

Flexibility and Comparability within a System

Susan Lyons and Joe Anderson



Agenda

- ⌘ Defining comparability in an era of flexibility (5 minutes)
- ⌘ Overview of options for flexibility in assessments under ESSA and considerations for score comparability (15 minutes)
- ⌘ Discussion Group Activity for high school assessment option (25 minutes)
- ⌘ Extra time for states to meet to share ideas and questions about assessment flexibility options (20 minutes)
- ⌘ Closing thoughts and next steps (5 minutes)

Defining Comparability

“When comparability exists, scores from different testing conditions can be used interchangeably” (Bennet, 2003, p. 2)

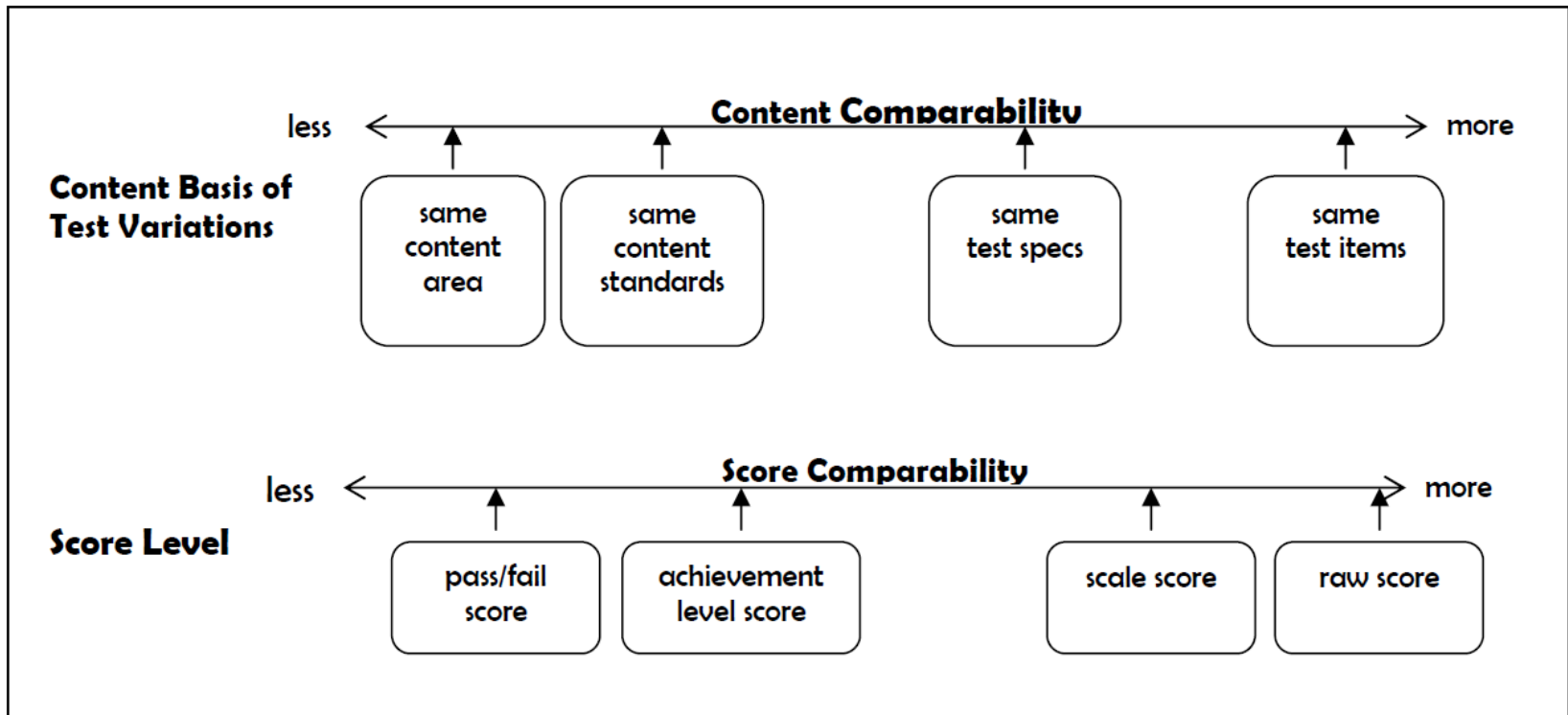
Comparability is:

- ⌘ Score-based
- ⌘ An evidence-based claim
- ⌘ A continuum

Comparability is **NOT**:

- ⌘ Necessarily at odds with flexibility (e.g., accommodations)
- ⌘ A coefficient
- ⌘ The same for every assessment (e.g., evidence required will differ)

Comparability Continuum



(Winter, 2010)

ESSA Opportunities for Flexibility in Assessments

ESSA requires state high-quality assessments in reading/language arts and math in grades 3-8 and once in high school (and grade-span tests in science). Adaptive assessment is permitted. The requirement for 95% participation is maintained.

Flexibility options for assessments offered under ESSA include:

1. Interim assessment option (see ESSA § 1111(b)(2)(B)(viii))
2. Locally-selected, nationally-recognized high school assessments (see ESSA § 1111(b)(2)(H))
3. Advanced mathematics assessment for eighth grade (see ESSA § 1111(b)(2)(C))
4. A pilot for innovative assessment and accountability models (up to seven states in the first three years) (see ESSA § 1204)



Increased Flexibility in Accountability

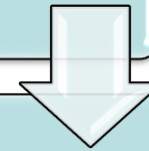
Though limited, there is increased flexibility for states in the statewide accountability system, examples include:

- ⌘ School quality/student success indicator
- ⌘ Goal Setting
- ⌘ Consequences
- ⌘ Reporting Systems

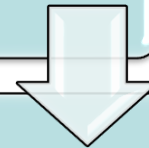


Alignment with Theory of Action

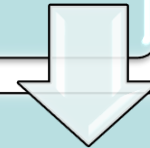
What is my state vision and overall educational goals?



How will we achieve those goals? What are the mechanisms for change?



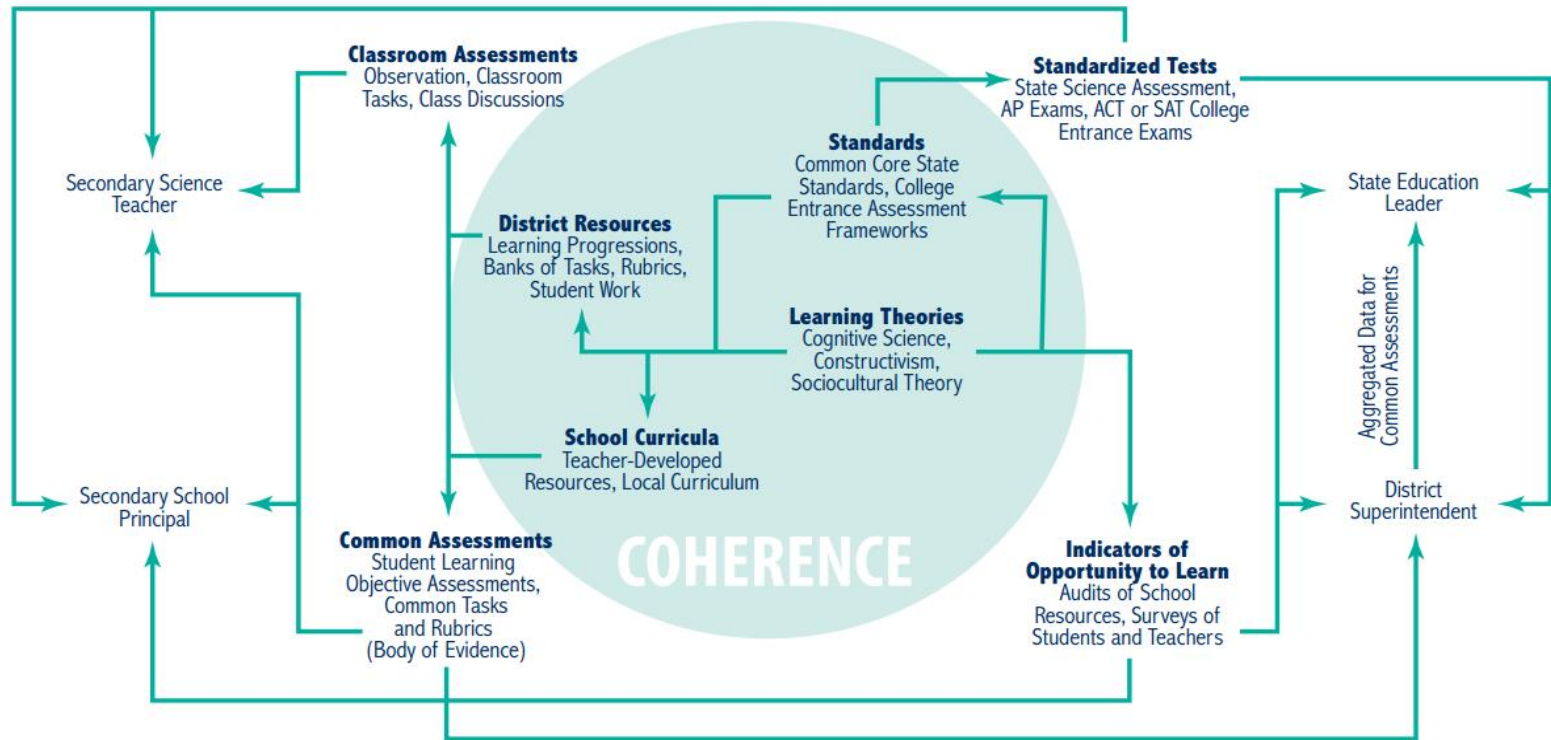
What role does the assessment and accountability system play in achieving those goals?



Will local flexibility in assessments for accountability improve the likelihood of achieving the desired outcome?

Balanced Assessment System

ESSA is an opportunity to use components of a balanced assessment system to inform summative student scores and school accountability.



(Chattergoon & Marion, 2016)

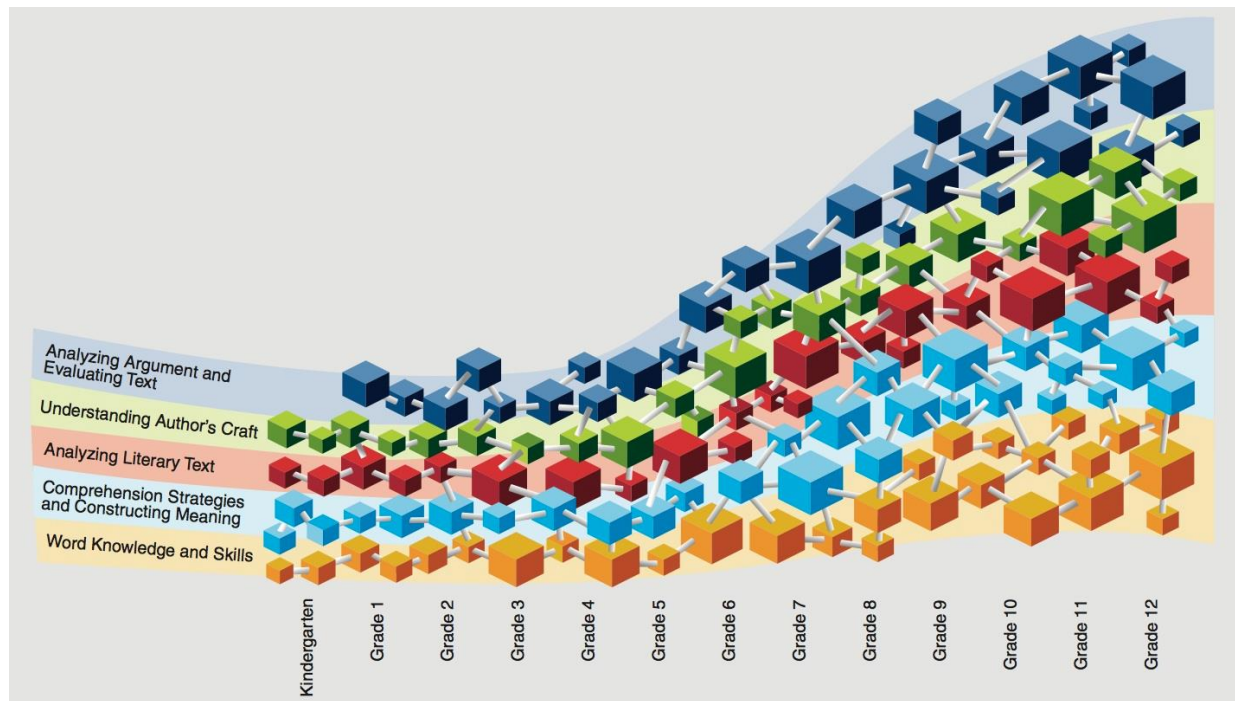
Balanced Assessment System Design Considerations?

- ⌘ What are the purposes you intend to serve with the assessment system?
- ⌘ What system of assessments make sense to support at the state level versus the local level?
- ⌘ Will local assessment information inform state accountability and reporting?
- ⌘ What flexibility opportunities will my state take advantage of to support the state assessment and accountability system?
 1. Interim assessment option (see ESSA § 1111(b)(2)(B)(viii))
 2. Locally-selected, nationally-recognized high school assessments (see ESSA § 1111(b)(2)(H))
 3. Advanced mathematics assessment for eighth grade (see ESSA § 1111(b)(2)(C))
 4. A pilot for innovative assessment and accountability models (up to seven states in the first three years) (see ESSA § 1204)

1. Interim Assessment Option

⌘ Example design options

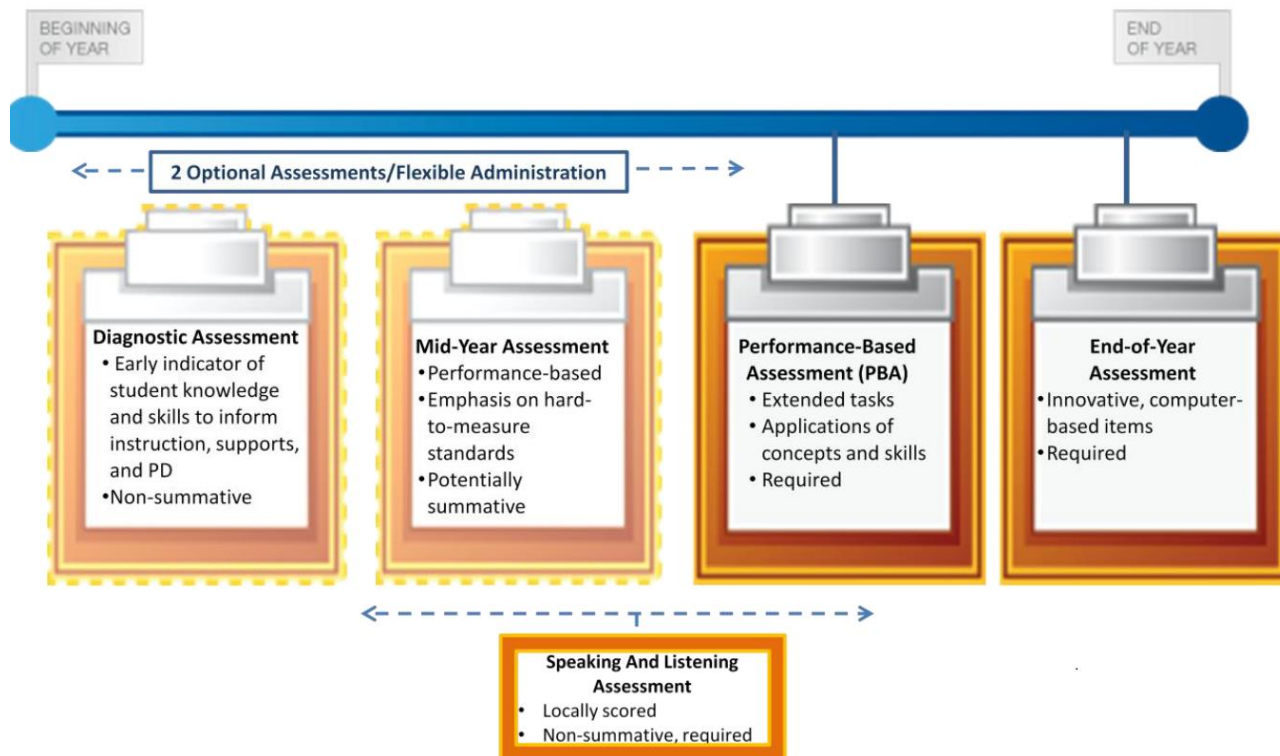
- Multiple mini-summative assessments throughout the year assessing different standards along a learning progression



1. Interim Assessment Option

⌘ Example design options (cont.)

- One or multiple common performance assessments along with shorter summative at end of year



2. Local option for High School – State review process

Does local flexibility at the high school level align with the state vision and goals?

Yes

Key Questions for Designing Review Process

- ✓ How would the state accountability system maintain comparability? Would the nationally-recognized assessment be willing to embed state assessment items to help equate the two scales?
- ✓ Is the assessment aligned to the state content standards? How similar are the blueprints? Would a state need to change their content standards?
- ✓ Is the district willing to pay for the nationally-recognized assessment?

No

Key Questions for Designing Review Process

- ✓ What benefits does the additional assessment bring to the district and its stakeholders?
- ✓ Does this assessment pose a threat to fulfilling the intended purposes of the accountability system? How can the purposes be protected while allowing for local flexibility?
- ✓ What are the practical implications of the accountability results? If the stakes are low, there may be limited risk to misclassifications resulting from multiple assessments options within the state.

3. Eighth Grade Exception

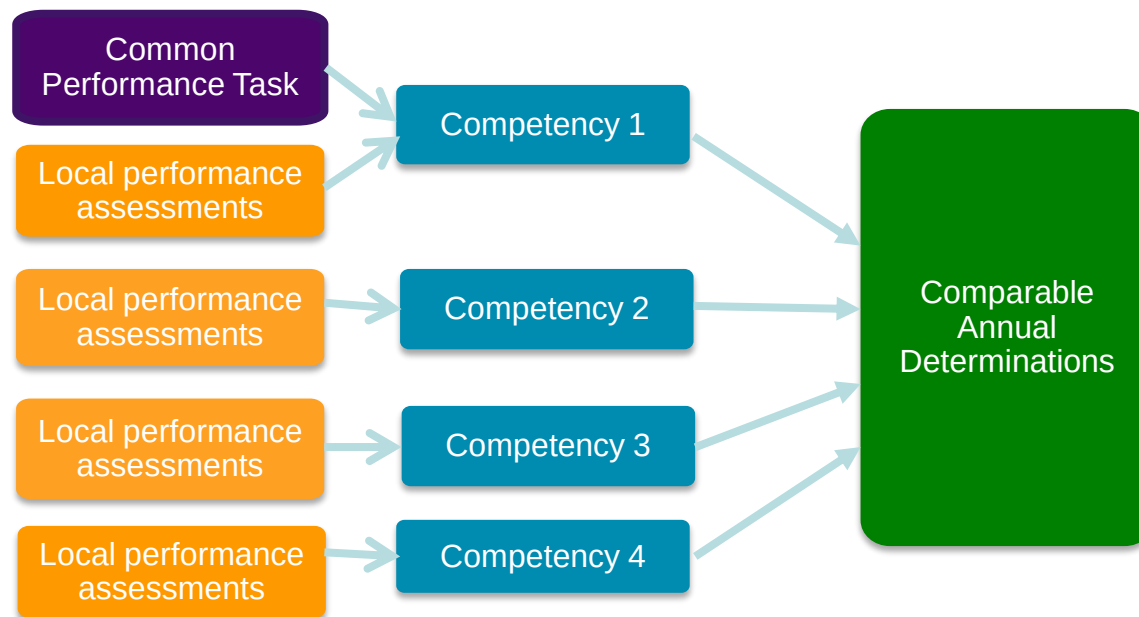
Anticipated Challenges with the Grade 8 Option for Course-Specific Assessment:

- ⌘ How will the state ensure all students in the state will have an opportunity to take (and be ready for) advanced middle school math? All students will need to have the opportunity to take more advanced math courses in middle school.
- ⌘ How will the state include the high school EOY assessments in the middle school accountability system? The performance standards will be, by definition, not comparable.

4. Demonstration Authority

⌘ Example design option

- Multiple *local* performance tasks used to inform competency scores and student annual determinations



4. Demonstration Authority

Comparability Considerations

1. Evaluating **assessment quality and content coverage.**
2. Setting **performance standards that are aligned with common statewide achievement level descriptors.** States may also want to consider how they can validate their performance standards. Examples of validation techniques include engaging in multiple standard setting approaches (e.g., contrasting groups and body of work) and including the state standardized assessment at select grade levels.
3. Engaging in **one or more auditing techniques.** Though claims of comparability are held at the annual determination level, the annual determinations will likely be made on the basis of information gathered using multiple local assessments. Auditing the comparability of scores on those local assessments will contribute to the evidence supporting the comparability of annual determinations. In order to audit the comparability of local assessments the state may want to plan to include one or more common assessments across pilot districts.

4. Demonstration Authority

Comparability Considerations

1. Evaluating **assessment quality and content coverage**. States will want to have processes for monitoring the quality and coverage of local assessment systems to ensure they provide accurate information about student achievement for all of the grade level content (e.g., the assessments measure the full range of the content standards).
2. Setting **performance standards that are aligned with common statewide achievement level descriptors**.
3. Engaging in **one or more auditing techniques**. Though claims of comparability are held at the annual determination level, the annual determinations will likely be made on the basis of information gathered using multiple local assessments. Auditing the comparability of scores on those local assessments will contribute to the evidence supporting the comparability of annual determinations. In order to audit the comparability of local assessments the state may want to plan to include one or more common assessments across pilot districts.

4. Demonstration Authority

Comparability Considerations

1. Evaluating **assessment quality and content coverage**. States will want to have processes for monitoring the quality and coverage of local assessment systems to ensure they provide accurate information about student achievement for all of the grade level content (e.g., the assessments measure the full range of the content standards).
2. Setting **performance standards that are aligned with common statewide achievement level descriptors**. States may also want to consider how they can validate their performance standards. Examples of validation techniques include engaging in multiple standard setting approaches (e.g., contrasting groups and body of work) and including the state standardized assessment at select grade levels.
3. Engaging in **one or more auditing techniques**.

4. Comparability by Design

Within-District Comparability in Scoring

- Within-district calibration with annotated anchor papers
- Checks on inter-rater consistency

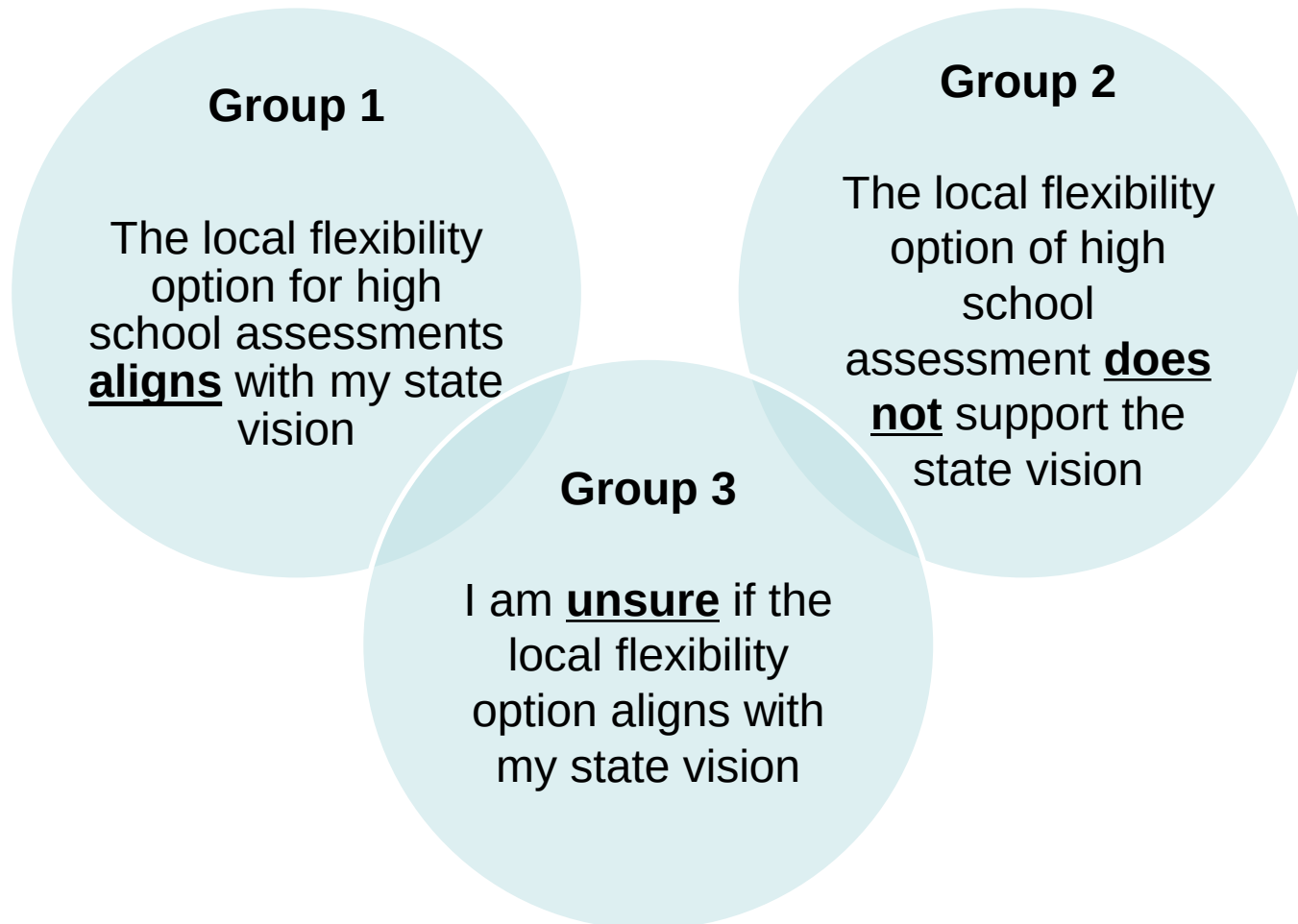
Cross-District Comparability in Evaluating Student Work

- Consensus Approach –*common tasks*
- Rank Ordering Approach –*local tasks*

Comparability of Annual Determinations

- Common ALDs and body of work performance standards validation
- External standardized assessment at select grade levels

Discussion Group Activity – High School Assessment



Closing Thoughts

1. What is my new insight or “biggest aha” about my state’s opportunity as we design our assessment and accountability plan under ESSA?
2. What is my biggest wonder?
3. What is the next step I plan to take to move my SEA forward on this issue?



References

- ⌘ Bennett, R.E. (2003). *Online assessment and the comparability of score meaning* (ETS-RM-03-05). Princeton, NJ: Educational Testing Service.
- ⌘ Chattergoon, R., & Marion, S. (2016). Not as Easy as It Sounds: Designing a Balanced Assessment System. *State Education Standard*, 16(1), 6-9.
- ⌘ Winter, P. (2010). Comparability and test variations. In P. Winter (Ed.), *Evaluating the comparability of scores from achievement test variations* (pp. 1-11). Washington, DC: Council of Chief State School Officers.