OFFICE OF ELEMENTARY AND SECONDARY EDUCATION

SCHOOL SUPPORT AND ACCOUNTABILITY

2023 State Assessment Conference
3F. Meeting Peer Review Requirements in IADA
September 27, 2023 from 12:45-2:00pm

Focus Area: 3

3F. Meeting the Requirements of Peer Review in the IADA

This session will provide participants with an overview of the tensions and challenges between the peer review requirements and IADA programs, as well as goals and strategies for producing and explaining compelling peer review evidence.

Scott Marion, Center for Assessment, smarion@nciea.org

Carla Evans, Center for Assessment, cevans@nciea.org

Meagan Karvonen, ATLAS, karvonen@ku.edu

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

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 Innovative Assessment Demonstration Authority (IADA).
- The observations and opinions of the session presenters are their own.

FOCUS AREA 3 SESSIONS

Day	Time	Session #. Title	Presenters
Sept 26	11:00-12:15pm	3A. Basics of IADA	Scott Marion Carla Evans
	1:30-2:45pm	3B. Lessons Learned about the Implementation of IADA	Scott Marion Carla Evans *Kinge Mbella (NC)* *Thomas Lambert (LA)* *Sam Ribnick (MA)* *Allison Timberlake (GA)*
	3:00-4:15pm	3C. Planning and Implementation in IADA	Scott Marion Carla Evans
Sept 27	8:45-10:00am	3D. Addressing Comparability in IADA [Repeats in 3G timeslot]	Scott Marion Carla Evans
	10:15-11:30am	3E. Including all Students in the IADA	Sheryl Lazarus Meagan Karvonen *Kinge Mbella (NC)* *Thomas Lambert (LA)* *Sam Ribnick (MA)*
	12:45-2:00pm	3F. Meeting the Requirements of Peer Review in the IADA	Scott Marion Carla Evans Meagan Karvonen Phoebe Winter
	2:15-3:30pm	3G. Addressing Comparability in IADA [Repeat from 3D]	Scott Marion Carla Evans

OVERVIEW

Time	Topic
12:45-12:50	Welcome, Introductions, & Overview
12:50-1:00	Warm-Up Question
1:00-1:05	Transitioning from IADA to Statewide: "Orderly Transition"
1:05-1:20	Pivotal Peer Review Issues for IADA
1:20-1:55	Application to State Examples
1:55-2:00	Wrap-Up

Warm-up Question

What questions do you want to make sure we discuss around **meeting peer review requirements in IADA**?

- Take a minute to jot down some ideas, then we will ask you to raise your hand and let us know.

Transitioning from IADA to Statewide:
"Orderly Transition"

PLANNING AN ORDERLY TRANSITION FROM LEGACY ASSESSMENT PROGRAMS

What is an orderly transition?

TIMELINE FOR AN IADA APPLICATION/IMPLEMENTATION

Application

- State or Consortium of States Notify ED of Intent to Apply (not required, but desirable)
- State or Consortium of States Apply
- ED provides feedback of application review within 90 days
- State(s) given opportunity to address feedback
- ED grants IADA to State(s)

Implementation/Demonstration Period (Five Years; with potential for additional time)

- States provide ED updates on progress through periodic monitoring calls (2 or 3 times per year) and Annual Performance Report (once per year)
- States may request an additional two-year extension beyond five years
- States may also request additional one-year extensions beyond the initial two-year extension
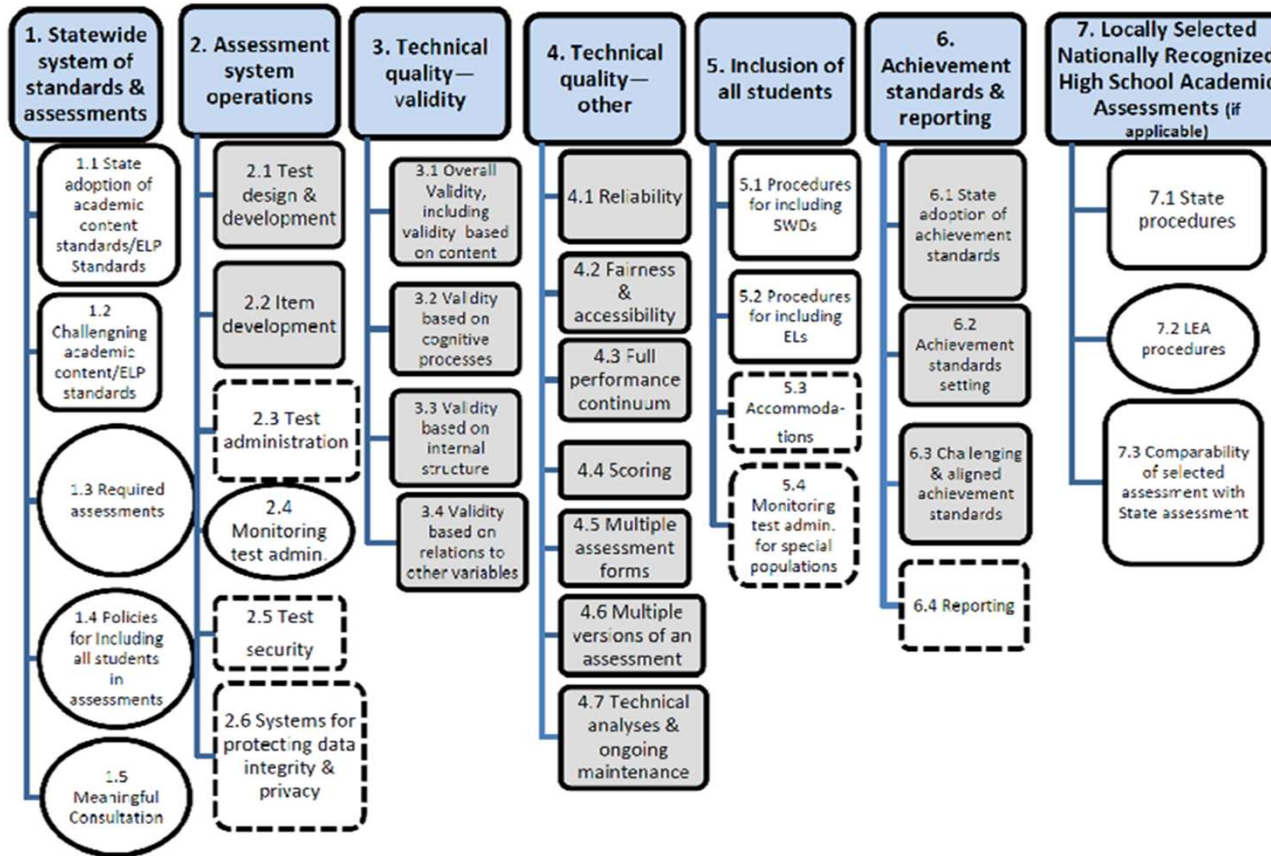
Transition to Statewide Use

- At end of demonstration period, State(s) either transition to use IADA pilot assessment as THE statewide assessment, OR
- Return to use of 'regular' Statewide assessment in all schools
- **If States transition to IADA pilot as State assessment, they submit evidence of that assessment for the Department's peer review after the first operational year Statewide**

Pivotal Peer Review Issues for IADA

II – CRITICAL ELEMENTS FOR STATE ASSESSMENT PEER REVIEW

1. Statewide system of standards & assessments
- 1.1 State adoption of academic content standards/ELP Standards
- 1.2 Challengning academic content/ELP standards
- 1.3 Required assessments
- 1.4 Policies for Including all students in assessments
- 1.5 Meaningful Consultation

2. Assessment system operations
- 2.1 Test design & development
- 2.2 Item development
- 2.3 Test administration
- 2.4 Monitoring test admin.
- 2.5 Test security
- 2.6 Systems for protecting data integrity & privacy

3. Technical quality— validity
- 3.1 Overall Validity, including validity based on content
- 3.2 Valdity based on cognitive processes
- 3.3 Valdity based on internal structure
- 3.4 Validity based on relations to other variables

4. Technical quality— other
- 4.1 Reliability
- 4.2 Fairness & accessibility
- 4.3 Full performance continuum
- 4.4 Scoring
- 4.5 Multiple assessment forms
- 4.6 Multiple versions of an assessment
- 4.7 Technical analyses & ongoing maintenance

5. Inclusion of all students
- 5.1 Procedures for including SWDs
- 5.2 Procedures for including ELs
- 5.3 Accommoda-tions
- 5.4 Monitoring test admin. for special populations

6. Achievement standards & reporting
- 6.1 State adoption of achievement standards
- 6.2 Achievement standards setting
- 6.3 Chalenging & aligned achievement standards
- 6.4 Reporting

7. Locally Selected Nationally Recognized High School Academic Assessments (if applicable)
- 7.1 State procedures
- 7.2 LEA procedures
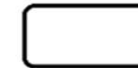- 7.3 Comparability of selected assessment with State assessment

Key

- ⬭ Critical elements primarily checked by Department staff (e.g., Critical Element 1.3)
- ▭ Critical elements likely addressed by coordinated evidence for all States administering the same assessments (e.g., Critical Element 2.1).
- ▢ Critical elements likely addressed with State-specific evidence (e.g., Critical Element 5.1).
- ⬚ Critical elements likely addressed by both State-specific evidence and

Alignment Examples

- Build alignment considerations into test design and development

- Make sure alignment requirements match your purpose and uses

  o Performance Assessments
    - Scoring rules are aligned to target standards
    - Variations in forms/tasks maintain target standards

  o Through-Year Assessments and Instructionally Embedded Assessments
    - The overall blueprint guides scoring and reporting for each component
    - Variations in administration across students produce summative results that meet the overall blueprint

  o Systems with Pre-Summative Components (not part of peer review)
    - Components used for summative routing are aligned to support accurate routing decisions
    - Pre-summative components administered to inform student instruction are explicitly aligned to state standards

Reliability Examples

- Performance Assessments
  - Interrater training and agreement
  - Reliability of scores across and within variations

- Through-Year Assessments
  - Reliability of each component
  - Effects of weighting on total score reliability

- Instructionally Embedded Assessments
  - Effects of re-takes on component reliability
  - Effects of various pathways on reliability

Fairness and Comparability Examples

- Accessibility

- Comparable content and scores

  - Performance Assessments
    - Timing in instruction
    - Administration conditions

  - Through-Year Assessments
    - Fairness by design
    - Equivalent instruction

  - Instructionally Embedded Assessments
    - Timing in relation to instruction
    - Consistency in retake decisions

Interpreting Summative Scores

- Score interpretations match system intent

- Checking design decisions against theory of action

- Building in plans for needed revisions

- Understanding how system characteristics (e.g., weighting) affect interpretations

- Clearly explaining how scores can be interpreted and **why**

Application to State Examples

ACTIVITY OVERVIEW

1. We will present an overview of the NH PACE system (key design features)
2. Activity & Discussion: Participants will then consider the key features of NH PACE and the peer review elements to identify which of the peer review elements would be particularly challenging to meet.
3. Repeat steps above for
   - Louisiana
   - Georgia MAP & NAAVY
   - North Carolina
   - Massachusetts Science

NEW HAMPSHIRE PACE ASSESSMENT SYSTEM

State summative assessment in select grades

Local performance assessments → Competency 1

Local performance assessments → Competency 2

Local performance assessments → Competency 3

Local performance assessments → Competency 4

PACE Common Performance Task

District-Level Competency Scores

PACE Comparable Annual Determinations
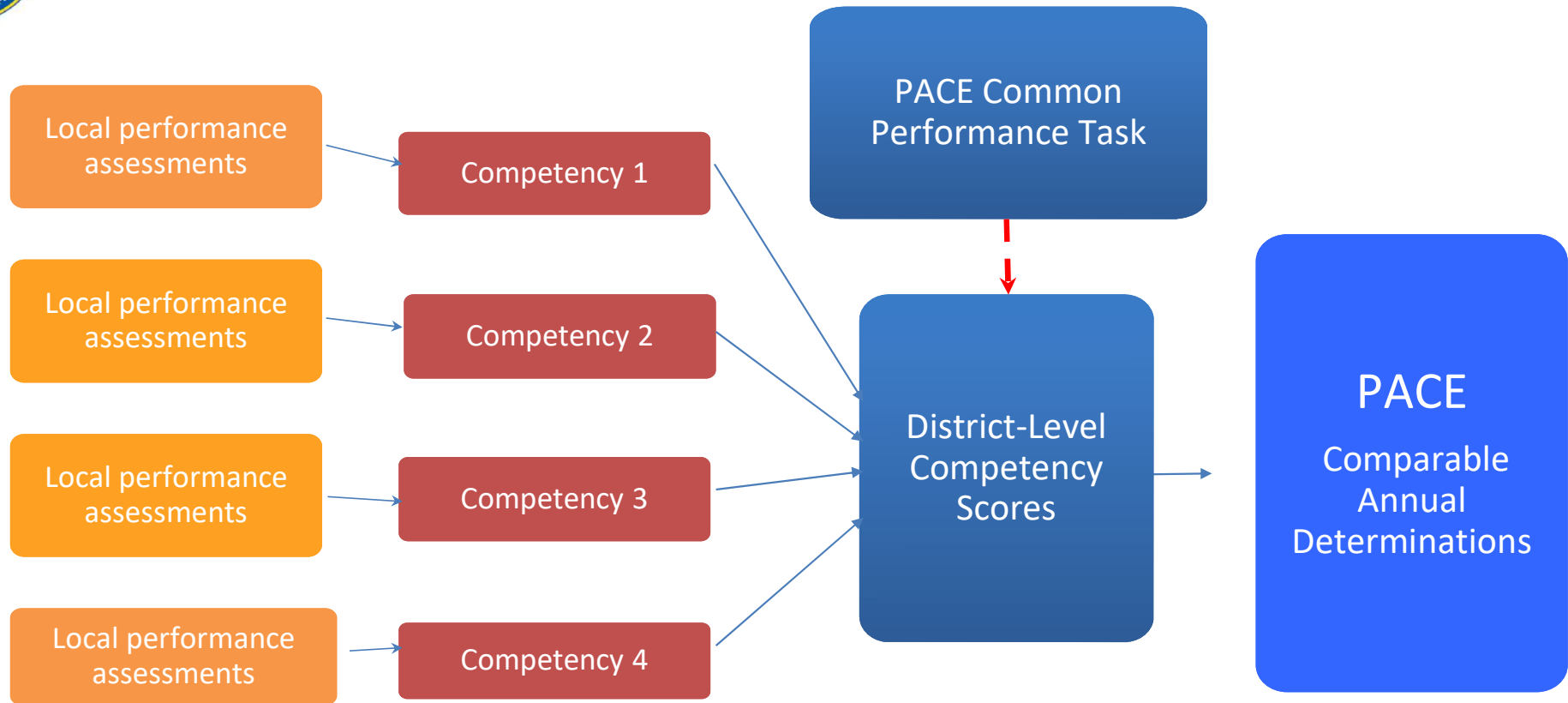
NH PACE: ACTIVITY & DISCUSSION

Which of the peer review elements would be particularly challenging to meet given the key innovative assessment pilot design features?
- Turn and talk (2-3 min) followed by a whole-group discussion/share-out (5 min)

1. **Alignment**: How would NHDOE show alignment between every local classroom's assessment system and the breadth and depth of state content standards?
2. **Reliability**: Because the innovative assessment system was based on local classroom assessments graded by classroom teachers and teacher judgments, how would the NHDOE show the reliability of the teacher judgments and the reliability of student achievement scores given the non-standardization in design, implementation, and administration?
3. **Comparability: A**gain, because of total local control except for the one common performance task, how would the NHDOE provide evidence that there were comparable judgments within schools about the quality/proficiency of student work; comparable judgments across participating LEAs in the pilot; and comparable judgments across the two state assessment systems (NHSAS and PACE)?

Louisiana's Innovative Assessment Pilot

Created and scaled quality, text- and content-rich curriculum (ELA Guidebooks) aligned to LA State Standards

Built assessments with a focus on quality (LEAP 2025)

Assessment system is comprised of end-of-unit assessment tied to units of instruction

Supported implementation of quality curriculum with aligned PD

PROFESSIONAL DEVELOPMENT

STANDARDS & CURRICULUM

ASSESSMENT

LA IAP: ACTIVITY & DISCUSSION

Which of the peer review elements would be particularly challenging to meet given the key innovative assessment pilot design features?
- Turn and talk (2-3 min) followed by a whole-group discussion/share-out (5 min)

1. **Technical quality--Creating a summative score:** How are they going to use information from the various three timepoints (because they are) for the through-year assessment as part of a student's summative determination of proficiency and provide information that students are not differentially affected because they may read a book in class and take that unit assessment rather than another, or that SWDs or ELLS are not differentially affected when information from earlier in the year is incorporated into EOY summative score?
2. **Accommodations:** Also, need to show evidence about how they are providing accommodations for SWDs and ELLs as part of design and administration since these are more like classroom-embedded assessments based on 'hot' reads.
3. **Alignment:** Louisiana could face alignment issues depending on the standards coverage of each of their TY components.

North Carolina's "extended" multi-stage adaptive test
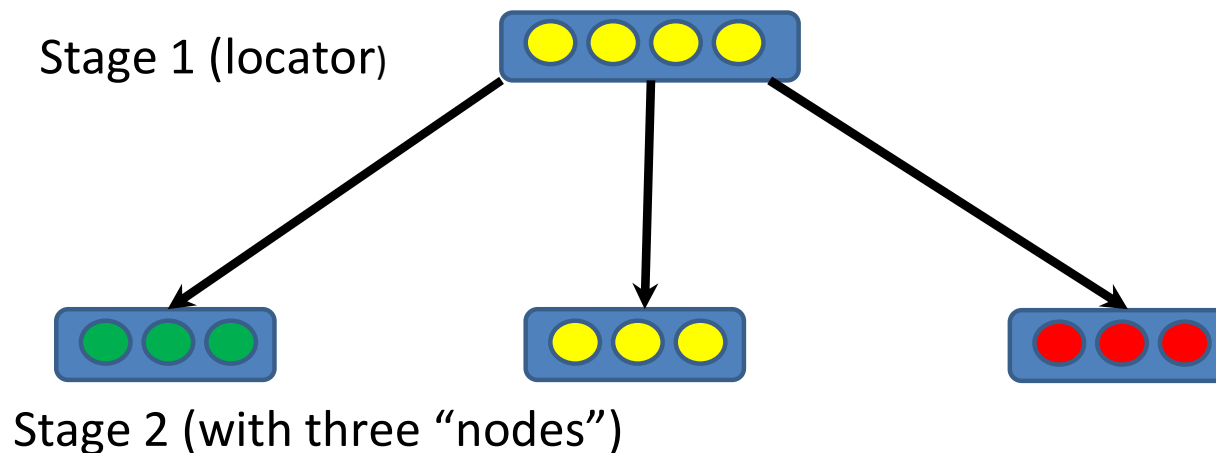
Building on the foundation of NC Check-Ins, an assessment administered as an interim that provides formative student-level information, the North Carolina Department of Public Instruction is piloting an innovative assessment that will optimize the purposes of formative, interim, and summative into one assessment.

Stage 1 (locator)

Stage 2 (with three "nodes")

NC has proposed using the results from the fall and winter interim tests as "stage 1." Basically, an optional through year assessment system with the state test at the end of the year holding all the weight

NC IAP: ACTIVITY & DISCUSSION

Which of the peer review elements would be particularly challenging to meet given the key innovative assessment pilot design features?
- Turn and talk (2-3 min) followed by a whole-group discussion/share-out (5 min)

Very little peer review issues because the end of year test is basically the same as before, except earlier information can inform a student's placement in the multi-stage adaptive so they **just need to show that that placement doesn't prevent a student from going all the way up and all the way down the achievement levels** and that they meet the requirements related to all the other typical peer review elements.

Georgia: Piloted two systems

Georgia took an unusual approach to the IADA. They ran two pilots, even though the law is clear that by the end of the demonstration period, there can be only one state assessment system.

- NAAVY—standards-based assessment happening over the course of the school year using diagnostic classification modeling (DCM) to define "mastery" on assessments tied to each standard in math and ELA (20-30 standards per content area per grade)
- GMAP—based on NWEA's Measures of Academic Progress (MAP), this is somewhat like the NC model involving testing 3x/year, potentially using the multi-stage design to replace the single summative test.

GA IAP: ACTIVITY & DISCUSSION

Which of the peer review elements would be particularly challenging to meet given the key innovative assessment pilot design features?
- Turn and talk (2-3 min) followed by a whole-group discussion/share-out (5 min)

1. **GMAP:** If we assume they would have relied on the earlier administrations being optional then the last administration would be used to produce student summative determinations. *If that is true*, and they were not using within-year information, then the same peer review issues with using a test that is not necessarily aligned to the GA state content standards or ALDs would be present. Also, we are not sure how much MAP has the same accessibility and accommodations abilities as the GA state test or peer review expectations.

2. **NAAVY:** Issues for peer review would likely be around alignment (did each student take enough assessments to provide a valid and reliable estimate of their achievement on the full breadth and depth of the state content standards), what was their evidence of reliability of student scores especially as they were collecting 'classroom type' information over the school year where students hopefully were improving their performance, to what extent did all students (including SWDs and ELLS) have the same opportunity to score in the highest achievement levels since earlier in the year information is included in the final determination of proficiency, etc.

Massachusetts

- Focuses on science
- Will maintain a shortened (½ length) MCAS throughout the pilot
- Developing innovative technology-enhanced tasks—
- Engaging in rapid prototyping to refine the tasks
- Tied to the Kaleidoscope Collective for Learning
  - The Kaleidoscope Collective is a pilot program created in November 2019 to nurture Deeper Learning in Massachusetts public schools.
  - Kaleidoscope builds partnerships with educators and leaders through professional development, coaching, guidance, tools, and resources -- all built on a foundation of educational equity.

MA IAP: ACTIVITY & DISCUSSION

Which of the peer review elements would be particularly challenging to meet given the key innovative assessment pilot design features?
- Turn and talk (2-3 min) followed by a whole-group discussion/share-out (5 min)

- **Alignment & Comparability:** No distinct issues with alignment or comparability due to linking/anchor mini-MCAS.
- **Reliability:** Need to provide evidence about how they are using the performance task within a student's score and how reliable that information is given the known generalizability issues with a limited set of performance tasks.

Summary: IADA Design Features & Meeting Peer Review
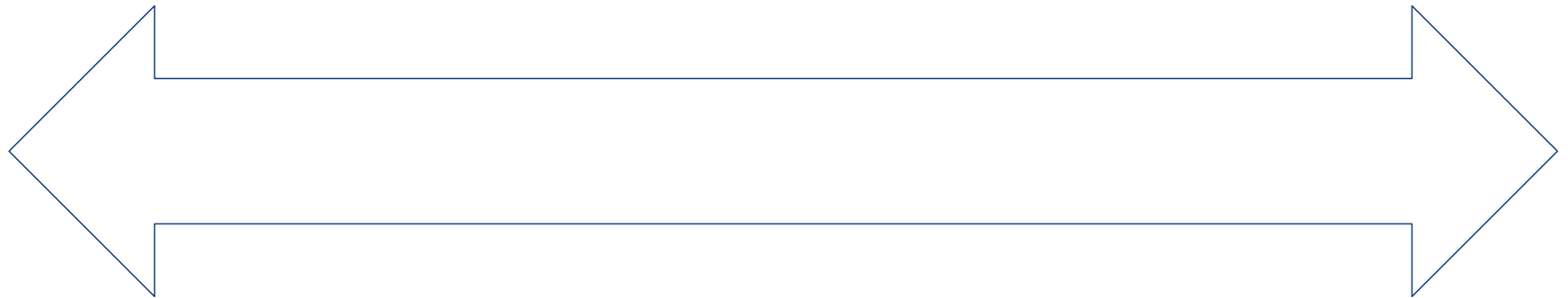
Flexible
Locally Selected
Unstandardized

Inflexible
Common
Standardized

QUESTIONS?

FOCUS AREA 3 SESSIONS

Day	Time	Session #. Title	Presenters
Sept 26	11:00-12:15pm	3A. Basics of IADA	Scott Marion Carla Evans
	1:30-2:45pm	3B. Lessons Learned about the Implementation of IADA	Scott Marion Carla Evans *Kinge Mbella (NC)* *Thomas Lambert (LA)* *Sam Ribnick (MA)* *Allison Timberlake (GA)*
	3:00-4:15pm	3C. Planning and Implementation in IADA	Scott Marion Carla Evans
Sept 27	8:45-10:00am	3D. Addressing Comparability in IADA [Repeats in 3G timeslot]	Scott Marion Carla Evans
	10:15-11:30am	3E. Including all Students in the IADA	Sheryl Lazarus Meagan Karvonen *Kinge Mbella (NC)* *Thomas Lambert (LA)* *Sam Ribnick (MA)*
	12:45-2:00pm	3F. Meeting the Requirements of Peer Review in the IADA	Scott Marion Carla Evans Meagan Karvonen Phoebe Winter
	2:15-3:30pm	3G. Addressing Comparability in IADA [Repeat from 3D]	Scott Marion Carla Evans

Thank You!

3F_Meeting Peer Review Requirements in IADA_092723   **U.S. Department of Education 2023 State Assessment Conference**