New York State
EDUCATION DEPARTMENT

Knowledge › Skill › Opportunity

NYSED Evaluation Workgroup

Webinar #1

Welcome!

Agenda

- Welcome and Introductions
- Objectives and Agenda
- 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?

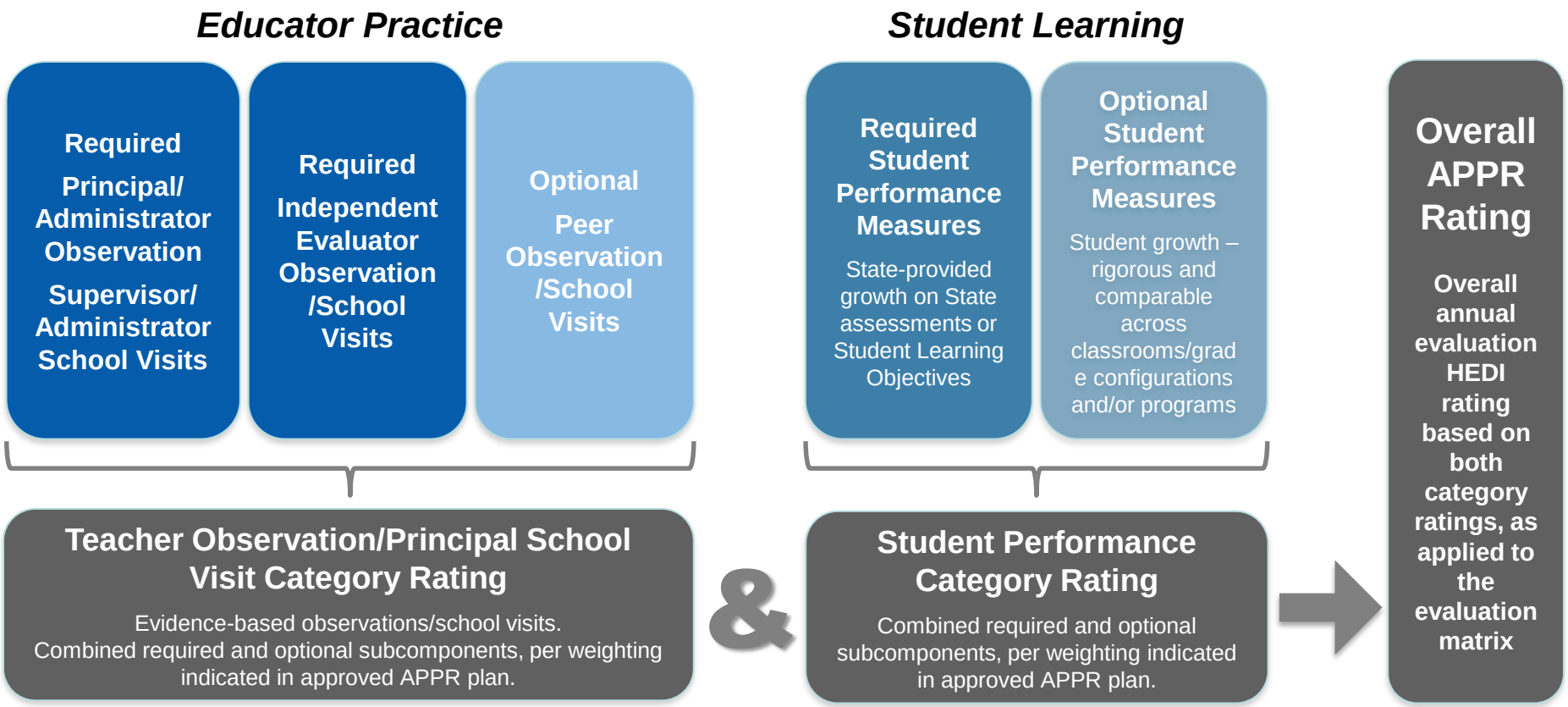
Overview of Education Law §3012-d Requirements for Student Performance Measures

Education Law §3012-d 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

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 and how did we get here in teacher/principal evaluation?

- Nearly all states (following the lead of districts) began designing & implementing new teacher/principal evaluations in the mid 2000s due to:
  - Large variation across (and within) states and districts on what was required
  - Little/no differentiation in ratings (Widget Effect) + interest in modifying compensation schedules
  - Minimal/no consideration for student outcomes
  - Minimal feedback to educators for improvement
- Federal policy (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 (Danielson framework, etc.), and (b) student outcomes.

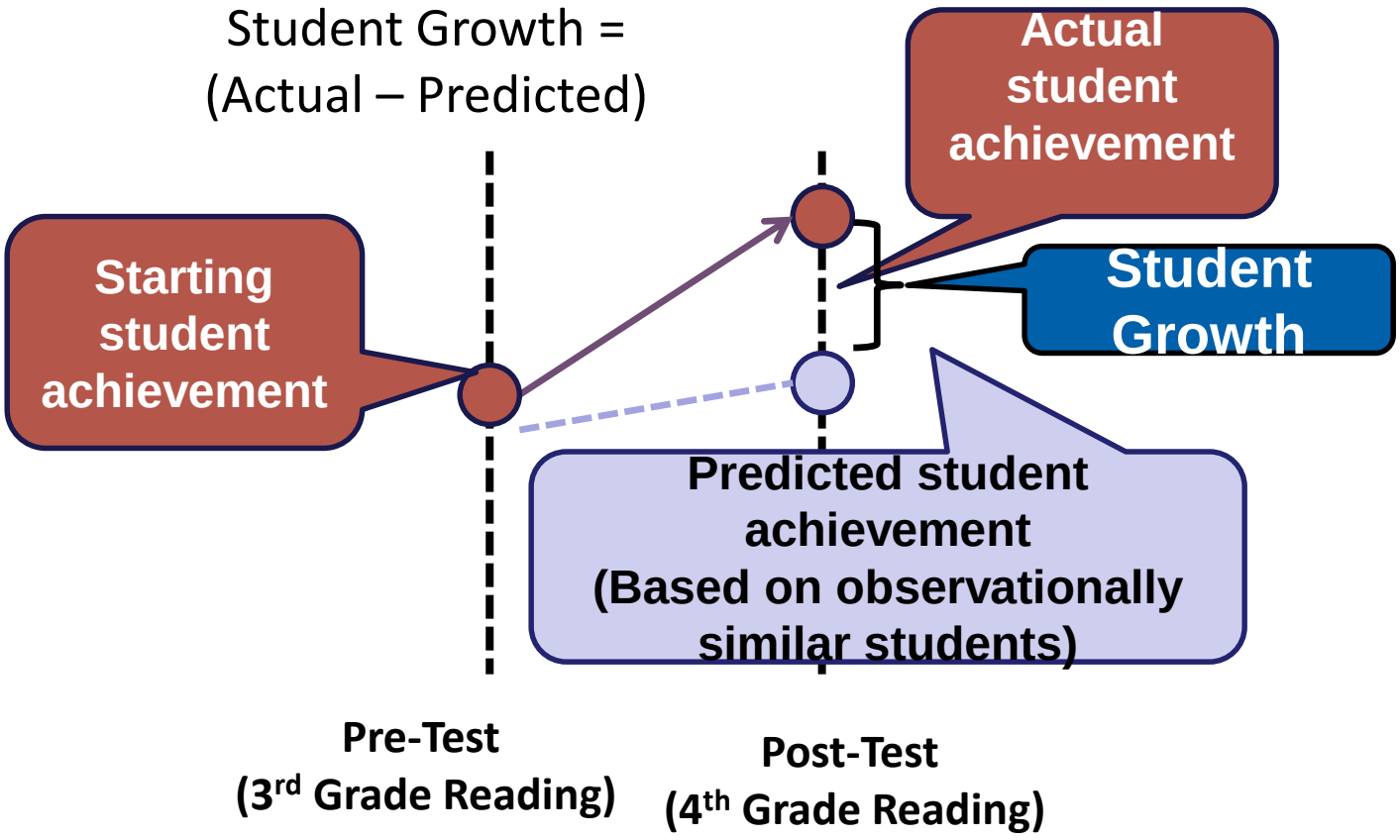
- States/districts also developed pie charts specifying weights for observation and student growth (50/50, etc.).

- Two main categories of student growth: statistical models (for tested grades/subjects) and student learning objectives (for NTGS); sometimes both were required or encouraged

Student growth models: Description and Purpose

- Student growth models 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.
- 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)

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 (usually) a set of at least some student characteristics (SES, SpEd, ELL, gender, race/ethnicity) at the student, classroom, and/or school level (for a more "level playing field")

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 (Common Core, 3 tests in 3 years) + concerns about standardized testing, generally (redundancy, opt-out movement, etc.)
  - Statistical growth models typically only cover ~30% of teachers
  - Narrowing curriculum/teaching to the test?
  - Low student motivation/high measurement error

Student growth models: Concerns/ Policy Issues (cont'd)

- High level of data complexity (and sometimes educator/administrator time) needed to calculate statistical models accurately (student/teacher linkage)

Student growth models: Concerns/ Policy Issues (cont'd)

- Reliability of value-added measures
  - How reliable for what purpose, and compared to what?
- Causal attribution (Do models accurately capture teachers' true effect on student performance)?
  - Students aren't randomly assigned to teachers; how big a deal is this?
- School growth = Principal growth? Probably not right away…

Student growth models: Concerns/ Policy Issues (cont'd)

- Models can't explain why a particular teacher's students scored better than expected (not diagnostic), or better than the teacher across the hall…so this measure is of limited use in a feedback-oriented system.
- Models are fundamentally different: a normative measure (me compared to teachers with similar students), whereas most other EE measures (esp. observations) are criterion-referenced (most educators can, and usually do, get high ratings).

Student Learning Objectives: Purpose and Opportunities/Challenges

- Created to provide student growth measure for NTGS
- Selected districts (Austin, etc.) 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 (esp. for non-core subjects) + lack of 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 & growth targets; review of results
  - Provide one answer to non-tested grades/subjects (NTGS) issue
  - Can be written by all educators (in theory)
  - Provide buy-in and ownership (potentially more so than statistical growth)

Themes from Student Learning Objectives Implementation (Cont'd)

- **Significant emerging challenges around:**
  - **Assessments:**
    - Not enough high-quality assessments (esp. for non-core subjects)
    - Resistance to having lists of approved assessments
  - **Growth Targets and Scoring:**
    - Lots of "educated guessing": growth targets not informed by data
    - 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

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 the kind of (standardized) tests that may not feel connected to teacher practice…and may deprive teachers of the long-term benefits of developing/refining their own assessments (individually or in teams)
  - 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 & Scoring

- **Tradeoffs of more/less structure:**
  - State-provided (data-informed) growth targets can eliminate much of the guesswork that would otherwise fall to teachers and evaluators
  - Educators are likely to resist “formulas” that inevitably will have cut points viewed as arbitrary (you get a 4 if at least 80% of your students make X amount of growth)
  - States can also make the scoring process less structured/more organic (esp. if SLOs don’t occupy a fixed weight) 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/schools need to implement a high-quality SLO process?**
- **Training: initial and ongoing**
- **Resources: process guides, sample SLOs, etc**.
  - States and districts almost always over-estimate how *much* and how *well* their communication percolates down, and how much buy-in actually exists

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 (esp. at HS level) who may also be a full-time teacher
  - How will longer-term capacity be built?
    - How to build principal capacity as evaluators of SLOs?
    - How to build SLO expertise into teacher and principal training programs (EPPs/IHEs)

Key Questions (cont'd)

- **Don't forget about data quality: which SLO data will be entered (by whom and when) into which platforms, and how will the data get integrated with other data sources to produce overall ratings?**
  - Many initiatives have been sunk by clunky technology…

Big Picture

- Critical for all stakeholders to remember that all potential measures of educator effectiveness (observations, different forms of student growth) have tradeoffs (sometimes technical, more often policy-based) – as would be the case if all of this went away and were replaced by what existed previously

- Despite all the (justified) attention to challenges with student growth measures, there are parallel challenges with using observations for educator evaluation…

Check for Understanding

In theory, 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?

Check for Understanding

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?

Check for Understanding

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

- **Overarching question: What would an ideal student learning component look like in an APPR evaluation system?**
  - **What would you change about:**
    - The purpose of the student learning component
    - The components of the student performance category
    - Processes of the student performance category
    - The use of data from the student performance category

Join Break Out Sessions

- [insert instructions for joining break out session in webinar]
- Appoint one person to share the groups comments, issues, recommendations.

Let's Share

- **Share 1-2 comments, issues, or recommendations that your group discussed regarding what an ideal student learning component would look like in an APPR evaluation system?**

Moving Forward

- **How can we move from the current student learning component to the ideal student learning component?**