Westat

Improving lives through research™

New York State Department of Education Educator Evaluation Workgroup: Summary Report

Authors

Mary Wolfson
Clarissa McKithen
Darcy Pietryka

May 17, 2019

Prepared for:
New York State Department of Education
89 Washington Avenue
Albany, New York 212234

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Contents

Appendix

New York State Department of Educator Evaluation Workgroup: Summary Report

The New York State Department of Education (NYSED) contracted Westat to assemble and facilitate three in-person and three webinar-based Evaluation Workgroup sessions. The Evaluation Workgroup was comprised of educators and other stakeholders representing districts and Boards of Cooperative Educational Services (BOCES) from across the state as follows: teachers (36 percent), administrators (31 percent), superintendents and other district-level administrators (17 percent), and other stakeholders (largely union representatives, 17 percent).[1] The charge of the workgroup was to provide recommendations to improve the Annual Professional Performance Review (APPR), the current educator evaluation system in New York.

This report summarizes the processes, inputs, and recommendations of the NYSED Evaluation Workgroup. First, we describe processes for each workgroup session in detail. This is followed by a summary of the synthesized feedback and recommendations generated by the workgroup across sessions. Finally, the inputs for each workgroup session are presented in the appendices.

Evaluation Workgroup Processes

Session 1: The first Evaluation Workgroup session was an all-day, in-person session held on November 16, 2018. The objectives for the session were as follows:

- Establish a shared understanding of the legislation around educator evaluation in New York;
- Establish a shared understanding of the purpose and roles of the NYSED Evaluation Workgroup;
- Provide feedback and recommendations on the purpose and use of evaluation system;
- Identify barriers to an ideal evaluation system, to inform future workgroup sessions;
- Learn about personalized professional development and its connection to the evaluation system;
- Discuss how data inform personalized professional development;

[1] Percentage totals do not equal 100 because the figures were rounded up.

- Share thoughts on perceptions and uses of personalized professional development; and
- Discuss next steps.

At the start of the session, representatives from NYSED provided participants with an overview of the APPR implementation timeline; key components of the APPR (i.e., educator practice and student learning) per state legislation; the intended purposes of the evaluation system (i.e., informing employment decisions and teacher and principal development); and the 4-year transition plan for the APPR. Having established a shared understanding of the current legislation around educator evaluation, the workgroup transitioned to establishing a shared understanding of its purpose, which was to create recommendations on educator evaluation to guide implementation and inform policy, regulations, and/or legislative changes. The workgroup also discussed Westat's role, which was to facilitate the workgroup sessions, serve as thought partners to the workgroup, and summarize the workgroup discussions and recommendations. The morning of session 1 concluded with two small group discussions. During the first small group discussion, table teams discussed, documented, and shared their group's perspective on the purposes and uses of an ideal educator evaluation system. During the second small group discussion, table teams discussed, documented, and shared their group's perspective on current and/or potential barriers to an ideal educator evaluation system. Feedback from the small group discussions was captured on small group discussion record sheets that were provided to each table team. These were collected by Westat at the conclusion of the morning and were used to inform the summary of feedback and recommendations presented in the next section.

In the afternoon, two educators from Broward County Public Schools were scheduled to give a presentation on personalized professional development plans, to provide an overview of the benefits of professional development plans, components of a personalized plan, and factors to be considered when developing plans for participants. This presentation was designed to inform a second set of small group discussions in the afternoon focused on how student data drive personalized professional development, based on workgroup members' experiences with the current evaluation system, and recommendations on how to improve both the perception and effectiveness of personalized professional development plans. However, the workgroup chose to extend the discussion from the morning based on participants' expressed interest in continuing the dialogue around perspectives on the purposes and uses of an ideal educator evaluation system and the current and/or potential barriers to such a system. Workgroup members therefore returned to their table teams in the afternoon to continue their small group discussions. Feedback from the afternoon discussion was captured on chart paper that each table team received to summarize its feedback. These pages were subsequently transcribed by Westat and used to inform the summary of feedback and recommendations presented in the next section. The first session concluded

with a discussion of next steps, which included an Evaluation Workgroup webinar in December and another in-person meeting in January.

Session 2: On December 20, 2018, Westat facilitated the second Evaluation Workgroup session. This session, which was webinar based, focused specifically on the student learning component of the APPR. This session served as a segue into the third workgroup session, but also gave participants an additional opportunity to clarify feedback provided during the first workgroup session. The objectives for session 2 were as follows:

- To understand the inputs provided during session 1 and how they informed the planning for the remaining sessions;
- To know the requirements for student performance measures specified in Education Law §3012-d;
- To know the history, components, opportunities, and challenges around student learning objectives and student growth measures; and
- To discuss what an ideal student learning component would look like in an APPR evaluation system.

This 90-minute session began with a summary of the feedback provided in the first session. To ensure the accuracy and clarity of the feedback summary, participants were given the opportunity to comment on ideas that were not captured and/or not captured accurately. Next, representatives from NYSED provided an overview of the requirements for student performance measures specified in Education Law §3012-d. This information was followed by a presentation on the history, components, opportunities, and challenges associated with student learning objectives and student growth measures from the perspective of a national expert. The workgroup was then divided into virtual break-out rooms that included approximately five workgroup participants, a representative from NYSED, and a Westat facilitator. The break-out groups discussed what an ideal student learning component would look like in an APPR evaluation system. During the discussion, group members were probed around what they would change about the purpose of the student learning component, the components of the student performance category, the processes of the student performance category, and the use of data from the student performance category. These discussions were audio recorded and transcribed by Westat and used to inform the summary of feedback and recommendations presented in the next section. Session 2 concluded with a discussion of next steps, which included a high-level overview of the next session (to be held in person).

Session 3: The third session took place on January 10, 2018. This all-day, in-person session was designed to focus on student performance measures and technical requirements for the inclusion of assessments in evaluation. Major objectives for session 3 were to receive input from participants on the ideal student learning component for educator evaluation and to gather recommendations on ways to improve the existing student learning component. However, in response to the workgroup's interests, much of the conversation focused on participants' perspectives on the ideal student learning component, with particular emphasis on NYSED's, or the state education agency's (SEA's), and the local education agencies' (LEAs') respective roles in the implementation of an ideal evaluation system. Similar to the first session, the morning began with the sharing of foundational information on the student learning component requirements specified in the legislation, which was followed by a small group discussion around the ideal student learning component. Feedback from the small group discussions was captured on small group discussion record sheets that were provided to each table team. These were collected by Westat at the conclusion of the morning and were used to inform the summary of feedback and recommendations presented in the next section. In the afternoon, workgroup members returned to their table teams to continue their small group discussions. These small group discussions were more focused on the role of the SEA and LEAs in the implementation of an ideal evaluation system. Feedback from the afternoon discussion was captured on chart paper that each table team received to summarize its feedback. These pages were subsequently transcribed by Westat and used to inform the summary of feedback and recommendations presented in the next section.

Evaluation Workgroup Feedback and Recommendations

The Purpose and Use of an Ideal Educator Evaluation System

An ideal evaluation system is one that centers both the teachers' and students' learning and development.

Based on the discussions over the course of the sessions, it seems that the purpose and use of an ideal evaluation system overlap. The overarching theme throughout these discussions was that ***an*** ideal evaluation system is one that centers both the teachers' and students' learning and development. Specifically, educators seek an evaluation system that ***enhances teaching practice,*** especially through professional learning and growth opportunities and through meaningful, unbiased feedback from evaluators. Educators also believe that the ideal evaluation system ***encourages thoughtful self-reflection*** that drives growth and improves practice, while simultaneously supporting collaboration

(e.g., among educators and/or between practitioners and administrators/evaluators), the sharing of best practices, and opportunities for practitioners to assist one another.

Primarily, though, the ideal evaluation system ***benefits students*** by improving the quality of instruction they receive and enhancing their overall educational experience, as opposed to just improving their test scores. Such an evaluation system emphasizes equity and seeks to provide all students with access to high-quality teaching and learning. At the same time, the system is sensitive and responsive to the various factors (e.g., community, income, familial) that might affect students' learning.

Current and/or Potential Barriers to an Ideal Educator Evaluation System

During the workgroup sessions, participants also discussed some of the barriers that prevent the current evaluation system from being an ideal system. These discussions were organized around the two main components of educator evaluation in New York: (1) student learning and (2) observation of educator practice. Participants discussed each component and identified specific challenges and barriers related to each. This section summarizes these group conversations.

Student Learning

While a minority of workgroup members expressed the belief that the student learning component should not be linked to teacher and principal evaluation and called for its elimination from the evaluation system, the majority expressed concerns that should be addressed to improve the student learning component. Both groups voiced concerns about the state assessments that inform the student learning component.

The workgroup members who advocated for the elimination of the student learning component felt that assessments, overall, are not reliable measures of student learning. The workgroup members who supported the use of assessments expressed the belief that state assessments are helpful tools to analyze student performance and address problem areas, but they did not think that state assessments are ideal for high-stakes teacher and principal evaluations. Further, some workgroup members expressed concern with the weight (50 percent) assigned to the student learning evaluation component. Some workgroup members commented that such a heavy weight raises stakes even further and may result in teachers focusing too much on scores rather than on "what they're meant to do, which is to help all students."

Regardless of whether they supported the student learning component of educator evaluation or advocated for its elimination, all workgroup participants shared concerns with the current assessment system. First, they noted their concern that state assessments may not accurately reflect the complex nature of teaching

and learning, explaining that the assessments do not necessarily capture the progress made between the beginning and end of a school year (especially for students performing far below grade level). Some participants also suggested that NYSED extend its current phase-in process (e.g., for an additional 3 years), which may give the state additional time to develop "exams that educators and students can trust." These beliefs about assessments surfaced in specific concerns with both the student growth model and student learning objectives.

Student Growth: The workgroup's expressed lack of confidence in the state assessment system to capture the complete nature of learning led to the recommendation that the student learning component, if it must be a part of the system, be "contextualized" (i.e., that it take into account students' unique circumstances). Specifically, workgroup participants expressed some concern that the student growth measure does not take into consideration student characteristics (e.g., English language learner, disability status, poverty, student mobility) and other external factors outside of a teacher's control that might impact student achievement. Some educators are under the impression that, as a result, teachers of certain student populations are less likely to receive favorable scores on the student learning component, which may discourage teachers from teaching these students. However, these concerns and those listed above signify some miscommunication about the evaluation system, given that the state's student growth model includes considerations for prior academic history, disability status, poverty status, and status as an English language learner, and suggest that further engagement and learning about the state's student growth model could improve educators' understanding of one part of the student learning component.

Despite workgroup members' concerns that the state's student growth model does not take into account student characteristics, the model includes considerations for prior academic history, disability status, poverty status, and status as an English language learner.

Student Learning Objectives: Workgroup participants had similar ideas about student learning objectives (SLOs) and expressed reservations about the validity (e.g., lack of research indicating effectiveness), methodology, and usefulness of SLOs, particularly in the APPR system. They explained that SLOs may be better suited for progress analysis and educator planning. For some participants, the SLO process, as it stands, forces educators to focus on easily quantifiable learning targets, rather than richer, more complex targets that promote student learning, and focuses too much on getting an effective rating rather than improving teaching practice and/or student performance. Such practice may also be related to educators' contention that they do not have enough knowledge about students at the start of the school year to set useful SLO targets. As a result, the goal-setting process feels arbitrary for some teachers and does not

help them reflect on practice. Workgroup participants also noted that SLOs present challenges for districts with high student turnover, since the cohort for which the goals are set at the start of the year is likely to change significantly by the end of the year.

To address the perceived barriers listed above, workgroup participants offered various suggestions to establish a more ideal student learning component, as follows:

- Inclusion of multiple measures of student growth and achievement, which might include portfolios, special projects, and other evidence of academic learning among students, and
- Focus on educators' inputs and processes to help student learning and growth rather than the outcomes that students produce ("What is a teacher giving a child despite what the child is able to produce?"), placing greater emphasis on factors such as student engagement and classroom/school climate.

> *"The SLO process, as it stands, forces educators to focus on easily quantifiable learning targets, rather than richer, more complex targets that promote student learning."*
>
> **Evaluation Workgroup Participant**

Observation/School Visits

Workgroup participants seemed generally satisfied with the observation component of the evaluation system; however, some participants requested that educators have more flexibility in exemplifying their teaching practice. This suggestion may be in response to NYSED's list of "prohibited elements" for teacher and principal evaluation, including lesson plans, portfolios and other artifacts, professional goal setting, and parent and student feedback, as outlined in the APPR Guidelines (see Appendix). Also, in earlier evaluation systems, tenured teachers were allowed to both have traditional observations and use artifacts and goal setting to demonstrate effective practice, which workgroup participants believe empowered teachers and addressed some administrators' concerns about lack of expertise to evaluate all teachers. Participants also explained that some aspects of professional performance cannot be captured in an observation. Therefore, it is important to incorporate other sources of evidence that are more reflective of the entirety of the work that educators do in their buildings and classrooms.

> *The majority of evaluation workgroup participants also advocated for more differentiation in the frequency of observations, particularly for experienced and proficient educators, since less time spent observing these teachers provides greater time to support teachers who need it most.*

A large proportion of evaluation workgroup participants also advocated for more differentiation in the frequency of observations, particularly for experienced and proficient educators, since less time spent observing these teachers provides greater time to support teachers who need it most. However, it should be noted that the duration and frequency of observations and schools visits is within local discretion (see APPR Guidelines in Appendix).

Evaluator training on the observation component of educator evaluation also surfaced as an area for improvement within the evaluation system. Workgroup participants highlighted a need for comprehensive training of evaluators, particularly to help them observe classrooms made up of high-need populations (e.g., English language learners or students with disabilities) and those implementing Next Generation Learning Standards. Participants also recommended enhancements to other fundamental components of observation training, such as inter-rater reliability and observer calibration. In addition, workgroup participants mentioned a need for more accessible training. For example, in some parts of the state, administrators hired during the school year have to wait until the next training cycle to receive training, and all administrators have to travel off site for training. Virtual training and recertification for evaluators was discussed as a solution to increase the accessibility of evaluator training. Finally, in addition to training for evaluators, participants mentioned that educators need training on the evaluation system and the standards used for their evaluations so that they can be informed participants in the evaluation process.

Educators need training on the evaluation system and the standards used for their evaluations so that they can be informed participants in the evaluation process.

The Role of the SEA and LEAs in an Ideal Educator Evaluation System

The majority of workgroup participants expressed the belief that, in an ideal evaluation system, LEAs would have greater control over the development and implementation of the system. Educators argued that, given the current lack of trust in the existing state-level system, local control might restore trust and confidence. Local control would also give districts the opportunity to develop a system that reflects the needs and values of their respective learning communities and would allow those who are being evaluated to provide input on the system. Workgroup participants clarified, however, that in an ideal system the state would also play a role.

During the second in-person session, participants worked in groups to identify the ideal roles of each entity – SEA and LEA. Broadly speaking, educators believed that LEAs should determine the appropriate

measures for the evaluation system, based on local needs; select the assessments/evidence of student learning and rubrics of professional practice; determine the weight of the student learning component (and other components) of the evaluation system; and decide how to use the evaluation data and make decisions based on the data (e.g., human capital management decisions). The ideal role of the SEA, on the other hand, would be to establish and oversee districts' implementation of teaching, leading, and learning standards and policies; manage the approval of assessments and rubrics for evaluation purposes; monitor districts' established evaluation systems; provide technical assistance and training to help districts successfully implement their evaluation systems; and share state-collected data for districts' use (e.g., for analysis and comparison).

Conclusion and Next Steps

Shortly after the third Evaluation Workgroup session, the New York State Legislature passed updated educator evaluation legislation that was subsequently signed into law by the Governor. Based on this shift in the policy context, the workgroup was paused and subsequently disbanded. Given that this new legislation took effect upon signature by the Governor, NYSED is currently updating its educator evaluator regulations to support LEAs in implementing the New York educator evaluation system as specified in the new legislation.

The feedback and recommendations provided by the Evaluation Workgroup represent valuable insights into concerns about the educator evaluation system as it existed under Education Law §3012-d. While this legislation has been updated, many of the insights voiced by the Evaluation Workgroup are still valuable to school and district leaders who are seeking to improve their implementation of the educator evaluation system under the new legislation. The value of this summary report is in the insight it provides into educators' experience with the components, processes, and dynamics that remain central to educator evaluation.

Appendix A

Evaluation Workgroup Session 1

New York State
EDUCATION DEPARTMENT
Knowledge › Skill › Opportunity

NYSED Evaluation Workgroup

Session #1: Appropriate Purposes and Uses for Evaluation

Welcome!

APPR Timeline

Timeline

New York State's Evaluation System

2010:

- Governor signs Chapter 103 of the Laws of 2010; adding §3012-c, which establishes a comprehensive evaluation system for teachers and principals, effective July 1, 2010.
- USDE announces that New York is selected for a RTTT award of approximately $700M.

2011-12:

- First year of State-provided growth score results for all 4-8 ELA and math teachers and their building principals.
- Evaluations for teachers and principals are conducted in some NYS districts (e.g., School Improvement Grant and Teacher Incentive Fund recipients).
- Evaluation Law is revised. Governor signs the bill into law on March 27, 2012 (Chapter 21 of the Laws of 2012). Board of Regents adopts emergency regulations to conform to the major 2012 legislative changes.

2012-13:

- All NYS districts must have an approved APPR plan by January 17, 2013 or risk state aid increases.
- Evaluations for teachers and principals are done in all districts except for NYC. NYC is required by law to have a State-imposed evaluation plan.
- Legislature further amends the Evaluation Law (Part A of Chapter 57 of the Laws of 2013).

2013-14:

- Second year of evaluations for all districts in NYS, except NYC. First year for NYC.
- Legislature further amends the Evaluation Law (Chapter 56 of the Laws of 2014)

Timeline

New York State's Evaluation System (cont.)

2014-15:

- Governor signs Chapter 56 of the Laws of 2015, establishing a revised evaluation system for teachers and principals (Education Law §3012-d).
- All districts are required to have an approved APPR plan under the new statute by November 15, 2015 or to have an approved Hardship Waiver.

2015-16:

- 18% (n=122) of districts have approved plans under Education Law §3012-d; 82% (n=567) remain under Education Law §3012-c with an approved Hardship Waiver.
- All districts must have an approved APPR plan under Education Law §3012-d by December 31, 2016.
- At its December meeting, the Board of Regents adopts a transition period during which time the results of the grades 3-8 ELA/math State assessments and any State-provided growth scores are to be used for advisory purposes only. Separate transition evaluations that exclude these measures will be provided to affected educators.

2016-17:

- First full year of implementation of Education Law §3012-d.

2017-18:

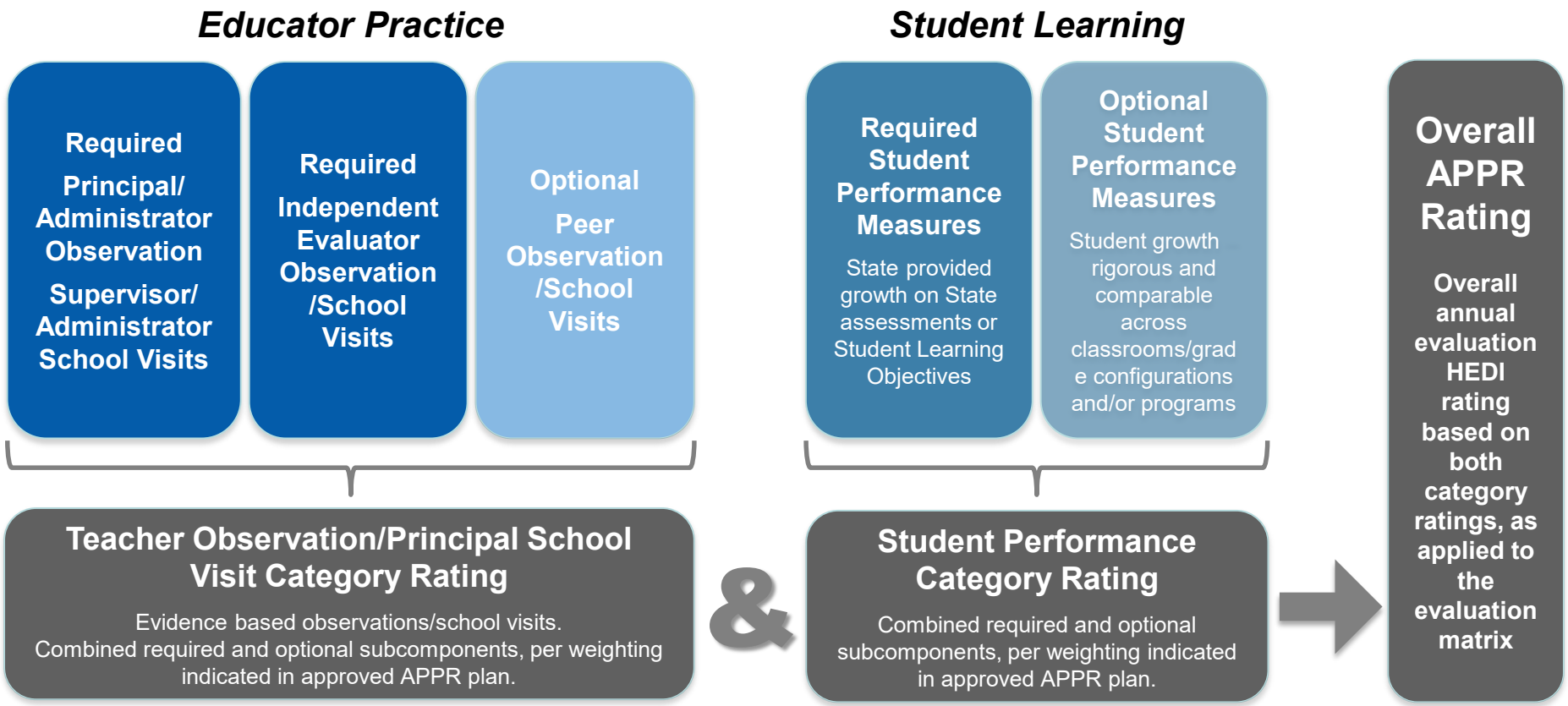
- Second full year of implementation of Education Law §3012-d.
- Commissioner's APPR Survey released to field.
- On November 5, 2018, Board of Regents announces additional one-year extension of APPR Transition Period (through June 2020).
- Regulations will be brought to Board of Regents at its December meeting.

Education Law §3012-d APPR Components

Education Law §3012-d

Components of the APPR Evaluation System

- Evaluations include educator practice and student learning measures
- Measures result in a single overall educator effectiveness rating



Education Law §3012-d

Components of the APPR Evaluation System

- **The overall APPR rating is determined by the statutory matrix:**

		Observation/School Visit			
		Highly Effective (H)	Effective (E)	Developing (D)	Ineffective (I)
Student Performance	Highly Effective (H)	H	H	E	D
	Effective (E)	H	E	E	D
	Developing (D)	E	E	D	I
	Ineffective (I)	D*	D*	I	I

Overview of APPR Transition Period Regulations (2015-16 through 2018-19 school years)

The Governor's Common Core Task Force "unanimously affirms the importance of adopting and maintaining high educational standards and rigorous performance measures to increase the competitive standing of, and therefore the opportunities for, all our students."

Richard Parsons, Chair, Common Core Task Force

Many of the recommendations in the Task Force report reflect areas the Board and Department have discussed and are taking action on.

- **The Task Force's charge did not include teacher and principal evaluation. However, as discussed, the Task Force advanced a recommendation stating "results from assessments aligned to the current Common Core Standards, as well as the updated standards, shall only be advisory and not be used to evaluate the performance of individual teachers and students."**

- **Reflective of statewide stakeholder input.**

For teachers and principals, the regulations:

- **Ensure that there will be no consequences for teachers and principals related to 3-8 ELA and mathematics state assessments and no growth score on Regents exams until the start of the 2019-2020 school year.**

- **Prohibit the use of results from the 3-8 state assessments for use in evaluating the performance of individual teachers, principals or students.**

- **Provide for a four year transition period for annual professional performance reviews (APPRs) while the State completes the transition to higher learning standards.**

- **During the transition period, transition scores and HEDI ratings will replace the scores and HEDI ratings for teachers and principals whose HEDI scores are based, in whole or in part, on State assessments in grades 3-8 ELA or mathematics (including where State-provided growth scores are used) or on State-provided growth scores on Regents examinations until the State's new system is fully phased in.**

- **For purposes of public reporting of aggregate data and disclosure to parents on request the original composite score and rating and the transition composite score and rating must be reported with an explanation of such transition composite score and rating.**

Why are we here?

Role of evaluation workgroup

- **Create recommendations on educator evaluation to**
  1. **Guide implementation**
  2. **Inform policy, regulations, and/or legislative changes**

- **Meeting dates**
  1. **Friday, November 16, 2018: Overall evaluation system**
  2. **Thursday, January 10, 2019: Student performance measures**
  3. **Thursday, March 7, 2019: Observation and other measures**

Why are we here?

Role of evaluation workgroup

- **Webinar dates**
    1. **December 20, 2018: Webinar Session #2**
    2. **February 14, 2019: Webinar Session #3**
    3. **March 27, 2019: Final Webinar**

Why is Westat here?

- **Facilitate workgroup meetings**
- **Serve as thought partners**
- **Provide summaries of workgroup discussions and recommendations**

Let's Play Ball!

- **Say hello!**
  - **Name**
  - **LEA/Organization**
  - **Role**
- **Establishing norms for the day**

Session #1 Objectives

1. Share roles of the NYSED Evaluation Workgroup
2. Provide feedback and recommendations on purpose and use of evaluation system
3. Understand intended purpose and use of evaluation
4. Discuss and give feedback on each component of evaluation system to inform future meetings
5. Learn about personalized PD and its connection to evaluation system
6. Discuss how data informs personalized PD
7. Share thoughts on perceptions and uses of personalized PD
8. Discuss next steps

Agenda

Time	Activity
9:00	Welcome and introductions
9:45	Consider purposes and uses of evaluation
10:30	Reflect on experiences with evaluation and identify focus areas for future meetings
11:30	Lunch
12:30	Discuss personalized PD and data sources
2:45	Share next steps

Agenda

Time	Activity
9:00	Welcome and introductions
9:45	Consider purposes and uses of evaluation
10:30	Reflect on experiences with evaluation and identify focus areas for future meetings
11:30	Lunch
12:30	Discuss personalized PD and data sources
2:45	Share next steps

Partner Conversations and Share Out (12 min)

- **In your ideal evaluation system – conceptually, what would be the most important *purpose*? What would be the most important *use*?**

Current Evaluation System (8 min)

Two major purposes

1. Employment decisions
2. Teacher and principal development

Current Evaluation System

- **What does the legislation say?**
  - **Evaluation system = Student performance + Observation**
    - Student performance – Must use state-provided growth score, if available; otherwise, student learning objective
      - *Note optional second subcomponent may be used*
    - Observation – Conducted by supervisor and trained evaluator external to the school
      - *Note option to use trained peer rated effective or higher*
  - **Weighting and scoring ranges of components – Must be transparent at the beginning on the school year**
  - **Matrix – Must be used to determine composite score**
  - **Prohibited elements – Cannot be used in evaluation system**
  - **Student assignments – May not assign a student to two ineffective teachers for two consecutive years**

Group Discussion (15 min)

- **Overarching question: How do your experiences with the evaluation system match the legislation's description?**
  - **Think about the formal and informal ways the evaluation system has shaped your experiences as an educator in NY.**
    - At the administrator level, how has this impacted any employment decisions?
    - How have your evaluations played a role, either formally or informally, in shaping your professional learning plans?
    - What other components do you think should be a part of an evaluation system?

Group Discussion (10 min)

- **In your ideal evaluation system – what would you add or change about:**
  - **The overall system purpose**
  - **The use of data from the system**
  - **The system itself (components, processes, etc.)**

  - ***Note: Future discussions and recommendations will be grounded by the shared agreement related to these***

Agenda

Time	Activity
9:00	Welcome and introductions
9:45	Consider purposes and uses of evaluation
10:30	Reflect on experiences with evaluation and identify focus areas for future meetings
11:30	Lunch
12:30	Discuss personalized PD and data sources
2:45	Share next steps

The Nitty Gritty of the Evaluation System: Group Conversation (15 min)

- **From your own experiences, how has your region/district/school implemented the evaluation system? Think about:**
  - **How the new system was communicated and rolled out**
  - **Who was involved in each phase of design and rollout**
  - **How staff were trained**

Note: This conversation should focus on the whole system, as we will have 30 minutes for table conversation to talk more specifically about observations and student performance.

Table Talk Part 1: Observations (15 min)

What issues or decision points would you like the Evaluation Workgroup to discuss and provide recommendations on related to the evaluation system's use of OBSERVATIONS?

Consider your thoughts on:

- **The role of**
  - **Trained administrators**
  - **Independent trained evaluators**
  - **Trained peer teachers**
- **Training for observers and staff observed**
- **How data are used**
- **Follow-up**

Table Talk Part 2 (15 min)

What issues or decision points would you like the Evaluation Workgroup to continue to discuss and provide recommendations on related to the evaluation system's use of STUDENT PERFORMANCE?

Consider your thoughts on:

- **State-provided growth scores**
- **Student learning objectives**
- **Locally selected measures of student growth**

Let's Share (15 min)

- **Share one issue or decision point that your table discussed to have the Evaluation Workgroup continue to collaborate on and provide recommendations on related to the evaluation system's use of**
  - **Observations, and**
  - **Student performance**

Agenda

Time	Activity
9:00	Welcome and introductions
9:45	Consider purposes and uses of evaluation
10:30	Reflect on experiences with evaluation and identify focus areas for future meetings
11:30	Lunch
12:30	Discuss personalized PD and data sources
2:45	Share next steps

Agenda

Time	Activity
9:00	Welcome and introductions
9:45	Consider purposes and uses of evaluation
10:30	Reflect on experiences with evaluation and identify focus areas for future meetings
11:30	Lunch
12:30	Discuss personalized PD and data sources
2:45	Share next steps


Personalized Professional Development Plans

Learning Outcomes

- Increase participant's understanding of personalized professional development plans.
- Provide relevant research in order for participants to gain a better understanding of the benefits of personalized professional development plans.
- Explore a variety of data types that may be factored when constructing personalized professional development plans.

The premise of andragogy (adult learning theory) tells us that adults want to have professional development opportunities that work for their individual learning styles, delivered in a timely fashion, fully supported in the implementation of the learning and not in a "one shot" approach.

Overview Of Personalized Professional Learning Plan

Research shows that when it is well designed and well implemented, professional learning has the power to strengthen practice and improve student learning.

January 2017 | A New Vision for Professional Learning Learning Forward EducationCounsel

Questions To Consider

ESSENTIAL QUESTIONS FOR PROFESSIONAL DEVELOPMENT:

1. **Why is professional development important?**
2. **How do we make professional development meaningful to each individual?**
3. **How do we ensure that we are meeting the needs of the diverse learning styles of the adults?**
4. **How do we create a safe environment for risk-taking, collaboration, and support for adult learning?**

Questions To Consider

ESSENTIAL QUESTIONS FOR PROFESSIONAL DEVELOPMENT:

5. **How do we continue to learn and grow to keep adapting and evolving in the professional development arena for the adult learners?**
6. **How do we ensure that all professional development is based on the needs of the whole child?**
7. **How do we ensure that professional development aligns with the goals of the district and schools?**

Personalized Learning Plan Components

- Multiple methods of teaching
- Development of professional knowledge and skills
- Focus on strengths and weaknesses
- Deepening understanding both content and research based approaches to teaching

Personalized Learning Plan Components

- Recognizes teachers as professionals
- Promote individualized improvement
- Must be learner centered
- Opportunity for feedback

Factors to Consider in Developing Personalized Development Plans

- Learning climate and culture
- Structure and content of the professional learning
- Timing, duration, and frequency of professional learning
- Use of feedback and data

What It's Not

- One shot professional development
- One shoe fits all
- Evaluative
- An "I got you"

Round Table Activity

What The Research Says

Enables educators to directly explore, discover, and adapt strategies needed to meet their individual needs.

It allows educators to practice new instructional strategies, and apply knew knowledge immediately in a classroom-based, or embedded learning environment, to address adult learning needs.

"What is Personalized professional Learning?"
By Kristi Meeuwse and Diane Mason June 28, 2017

Advice To Adult Learners

- Set a cooperative learning climate.
- Create mechanisms for mutual planning.
- Arrange for a diagnosis of learner needs and interests.
- Enable the formulation of learning objectives based on the diagnosed needs and interests.
- Design sequential activities for achieving the objectives.
- Execute the design by selecting methods, materials, and resources.
- Evaluate the quality of the learning experience while re-diagnosing needs for further learning.

Adapted from: Knowles, M. (1970). The Modern Practice of Adult Education: From Pedagogy to Andragogy.

Levels for Consideration when Evaluating Personalized Professional Development Plans

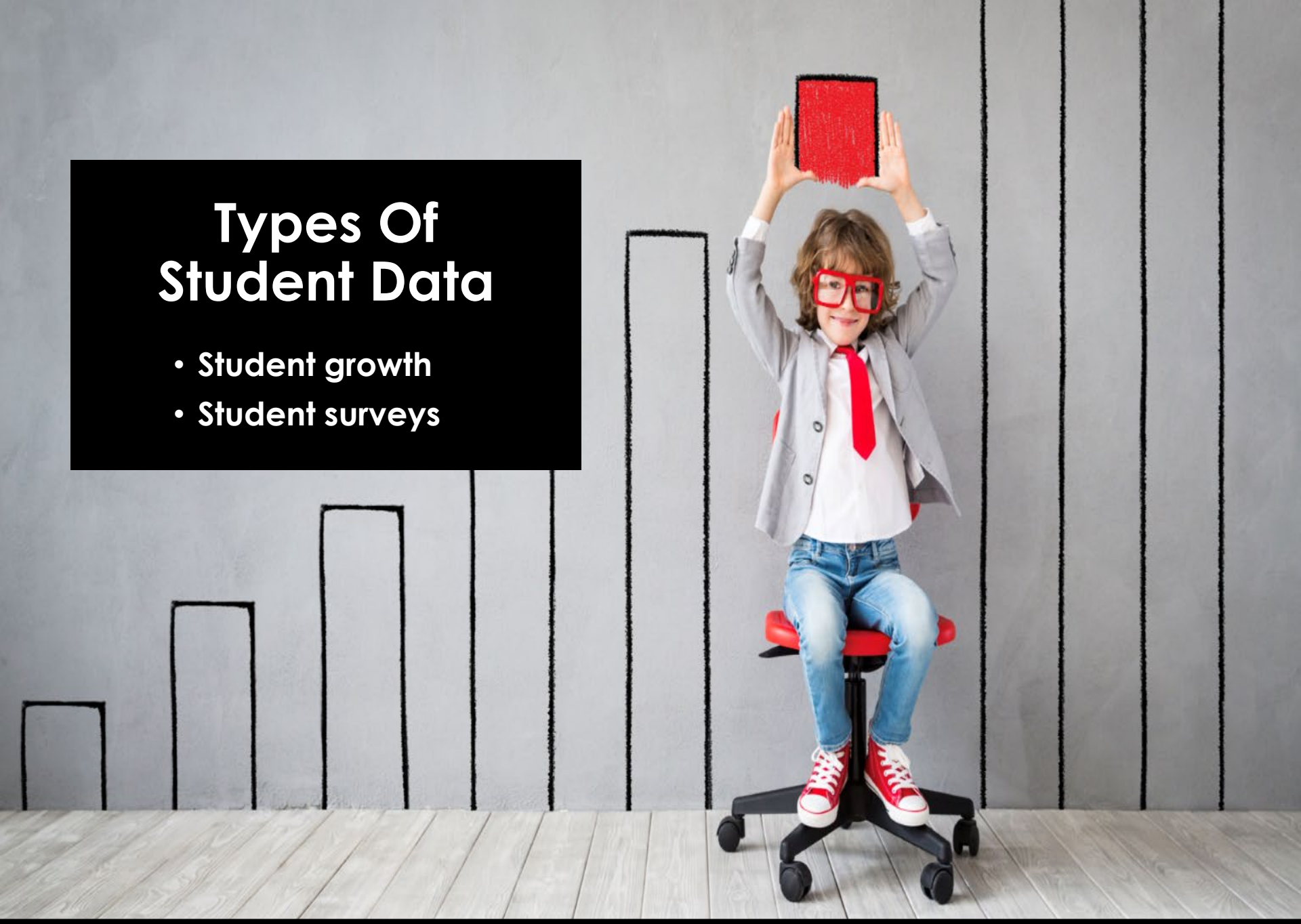
LEVELS	PURPOSE	BENEFITS
REACTION	Measures how those who participate in professional development programs react to it	1. Helps improve future training. 2. Creates trust in participants. 3. Quantitative information useful to managers during Observation Feedback 4. Establishes standards of performance
LEARNING	This level determines if the professional development program has: changed attitudes; improved knowledge; increased skills.	1. Measures effectiveness of instruction. 2. Measures specific learning (information, attitudes, skills). 3. Results = changes in instruction
CHANGE IN BEHAVIOR	Determines the extent to which behavior has changed as a result of the professional development program	1. Intrinsic rewards: self-esteem, empowerment if successful. 2. Extrinsic rewards: praise, promotion, salary .
RESULTS	What final results occurred because participants attended the professional development program?	1. Measurable increases in quality: teamwork; morale, safety.

Using Data To Inform Personalized Professional Learning Plans

Data-driven includes using a variety of sources and types of student, educator, and system data to identify learning needs, set goals, plan, assess, and evaluate professional learning, preferably in a cycle of ongoing learning and improvement.

Types Of Student Data

- Student growth
- Student surveys



Types of Assessment Data

- Formative
- Common assessments
- State
- Performance based assessments

Types Of Instructional Practice Data

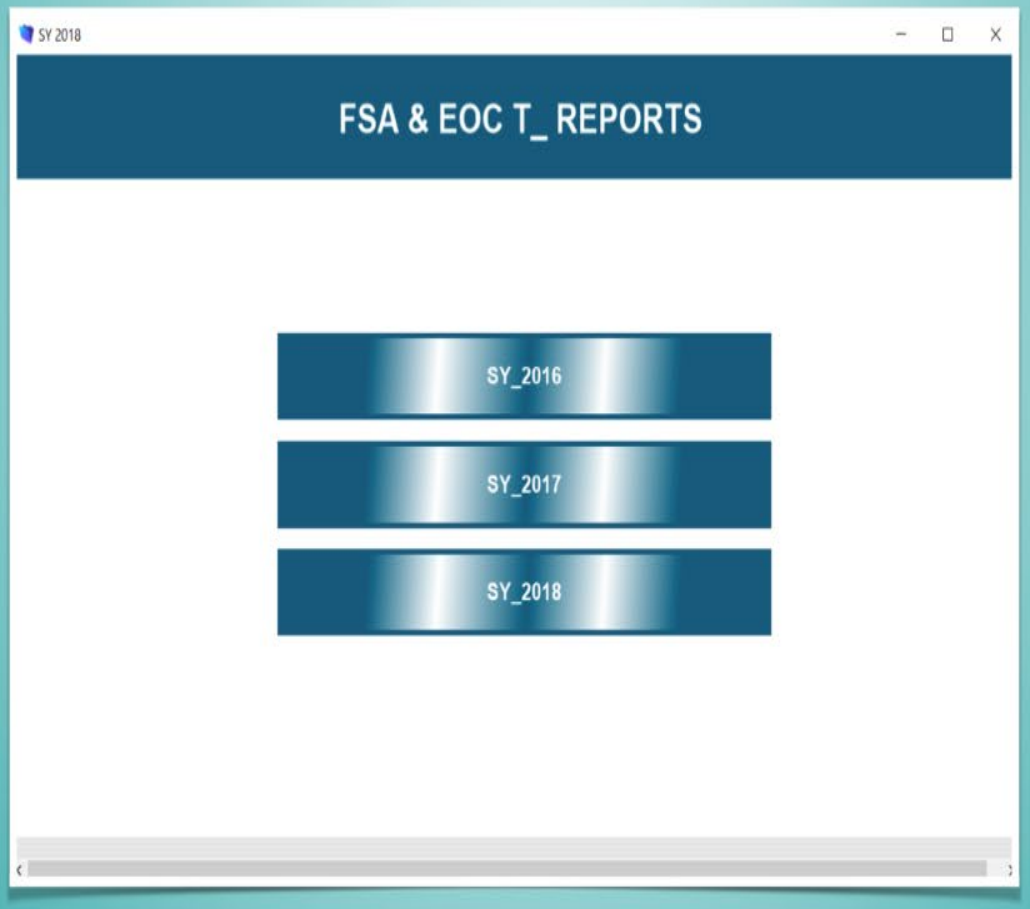
- School-wide growth
- Deliberate practice
- Results based on teacher performance
- Specific topics within the evaluation rubric


Types Of Content
Standard Data
• Tested categories
• Trend Data

Potential Elements for Personalized Professional Development Plans

- Define specific standard-based goals
- Time period (such as 12-month)
- Identify areas that can exhibit growth

Using Data To Drive Personalized Learning Plans



SY 2018

File Edit View Insert Format Records Scripts Window Help

Algebra 1 2018 Report by Cadre Director

LEGEND Algebra 1_Read_Scale_Score	
Level 1	425-486
Level 2	487-496
Level 3	497-517
Level 4	518-531
Level 5	532-575

LEGEND
Algebra 1_Ach_Level_Mean (2 Groups)
3+ - 3, 4, 5
<3 - 1, 2

Mean Percentage Earned by Content Area
Red Cells - Lower than 30%
Green Cells - Higher than 70%

FIND HOME

Number of Teachers (531)		
5	4	7
Percentage of Assessment		
41%	40%	19%

(<30%)

Cadre Director	School#	School	P#	Teacher Name	Course Title	Class Enrollment	ALGEBRA 1_Scale_Score_mean	ALGEBRA 1_Ach_level_mean 1-5	Algebra and Modeling	Functions and Modeling	Statistics and the Number System
									Mean % Earned by Content Area		
			00098831		ALGEBRA 1 HON	22	485.64	<3	21%	32%	15%
			00098831		ALGEBRA 1 HON	36	523.50	3+	53%	58%	31%
			00098831		ALGEBRA 1 HON	118	513.69	3+	46%	46%	23%

Professional Development Plans

SY_2017

File Edit View Insert Format Records Scripts Window Help

LEGEND ALGEBRA 1	
Level 1	425-486
Level 2	487-496
Level 3	497-517
Level 4	518-531
Level 5	532-575

Algebra 1 2017 Report by Cadre Director

FIND

Number of Algebra 1 Teachers (333)			
(Red <30%)	3	3	3

Mean Points Earned by Content Area			
Algebra_1_AchLevel_mean	Algebra and Modeling	Functions and Modeling	Statistics and the Number System
Number of Points Possible			
1-5	41%	40%	19%

Cadre Director	School#	School	P#	Teacher Name	Course Title	Class Enrollment	ALGEBRA_1_Scale Score_mean				
			00108821		ALGEBRA 1	91	499.77	2.56	31%	31%	22%
			00098831		ALGEBRA 1 HON	86	521.35	3.77	53%	47%	33%

SY_2016

FIND

LEGEND
ALGEBRA 1

Level 1	425-486
Level 2	487-496
Level 3	497-517
Level 4	518-531
Level 5	532-575

Algebra 1 2016 Report by Cadre Director

Number of Algebra 1 Teachers (381)			
(Red <30%)	0	0	1

Mean Points Earned by Content Area			
Algebra_1_AchLevel_mean	Algebra and Modeling	Functions and Modeling	Statistics and the Number System
Number of Points Possible			
1-5	40.5%	40.5%	19%

Cadre Director	School#	School	P#	Teacher Name	Course Title	Class Enrollment	ALGEBRA_1_Score_mean	Algebra_1_AchLevel_mean	Algebra and Modeling	Functions and Modeling	Statistics and the Number System
			00098831		ALGEBRA 1	70	510.74	3.23	37%	32%	28%
			00098831		ALGEBRA 1 HON	60	525.57	3.97	53%	43%	41%

Potential Elements for Personalized Professional Development Plans

- Timeline for achieving growth
- Manner in which growth will be assessed
- Activities to support growth in identified areas
- Artifacts that serve as benchmarks of growth

Potential Elements for Personalized Professional Development Plans

Final Stage should include:

- Meeting with supervisor
- Outline of Additional support
- Review plan and artifacts

Guided Discussion

What do you recommend to improve both the perception and effectiveness of personalized professional development plans?

Agenda

Time	Activity
9:00	Welcome and introductions
9:45	Consider purposes and uses of evaluation
10:30	Reflect on experiences with evaluation and identify focus areas for future meetings
11:30	Lunch
12:30	Discuss personalized PD and data sources
2:45	Share next steps

Appendix B

Summary of 2015 Evaluation Changes

Evaluation – Summary of Changes

- **The enacted budget created requirements and options for a new evaluation system administered by the Department in accordance with Commissioner's Regulations promulgated by the Board.**
- **The new evaluation system is comprised of two components that determine each educator's rating:**
  - **Student performance:** Requires the use of a state-provided growth score, if available; otherwise requires the use of a student learning objective (SLO). SLOs must use State assessments, as available.
    - If added by local collective bargaining, an optional second subcomponent could be used, comprised of an additional state-provided growth score on a state test or a growth score from a state-designed supplemental assessment calculated using a state-provided or approved growth model. These state-designed supplemental assessments include those developed, designed, purchased, or acquired by SED.
  - **Observations:** Requires observations by a supervisor and an independent evaluator from outside the school building.
    - Districts also have the option of having observations conducted by a trained peer who has been rated Effective or Highly Effective.

Evaluation – Summary of Changes

The statute mandates the "matrix" below to determine a teacher's composite score based on the two categories of the evaluation (see §3012-d (5)):

		Observation			
		Highly Effective (H)	Effective (E)	Developing (D)	Ineffective (I)
Student Performance	Highly Effective (H)	H	H	E	D
	Effective (E)	H	E	E	D
	Developing (D)	E	E	D	I
	Ineffective (I)	D*	D*	I	I

*** If a teacher is rated Ineffective on the Student Performance category, and a State-designed supplemental assessment was included as an optional subcomponent of the Student Performance category, the teacher can be rated no higher than Ineffective overall (see §3012-d (5)(a) and (7)).**

Evaluation – Summary of Changes

- The statute prohibits certain elements from being used as part of an evaluation, including:
  - Lesson plans, artifacts of teacher practice, and student portfolios;
  - Instruments of parent or student feedback;
  - Use of professional goal-setting;
  - Locally developed assessments not approved by the Department as a state-designated supplemental assessment; and
  - Growth or achievement targets that do not meet minimum standards established by the regulations of the commissioner.
- Districts will be prohibited from assigning a student to two Ineffective teachers for two consecutive school years.

Appendix C

Uses of Annual Evaluations

What are annual evaluations used for?

According to the NY law:

Evaluations shall be a **SIGNIFICANT FACTOR** for **EMPLOYMENT DECISIONS**, including but not limited to:

- promotion,
- retention,
- tenure determination,
- termination, and
- supplemental compensation.

Such evaluations shall also be a **SIGNIFICANT FACTOR** in **TEACHER AND PRINCIPAL DEVELOPMENT**, including but not limited to:

- coaching,
- induction support, and
- differentiated professional development.

Appendix D

Evaluation Workgroup Session 2

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

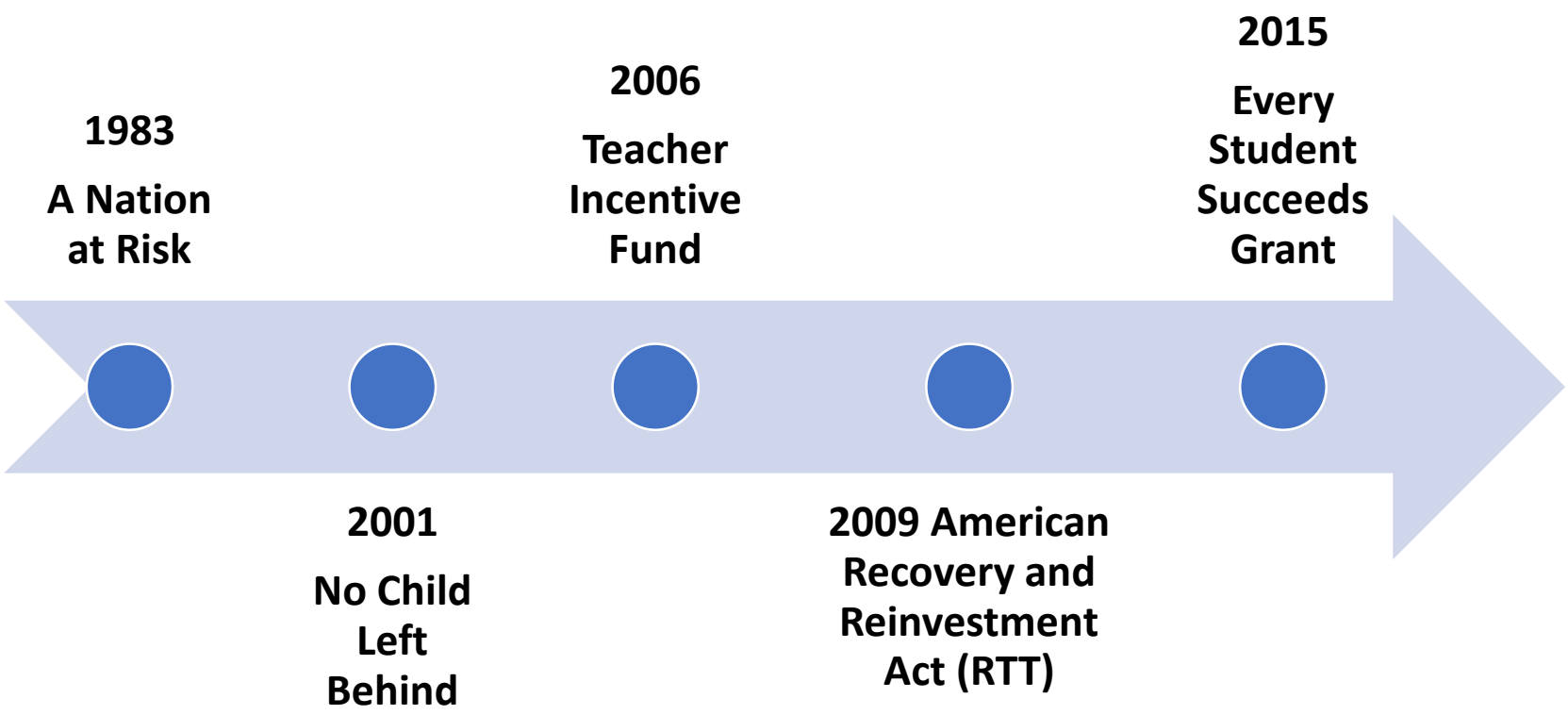
Time	Topic
9:00 am	Welcome and introductions
9:15 am	Reviewing inputs from Webinar #1
9:45 am	The ideal student learning component
11:30 am	**Lunch**
12:30 pm	How can the consistency in the implementation of SLOs across LEAs and schools be improved?
1:30 pm	How can the assessment quality be balanced 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

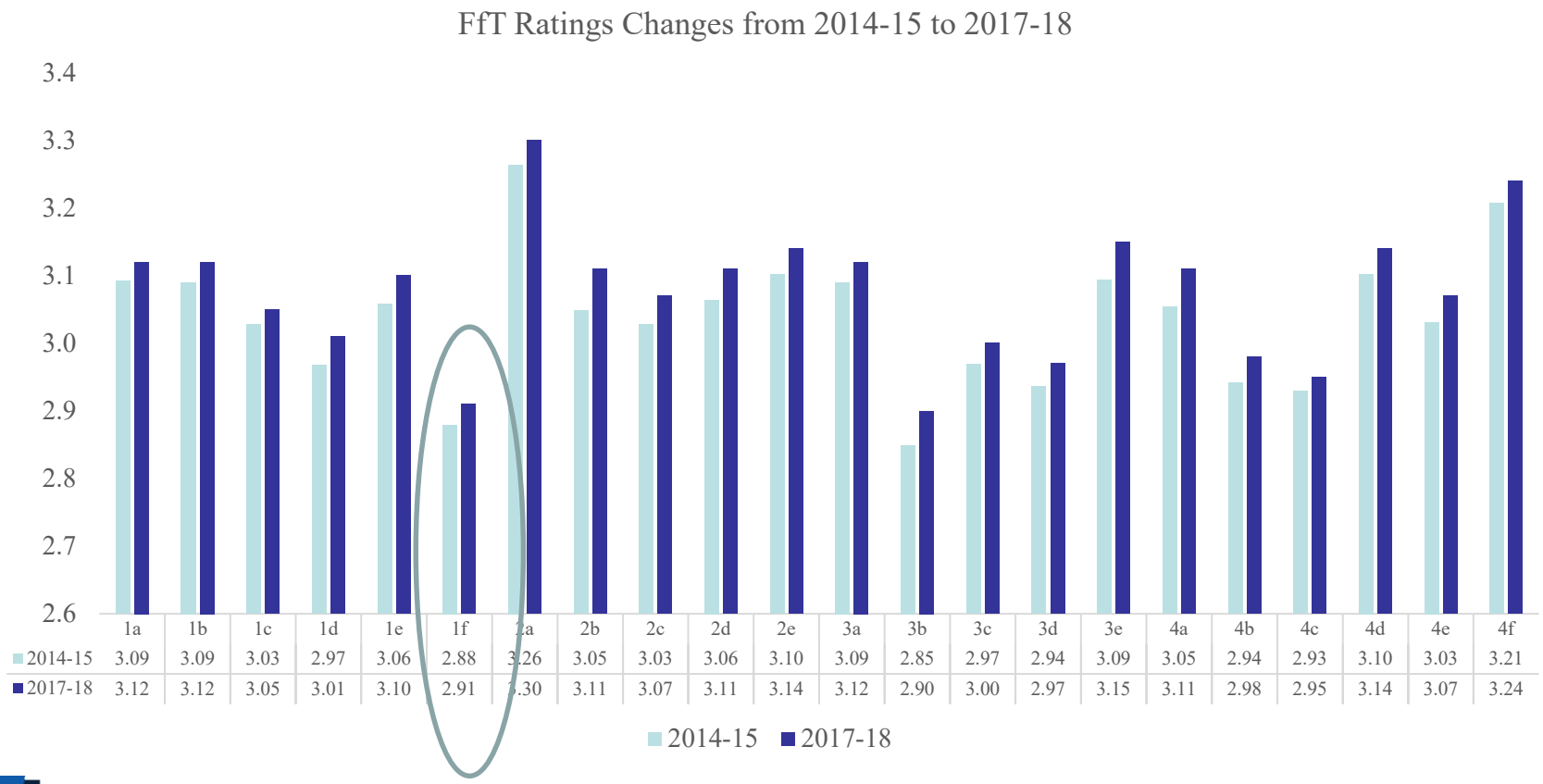
- **High-quality assessments must enable ALL students to demonstrate their knowledge (UDL principles):**
  - Precisely-defined constructs
  - Clear instructions
  - Maximum readability
  - Allowable accommodations for SwD and ELL
  - Items free of bias (DIF analysis)

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

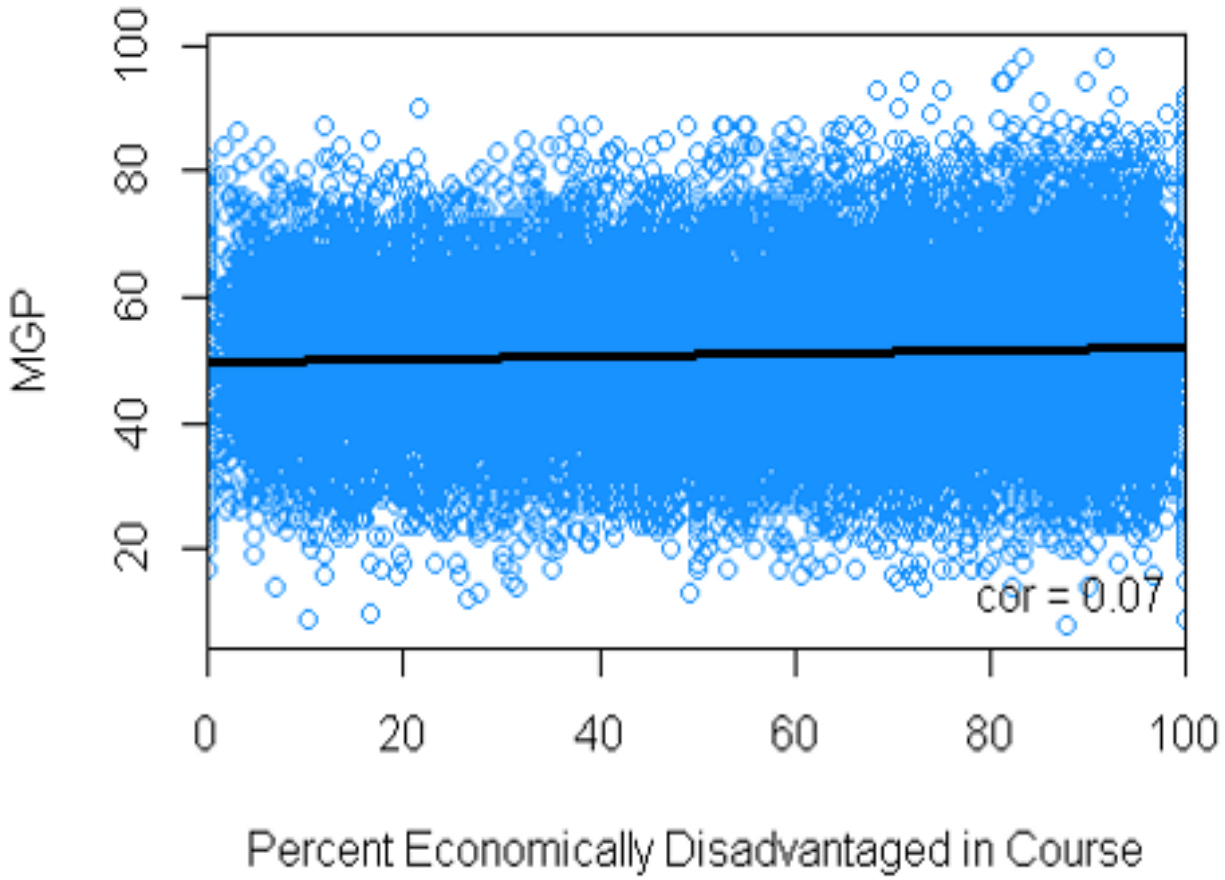
Statistical Growth Models (In Brief)

- 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)

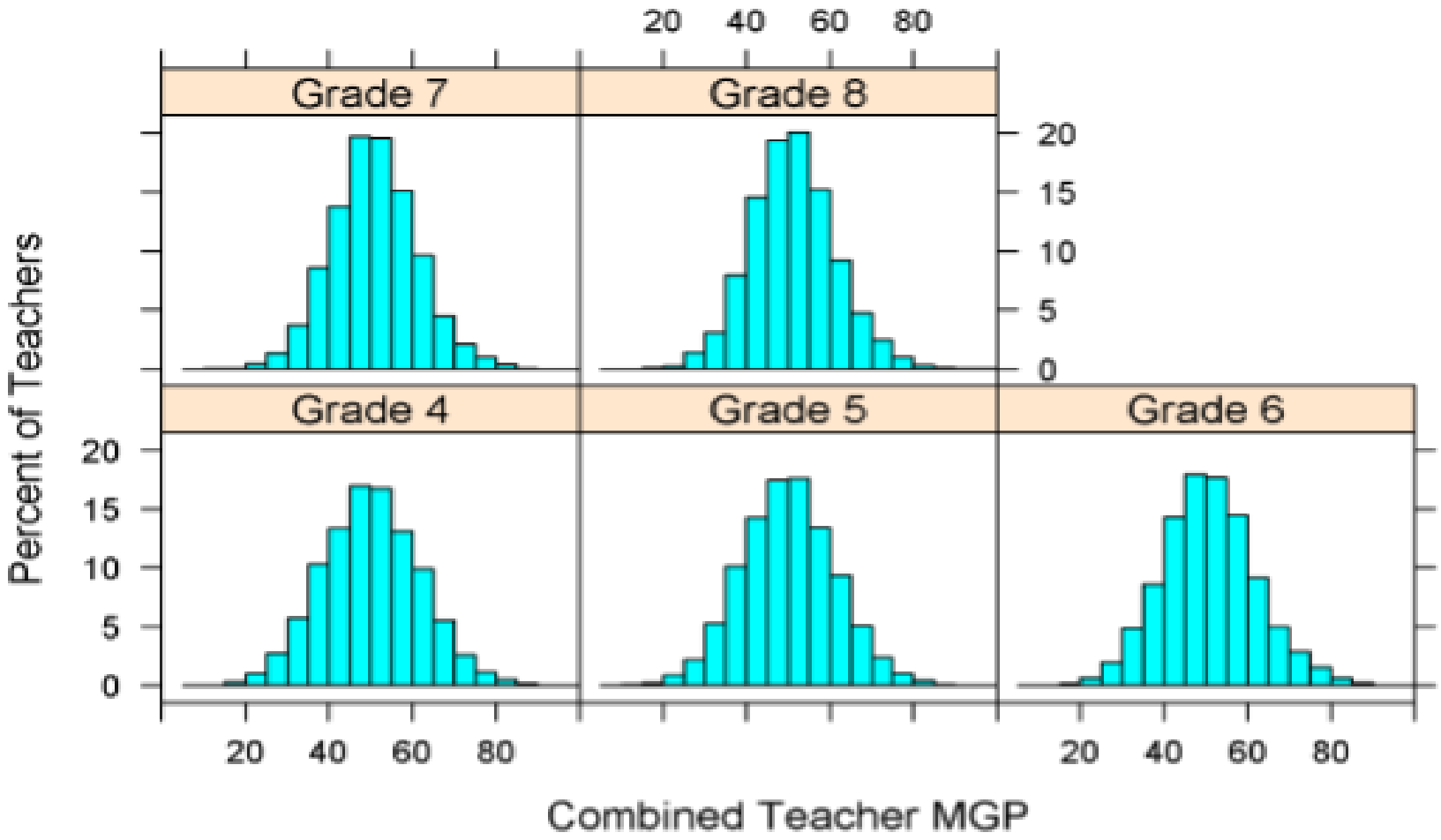
Figure 11. Relationship of Grades 4–8 Teacher MGP Scores to the Percentage of Economically Disadvantaged Students or Course



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



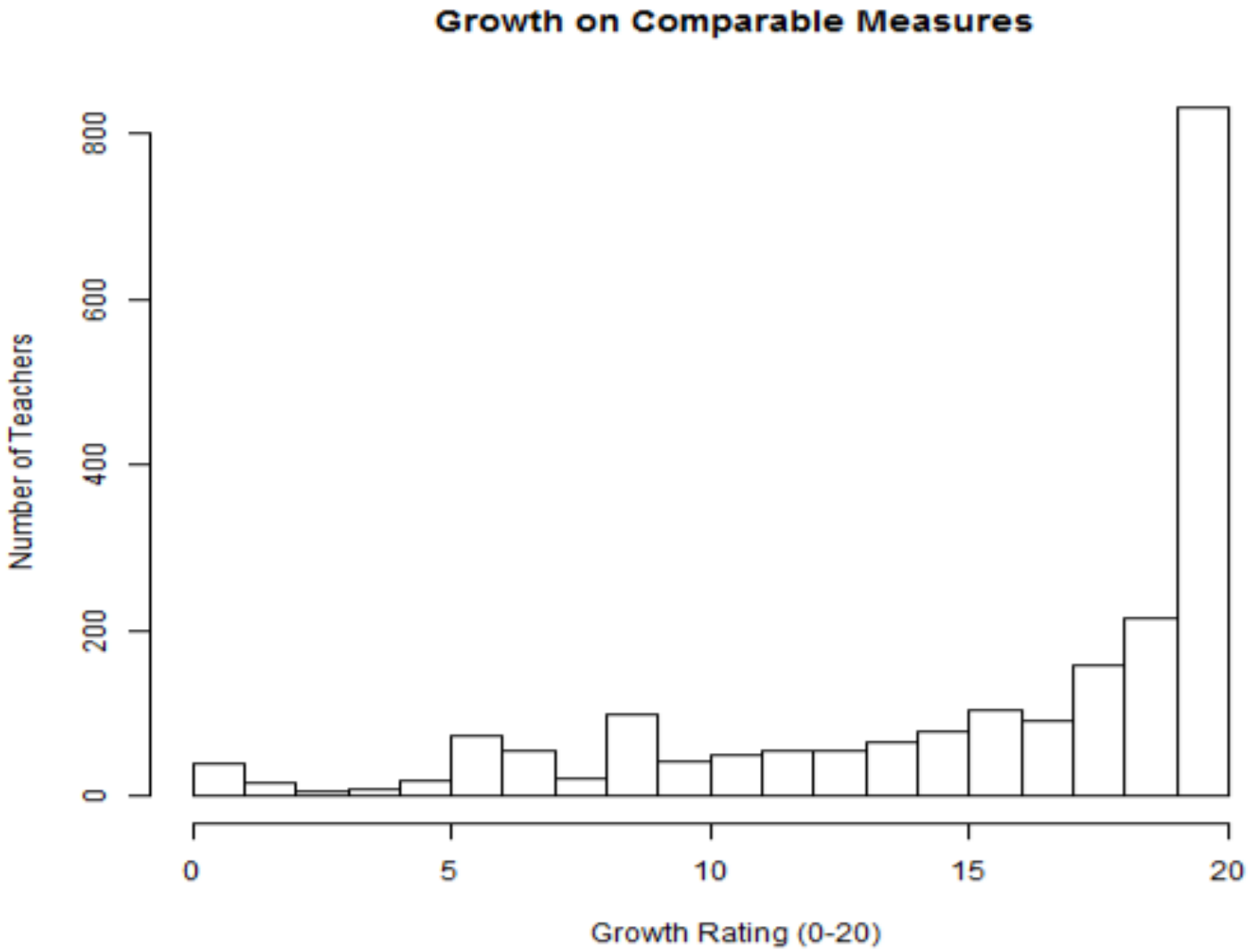


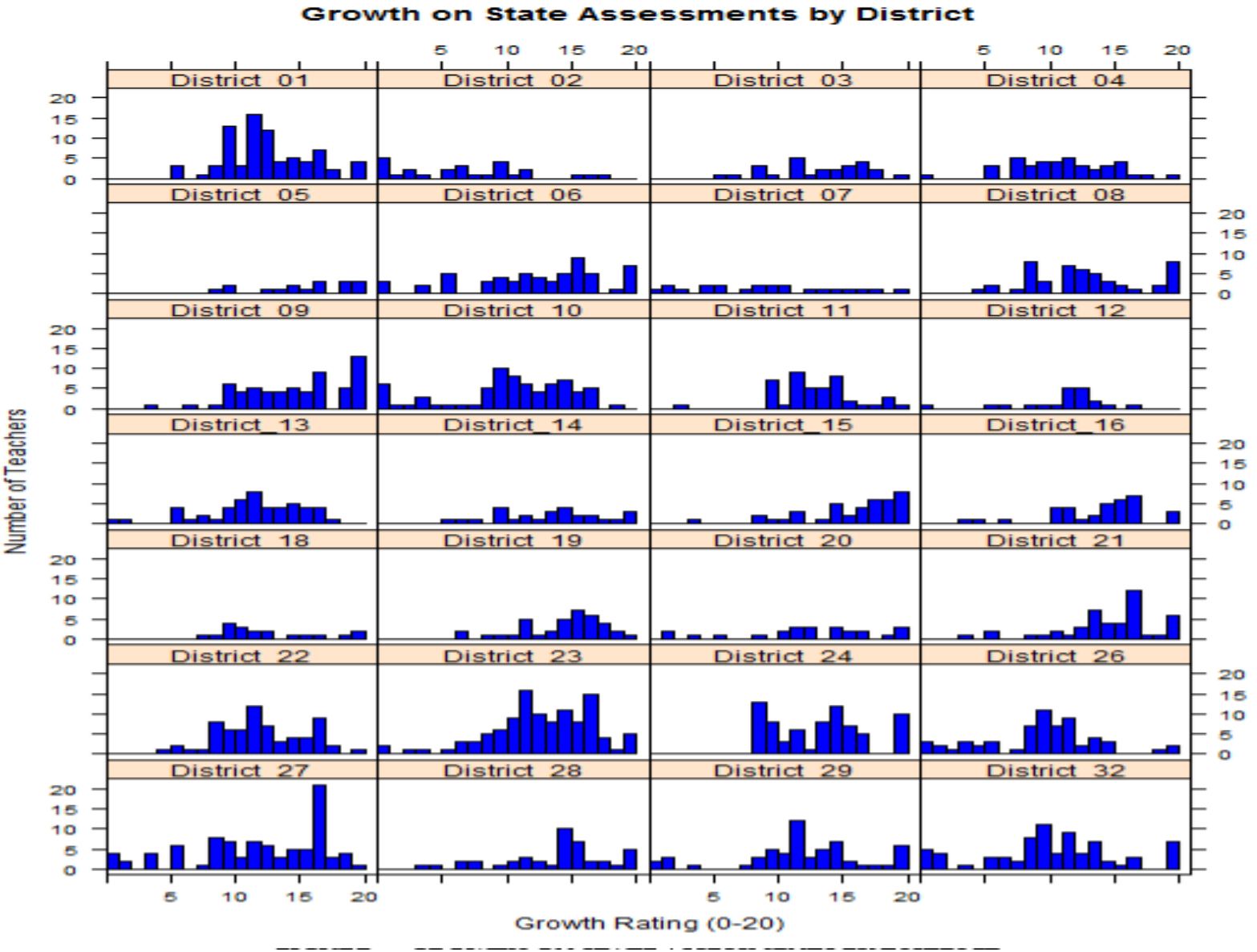
FIGURE 12: DISTRIBUTION OF GROWTH RATINGS ON COMPARABLE MEASURES MODELS

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)

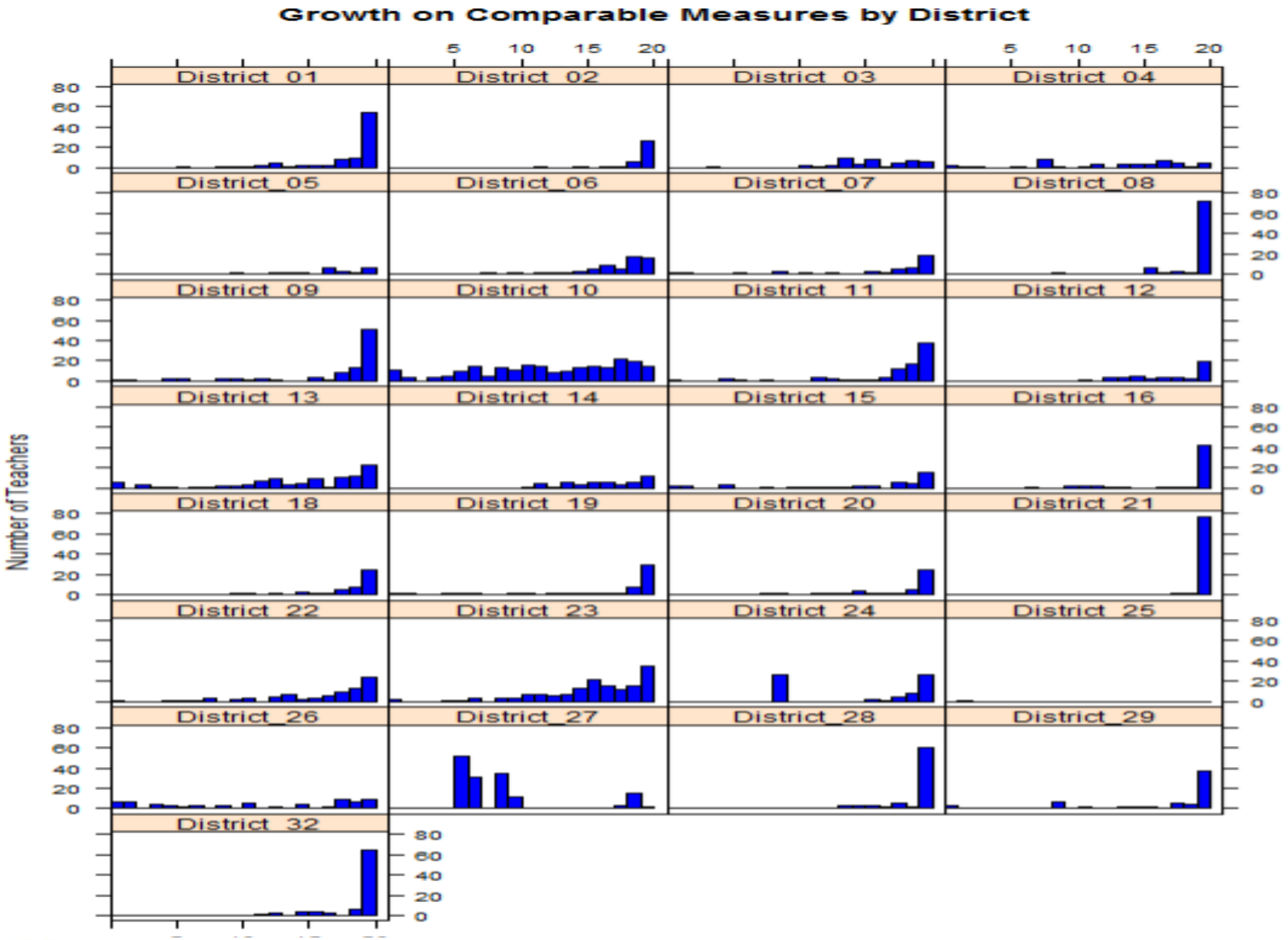

Growth on Comparable Measures by District
Number of Teachers
District_01
District_02
District_03
District_04
District_05
District_06
District_07
District_08
District_09
District_10
District_11
District_12
District_13
District_14
District_15
District_16
District_18
District_19
District_20
District_21
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

Appendix E

Evaluation Workgroup Session 3

New York State
EDUCATION DEPARTMENT
Knowledge › Skill › Opportunity

NYSED Evaluation Workgroup

Webinar #1

Welcome!

Agenda

- Welcome and Introductions
- Agenda and Objectives
- Session #1 Debrief
- Current practices and policies regarding SLOs and growth models in NY
- History and Review of Student Growth Models and SLOs
- Break-out group discussion and share out of ideal student learning component of evaluation system
- Closing and Next Steps

Objectives

- Understand the inputs provided during Session 1 and how they have informed the planning for the remaining two in person sessions
- Know the requirements for student performance measures specified in Education Law §3012-d
- Know history, components, opportunities and challenges around student learning objectives and student growth measures
- Discuss what an ideal student learning component would look like in an APPR evaluation system

Session #1 Debrief

Questions Addressed in Session 1

- In your ideal evaluation system – conceptually, what would be the most important ***purpose****?* What would be the most important ***use****?*

- What are the current and/or potential barriers to an ideal educator evaluation system?

Purposes and Uses of an Ideal Educator Evaluation System

Educators seek an evaluation system that enhances teaching practice, especially through professional learning and growth opportunities, as well as via meaningful, unbiased feedback from evaluators.

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

Barriers to an ideal system…

- **Educator Practice**
  - Frequency of observations is the same for all teachers regardless of performance level or experience
  - Some aspects of professional performance cannot be captured through observation
  - Educators do not have option to choose traditional observation or other activities that exemplify teaching
  - Observations must be growth oriented, and not punitive
  - Evaluators must be trained for observing classrooms with different types of students (e.g., English learners, students with disabilities), and understand how Learning Standards are implemented in the classroom

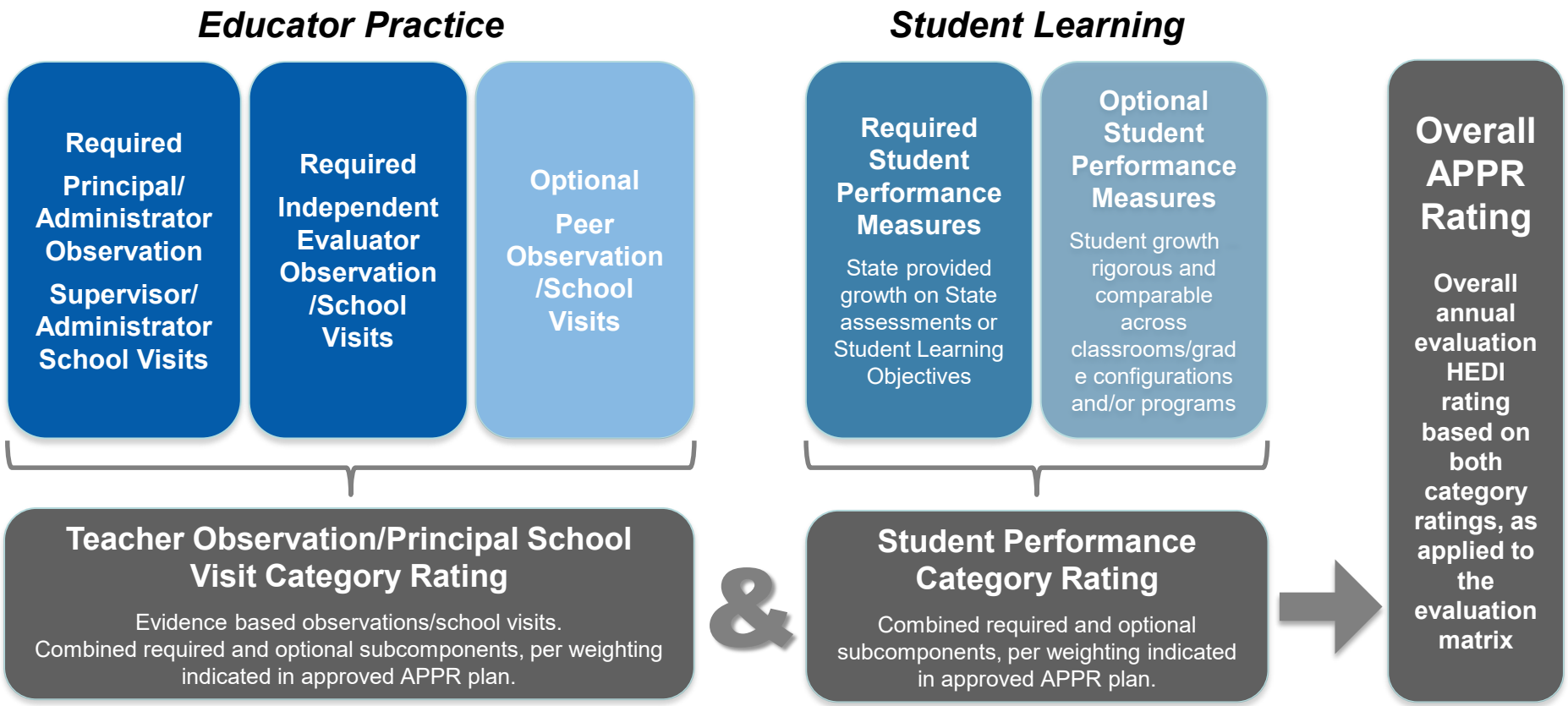
Session 2: Student Learning	Session 3: Educator Practice
What would an ideal student learning component look like in an APPR evaluation system?	***What would an ideal educator practice component look like in an APPR evaluation system?***
• How can the consistency in the implementation of student performance measures across LEAs and schools be improved? • Should the requirements for inclusion of local assessments be revised? If so, how? • Would changes to the assessments used for student growth improve the quality of student growth measures? If so, what changes? • What other options for student growth would the workgroup recommend for teachers of untested grades and subjects?	• How can the observation system be more responsive and flexible? • How can the observation system support reflection and growth? • Should the requirements for new teachers be revised to be less punitive? If so, how? • What other measures of professional practice would the group recommend?

Overview of Education Law §3012-d Requirements for Student Performance Measures

Education Law §3012-d

Components of the APPR Evaluation System

- Evaluations include educator practice and student learning measures
- Measures result in a single overall educator effectiveness rating



Education Law §3012-d

Components of the APPR Evaluation System

- **The overall APPR rating is determined by the statutory matrix:**

		Observation/School Visit			
		Highly Effective (H)	Effective (E)	Developing (D)	Ineffective (I)
Student Performance	Highly Effective (H)	H	H	E	D
	Effective (E)	H	E	E	D
	Developing (D)	E	E	D	I
	Ineffective (I)	D*	D*	I	I

Required and Optional Student Performance Measures – Education Law §3012-d

Education Law §3012-d

Student Performance Requirements

Required Measures

- Teachers of grades 4-8 ELA and math, principals of buildings covering these grade levels, and high school principals (all of grades 9-12) **receive a State-provided growth score**.
  - Statistical growth score calculated based on students' ELA and math State assessment results in the current year compared to similar students.
  - The term "similar students" in this context means not just students with the same academic history, but also students with the same demographic characteristics (i.e., English language learner (ELL), economic disadvantage, or disability (SWD) status).
  - HS principals have an additional measure based on the growth in Regents examinations passed.

Education Law §3012-d

Student Performance Requirements

Required Measures

- All other teachers and principals have Student Learning Objectives (SLOs).
- An SLO is an academic growth goal set for an educator's students at the start of a course.
  - Represents the most important learning that is aligned to learning standards, as well as other school and district/BOCES priorities.
  - SLO growth targets must be specific and measurable, based on available prior student learning data. This baseline data may come from a variety of sources including pre-tests/pre-assessments and a student's prior academic history.
  - Educators' scores are based upon the degree to which the goals were attained, as evidenced by student academic performance at the end of the course.

Education Law §3012-d

Student Performance Requirements

Required Measures

- Where a course or grade level ends in a State-created or administered assessment, the Education Law requires that that assessment be used as the evidence for the SLO (e.g., grade 8 science, Regents courses, NYSAA, NYSESLAT).
- The required student performance measures must cover the majority of a teacher's students across all the courses/grades they teach.
- For principals, at least 30% of students enrolled in the building must be covered by the required measures.
- Some educators have a mix of State-provided growth scores and SLOs.

Education Law §3012-d

Student Performance Requirements

Required Measures

- The required student performance measures must cover the majority of a teacher's students across all the courses/grades they teach.
- For principals, at least 30% of students enrolled in the building must be covered by the required measures.
- Some educators have a mix of State-provided growth scores and SLOs.
- Each measure assigns a score from 0-20, and the overall score corresponds to a rating of Highly Effective, Effective, Developing, or Ineffective (HEDI).

Education Law §3012-d

Student Performance Requirements

Optional Measures

- In addition to State-provided growth scores and/or SLOs, all school districts and BOCES also have the option to collectively bargain additional, "optional" student performance measures under the law.
- Under Education Law §3012-d, this second measure must be:
  - A second State-provided growth score based on a State-created or administered assessment; or
  - A growth score based on a State-approved assessment calculated using a State-approved growth model.
- Each measure assigns a score from 0-20, and the overall score corresponds to a rating of Highly Effective, Effective, Developing, or Ineffective (HEDI).

APPR Transition Period Requirements

- For the 2015–16 through 2018–19 school years, educators whose evaluations are to be based on the grades 3-8 ELA and math State tests and/or State-provided growth scores receive an “original” evaluation that includes these measures. This evaluation is for advisory purposes only.
- These educators also receive a “transition” evaluation that excludes those required or optional student performance measures and instead uses the results of one or more Alternate SLOs based on locally-determined assessments. This evaluation is used for all employment-related decisions.
- The Board of Regents will vote in April on proposed regulations to extend the Transition Period through the 2019–20 school year.

History and Review of Student Growth Models and Student Learning Objectives (SLOs)

Where are we with educator evaluation and how did we get here?

- Nearly all states began designing and implementing new teacher and/or principal evaluations in the mid 2000s due to:
  - Large variation across and within states and districts on what was required
  - Little to no differentiation in ratings and interest in modifying compensation schedules
  - Minimal to no consideration for student outcomes
  - Minimal feedback to educators for improvement
- Federal policy (e.g., RTTT, ESEA waivers, etc.) required "significant emphasis" on student growth; specific definition left up to states

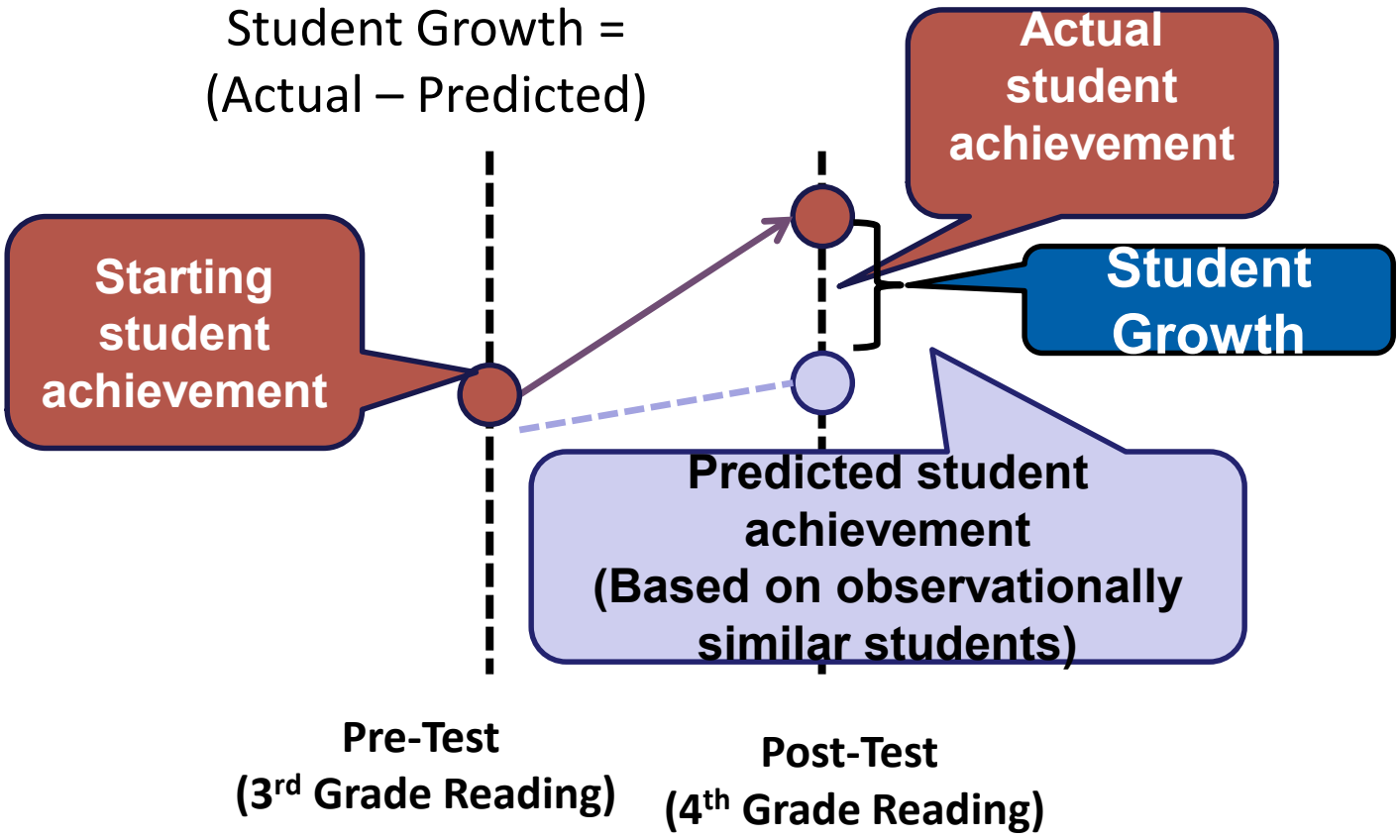
Quick Review (cont'd)

- In response, most states and districts adopted a combination of (a) professional practice measures grounded in research on effective teaching (e.g., Danielson framework, etc.), and (b) student outcomes.
- States/districts also developed pie charts specifying weights for observation and student growth.
- Two main categories of student growth: statistical models and student learning objectives.

Student Growth Models: Description and Purpose

- Student growth models measure the contribution of schooling at various levels (e.g., 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.
- Goal: provide information on what different levels of education can and should control (e.g., improved achievement for all students), but factor out what they can't control (e.g., student characteristics and out of school factors)

Student Growth: A Visual Representation



Student Growth Models: Key Takeaways

- Student growth models measure growth, not attainment
- Student growth models measure the *entire range of growth*, not just movement across proficiency categories
- Student growth models use statistical controls for students' prior achievement and typically a set of at least some student characteristics (e.g., gender, race/ethnicity, ELL status) at the student, classroom, and/or school level

Poll

By taking into account factors that are outside of the teachers' control, student growth measures provide a fairer measure of teacher contributions to growth than attainment measures.

- Strongly Disagree
- Disagree
- No opinion
- Agree
- Strongly Agree

Comments, questions, wonderings or other thoughts?

Student Growth Models: Concerns/Policy Issues

- Lack of buy-in and understanding among educators
- Complexity & understandability of student growth models

The student growth model is defined by four equations:

$$\text{Student achievement: } y_{1i} = \zeta + \lambda y_{0i} + \lambda^{alt} y_{0i}^{alt} + \beta X_i + \alpha S_i + e_i \quad (1)$$

$$\text{Posttest measurement error: } Y_{1i} = y_{1i} + v_{1i} \quad (2)$$

$$\text{Same-subject pretest measurement error: } Y_{0i} = y_{0i} + v_{0i} \quad (3)$$

$$\text{Other-subject pretest measurement error: } Y_{0i}^{alt} = y_{0i}^{alt} + v_{0i}^{alt} \quad (4)$$

where:

- y_{1i} is true post achievement;
- y_{0i} and y_{0i}^{alt} are true prior achievement in the same subject and in the other subject (math in the ELA model, ELA in the math model), with slope parameters λ and λ^{alt};
- X_i is a vector of characteristics of student i, with slope parameter vector β;
- S_i is a vector of indicators for school;
- α is a vector of school effects;
- e_i is the error in predicting post achievement given the explanatory variables included in the model;
- Y_{1i} is measured post achievement;
- v_{1i} is measurement error in post achievement;
- Y_{0i} and Y_{0i}^{alt} are measured prior achievement; and
- v_{0i} and v_{0i}^{alt} are measurement error in prior achievement.

Student Growth Models: Concerns/Policy Issues (cont'd)

- **Stakeholder concerns about state tests:**
  - Year-to-year changes in state assessment systems and content standards and general concerns about standardized testing
  - Statistical growth models typically only cover ~30% of teachers
  - Narrowing curriculum
  - Low student motivation

Student Growth Models: Concerns/Policy Issues (cont'd)

- High level of data complexity and effort needed to calculate statistical models accurately
- Reliability of value-added measures
- Causal attribution (i.e., Do models accurately capture teachers' true effect on student performance?)
- Use of school growth to indicate principal growth

Student Growth Models: Concerns/Policy Issues (cont'd)

- Models cannot explain why a particular teacher's students scored better than expected, so this measure is of limited use in a feedback-oriented system.
- Models are fundamentally different: a normative measure, whereas most other educator effectiveness measures are criterion-referenced.

Student Learning Objectives: Purpose and Opportunities/Challenges

- Created to provide student growth measure for non-tested grades and subjects (NTGS)
- Selected districts were using prior to national use
- Opportunities:
  - High face validity and buy-in for educator-developed growth measures
  - Can promote greater collaboration
- Challenges:
  - Lack of high-quality assessments and technical rigor with many teacher-developed assessments
  - Take substantial time and collaboration to do well
  - Because the SLO is used as part of high stakes evaluation, there is a potential incentive to set lower goals

Themes from Student Learning Objectives Implementation

- **On the plus side, SLOs:**
  - Represent good professional practice: collaborative review of data to determine areas of student need; discussion of strategies, evidence sources and growth targets; review of results
  - Provide one answer to NTGS issue
  - Can be written by all educators
  - Provide buy-in and ownership

Themes from Student Learning Objectives Implementation (Cont'd)

- **Significant emerging challenges around:**
  - **Assessments:**
    - Not enough high-quality assessments
    - Resistance to having lists of approved assessments
  - **Growth Targets and Scoring:**
    - Growth targets not necessarily informed by data
    - Potential incentive to set low targets
    - Scoring not consistent or comparable
  - **Training and Support:**
    - Not enough time for educators or evaluators to collaborate
    - Inadequate training on assessment development

Poll

In theory, student learning objectives represent good professional practice: collaborative review of data to determine areas of student need; discussion of strategies, evidence sources & growth targets; review of results.

- Strongly Disagree
- Disagree
- No opinion
- Agree
- Strongly Agree

Comments, questions, wonderings or other thoughts?

Key Student Learning Objectives
Decision Points for Policymakers

More Structure

Less Structure

Assessments/Evidence Sources
Growth Targets
Scoring Rubric

Decision Point 1: Assessments

- **Tradeoffs of providing educators with more/less structure:**
  - Having approved lists of assessments for use as SLO evidence sources provides a “floor” of minimal technical quality, and saves teachers considerable time
  - Approved assessment lists will likely include standardized tests that may not feel connected to teacher practice and may deprive teachers of the long-term benefits of developing and refining their own assessments
  - Alternatively, states can allow educators to use their own assessments, which will create more buy-in, but these assessments may not have the same technical qualities.

Decision Point 2: Growth Targets and Scoring

- **Tradeoffs of more/less structure:**
  - State-provided and data-informed growth targets can eliminate much of the guesswork that would otherwise fall to educators and evaluators
  - Educators are likely to resist “formulas” that inevitably will have cut points viewed as arbitrary
  - States can also make the scoring process less structured with the main tradeoff here being that the scoring/rating process can become subjective

Additional Policy Considerations

- **What kinds of resources and supports do districts and schools need to implement a high-quality SLO process?**
  - Training: Initial and ongoing
  - Resources: Process guides, sample SLOs, etc.

Additional Policy Considerations (cont'd)

- **How will scale-up of training take place?**
  - Train-the-trainers models have benefits, but must ensure consistency and enough time
  - Concern about having only one trainer in a school/district who may also be a full-time teacher
  - How will longer-term capacity be built?

Key Questions (cont'd)

- **Do not forget about data quality: Which SLO data will be entered into which platforms, and how will the data get integrated with other data sources to produce overall ratings?**

Big Picture

- Critical for all stakeholders to remember that all potential measures of educator effectiveness have tradeoffs
- There are challenges with student growth measures, but there are parallel challenges with teacher practice measures.

Poll

To the extent that the consistency of implementation of SLOs can be improved, and better student growth measure options for teachers in untested grades and subjects can be offered, the student outcomes component of APPR could provide valuable information for educators regarding their effectiveness.

- Strongly Disagree
- Disagree
- No Opinion
- Agree
- Strongly Agree

Comments, questions, wonderings or other thoughts?

Break-Out Group Discussion

Break-Out Discussion

- **What would an ideal student performance component look like in an APPR evaluation system?**
  - **What would you change about:**
    - The purpose of the student performance component
    - The required and optional measures of the student performance component
    - Processes of the student performance component
    - The use of data from the student performance component
  - **How can we move from the current student performance component to the ideal student performance component?**

Closing and Next Steps

- Thank you!!!
- Your inputs from today's webinar will be used to inform the planning for our next in person session
- January 10th.

Appendix F

NYSED Form H

STUDENT ASSESSMENTS FOR TEACHER AND PRINCIPAL EVALUATION

FORM *H*

APPLICANT CERTIFICATION FORM

ASSESSMENTS FOR USE WITH STUDENT LEARNING OBJECTIVES

Please read each of the items below and check the corresponding box to ensure the fulfillment of the technical criteria.

PLEASE SUBMIT ONE "FORM H" FOR EACH APPLICANT. CO-APPLICANTS SHOULD SUBMIT SEPARATE FORMS.

The Applicant makes the following assurances:

Assurance	Check each box:
The assessment is rigorous, meaning that it is aligned to the New York State learning standards or, in instances where there are no such learning standards that apply to a subject/grade level, alignment to research-based learning standards.	☐
To the extent practicable, the assessment must be valid and reliable as defined by the Standards of Educational and Psychological Testing.	☐
The assessment can be used to measure one year's expected growth for individual students.	☐
For K-2 assessments, the assessment is not a "Traditional Standardized Assessment" as defined in Section 1.3 of this RFQ.	☐
For assessments previously used under Education Law §3012-c, the assessment results in differentiated student-level performance. If the assessment has not produced differentiated results in prior school years, the applicant assures that the lack of differentiation is justified by equivalently consistent student results based on other measures of student achievement.	☐
For assessments not previously used in teacher/principal evaluation, the applicant has a plan for collecting evidence of differentiated student results such that the evidence will be available by the end of each school year.	☐
At the end of each school year, the applicant will collect evidence demonstrating that the assessment has produced differentiated student-level results and will provide such evidence to the Department upon request.[10]	☐

[10] Please note, pursuant to Section 2.3 of this RFQ, an assessment may be removed from the approved list if such assessment does not comply with one or more of the criteria for approval set forth in this RFQ

Appendix G

NYSED slo-template-3012-d

New York State Student Learning Objective Template

All SLOs MUST include the following basic components:	
Population	*These are the students assigned to the course section(s) in this SLO - all students who are assigned to the course section(s) must be included in the SLO. (Full class rosters of all students must be provided for all included course sections.)*
Learning Content	*What is being taught over the instructional period covered? Common Core/National/State standards? Will this goal apply to all standards applicable to a course or to specific priority standards?*
Interval of Instructional Time	*What is the instructional period covered (if not a year, rationale for semester/quarter/etc.)?*
Evidence	*What specific State-developed or State-approved assessment(s) will be used to measure this goal? The assessment selected must align to the learning content of the course. State assessments (including Regents examinations, Regents equivalents, and/or any State- approved equivalents) must be used as evidence if one of the courses required to have an SLO has a State assessment.*
Baseline	*What is the starting level of students' knowledge of the learning content at the beginning of the instructional period?*

Target(s)	*What is the expected outcome (target) of students' level of knowledge of the learning content at the end of the instructional period? (All targets must include a minimum of one year of expected academic growth and all targets must be approved by the superintendent or another trained administrator serving as his or her designee.)*
HEDI Scoring	*Districts and BOCES must use the State-determined scoring ranges to determine final scores and HEDI ratings.*

HIGHLY EFFECTIVE			EFFECTIVE			DEVEL-OPING		INEFFECTIVE												
20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
97-100 %	93-96%	90-92%	85-89%	80-84%	75-79%	67-74%	60-66%	55-59 %	49-54%	44-48%	39-43%	34-38%	29-33%	25-28%	21-24%	17-20%	13-16%	9-12%	5-8%	0-4%

Rationale	*Describe the reasoning behind the choices regarding the components of the SLO and how the SLO will be used together with instructional practices to prepare students for future growth and development in subsequent grades/courses, as well as college and career readiness.*