



**OFFICE OF ELEMENTARY AND SECONDARY EDUCATION**

**SCHOOL SUPPORT AND ACCOUNTABILITY**  
**2023 State Assessment Conference**  
**1G: Preparing for Peer Review**  
September 27, 2023

1



**FOCUS AREA: ASSESSMENT PEER REVIEW**

**1G - Preparing for Peer Review**

Meagan Karvonen, ATLAS, karvonen@ku.edu

Zach Warner, NYS Ed Dept, zachary.warner@nysed.gov

Nathan Dadey, Center for Assessment, ndadey@nceia.org

Phoebe Winter, Independent Consultant, phoebe.winter@outlook.com

Brooke Nash, ATLAS, bnash@ku.edu

Chris Rozunick, Texas Educational Agency, Christine.Rozunick@tea.texas.gov

2



## A note About this Conference/Session

- The purpose of this conference/session is to provide an opportunity for State education agency (SEA) staff to interact and engage with relevant experts and other SEA staff about the Department's assessment peer review.
- The observations and opinions of the session presenters are their own.

3

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

3



## Session Overview

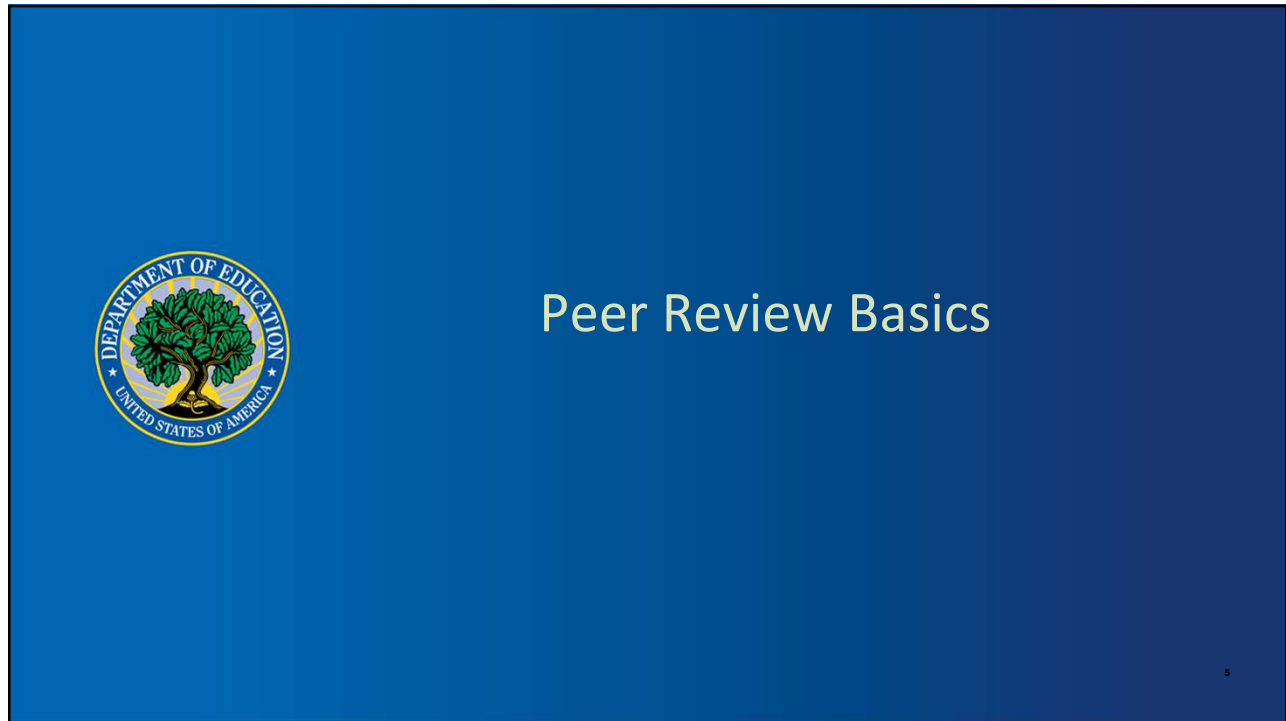
- Peer review basics
- Key messages from breakouts (top 10?)
- Illustrations from peer review submissions
- Deeper dive: discussion of some critical elements
- Q&A

4

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

4



5

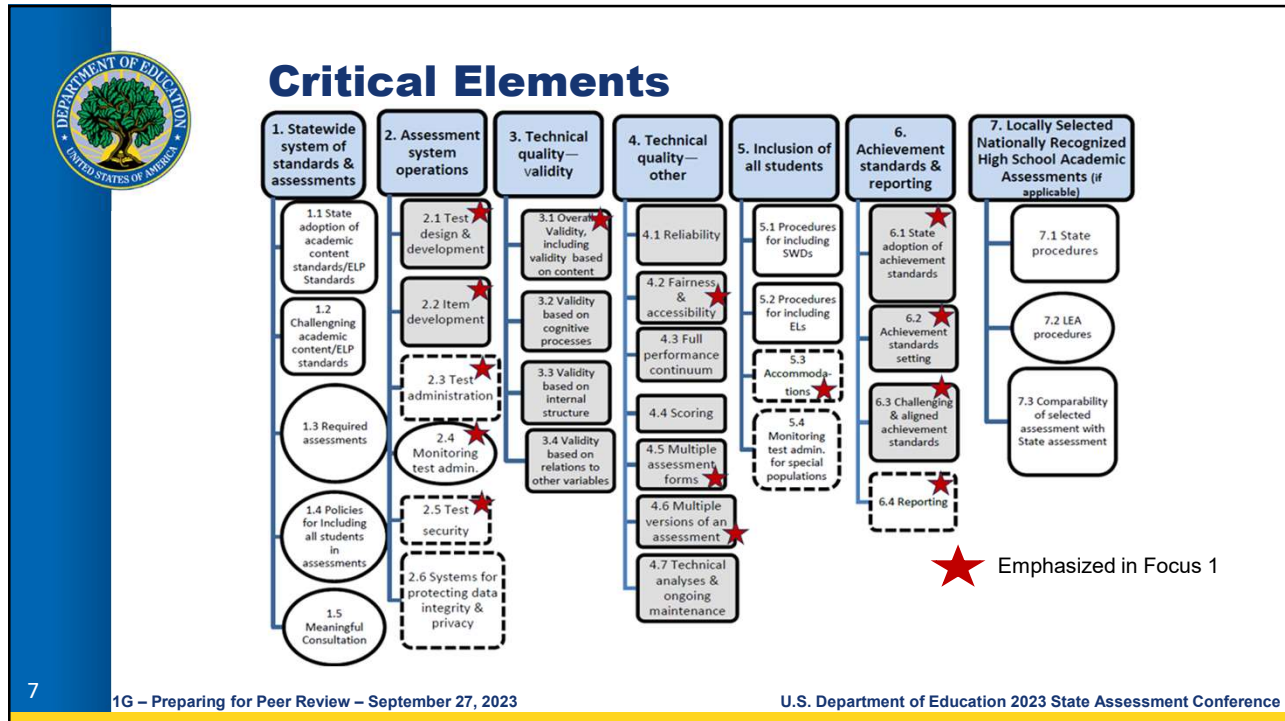
A white slide with a blue vertical bar on the left containing the U.S. Department of Education seal. The title "Peer Review Basics" is in bold blue text. Below it is a bulleted list of four items. The second item includes a blue hyperlink. At the bottom, there is a footer with the number "6", the text "1G – Preparing for Peer Review – September 27, 2023", and "U.S. Department of Education 2023 State Assessment Conference".

**Peer Review Basics**

- Peer review guidance (2018) is the ultimate source
- Available on conference website:  
<https://apps1.seiservices.com/2023SSA/Materials.aspx>
- Initial submission is typically within a year of first operational administration or based on USDE schedule
- Submissions consist of cover packet, indices, and sources of evidence

6 1G – Preparing for Peer Review – September 27, 2023 U.S. Department of Education 2023 State Assessment Conference

6



7

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

7

**Interpreting Requirements**

Critical Element 4.4 – Scoring	Examples of Evidence
<p>The State has established and documented standardized scoring procedures and protocols for its assessments (and for ELP assessments, any applicable domain or component sub-tests) that are designed to produce reliable and meaningful results, facilitate valid score interpretations, and report assessment results in terms of the State's:</p> <ul style="list-style-type: none"> <li>(1) <u>academic achievement standards</u>; or</li> <li>(2) <u>ELP standards</u>.</li> </ul> <p>For ELP assessments, if an English learner has a disability that precludes assessment of the student in one or more of the required domains/components (listening, speaking, reading, and writing) such that there are no appropriate accommodations for the affected domain(s)/component(s), the State must provide a description of how it will ensure that the student is assessed in the remaining domain(s)/component(s) in</p>	<p>Evidence to support this critical element for the State's general academic and ELP assessments. AA-AAAS, and AELPA includes:</p> <ul style="list-style-type: none"> <li>A chapter on scoring in a technical report for the assessments or other documentation that describes scoring procedures, including: <ul style="list-style-type: none"> <li>Procedures for constructing scales used for reporting scores and the rationale for these procedures;</li> <li>Procedures for combining scores to produce any composite scores that are reported and used, and the rationale for these procedures;</li> </ul> </li> <li>If the State uses a vertical reporting scale, procedures for linking and equating across grade-spans and/or test forms; and procedures for examining the stability of the vertical scale over time;</li> <li>Scale, measurement error, and descriptions of test scores;</li> <li>For scoring involving human judgment, including scoring conducted by test administrators, or local and school staff: <ul style="list-style-type: none"> <li>Evidence that the scoring of constructed-response items and performance tasks includes adequate procedures and criteria for ensuring and documenting inter-rater reliability (e.g., clear scoring rubrics, adequate training for and qualifying of raters, evaluation of inter-rater reliability, and documentation of quality control procedures);</li> <li>Results of inter-rater reliability of scores on constructed-response items and performance tasks;</li> </ul> </li> <li>For scoring of technology-enhanced items: <ul style="list-style-type: none"> <li>Evidence that the scoring procedures are working as intended across multiple test administration scenarios;</li> </ul> </li> <li>For machine scoring of constructed-response items: <ul style="list-style-type: none"> <li>Evidence that the scoring algorithm and procedures are appropriate, such as descriptions of development and calibration, validation procedures, monitoring, and quality control procedures;</li> <li>Evidence that machine scoring produces scores that are comparable to those produced by human scorers, such as rater agreement rates for human- and machine-scored samples of responses (e.g., by student</li> </ul> </li> </ul>

Actual requirement

Not a requirement.  
A helpful but non-exhaustive list of **examples** that **could** address the Critical Element.

8

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

8



## Key Messages from Breakout Sessions (Top Ten Takeaways)

9



### 1. “Yes” is possible.

- It is possible to use multiple (innovative) approaches and meet the requirements in the critical elements.
  - The submission probably won't look like your traditional summative submission.
  - Success is dependent on coherence across decisions while building the system (and a few other things).

10



## 2. “Yes” requires coherence.

- Success is dependent on coherence across decisions within the assessment system.
- The sequence of decisions won’t always follow best practice.
- In the end, the system still needs to operate as a whole (coherence!) and fulfill the purposes and intended uses.



## 3. Know what questions to answer, and when.

- How to plan for backward design? e.g.,
  - Purposes and intended uses
  - Reporting metrics and intended interpretations to support intended uses
  - Test design that supports the desired score reporting
  - Etc.

Consider using breakout slide decks to see the relationships between decisions



#### 4. Confirm the boundaries of peer review requirements.

- What evidence supports summative uses?
- What might peers expect that is out of bounds?
- Consult USDE long before the submission deadline.

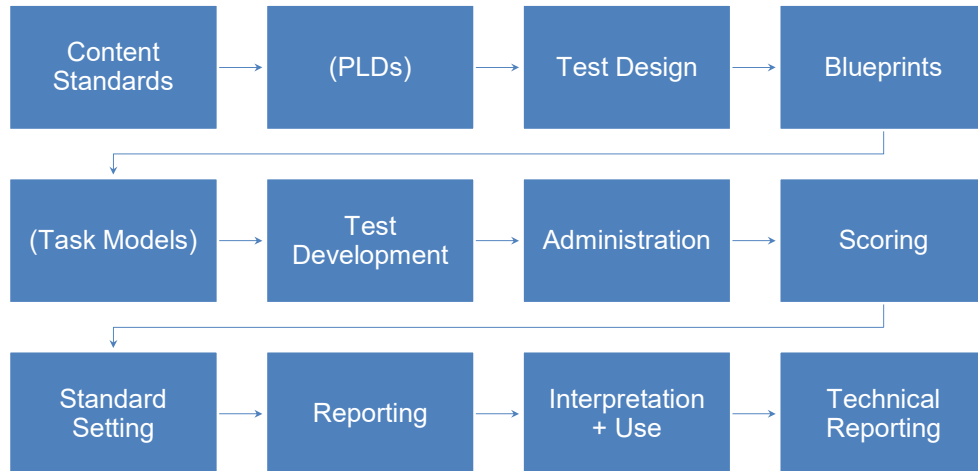


#### 5. Plan for peer review from day 2.

- Identify who is responsible for producing what evidence and who is responsible for writing the responses.
  - Consider including in RFP, with details about tasks & timelines
- Create (and update as needed) a peer review response plan.
  - Identify who is accountable
- Collect and organize evidence throughout stages of assessment system development.
  - Table of CEs, expected sources of evidence, status column
  - Update as you go; use findings to correct course or gather more evidence
  - Challenge yourself to draft responses along the way (don't procrastinate)
- Leave time to refine, synthesize evidence.



## Common Phases in Assessment Development



15

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

15



## 6. Present logical evidence.

- Think (and write) in terms of chains of evidence.
- Example from session 1B:
  - Test design differentially weights standards → blueprint structure and test development reflect the weighting → alignment criteria and evaluation reflect the weighting
- In the peer review submission:
  - Describe intended content relationships.
  - Describe procedural evidence.
  - Conduct an external alignment study using a design and criteria appropriate for the assessment.
  - Provide evidence of how the state interprets and responds to findings.

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

16





## 7. Educate your peers.

- Don't assume peers deeply understand your assessment design. Educate them.
  - Include a succinct statement that “answers the question” the critical element is asking or provides background needed to evaluate the evidence.
  - Leave “bread crumbs” in the index responses to cross-reference critical elements.
  - Explain atypical evidence.
  - Strive for coherence.
- Consider where and how to provide a succinct summary of the program (there are options)

17

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

17



Provide  
background  
needed to  
evaluate the  
evidence

### Notes

#### Validity Framework and Overall Evaluation

The DLM validity framework is based in the project's theory of action (1.a.i), developed with state partners. **There are four propositions to support the intended uses and interpretations of DLM scores:**

1. Scores represent what students know and can do.
2. Achievement level descriptors provide useful information about student achievement.
3. Inferences regarding student achievement, progress, and growth can be drawn at the conceptual area level.
4. Assessment scores provide useful information to guide instructional decisions.

**Summative scores from DLM assessments are intended for use for several purposes (1.b.vi):**

1. Reporting achievement and growth within the taught content aligned to grade-level content standards to a variety of audiences including educators and parents
2. Inclusion in state accountability models to evaluate school and district performance
3. Planning instructional priorities and program improvements for the following school year

Technical documentation of evidence supporting the validity of score interpretation and use includes material included in and referenced in chapters throughout the 2014-2015 Dynamic Learning Maps Technical Manual. The Manual addresses the design and development of the assessment, alignment of standards and test content, test administration, and test scores and reports. Evidence is presented related to content, response process, internal structure, relationships to other variables, and consequences (1.b.i). Evaluation of the evidence for overall validity of score interpretation and use is described for each proposition and related assumptions (1.b.ii), and is summarized in Chapter 11 of the Manual. Evaluation results indicate general support for the propositions and intended uses of summative results (1.b.iii), appropriate for the first year of a new assessment system. Additional validity studies are planned and in progress (1.b.iv) and additional procedural evidence is being collected as part of the consortium's continuous improvement process (1.b.v).

18

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

18



## Help peers interpret atypical evidence

Scoring is conducted at the linkage level within each Essential Element (EE), and an overall performance level is reported based on the total number of linkage levels mastered in the content area. Results are also reported for each EE and at the conceptual area level. Conceptual areas contain groups of related EE (content standards). Based on the diagnostic classification model, reliability evidence is provided at three levels:

- Content-area (performance-level) reliability provides reliability evidence for the total number of linkage levels mastered across all EEs for a given content area, which is analogous to total score reliability in Classical Test Theory (CTT)- or Item Response Theory (IRT)-based models. Estimates were calculated for each grade level in each content area, as demonstrated by the correlation between true and estimated number of linkage levels mastered. Values ranged from .909 to .965, indicating generally consistent measurement at the content area level (1.c.i).
- EE reliability provides reliability evidence for the number of linkage levels mastered within a single EE. Estimates were calculated for the 255 EEs across both content areas. EE reliability statistics are at a finer grain size than conceptual area because each conceptual area contains multiple EE. In this sense, conceptual areas are like strands and conceptual area results are like sub-scores. While conceptual area reliability estimates are planned for future analysis, EE reliability statistics provide evidence of consistency at the content standard level. Results from the Pearson correlation between true and observed values indicated that for 77.8% of EEs, the correlation was  $\geq .75$  (1.cii).
- Linkage Level reliability provides reliability evidence for the classification accuracy of each of the 1,275 individual linkage levels across both content areas. Although at a larger grain size than item-level reliability statistics in CTT or IRT-based models, the linkage level is the smallest reported unit in a diagnostic classification model scoring system (1.a.i, 1.b.i). Results of the tetrachoric correlation between true and observed mastery status indicated that for 82.2% of linkage levels, the correlation was  $\geq .80$  (1.c.iii).

19

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

19



## 8. Make your case.

- Explain your rationale within the indices.
- Make sure you have correctly and completely cited the evidence.
  - Don't include extraneous or ambiguous evidence.
- Make sure the evidence says what you think it says.
  - Consider a critical friend review

20

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

20



## 9. There are different ways to make your case.

- ***Direct evidence from the assessment program***
- Other examples of similar methodologies
- Published research
- Evidence of TAC advice

21

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

21



## 10. Peer review is a state responsibility.

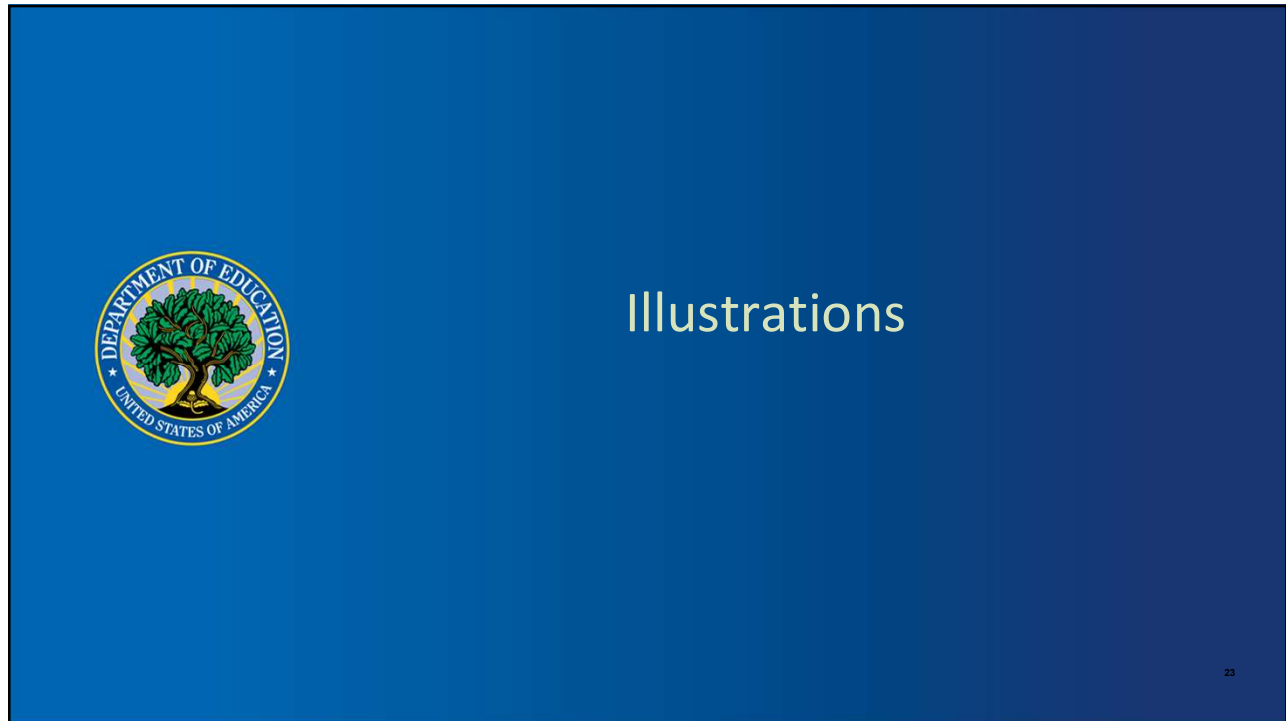
- Peer review yields a determination for the state (not a consortium, not a vendor).
- Contact School Support and Accountability for guidance.

22


1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

22



23



### **Examples: Making the Case**

- Two examples of using the index to frame the evidence that doesn't fit common peer assumptions
  - Language that educates
  - Example combination of evidence to make the case

24

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

24



# #1: NY Regents Exam, Depth + Breadth

Critical Element 2.1	Evidence	Notes
<p>The State's test design and test development process is well-suited for the content, is technically sound, aligns the assessments to <b>the depth and breadth of the State's academic content standards</b> for the grade that is being assessed and includes:</p>	<p>Test Maps</p>	<p>Each exam form is a sampling of the domain, as there are more learning standards to be measured than there are items that appear on an individual form.</p> <p>Therefore, it is necessary to consider multiple administrations when determining that the Regents Examinations effectively measure the full range of New York State's learning standards each year (as a reminder, Regents Exams are administered in January, June, and August each year).</p> <p>The item maps presented represent two of the three administrations of the Regents Examination in Algebra I offered in 20XX; each exam form measures both one set of anchor learning standards and another set of alternating learning standards.</p> <p>This allows New York State to measure the full range of learning standards through the course of the three Regents Examination forms administered in an academic year.</p>

25

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

25



## DLM Example: CE 2.1

### 2.1 Test Design and Development

Critical Element	Evidence	Notes
<p><b>2.1 – Test Design and Development</b></p> <p>The State's test design and test development process is well-suited for the content, is technically sound, aligns the assessments to the full range of the State's academic content standards, and includes:</p> <ul style="list-style-type: none"> <li>Statement(s) of the purposes of the assessments and the intended interpretations and uses of results</li> </ul>	<p>Several Technical Manual chapters/sections</p>	<ul style="list-style-type: none"> <li>Statement of purposes</li> <li>Explanation of content structures</li> <li>Intended uses</li> </ul>
<ul style="list-style-type: none"> <li>Test blueprints that describe the structure of each assessment in sufficient detail to support the development of assessments that are technically sound, measure the full range of the State's grade-level academic content standards, and support the intended interpretations and uses of the results</li> </ul>	<ul style="list-style-type: none"> <li>Several Technical Manual chapters/sections</li> <li>Content standards</li> <li>Blueprints</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of claims and content structures</li> <li>Intended assessment alignment and test structure</li> <li>Stakeholder involvement and processes for content standards development, including stakeholder evaluations of the process</li> <li>Rationale for how content standards ensure the assessment system meets the depth and breadth requirement</li> <li>Rationale for content standards prioritized for assessment and how blueprints address the breadth requirement</li> <li>Explanation of atypical content structure, intended alignment for ELA/mathematics in general, and exception for writing assessment</li> </ul>
<ul style="list-style-type: none"> <li>Processes to ensure that each assessment is tailored to the knowledge and skills included in the State's academic content standards, reflects appropriate inclusion of challenging content, and requires complex demonstrations or applications of knowledge and skills (i.e., higher-order thinking skills)</li> </ul>	<ul style="list-style-type: none"> <li>Several Technical Manual chapters/sections</li> <li>Section from Test Administration Manual</li> </ul>	<ul style="list-style-type: none"> <li>Explanation of DLM variant of Evidence-Centered Design approach</li> <li>Explanation of DLM variant of task templates</li> <li>Reminder of important parts of content structures and the goals behind the structure</li> </ul>
<ul style="list-style-type: none"> <li>If the State administers computer-adaptive assessments, the item pool and item selection procedures adequately support the test design</li> </ul>	<ul style="list-style-type: none"> <li>Several Technical Manual chapters/sections</li> </ul>	<ul style="list-style-type: none"> <li>Succinct reminder of content structure, test windows, test pools, test assignment procedures for each window</li> <li>Justification of sufficient coverage of the pool to meet operational needs</li> <li>Explanation for initial testlet assignment in adaptive spring portion</li> <li>Explanation of what is and is not subject to teacher choice, to ensure intended coverage and cognitive challenge</li> <li>Explanation of routing algorithm (similar to CAT)</li> </ul>

26

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

26





## DLM Example: CE 3.1

### 3.1 – Overall Validity, including Validity Based on Content

Critical Element	Evidence	Notes
<p><b>3.1 – Overall Validity, including Validity Based on Content</b></p> <p>The State has documented adequate overall validity evidence for its assessments.</p>	<p><b>Validity Framework and Overall Evaluation</b></p> <ul style="list-style-type: none"> <li>Several Technical Manual chapters/sections</li> </ul>	<p><b>Validity Framework and Overall Evaluation</b></p> <ul style="list-style-type: none"> <li>Brief summary of validity approach (theory of action, propositions, intended uses)</li> <li>Brief summary of synthesized validity evidence presented in culminating technical manual chapter</li> <li>Overall conclusion about sufficiency of existing evidence, plans for ongoing validity evaluation</li> </ul>
<p>The State's validity evidence includes evidence that the State's assessments measure the knowledge and skills specified in the State's academic content standards, including:</p> <ul style="list-style-type: none"> <li>For alternate assessments based on alternate academic achievement standards, the assessments show adequate linkage to the State's academic content standards in terms of content match (i.e., no unrelated content) and the breadth of content and cognitive complexity determined in test design to be appropriate for students with the most significant cognitive disabilities.</li> </ul>	<p><b>Measurement of Academic Content Standards</b></p> <ul style="list-style-type: none"> <li>Several Technical Manual chapters/sections</li> <li>Technical reports on external review</li> <li>External alignment study report</li> <li>Response to external alignment study</li> </ul>	<p><b>Measurement of Academic Content Standards</b></p> <ul style="list-style-type: none"> <li>Reminder that <u>score</u> interpretation is dependent on content structures (and what those content structures are)</li> <li>Procedural evidence about test development                             <ul style="list-style-type: none"> <li>Methods and stakeholder engagement in standards development</li> <li>Rationale for how content standards ensure the assessment system meets the depth and breadth requirement</li> <li>Explanation of how blueprint covers breadth and depth with teacher-driven flexibility</li> <li>Design and use of task templates</li> <li>Intended item alignment (content and cognitive complexity)</li> <li>Item writer training to promote alignment</li> <li>Pre-field testing alignment checks (methods and empirical evidence)</li> </ul> </li> <li>Empirical evidence:                             <ul style="list-style-type: none"> <li>Reminder of pre-FT external review criteria and results</li> <li>External alignment study purposes, alignment evaluation questions, and results</li> <li>Process for responding to external study findings and overall internal conclusions, including follow-up studies and refined procedures</li> <li>Evidence of blueprint coverage at the student level (aggregated data) and steps toward continuous improvement</li> </ul> </li> </ul>

27

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

27



## Deeper Dive: Discuss Critical Elements

28

28



## What's the goal?

- Think about how to interpret and respond to some critical elements, using approaches you are considering or pursuing
- Use table discussions to “stress test” the 2018 guidance
- Facilitators will share personal opinions and compile notes to share with USDE

29

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

29



## For each Critical Element:

- Think about the assessment program
- Read the requirement and the example evidence
- Discuss guiding questions at your table
- Share highlights from table discussion at the end of each cycle

30

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

30



## Guiding Questions

1. **What evidence makes sense** for this Critical Element, given the assessment you're thinking about?
2. Where are some **potential stumbling blocks** when thinking about how to respond to this critical element?
3. What **additional examples** would help you understand how to prepare a submission for this Critical Element?

31

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

31



## Step 1: Move yourself as needed.

At a table where nobody is thinking about multiple approaches?

Find new friends.

32

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

32





## Example 1: CE 2.1 (Test Design & Dev.)

### Critical Element 2.1 – Test Design and Development

	Examples of Evidence
<p>The State's test design and test development process is well-suited for the content, is technically sound, aligns the assessments to (1) <u>the depth and breadth of the State's academic content standards</u> for the grade that is being assessed, or (2) <u>the depth and breadth of the State's ELP standards</u>, and includes:</p> <ul style="list-style-type: none"> <li>Statement(s) of the purposes of the assessments and the intended interpretations and uses of results;</li> <li>Test blueprints that describe the structure of each assessment in sufficient detail to support the development of assessments that are technically sound, measure the depth and breadth of (1) <u>the State's grade-level academic content standards</u> or (2) <u>the State's ELP standards</u>, and support the intended interpretations and uses of the results.</li> <li>Processes to ensure that each academic assessment is tailored to the knowledge and skills included in <u>the State's academic content standards</u>, reflects appropriate inclusion of challenging content, and requires complex demonstrations or applications of knowledge and skills (i.e., higher-order thinking skills).</li> <li>Processes to ensure that the ELP assessment is tailored to the knowledge and skills included in <u>the State's ELP standards</u> and reflects</li> </ul>	<p>Evidence to support this critical element for all of the State's assessments includes:</p> <p>For the State's general <u>academic</u> content and <u>ELP</u> assessments:</p> <ul style="list-style-type: none"> <li>Relevant sections of State code or regulations, language from contract(s) for the State's academic and ELP assessments, test coordinator or test administrator manuals, or other relevant documentation that states the purposes of these assessments and the intended interpretations and uses of results;</li> <li>Test blueprints that: <ul style="list-style-type: none"> <li>Describe the structure of each <u>academic content</u> and <u>ELP</u> assessment in sufficient detail to support the development of a technically sound assessment, for example, in terms of the number of items, item types, the proportion of item types, response formats, range of item difficulties, types of scoring procedures, and applicable time limits;</li> <li>Align to either: (1) the depth and breadth of the State's <u>grade-level academic content standards</u> in terms of balance of content (i.e., knowledge, cognitive process, cognitive complexity); or (2) <u>the State's grade-level (or grade-band) ELP standards</u> in terms of content (i.e., knowledge and linguistic process), the depth and breadth of the State's grade-level/grade-band standards and balance of content; and documentation that the test design is tailored to the specific knowledge and linguistic skills in the State's ELP standards, and reflects academic language complexity appropriate for each grade-level/grade-band;</li> </ul> </li> <li>Documentation that the test design that is tailored to the specific knowledge and skills in: (1) <u>the State's academic content standards</u> (e.g., includes extended response items that require demonstration of writing skills if the State's reading/language arts academic content standards include writing) or (2) <u>the State's ELP standards</u> (e.g., includes speaking, listening, reading, and writing skills and tasks found in the standards);</li> <li>Documentation of the approaches the State uses to include challenging content and complex demonstrations or applications of knowledge and skills (i.e., items that assess higher-order thinking skills, such as item types appropriate to the content that require synthesizing and evaluating information and analytical text-based writing or multiple steps and student explanations of their work); for example, this could include test specifications or test blueprints that require a certain portion of the total score be based on item types that require complex demonstrations or applications of knowledge and skills and the rationale for that design.</li> </ul> <p><b>33</b> For the State's technology-based general assessments, in addition to the above:</p> <ul style="list-style-type: none"> <li>Evidence of the usability of the technology-based presentation of the assessments, including the usability of accessibility tools and features (e.g., embedded in test items or available as an accompaniment to the items), such as descriptions of conformance with established accessibility standards and best practices and usability studies;</li> </ul>

33

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

33



## Guiding Questions

1. What evidence makes sense for this Critical Element, given the assessment you're thinking about?
2. Where are some **potential stumbling blocks** when thinking about how to respond to this critical element?
3. What **additional examples** would help you understand how to prepare a submission for this Critical Element?

34

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

34



## Example 1: CE 2.1 (Test Design & Dev.)

### Critical Element 2.1 – Test Design and Development

	Examples of Evidence
<p>The State's test design and test development process is well-suited for the content, is technically sound, aligns the assessments to (1) <u>the depth and breadth of the State's academic content standards</u> for the grade that is being assessed, or (2) <u>the depth and breadth of the State's ELP standards</u>, and includes:</p> <ul style="list-style-type: none"> <li>Statement(s) of the purposes of the assessments and the intended interpretations and uses of results;</li> <li>Test blueprints that describe the structure of each assessment in sufficient detail to support the development of assessments that are technically sound, measure the depth and breadth of (1) <u>the State's grade-level academic content standards</u> or (2) <u>the State's ELP standards</u>, and support the intended interpretations and uses of the results.</li> <li>Processes to ensure that each academic assessment is tailored to the knowledge and skills included in <u>the State's academic content standards</u>, reflects appropriate inclusion of challenging content, and requires complex demonstrations or applications of knowledge and skills (i.e., higher-order thinking skills).</li> <li>Processes to ensure that the ELP assessment is tailored to the knowledge and skills included in <u>the State's ELP standards</u> and reflects</li> </ul>	<p>Evidence to support this critical element for all of the State's assessments includes:</p> <p>For the State's general <u>academic</u> content and <u>ELP</u> assessments:</p> <ul style="list-style-type: none"> <li>Relevant sections of State code or regulations, language from contract(s) for the State's academic and ELP assessments, test coordinator or test administrator manuals, or other relevant documentation that states the purposes of these assessments and the intended interpretations and uses of results;</li> <li>Test blueprints that: <ul style="list-style-type: none"> <li>Describe the structure of each <u>academic content</u> and <u>ELP</u> assessment in sufficient detail to support the development of a technically sound assessment, for example, in terms of the number of items, item types, the proportion of item types, response formats, range of item difficulties, types of scoring procedures, and applicable time limits;</li> <li>Align to either: (1) the depth and breadth of the State's <u>grade-level academic content standards</u> in terms of balance of content (i.e., knowledge, cognitive process, cognitive complexity); or (2) <u>the State's grade-level (or grade-band) ELP standards</u> in terms of content (i.e., knowledge and linguistic process), the depth and breadth of the State's grade-level/grade-band standards and balance of content; and documentation that the test design is tailored to the specific knowledge and linguistic skills in the State's ELP standards, and reflects academic language complexity appropriate for each grade-level/grade-band;</li> </ul> </li> <li>Documentation that the test design that is tailored to the specific knowledge and skills in: (1) <u>the State's academic content standards</u> (e.g., includes extended response items that require demonstration of writing skills if the State's reading/language arts academic content standards include writing) or (2) <u>the State's ELP standards</u> (e.g., includes speaking, listening, reading, and writing skills and tasks found in the standards);</li> <li>Documentation of the approaches the State uses to include challenging content and complex demonstrations or applications of knowledge and skills (i.e., items that assess higher-order thinking skills, such as item types appropriate to the content that require synthesizing and evaluating information and analytical text-based writing or multiple steps and student explanations of their work); for example, this could include test specifications or test blueprints that require a certain portion of the total score be based on item types that require complex demonstrations or applications of knowledge and skills and the rationale for that design.</li> </ul> <p><b>33</b> For the State's technology-based general assessments, in addition to the above:</p> <ul style="list-style-type: none"> <li>Evidence of the usability of the technology-based presentation of the assessments, including the usability of accessibility tools and features (e.g., embedded in test items or available as an accompaniment to the items), such as descriptions of conformance with established accessibility standards and best practices and usability studies;</li> </ul>

35

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

35



## Guiding Questions

1. What evidence makes sense for this Critical Element, given the assessment you're thinking about?
2. Where are some **potential stumbling blocks** when thinking about how to respond to this critical element?
3. What **additional examples** would help you understand how to prepare a submission for this Critical Element?

36

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

36



## Example 2: CE 4.2 (Fairness & Accessibility)

### Critical Element 4.2 – Fairness and Accessibility

	Examples of Evidence
<p><i>For all State academic and ELP assessments</i>, assessments should be developed, to the extent practicable, using the principles of universal design for learning (UDL) (see definition).</p>	<p>Evidence to support this critical element for the State's general academic content and ELP assessments, AA-AAAS, and AELPA includes:</p> <ul style="list-style-type: none"> <li>• Documentation of steps the State has taken in the design and development of its assessments, such as:               <ul style="list-style-type: none"> <li>○ Documentation describing approaches used in the design and development of the State's assessments (e.g., principles of UDL, language simplification, accessibility tools and features embedded in test items or available as an accompaniment to the items);</li> <li>○ Documentation of the approaches used for developing items;</li> <li>○ Documentation of procedures used for maximizing accessibility of items during the development process, such as guidelines for accessibility and accessibility tools and features included in item specifications;</li> <li>○ Description or examples of instructions provided to item writers and reviewers that address writing accessible items, available accessibility tools and features, and reviewing items for accessibility;</li> <li>○ Documentation of procedures for developing and reviewing items in alternative formats or substitute items and for ensuring these items conform with item specifications;</li> <li>○ Documentation of routine bias and sensitivity training for item writers and reviewers;</li> </ul> </li> </ul>
<p><i>For academic content assessments</i>, the State has taken reasonable and appropriate steps to ensure that its assessments are accessible to all students and fair across student groups in their design, development and analysis.</p>	
<p><i>For ELP assessments</i>, the State has taken</p>	
<p>reasonable and appropriate steps to ensure that its assessments are accessible to all EL students and fair across student groups, including ELs with disabilities, in their design, development, and analysis.</p>	<ul style="list-style-type: none"> <li>○ Documentation that for (1) <b>academic assessments</b>, experts in the assessment of students with disabilities, ELs and individuals familiar with the needs of other student populations in the State were involved in item development and review; or (2) <b>ELP assessments</b>, experts in language assessment, and in the assessment of ELs and students with disabilities, including ELs with hearing or vision impairments, were involved in item development and review;</li> <li>○ Descriptions of the processes used to write, review, and evaluate items for bias and sensitivity;</li> <li>○ Description of processes to evaluate items for bias during pilot and field testing;</li> <li>○ Evidence submitted under Critical Elements 2.1 – Test Design and Development and Critical Element 2.2 – Item Development;</li> <li>• Documentation of steps the State has taken in the analysis of its assessments, such as results of empirical analyses (e.g., differential item functioning (DIF) and differential test functioning (DTF) analyses) that identify possible bias or inconsistent interpretations of results across student groups.</li> </ul>

37

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

37



## Guiding Questions

1. What evidence makes sense for this Critical Element, given the assessment you're thinking about?
2. Where are some **potential stumbling blocks** when thinking about how to respond to this critical element?
3. What **additional examples** would help you understand how to prepare a submission for this Critical Element?

38

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

38



## Example 3: CE 6.4 (Reporting)

### Critical Element 6.4 – Reporting

The State reports its assessment results for all students assessed, and the reporting facilitates timely, appropriate, credible, and defensible interpretations and uses of those results by parents, educators, State officials, policymakers and other stakeholders, and the public.

The State reports to the public its assessment results on: (1) student academic achievement for all students and each student group at each achievement level and (2) English language proficiency for all ELs including the number and percentage of ELs attaining ELP.

For academic content assessments, the State reports assessment results, including itemized score analyses, to districts and schools so that parents, teachers, principals, and administrators can interpret the results and address the specific academic needs of students, and the State also provides interpretive guides to support appropriate uses of the assessment results.

- The State provides for the production and delivery of individual student interpretive, descriptive, and

### Examples of Evidence

Collectively, for the State's assessment system, evidence to support this critical element must demonstrate that the State's reporting system facilitates timely, appropriate, credible, and defensible interpretation and use of its assessment results.

Evidence to support this critical element for all the State's assessments includes:

- Evidence that the State reports to the public : (1) student academic achievement for all students and each student subgroup (at each proficiency level and the percentage of students not tested); or (2) English language proficiency for all ELs (including the number and percentage of ELs attaining ELP) after each test administration, such as:
  - State report(s) of assessment results (e.g., a State report card);
  - Appropriate interpretive documents provided in or with the State report(s) that addresses appropriate uses and limitations of the data (e.g., when comparisons across student groups of different sizes are and are not appropriate).
- Evidence that the State reports results for use in instruction, such as:
  - Written materials and other documentation such as interpretive guides from the State and from eligible entities;
  - Evidence that the State's reporting system includes supporting information to facilitate accurate interpretation of data for those who will receive and use its reports, such as information about the content and structure of assessments, intended purposes and uses of scores, and how the assessments are related to its (1) academic content standards; or (2) ELP standards;
  - Instructions for districts, schools, and teachers for access to assessment results, such as an electronic database of results;
  - Examples of reports of assessment results at the classroom, school, district and State levels provided to teachers, principals, and administrators that include itemized score analyses, results according to proficiency levels, performance level descriptors, and, as appropriate, other analyses that go beyond the total score (e.g., analysis of results by strand/domain/component);
  - Instructions for teachers, principals, and administrators on the appropriate interpretations and uses of results for students tested that include: the purpose and content of the assessments; assistance in interpreting the results; appropriate uses and limitations of the data; and information to allow use of the assessment results appropriately for addressing the specific academic needs of students, student groups, schools and districts.
  - Timeline that shows results are reported to districts, schools, and teachers in time to allow for the use of the

39

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

39



## Guiding Questions

1. What **evidence** makes sense for this Critical Element, given the assessment you're thinking about?
2. Where are some **potential stumbling blocks** when thinking about how to respond to this critical element?
3. What **additional examples** would help you understand how to prepare a submission for this Critical Element?

40

1G – Preparing for Peer Review – September 27, 2023

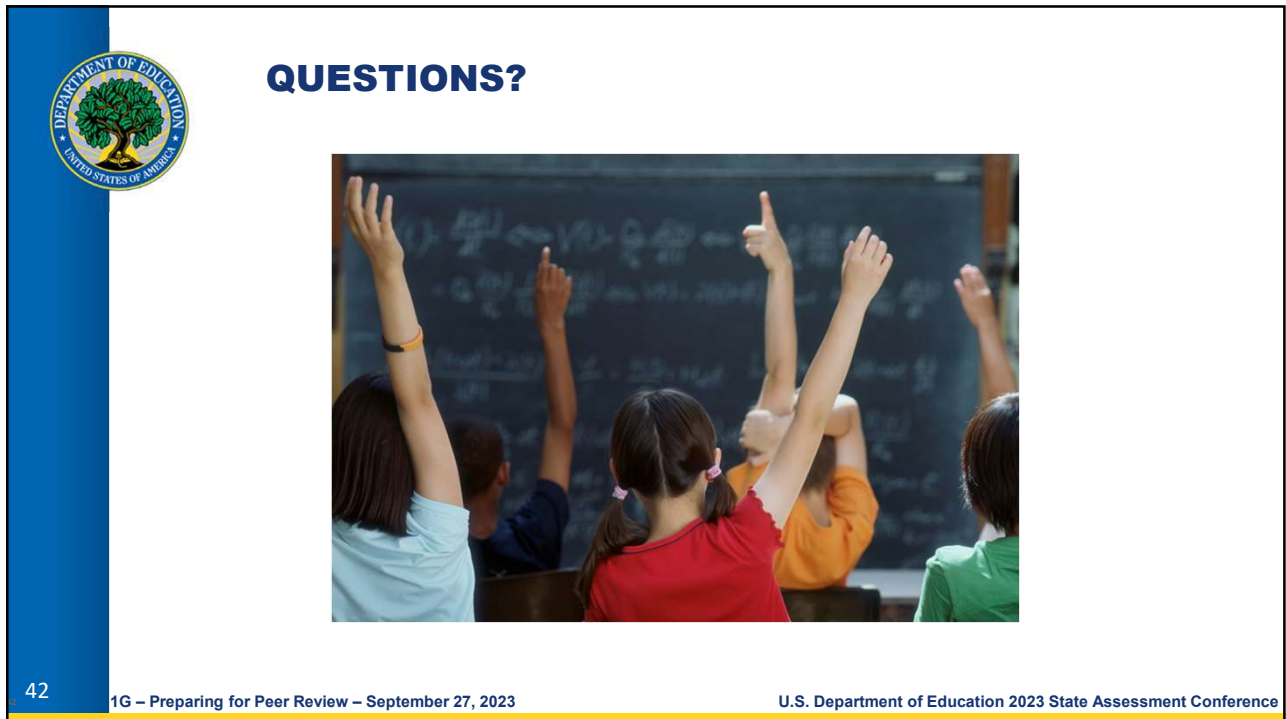
U.S. Department of Education 2023 State Assessment Conference

40





41



42



## STILL MORE QUESTIONS?

- Submit your questions using the QR code.
- We will provide the anonymous responses to USDE to inform their future thinking.



43

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

43



## Resources

- USDE peer review guidance (2018) – available on conference website
  - <https://apps1.seiservices.com/2023SSA/Materials.aspx>
- Center for Assessment’s Annotated Peer Review Guidance (2015)
  - <https://www.nciea.org/library/annotated-assessment-peer-review-guidance/>
- Center for Assessment’s Considerations for Through-Year Assessments (2023)
  - <https://www.nciea.org/library/through-year-assessment-ten-key-considerations/>

44

1G – Preparing for Peer Review – September 27, 2023

U.S. Department of Education 2023 State Assessment Conference

44



**Thank You!**