New York State
EDUCATION DEPARTMENT
Knowledge › Skill › Opportunity

# NYSED Evaluation Workgroup

## Session #2: Student Performance Measures & Technical Requirements for Inclusion of Assessments in Evaluation

# Welcome!

# Session #2 Objectives

1. Provide input on the ideal student learning component for educator evaluation
2. Review existing requirements for the student learning component of educator evaluation, and discuss ways to improve that component
3. Process feedback and assess consensus around recommendations
4. Discuss next steps

# Agenda

| Time | Topic |
|---|---|
| 9:00 am | Welcome and introductions |
| 9:15 am | Reviewing inputs from webinar |
| 9:45 am | The ideal student learning component |
| 11:30 am | Lunch |
| 12:30 pm | How can the consistency in the implementation of SLOs across LEAs and schools be improved? |
| 1:30 pm | How can the assessment quality be balanced with inclusion of additional assessments in SLOs? |
| 2:15 pm | Student Growth Measures |
| 2:45 pm | Closing & Next Steps |

# Group Norms

- Presume positive intentions
- Fully engage, active listening, and speaking
- No cross talk
- Respect for everyone's opinions and views. Open to all experiences and views.
- Talking piece (something physical to hold)—respect those who speak.
- Equal airtime
- Respectful of time
- No cell phones
- Be curious
- No need to bash the administration
- Ensure all stakeholders' voices are heard
- Subgroup work—no silos. Ensure knowledge is shared with everyone during group work.
- Keep children as the focus and at the center

# What is our purpose?

- **Provide and capture recommendations to improve educator evaluation system**
  - **Improvements to existing system**
  - **Components and measures of an ideal system**

- **In order to get there we will**
  - **Identify and surface barriers**
  - **Consider technical information provided by experts**
  - **Recommend solutions**

# What research tells us about the importance of student outcomes in evaluation

- **Teachers are the single most important school based, and principals are the second most influential factor and have a multiplicative effect related to student outcomes** (McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2000; Rowan, Correnti & Miller, 2002; Wright, Horn, & Sanders, 1997).
- **Students of teachers with higher teacher effectiveness estimates outperformed students of teachers with lower teacher effectiveness estimates** (Cantrell and Kane, 2013).
- **Students assigned to more effective teachers are more likely to attend college, attend higher- ranked colleges, earn higher salaries, live in higher SES neighborhoods, and save more for retirement** (Chetty, Friedman and Rockoff, 2011)

# Agenda

| Time | Topic |
|---|---|
| 9:00 am | Welcome and introductions |
| **9:15 am** | **Reviewing inputs from Webinar #1** |
| 9:45 am | The ideal student learning component |
| 11:30 am | Lunch |
| 12:30 pm | How can the consistency in the implementation of SLOs across LEAs and schools be improved? |
| 1:30 pm | How  can the assessment quality be balanced with inclusion of additional assessments in SLOs? |
| 2:15 pm | Student Growth Measures |
| 2:45 pm | Closing & Next Steps |

New York State
EDUCATION DEPARTMENT
Knowledge › Skill › Opportunity

# Reviewing Inputs from Webinar #1

## 9:15-9:45 am

# An ideal evaluation system….

- Informs professional growth and evaluation
- Requires thoughtful self-reflection
- Supports collaboration
- Benefits students
- Emphasizes equity
- Takes into account factors outside of the teachers' control that have been shown to influence learning

# Barriers to an ideal system…

- **Student Learning**
  - Assessments may not fully capture students' progress
  - Year to year changes in assessments make it difficult to understand student growth
  - Factors outside of the classroom teachers' control must be taken into account
  - SLOs are not always implemented to improve teaching practice or student performance
  - Teachers may not have enough knowledge about students at the start of the school year to set useful SLO targets
  - Teachers in schools with high student turnover may develop targets at the beginning of the year that aren't relevant to student in their classes at the end of the year

# Agenda

| Time | Topic |
|---|---|
| 9:00 am | Welcome and introductions |
| 9:15 am | Reviewing inputs from Webinar #1 |
| **9:45 am** | **The ideal student learning component** |
| 11:30 am | Lunch |
| 12:30 pm | How can the consistency in the implementation of SLOs across LEAs and schools be improved? |
| 1:30 pm | How  can the assessment quality be balanced with inclusion of additional assessments in SLOs? |
| 2:15 pm | Student Growth Measures |
| 2:45 pm | Closing & Next Steps |

New York State
EDUCATION DEPARTMENT
Knowledge › Skill › Opportunity

# The Ideal Student Learning Component

## 9:45-11:30 am

# Table Team Activity

**Premise:** You and your table team will represent a hypothetical LEA, which has been given local control over the student learning component of an evaluation system for teachers and principals.

**Guiding Question:** What would your ideal student learning component look like?

45 min table discussion

60 min whole group share out

# Agenda

| Time | Topic |
|---|---|
| 9:00 am | Welcome and introductions |
| 9:15 am | Reviewing inputs from Webinar #1 |
| 9:45 am | The ideal student learning component |
| **11:30 am** | **Lunch** |
| 12:30 pm | How can the consistency in the implementation of SLOs across LEAs and schools be improved? |
| 1:30 pm | How  can the assessment quality be balanced with inclusion of additional assessments in SLOs? |
| 2:15 pm | Student Growth Measures |
| 2:45 pm | Closing & Next Steps |


New York State
EDUCATION DEPARTMENT

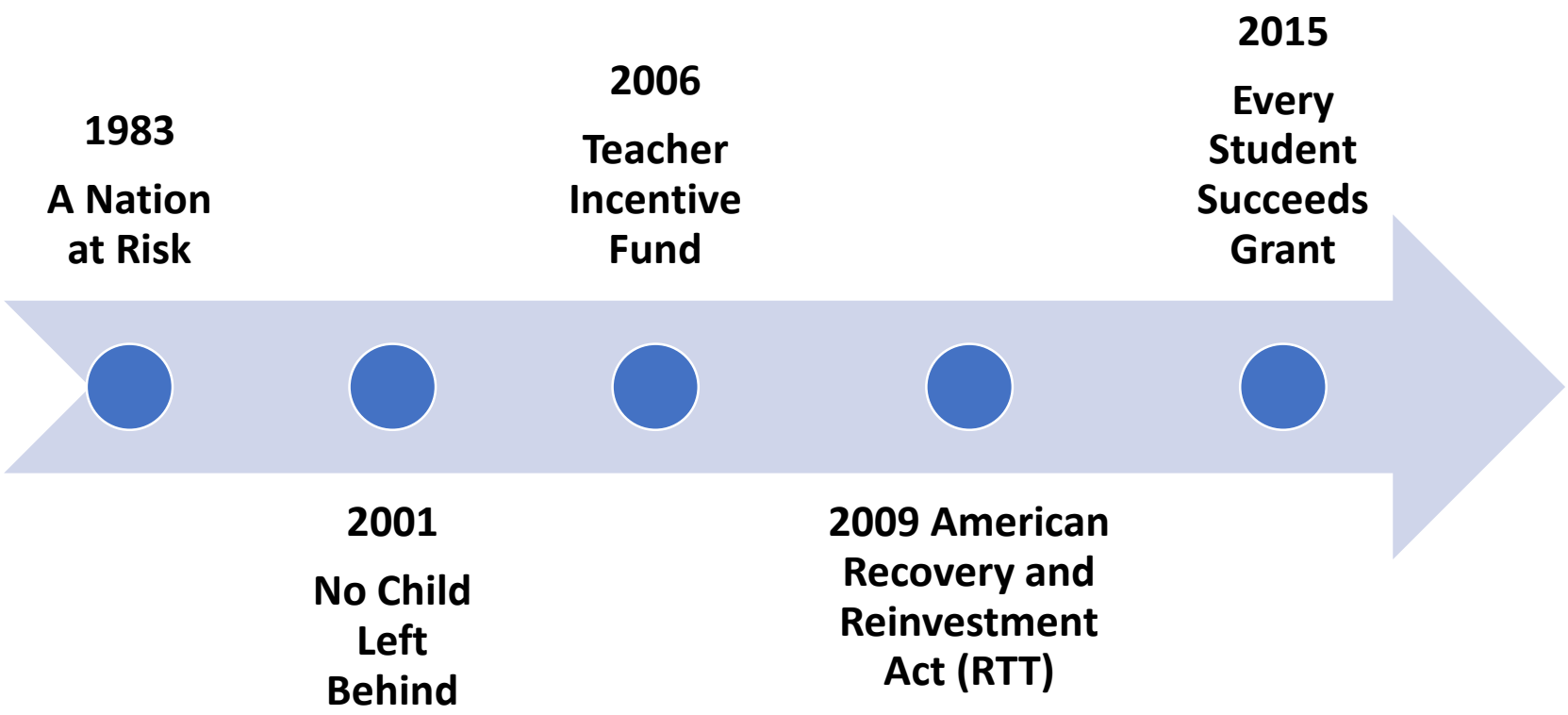
Knowledge › Skill › Opportunity

# How can the consistency in the implementation of SLOs across LEAs and schools be improved?

# The SLO Process

## 12:30-1:30am

# A Little History



# Just how widely used are SLO's?

- 25 states include a definition of SLO's in their teacher evaluation systems

# What types are there?

- SLO's for individual teachers in 23 states
- SLO's for teams of teachers or grade levels in 3 states, optional in 7
- Schoolwide SLO's required in 3 states and optional in 4 states

# Common Elements of State Definitions

| Element | Number of States |
|---|---|
| Measurable | 12 |
| Based on student growth and achievement | 16 |
| Aligned with state or local standards | 9 |
| Based on prior student learning data | 9 |
| Measure teacher impact on student learning | 4 |
| Aligned with course content | 4 |

# Assessments Used to Evaluate Student Learning Objectives

| Assessment type or feature | Number of states |
|---|---|
| National or state standardized assessment | 14 |
| District-wide or school-wide measures | 12 |
| Classroom-based measures | 12 |
| Test Vendor-developed content | 3 |
| Comparable across classrooms | 5 |
| Valid and reliable | 3 |
| Aligned with state standards | 2 |
| Rigorous | 2 |

# Why use SLO's?

- SLO process contains key aspects of good instruction: review of student data, goal setting, progress monitoring, reflection
- Can be applied in all subject and content areas
- Adaptable
- Encourage collaboration among teachers
- Promote reflective practice
- Provide teachers some ownership of how they are evaluated

# Basic SLO Process

# Implementation Strategies to Help Ensure Consistency and Rigor

- Provide exemplary SLOs across subject areas
- Approve assessments for use in SLOs
- Assessment literacy training
- Build Principal capacity to assess and provide feedback to improve SLO quality and rigor
- School or team-based goals (individual targets)
- Mid-year SLO review
- Student data use training
- Randomly sample SLOs for audit
- Consideration of SLO quality/rigor in scoring SLO

# Examples From Other States



# Table Talk #1 (25 minutes)

**Guiding Question: Thinking about the SLO process, how can the consistency in the implementation of SLOs across LEAs and schools be improved?**

- Are there implementation strategies that are currently not in use that you would recommend? Why?
- Are there improvements to the SLO template that you would recommend to improve the alignment of the template and the process?

# Group Discussion #1 (20 min)

- **Please share out your or your group's decisions and discuss the rationale.**

# Agenda

| Time | Topic |
|---|---|
| 9:00 am | Welcome and introductions |
| 9:15 am | Reviewing inputs from Webinar #1 |
| 9:45 am | The ideal student learning component |
| 11:30 am | Lunch |
| 12:30 pm | How can the consistency in the implementation of SLOs across LEAs and schools be improved? |
| **1:30 pm** | **How can the assessment quality be balanced with inclusion of additional assessments in SLOs?** |
| 2:15 pm | Student Growth Measures |
| 2:45 pm | Closing & Next Steps |

New York State
EDUCATION DEPARTMENT

Knowledge › Skill › Opportunity

**How can the assessment quality be balanced with inclusion of additional assessments in SLOs?**

**1:30-2:15 p.m.**

# Assessments in SLOs

- **Goal: have assessment choices available for all educators to use in SLOs with (a) high degree of ownership/buy-in; and (b) sufficient technical quality**
- **Current Reality:**
  - **Some grades/subjects have more (and better) assessments than others**
  - **Many teachers dislike the use of traditional standardized assessments in SLOs and prefer locally-developed options**
  - **Locally-developed and classroom assessments have greater buy-in but more technical challenges**
  - **Educator capacity around assessment development is often low, although can be built over the long term**

# Desired Technical Properties of Assessments (abridged)

- What are some key aspects of assessment quality and why do we have them?
- From the *Standards for Educational and Psychological Testing* (2014 edition):
- Part I: Foundations
  - Validity
  - Reliability/Precision and Errors of Measurement
  - Fairness in Testing
- Part II: Operations
  - Test Design and Development
  - Scores, Scales, Norms, Score Linking, & Cut Scores
  - Test Administration, Scoring, and Reporting

# Validity

- **Validity: the degree to which evidence and theory support the interpretation of test scores for their proposed uses**
- **Not a single statistic; an ongoing process**
  - **Documented alignment to content standards**
  - **Involvement of educators in item design and review**
  - **Varied set of items by level of cognitive complexity and item type**

# Reliability

- Reliability = precision/stability of results
- Would student scores change if:
  - They got a different set of items that purported to measure the same knowledge?
  - Someone else scored their assessments?
  - They took the same test another time?

# Fairness

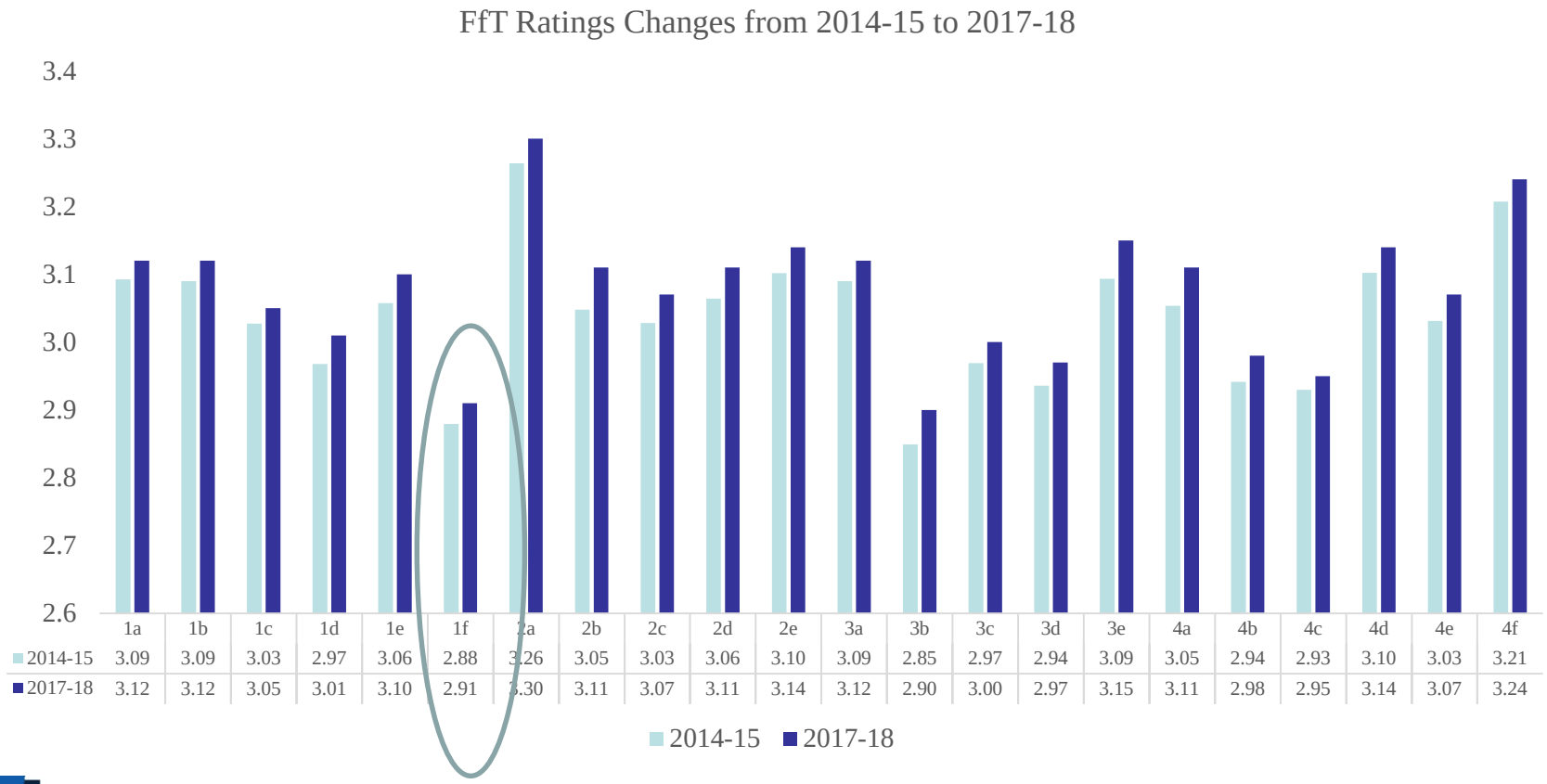
- **High-quality assessments must enable ALL students to demonstrate their knowledge (UDL principles):**
  - Precisely-defined constructs
  - Clear instructions
  - Maximum readability
  - Allowable accommodations for SwD and ELL
  - Items free of bias (DIF analysis)

# Test Design

- Is the test scaled or simple # correct?
- Under a pre/post arrangement (typical for many SLOs), are pre and post equated for difficulty?
- If cut scores exist, how were they established?
- Does the assessment contain enough items to accurately differentiate student knowledge (are there items for low, medium, and high performers)?
- What are reasonable (and ambitious) expectations for growth? How much do they vary based on students' starting point?

# Educator Capacity: Assessment Development

- **Most educators get relatively little training in this area…**




| | 1a | 1b | 1c | 1d | 1e | 1f | 2a | 2b | 2c | 2d | 2e | 3a | 3b | 3c | 3d | 3e | 4a | 4b | 4c | 4d | 4e | 4f |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2014-15 | 3.09 | 3.09 | 3.03 | 2.97 | 3.06 | 2.88 | 3.26 | 3.05 | 3.03 | 3.06 | 3.10 | 3.09 | 2.85 | 2.97 | 2.94 | 3.09 | 3.05 | 2.94 | 2.93 | 3.10 | 3.03 | 3.21 |
| 2017-18 | 3.12 | 3.12 | 3.05 | 3.01 | 3.10 | 2.91 | 3.30 | 3.11 | 3.07 | 3.11 | 3.14 | 3.12 | 2.90 | 3.00 | 2.97 | 3.15 | 3.11 | 2.98 | 2.95 | 3.14 | 3.07 | 3.24 |

# Key Decision for States/Districts

- Many (most?) educators prefer to use locally-developed or classroom assessments for SLOs.
- However, many of these assessments have low (or unknown) technical qualities, and improving educator capacity in this area is a long-term project.
- In the short term, states must weigh the tradeoffs of *greater educator buy-in* (from a more flexible approach to allowable assessments and how growth targets are set) vs. *ensuring minimal technical quality* (from approved assessment lists and pre-determined growth target formulas).

# Table Talk #2 (15 min)

**Guiding Question:** When considering the use of locally-developed and third-party assessments in educator evaluator systems, how can the State balance the need for assessments that meet certain technical criteria with the desire for LEAs to have flexibility in their approaches to evaluation?

# Group Discussion #2 (15 min)

- **Please share out your or your group's recommendations and discuss the rationale.**

# Agenda

| Time | Topic |
|---|---|
| 9:00 am | Welcome and introductions |
| 9:15 am | Reviewing inputs from Webinar #1 |
| 9:45 am | The ideal student learning component |
| 11:30 am | Lunch |
| 12:30 pm | How can the consistency in the implementation of SLOs across LEAs and schools be improved? |
| 1:30 pm | How can the assessment quality be balanced with inclusion of additional assessments in SLOs? |
| **2:15 pm** | **Student Growth Measures** |
| 2:45 pm | Closing & Next Steps |

New York State
EDUCATION DEPARTMENT
Knowledge › Skill › Opportunity

# Student Growth Measures

## 2:15 -2:45 pm

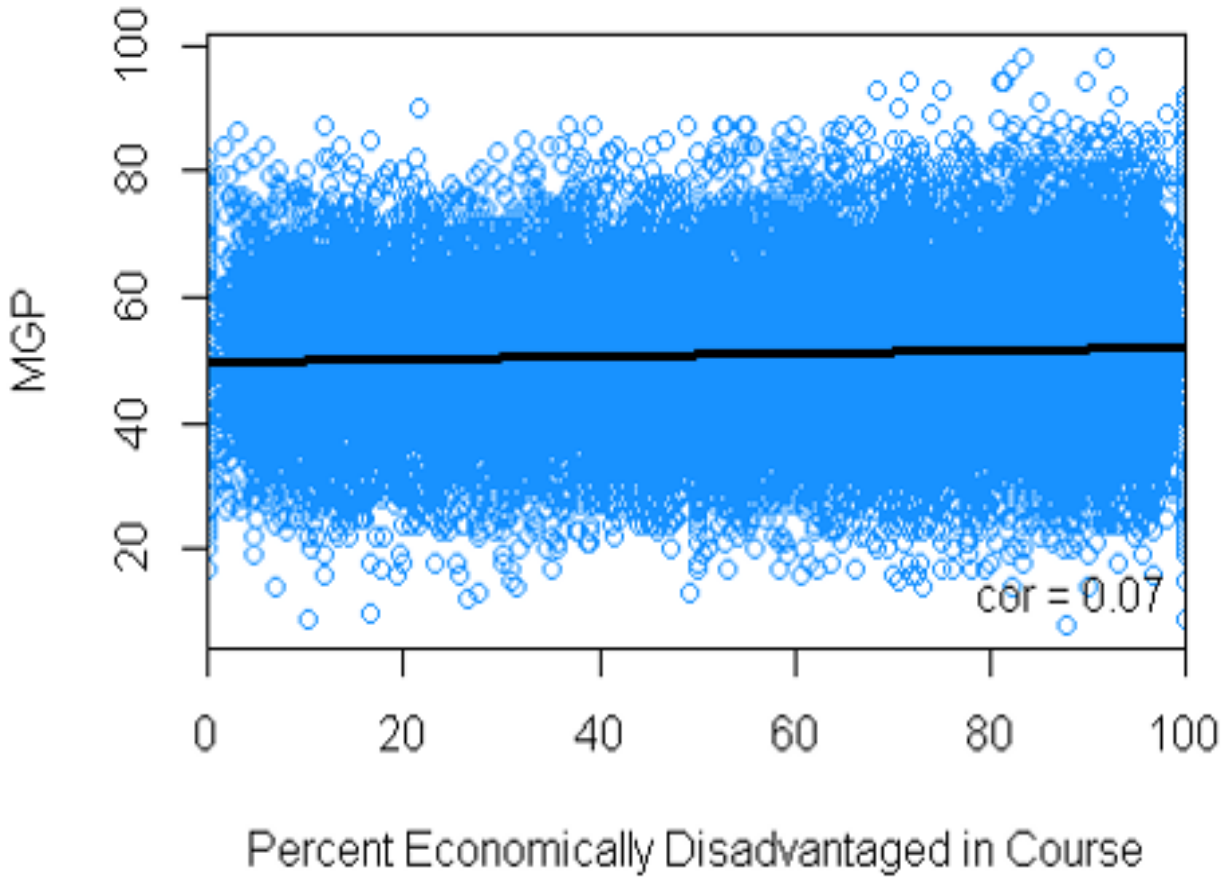
# Statistical Growth Models (In Brief)

- A **group of models d**esigned to measure the contribution of schooling at various levels (school, grade, classroom, etc.) to gains in student performance over time.
- Uses statistical techniques to separate the impact of schooling from other factors that may influence growth, but are generally beyond the control of schools/educators (prior achievement, EcDis, SpEd, ELL).
- Goal: provide information on what different levels of education (school, classroom, etc.) can and should control (improved achievement for all students), but factor out what they can't control (student characteristics and out of school factors)

# Selected Observations from NY Data

- Concern: student growth is influenced by factors that educators don't control (creating potential disincentives)
- Data show low correlations between classroom factors (poverty, SpEd, etc.) and SGPs; this means there is very little "penalty" for teaching these kinds of students (as would be the case with proficiency rates)

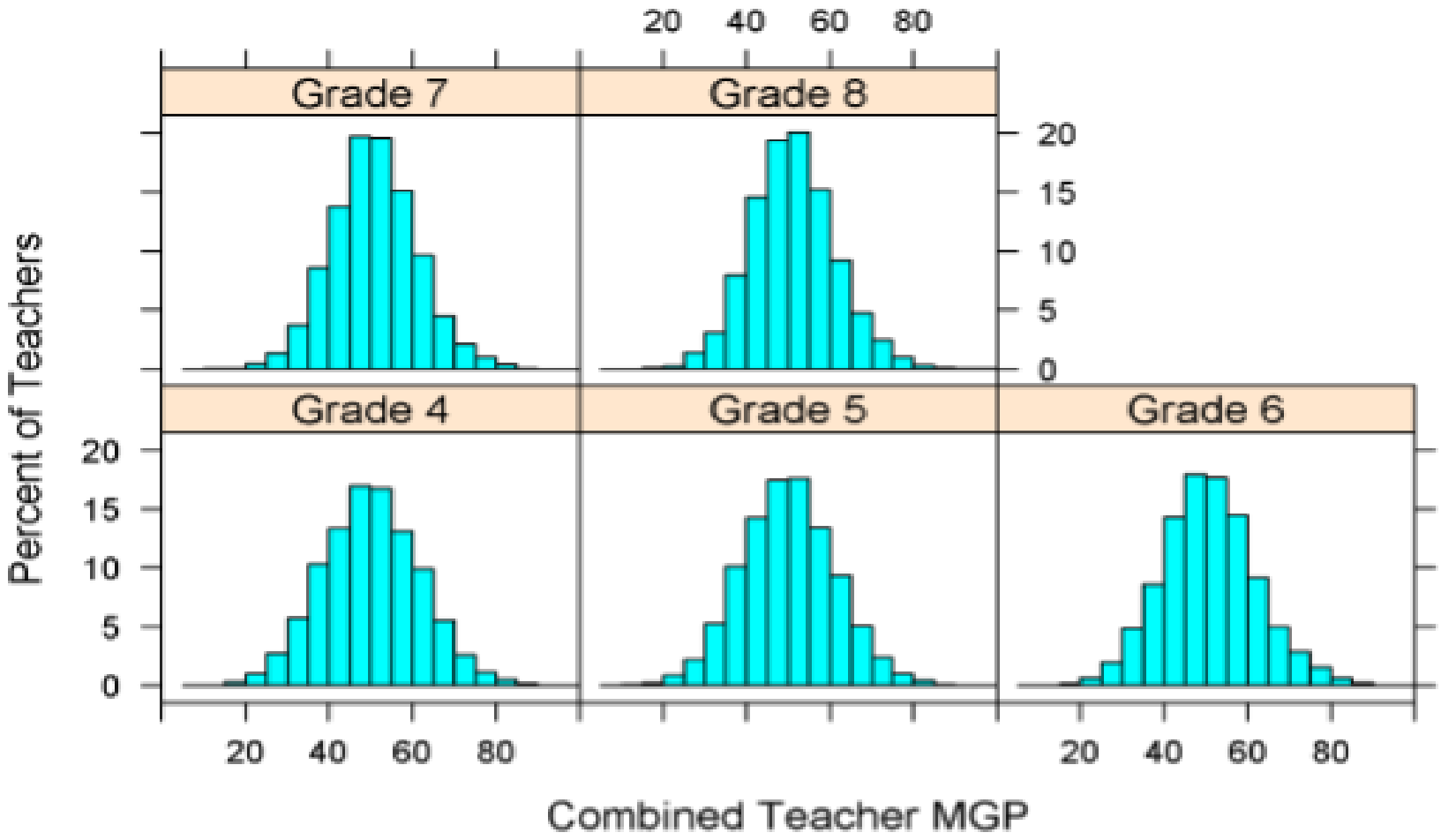
**Figure 11. Relationship of Grades 4–8 Teacher MGP Scores to the Percentage of Economically Disadvantaged Students or Course**



# Selected Observations (cont'd)

- Concern: distribution of Student Growth scores is fundamentally different from other measures
- Data confirm this to be true, although most educators are still Effective or Highly Effective on Student Growth

**Figure 5. Distribution of Grades 4–8 Teacher MGPs by Grade, Adjusted Model**



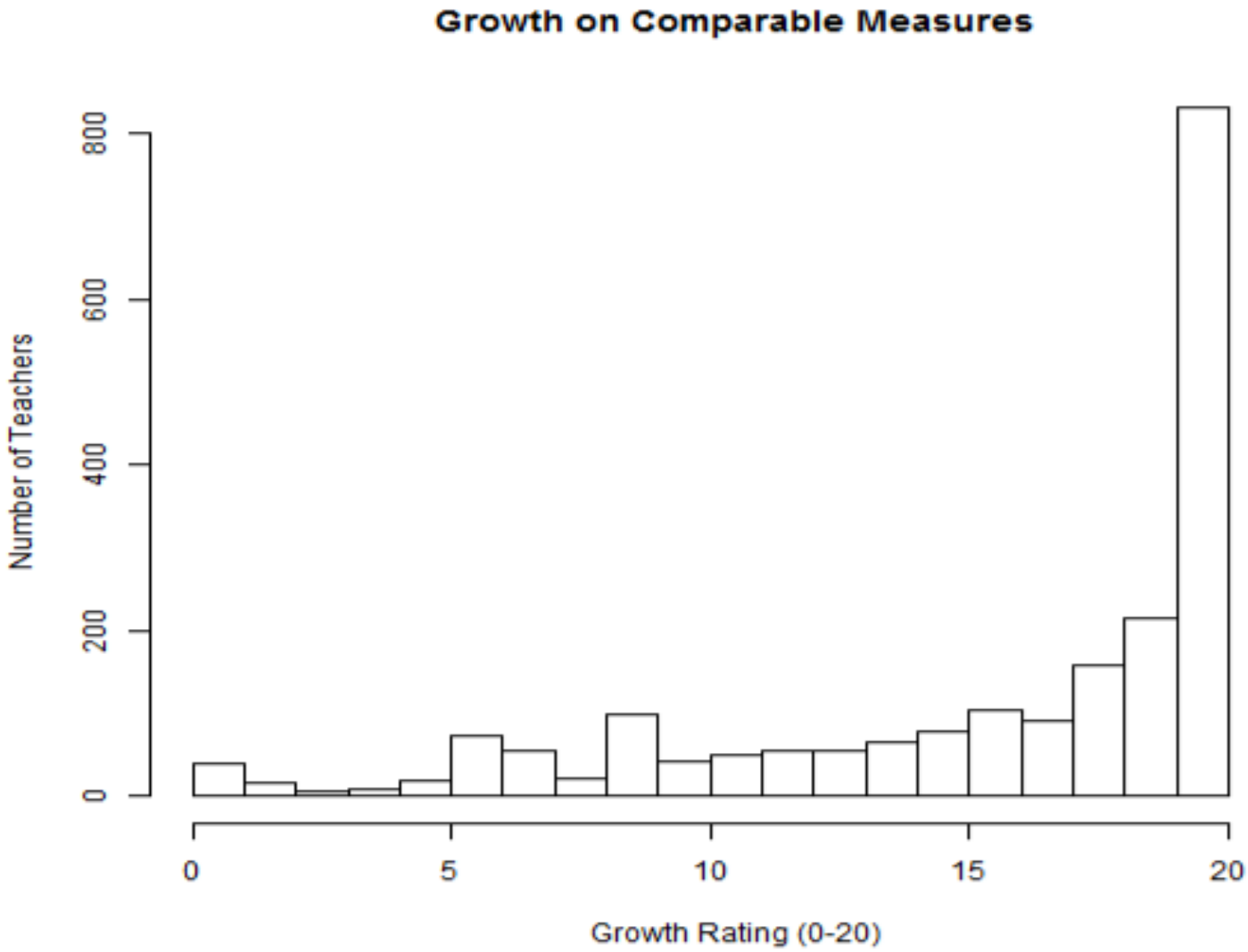


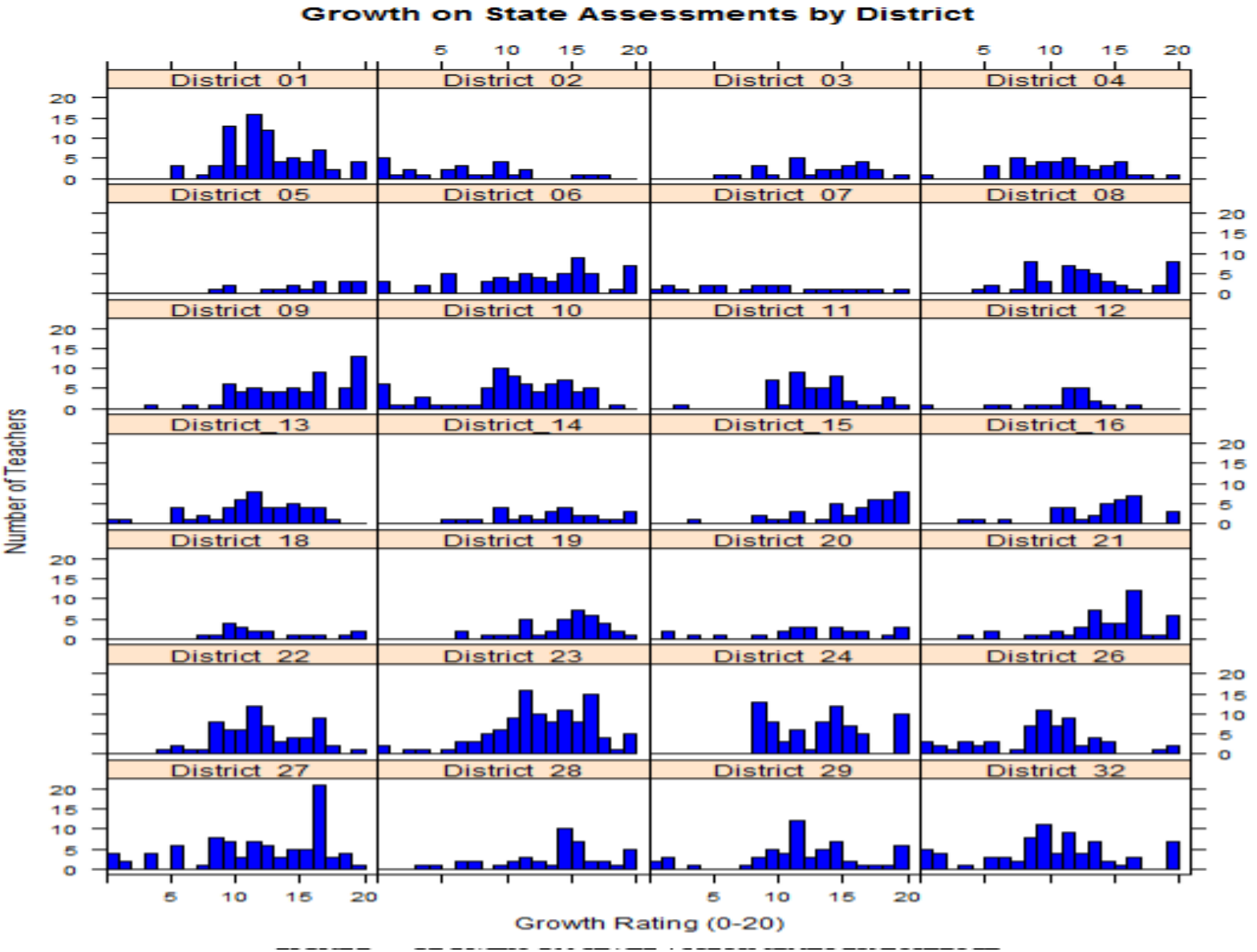

FIGURE 12: DISTRIBUTION OF GROWTH RATINGS ON COMPARABLE MEASURES MODELS

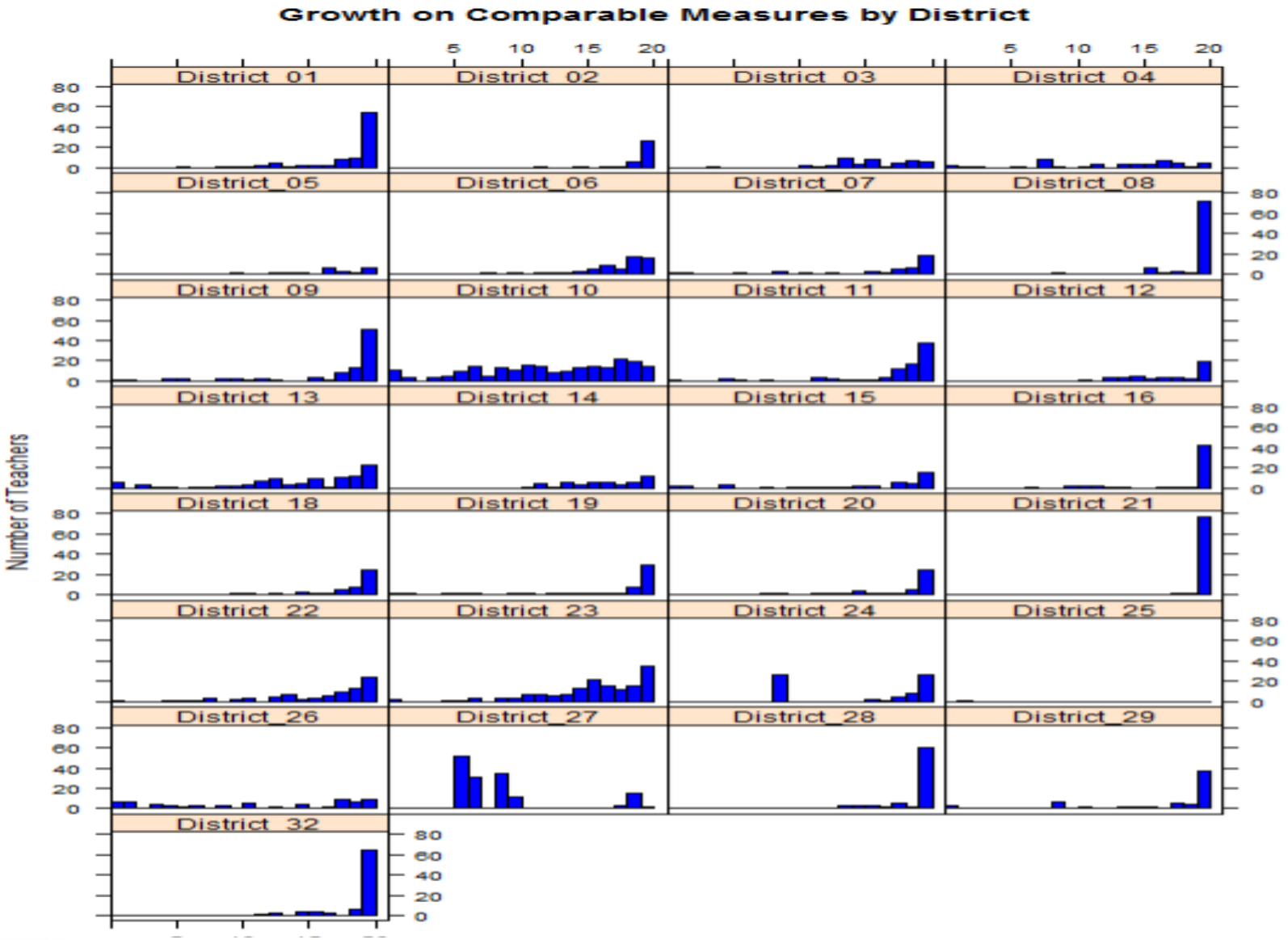
## *Growth Ratings for Grades 4–8*

This section describes the observed distribution of the growth ratings assigned using the rules described earlier in the results section. Table 15 shows the distribution for Grades 4–8 teachers, schools, and principals who serve students in Grades 4–8 (including, for instance, schools serving Grades 4–12) from 2012–13 to 2015–16.

**Table 15. Grades 4–8 Teacher, School, and Principal Growth Ratings**

| School Year | Educator Level | Highly Effective | Effective | Developing | Ineffective |
|---|---|---|---|---|---|
| 2012–13 | Teacher | 7% | 76% | 11% | 6% |
| | School | 9% | 75% | 9% | 7% |
| 2013–14 | Teacher | 8% | 77% | 10% | 6% |
| | Principal | 6% | 77% | 10% | 7% |
| | School | 7% | 76% | 10% | 7% |
| 2014–15 | Teacher | 7% | 77% | 11% | 5% |
| | Principal | 7% | 77% | 10% | 6% |
| | School | 7% | 76% | 11% | 6% |
| 2015–16 | Teacher | 8% | 76% | 11% | 5% |
| | Principal | 7% | 77% | 10% | 6% |
| | School | 6% | 78% | 8% | 7% |


Growth on State Assessments by District
District 01
District 02
District 03
District 04
District 05
District 06
District 07
District 08
District 09
District 10
District 11
District 12
District_13
District_14
District_15
District_16
District 18
District 19
District 20
District 21
District 22
District 23
District 24
District 26
District 27
District 28
District 29
District 32
Number of Teachers
Growth Rating (0-20)


Growth on Comparable Measures by District
Number of Teachers
District_01
District_02
District_03
District_04
District_05
District_06
District_07
District_08
District_09
District_10
District_11
District_12
District_13
District_14
District_15
District_16
District_18
District_19
District_20
District_21
District_22
District_23
District_24
District_25
District_26
District_27
District_28
District_29
District_32

# Table Talk #3 (15 min)

Consider the following two scenarios:

- **Scenario 1: Student Growth remains an advisory measure, or are removed entirely and permanently from educator evaluation**
  - If so: what (if anything) replace them as measures of student learning and educators' contributions to it (for example, SLOs)? Or, focus only on professional practice measures?
- **Scenario 2: Expand Student Growth to other assessments (not just state tests)**
  - This can be done, to some extent, with end-of-course exams (e.g., Hillsborough County)
  - Takes LOTS of time and $$ - and may not address other concerns (even the best assessment only measures a slice of what students know, etc.)
  - Almost impossible to not set some limits (approved assessment lists) – that is, there's no way to make multitudes of classroom assessments comparable (comparable item difficulty, ensuring alignment to standards, etc.).

# Guiding Questions

- How does each scenario address identified barriers?
- What concerns do you have with each scenario?
- Is there one scenario that your group recommends?

# Agenda

| Time | Topic |
|---|---|
| 9:00 am | Welcome and introductions |
| 9:15 am | Reviewing inputs from Webinar #1 |
| 9:45 am | The ideal student learning component |
| 11:30 am | Lunch |
| 12:30 pm | How can the consistency in the implementation of SLOs across LEAs and schools be improved? |
| 1:30 pm | How  can the assessment quality be balanced with inclusion of additional assessments in SLOs? |
| 2:15 pm | Student Growth Measures |
| **2:45 pm** | **Closing & Next Steps** |

# Closing

- Next steps: summarizing workgroup recommendations for student learning component of evaluation system
- Next topic: Educator practice component and other measures
- Next webinar: February 14th, 3:30-5:00pm
- Next in-person meeting: March 7th