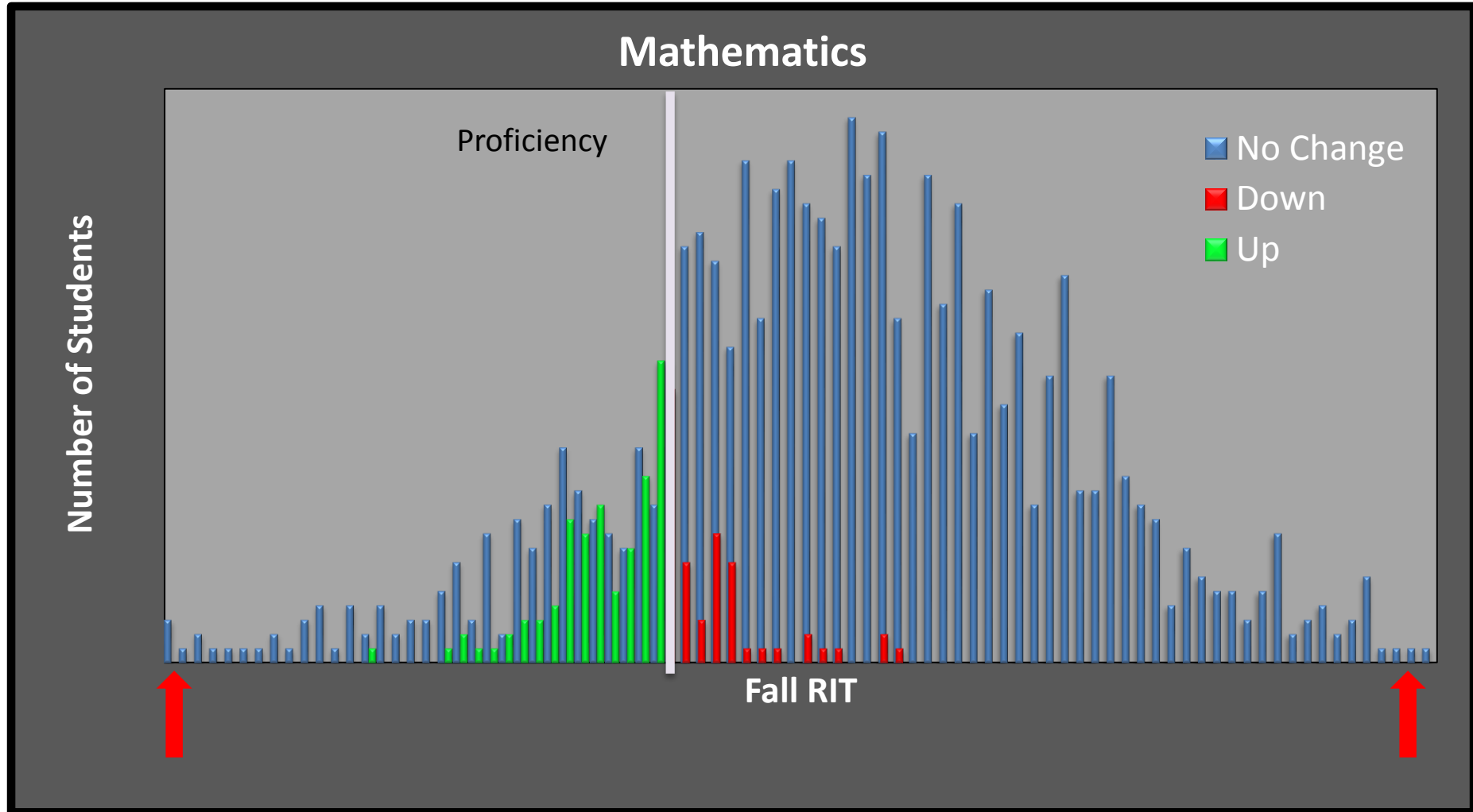


The further away a student is from “average” or “grade-level”, the less precision we have about estimates of student achievement



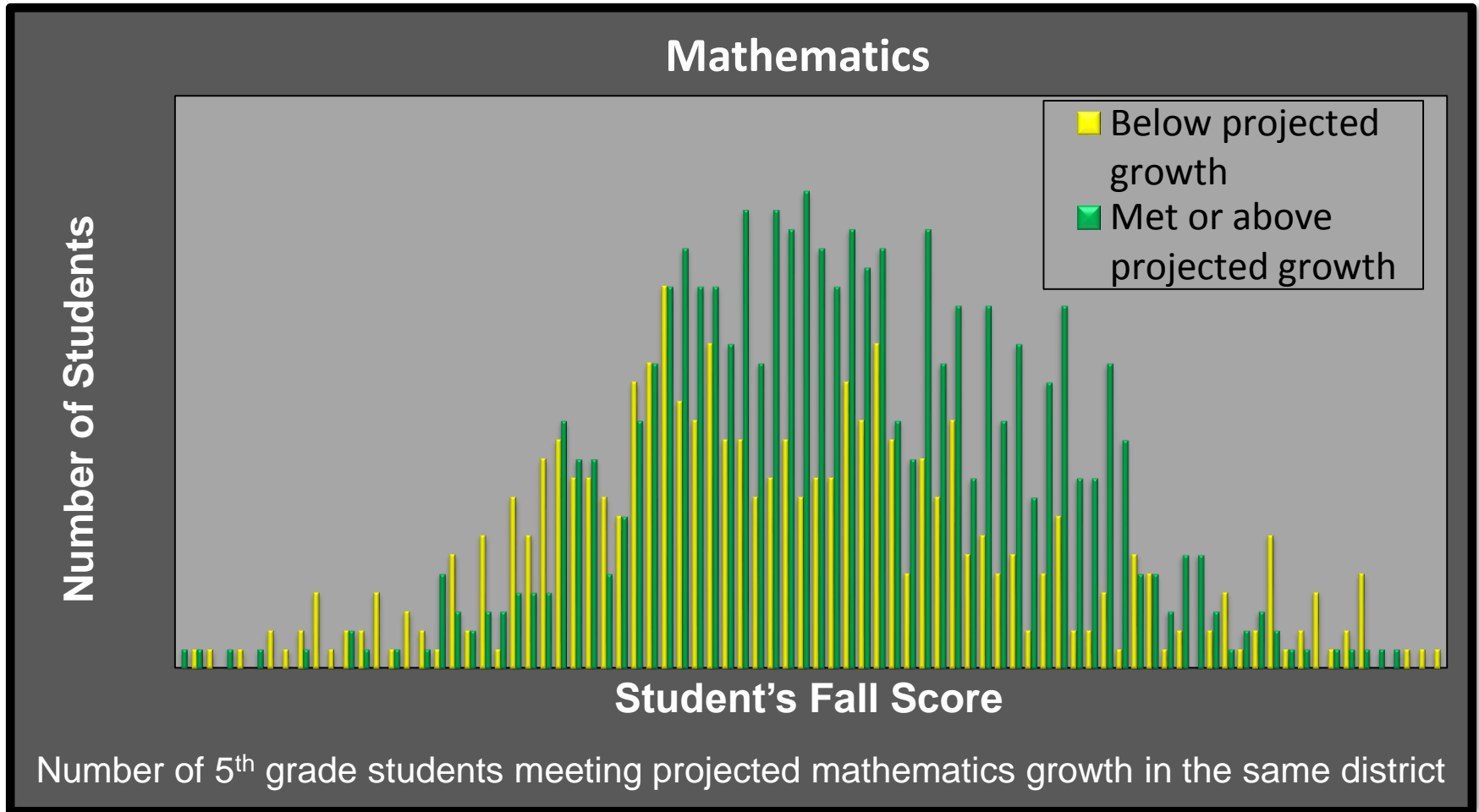
Narrow SLG Goal Focus (What Happens to Everything/Everyone Else?)

What Gets Measured and Attended to Really Does Matter



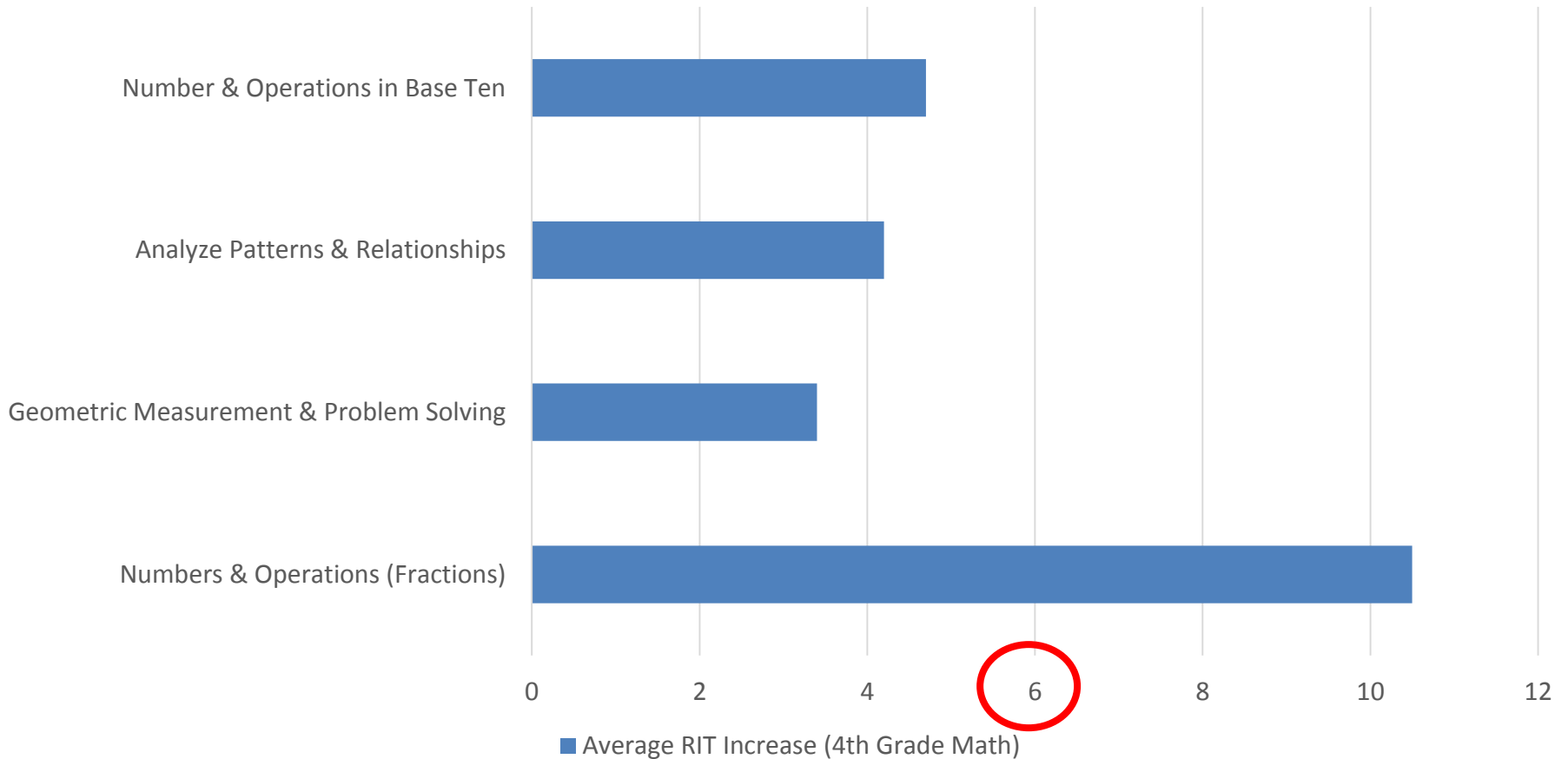
One district's change in 5th grade mathematics performance relative to the State proficiency cut scores

Measuring Growth Impacts All Students



Emphasis on Subject Sub-Goals

Average RIT Increase (4th Grade Math)

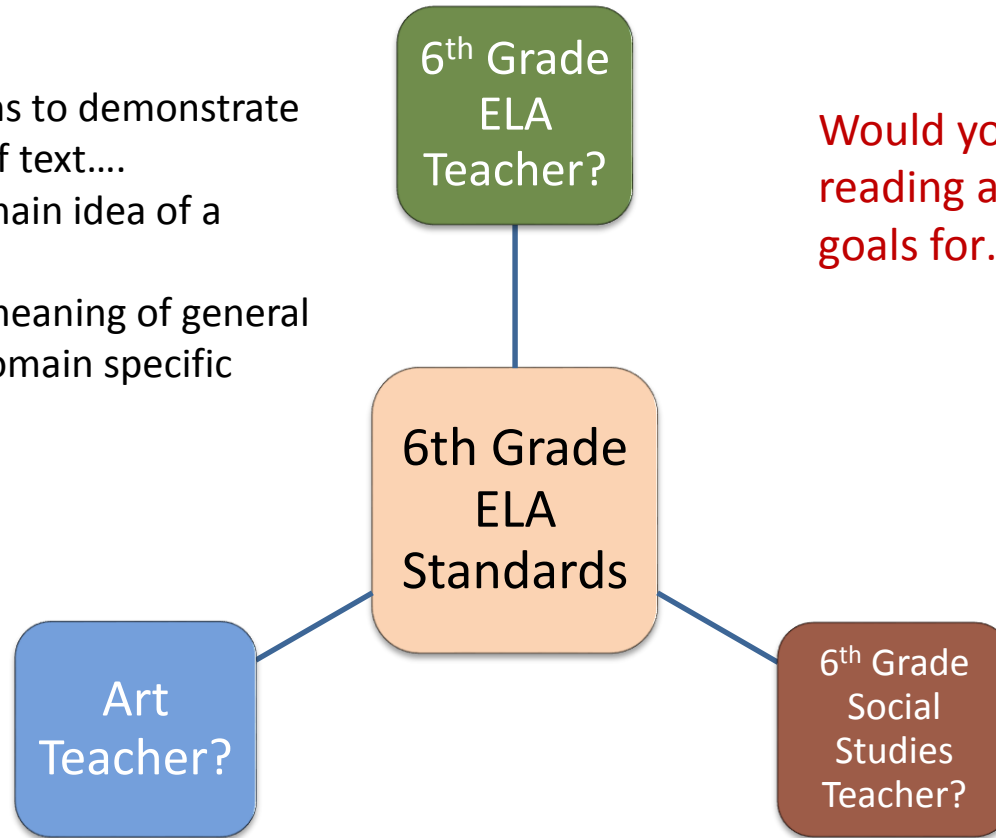




Alignment Between Content Taught & Content Assessed for SLG Goals

What is Measured Should be Aligned to What is Taught

1. Answer questions to demonstrate understanding of text....
2. Determine the main idea of a text....
3. Determine the meaning of general academic and domain specific words...



Would you use a general reading assessment for SLG goals for....

~30% of teachers teach in tested subjects and grades

What is Measured Should be Aligned to What is Taught

- Specific advanced content
 - HS teachers teaching discipline specific content
 - Especially 11th and 12th grade
 - MS teachers teaching HS content to advanced students
- Non-tested subjects
 - School-wide results more likely reflect “professional responsibility” rather than “effectiveness”
- Teachers providing remedial services

Florida Teachers Union Sues Over State Evaluation Law

By Christine Jordan Sexton - Apr 16, 2013 11:48 AM PT

    0 COMMENTS

 QUEUE

Florida's largest teachers union and its national affiliate sued the state seeking to overturn a law that ties teacher salaries to job performance, measured in part by standardized test scores.

The union, the Florida Education Association, and seven Florida teachers alleged the 2011 law violates their due process and equal protection rights under the U.S. Constitution because their grades are based on the test results of students they didn't teach, according to a complaint filed today in Gainesville federal court. The law also creates separate classes of public school teachers by eliminating tenure for newly hired teachers while maintaining tenure for older teachers, according to the plaintiffs.

The law was supported by Florida Governor [Rick Scott](#), a Republican who at the time said it would give "Florida the best educated workforces to compete in the 21st century economy."

John Tupps, a spokesman for Scott, also declined to immediately comment on the complaint. Named as defendants in the case are the Florida Board of Education and Florida Commissioner of Education Tony Bennett. Cheryl Etters, a spokeswoman for the board, also declined to immediately comment.

This is the second lawsuit the union has filed challenging the law. In the first case, filed in state court in Tallahassee, the union alleged the law violates teachers' collective bargaining rights and is unconstitutional. That case is pending.

Ratings

The law requires all school personnel to be evaluated annually using a scale of highly effective, effective, needs improvement or unsatisfactory. For classroom teachers, half of the evaluation must be based on student learning gains. For administrators, 40 percent of the evaluation must be based on learning gains and for non-classroom personnel, 30 percent of the evaluation is based on classroom gains.

"...because their grades are based on the test results of students they didn't teach..."

"...some of the plaintiffs teach orchestra, art, or health..."



Consideration of Context in SLG Goal Development

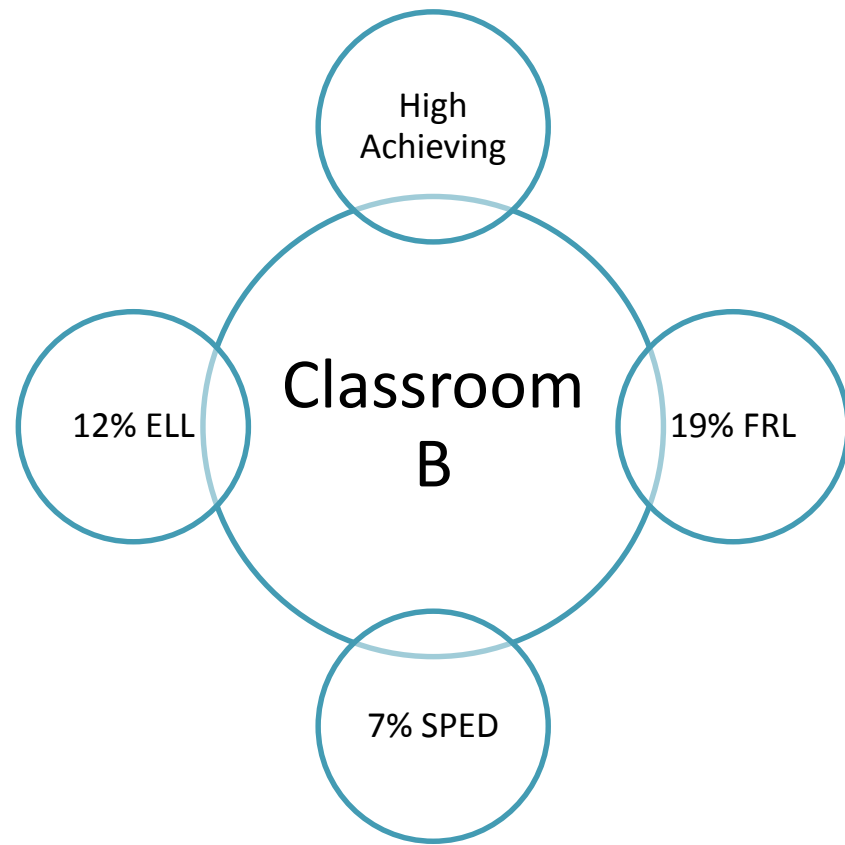
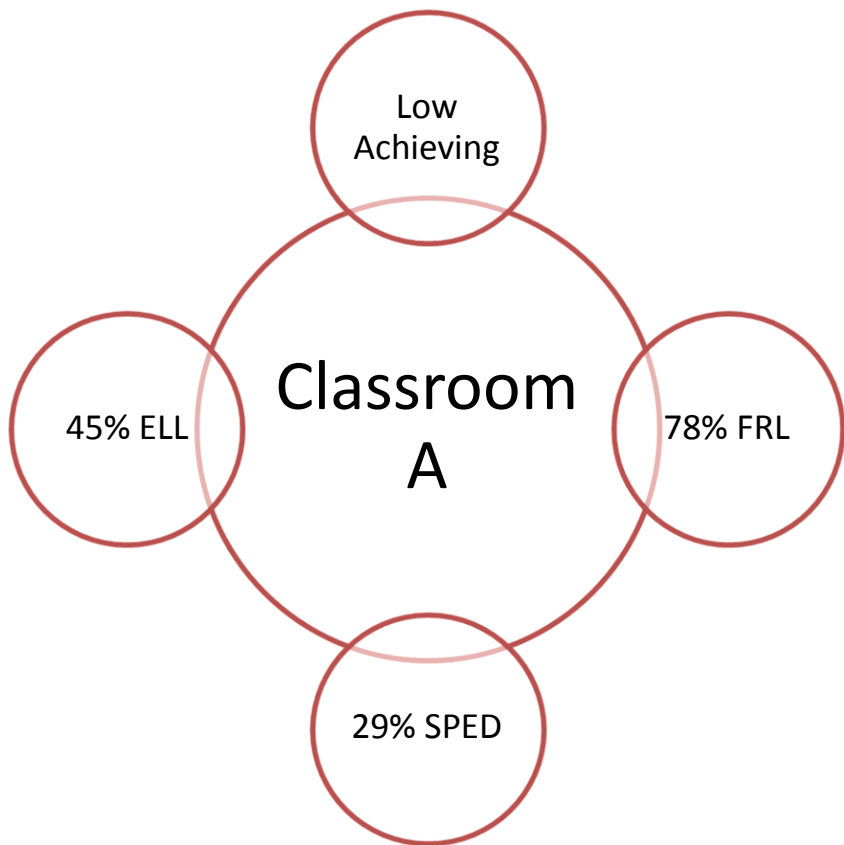
When Context Isn't Considered...

% of Students Meeting Growth Projections	New York Effectiveness Category
100%	Highly Effective
84%	Highly Effective
84%	Effective
72%	Effective
72%	Developing
60%	Developing
60%	Ineffective
0%	Ineffective

Typically
50%-55%



Comparison of Classrooms/Grades/Schools



Should we have the same expectations for these teachers?
Does different evaluation goals mean different learning expectations?
What implications might these differences have for teachers?

Need for Context

- **Historical context**

- How much growth have my students previously shown? What was the prior achievement level of my students?

- **Similar student context**

- How much growth do similar students show?

- **Classroom/school context**

- Do the students in my classroom/school have certain characteristics that may influence how much growth they show?

- **Goal context**

- Safety goal or stretch goal?

Context helps set goals that are fair and realistic for teachers, and keep student learning at the forefront

Final Questions

- Are you paying attention to the tests you are using...are they fair for teachers and students?
- Are SLG goals benefitting all students in all areas...or some students in some areas?
- Are evaluation expectations aligned with teaching expectations?
- Are equal students/teachers being treated equally...and vice versa?
 - One size doesn't always fit everyone



- Questions?
- Comments?
- Feedback?

Thanks!

Nate Jensen, Ph.D.

Research Scientist, NWEA

nate.jensen@nwea.org