



OFFICE OF ELEMENTARY AND SECONDARY EDUCATION

SCHOOL SUPPORT AND ACCOUNTABILITY

2023 State Assessment Conference

3D & 3G. Addressing Comparability in IADA

September 27, 2023 from 8:45-10:00am or 2:15-3:30pm



FOCUS AREA: 3 3D & 3G. Addressing Comparability in IADA

This session will provide participants with an overview of the challenges associated with the IADA comparability requirements. Drawing on some of the RFI responses and previous recommendations, we will have an in-depth discussion of options for meeting the IADA comparability requirements.

Scott Marion, Center for Assessment, smarion@nciea.org
Carla Evans, Center for Assessment, cevans@nciea.org



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 <i>Kinge Mbella (NC)</i> <i>Thomas Lambert (LA)</i> <i>Sam Ribnick (MA)</i> <i>Allison Timberlake (GA)</i>
	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 <i>Kinge Mbella (NC)</i> <i>Thomas Lambert (LA)</i> <i>Sam Ribnick (MA)</i>
	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

Session 3D Time (AM)	Session 3G Time (PM)	Topic
8:45-8:50	2:15-2:20	Welcome, Introductions, & Overview
8:50-9:00	2:20-2:30	Warm-Up Questions
9:00-9:40	2:30-3:10	Presentation
9:40-9:55	3:10-3:25	Discussion Questions Audience Q & A
9:55-10:00	3:25-3:30	Wrap-Up



WARM-UP QUESTIONS

Turn to a partner and discuss:

1. How do you define comparability?
2. What questions do you have about comparability in IADA? *[one person should be willing to share out questions that haven't already been said aloud]*



OVERVIEW

- What is comparability?
- Why do we care about comparability?
- What do the regulations say?
- What are the issues for IADA states?
- What are different ways of thinking about comparability? What would require regulatory changes, if anything?
- What are other ideas for addressing comparability in IADA?



WHAT IS COMPARABILITY?

*“Similar to validity, comparability is not an attribute of a test or test form, nor is it a yes/no decision. Instead, comparability relates to the degree to which the scores resulting from different assessment conditions can support the same inferences about what students know and can do... **Comparability is defined, therefore, as the degree to which the results of assessments intended to measure the same learning targets produce the same or similar inferences.**”*

(Evans & Lyons, 2017, p. 2)



WHY DO WE CARE ABOUT COMPARABILITY?

State assessments are used to hold schools accountable for the academic performance of their students (among other indicators).

Without comparability across assessment results, students and schools could be held to different performance expectations, which could quickly become unfair and inequitable.





WHAT DO THE REGULATIONS SAY?

Section § 200.105(b)(4)(ii) has been added to require that States' innovative assessment systems **generate results, including annual summative determinations**, that are valid, reliable, and comparable for all students and for each subgroup of students among participating schools and LEAs, which an SEA must annually determine as part of its evaluation plan described in § 200.106(e) (proposed § 200.78(e)).

Section 200.105(b)(4) has been revised to clarify that determinations of the comparability between the innovative and statewide assessment system **must be based on results, including annual summative determinations**, as defined in § 200.105(b)(7), that are generated for all students and for each subgroup of students.

Comparability must be re-evaluated every year.

Final Regulations for IADA: <https://www.gpo.gov/fdsys/pkg/FR-2016-12-08/pdf/2016-29126.pdf>



COMPARABILITY IN IADA

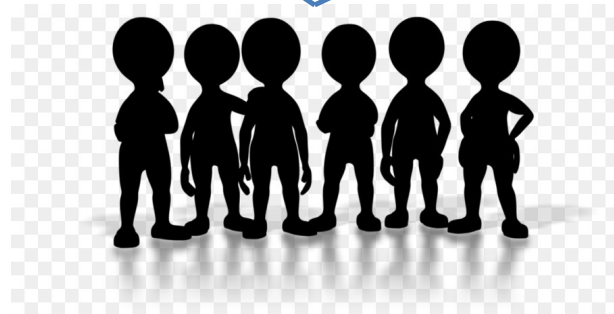
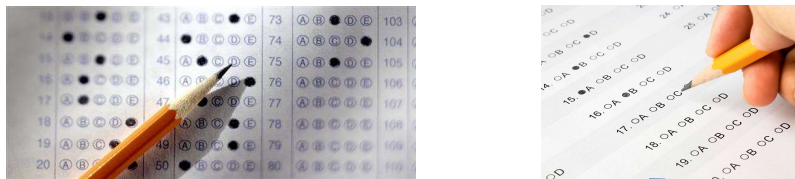
Although “comparable” is not defined in ESSA, the **IADA regulations** issued by the Department ask state education agencies to determine comparability of assessment results in **one of five defined ways**.

1. **Double testing** at least once per grade span in same subject
2. **Double testing** using a demographically representative sample of all students and subgroups at least once per grade span in same subject
3. **Linking or anchor sets** from state to innovative (items or performance tasks)
4. **Linking or anchor sets** from innovative to state (items or performance tasks)
5. **Alternative method:** “An alternative method for demonstrating comparability that an SEA can demonstrate will provide for an equally rigorous and statistically valid comparison between student performance on the innovative assessment and the statewide assessment, including for each subgroup of students...”

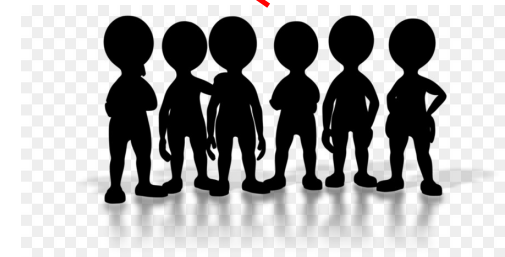
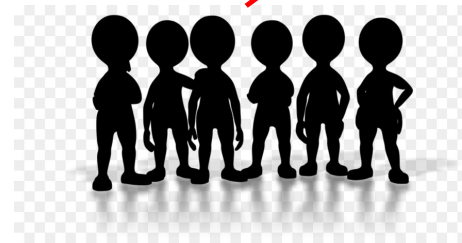
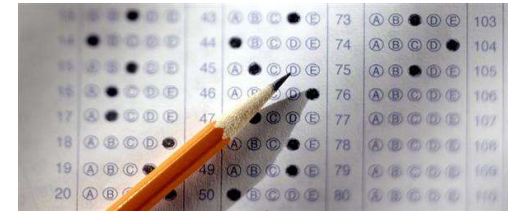


BASIC REQUIREMENTS FOR COMPARABILITY

Same Students Taking Two Different Test

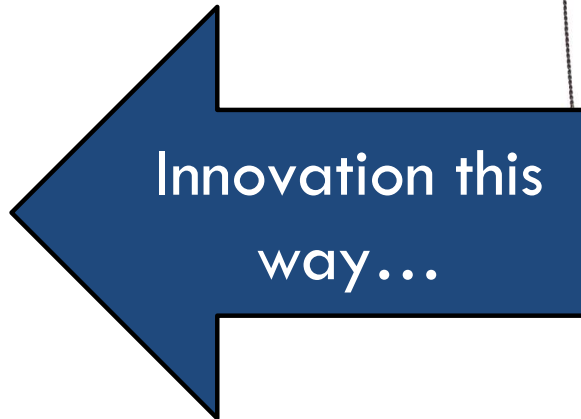


Different Students Taking The Same Test Questions





WHY AREN'T YOU BEING MORE INNOVATIVE?



Comparability
to Legacy
Assessment
Program

NOTE: States are free to establish new achievement standards for their new assessment once they have implemented the test Statewide. The comparability requirement is in effect only so long as the State continues to use the legacy assessment in non-IADA pilot schools.



WHAT ARE THE ISSUES AROUND COMPARABILITY IN IADA?

IADA Request for Information (RFI): April 2023

- Department sought feedback regarding:
 - Alternative methodologies to establish comparability
 - IADA planning timelines
 - Other barriers
- 8,861 comments received (8,800 duplicate comments).
- Comments available at:
<https://www.regulations.gov/docket/ED-2023-OESE-0043/comments?sortBy=postedDate&sortDirection=desc>
(hint—sort by newest to oldest to see the majority of unduplicated comments)



IADA RFI: COMPARABILITY

- Majority of comments encouraged the Department to more clearly define the “alternative method” for demonstrating comparability in 34 CFR § 200.105(b)(4)(i)(E).
- Several comments urged the Department to consider “comparability” to focus only on characteristics of the innovative assessment pilot.
 - In other words, only focusing on whether the IADA assessment was of high quality and also was aligned with the State’s academic content standards.
 - ESEA requires that the State evaluate the comparability of both content standard alignment and achievement standard alignment.



IADA RFI: COMPARABILITY (CONT.)

- One comment described a comparability evaluation that could be based upon **evidence of the alignment of both the innovative assessment and the statewide assessment to the content standards**; and **evidence of the consistency of achievement classifications across the two systems (achievement standards)**.
- Several comments encouraged the Department to **make no changes to the current regulations** regarding comparability in order to protect marginalized groups of students often negatively impacted by changes in statewide assessments.



IADA RFI: COMPARABILITY (CONT.)

- Several respondents argued that the IADA comparability requirements stifle innovations in State assessments.
- Comparability requirement was seen as forcing States to link their innovative assessment pilot to the current statewide assessment, which places unhelpful limitations on assessment design and development.



WHAT ARE THE ISSUES FOR IADA STATES?: THREATS TO REAL INNOVATION

Legitimate reasons for non-comparability:

1. To measure the state-defined learning targets more **efficiently** (e.g., reduced testing time);
2. To measure the learning targets more **flexibly** (e.g., when students are ready to demonstrate “mastery”);
3. To measure the learning targets more **deeply**; or
4. To measure targets more **completely** (e.g., listening, speaking, extended research, scientific investigations).

“Perfect agreement would be an indication of failure.” – Dr. Robert Brennan



WHAT ARE DIFFERENT WAYS OF THINKING ABOUT COMPARABILITY?

How Comparable Is Comparable Enough?

The comparability of test scores is a matter of degree. How comparable scores need to be for a specific test variation depends on how the test scores will be interpreted and used. We can think of the degree of comparability along two related dimensions, content and score level, as shown in Figure 1. Here content refers to both the subject matter and the degree of construct-relevant cognitive demand.

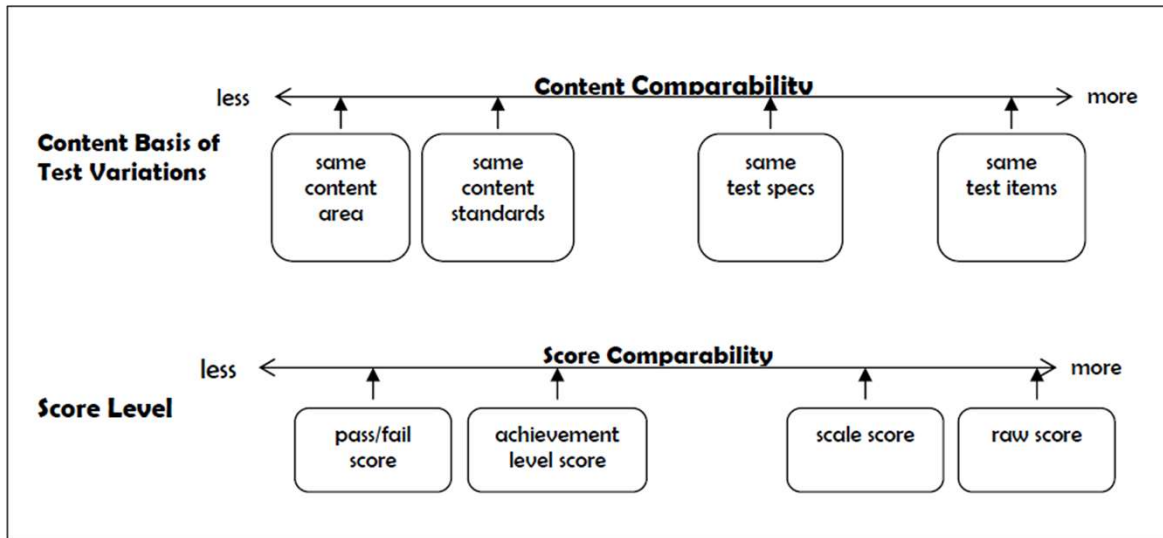


Figure 1. Comparability Continuum

Winter, P. C. (2010). *Evaluating the comparability of scores from achievement test variations*. Council of Chief State School Officers.

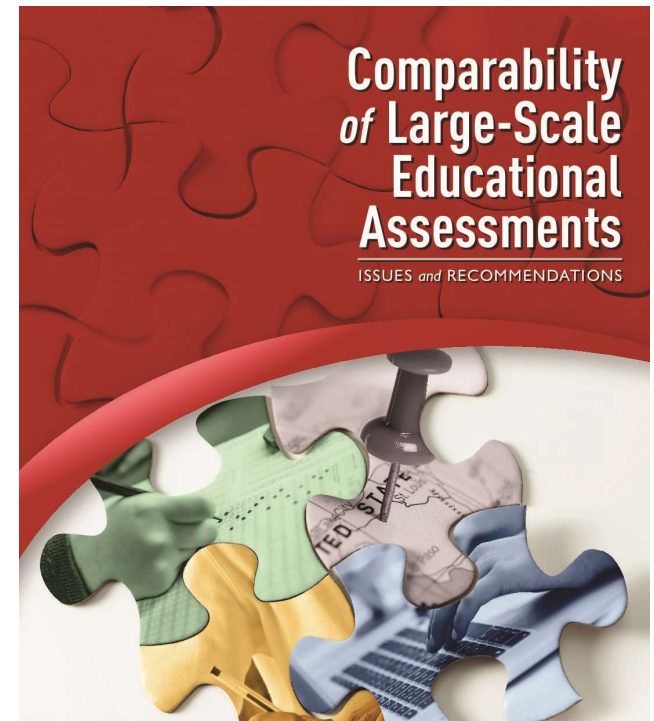


COMPARABILITY IS NOT JUST A CHALLENGE WITH IADA

Producing “**comparable annual determinations**” is a key ESSA requirement.

- Online vs. paper and pencil
- Test accommodations
- Computer adaptive tests
- Translations for English learners
- Alternative achievement standards

Remember, correlation does not mean comparable.





WHAT ARE CURRENT APPROACHES TO ADDRESSING COMPARABILITY IN IADA STATES?

Comparability Approaches:

- **Massachusetts** (mini-version of MCAS—linking set; scale score level comparability claims)
- **Louisiana** (linking set; scale score level comparability claims)
- **North Carolina** (comparability not a challenge because innovative through-year components are completely optional; only locate a student's starting position in multi-stage CAT)
- **New Hampshire** (multiple levels of comparability because mostly locally-selected assessments: within IADA district, across districts in IADA, across IADA and state assessment program; double test once per grade span; achievement level comparability)



WHAT ARE DIFFERENT WAYS OF THINKING ABOUT COMPARABILITY?

What are different ways of thinking about comparability?

Comparable to what?

- Achievement levels
- Scaled scores
- **State standards** (*comparability is reframed conceptually in terms of validity*)**
- **Across the students taking the IAP****
- Other...

****What would require regulatory changes—perhaps clarification under alternative methods?**



ANCHOR COMPARABILITY TO STANDARDS

(comparability is reframed conceptually in terms of validity)

In other words, comparability to the traditional state assessment limits innovation. The innovative assessment should be required to document that it is aligned to outcomes, in particular the depth and breadth of the state content standards. We suggest that the focus on comparability be reframed conceptually in terms of validity, that is the evidentiary basis for score interpretation and use of outcomes. Anchoring comparability to the existing test may limit the depth of thinking encouraged on the innovative assessment, because many state assessments do not measure deeper knowledge. Current IADA states face technical challenges to meet comparability requirements as they are now defined. Allowing a different, but very appropriate target (i.e., content standards) of comparability will remove this hurdle, so states do not have to restrict their innovative designs to align to traditional assessment systems. This approach to comparability still allows state leaders to use the data to inform federal accountability determinations in meaningful and appropriate ways. This more effectively supports innovation.

<https://www.regulations.gov/comment/ED-2023-OESE-0043-8858>





ANCHOR COMPARABILITY ACROSS STUDENTS TAKING THE ASSESSMENT

From: A for Arizona; Collaborative for Student Success; Education First; Great Leaders Strong Schools Indiana Department of Education; Learning Policy Institute; National Association of Charter School Authorizers (NACSA); New Meridian; NWEA; Our Turn; Teach Plus; The Education Trust; The Learning Agency.

We recommend a revision to the current interpretation of and guidance for comparability requirements such that the burden of demonstrating comparability does not require adherence to the legacy test, which is arguably the one a state believes is not meeting their needs or may be of lower quality. **New assessments should not be required to produce the same scores as existing tests**, but instead should provide strong evidence that the innovative test is of equal or higher quality than the legacy assessment and comparability be demonstrated only within the group of students taking the innovative assessment.

<https://www.regulations.gov/comment/ED-2023-OESE-0043-8861>



ONE NOTE ABOUT IADA COMPARABILITY

- When applying for the IADA, States **must describe a plan** to evaluate comparability once the IADA pilot is implemented.
- It is expected that this comparability plan **will have the potential** to satisfy IADA statute and regulations.
- It is **NOT expected** that States will have established the comparability of the IADA pilot at the time of the IADA application.



DISCUSSION QUESTIONS

What are other options for meeting IADA requirements?

More alternative methods?

What are your ideas and thoughts?



QUESTIONS?





ADDITIONAL RESOURCES

Paper

Docket ID ED-2016-OESE-0047
COMPARABILITY OPTIONS FOR STATES APPLYING FOR THE INNOVATIVE
ASSESSMENT AND ACCOUNTABILITY DEMONSTRATION AUTHORITY:
COMMENTS SUBMITTED TO THE UNITED STATES DEPARTMENT OF
EDUCATION REGARDING PROPOSED ESSA REGULATIONS¹

National Center for the Improvement of Educational Assessment

Susan Lyons and Scott Marion

September 7, 2016

Executive Summary and Key Policy Recommendations

John King, Secretary of Education, proposed new regulations under title I, part B of the Elementary and Secondary Education Act of 1965 (ESEA) to implement changes made to the ESEA by the Every Student Succeeds Act (ESSA) enacted on December 10, 2015, including the ability of the Secretary to provide demonstration authority to a State educational agency (SEA) to pilot an innovative assessment and use it for accountability and reporting purposes under title I, part A of the ESEA before scaling such an assessment statewide. This document is focused on the comparability requirements spelled out in §200.77 of the draft regulations in large part because this is one of the trickier issues for states to wrestle with and it was such a prominent feature of the proposed regulations.

As spelled out in the full document that follows, the recommendations contained herein are based on insightful contributions from some of the most prominent measurement, accountability, and innovation experts in the United States. The document provides a robust conceptualization of comparability and discusses how such a conceptualization should be applied to states proposing an innovative assessment and accountability system. We then provide a framework for designing options to evaluate comparability that considers the types of measures (items) and student sample used. As called for in §200.77, we offer more than a dozen potential approaches for evaluating comparability beyond the three proposed by ED in §200.77. We do not mean for this to be an exhaustive list, rather it should be considered a set of illustrative exemplars to highlight key aspects of the proposed framework.

<https://www.nciea.org/wp-content/uploads/2021/11/Comparability-Recommendations-for-Section-1204-Pilots.pdf>

Blogs

<https://www.nciea.org/blog/perfect-test-comparability-is-a-sign-of-failure/>



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 <i>Kinge Mbella (NC)</i> <i>Thomas Lambert (LA)</i> <i>Sam Ribnick (MA)</i> <i>Allison Timberlake (GA)</i>
	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 <i>Kinge Mbella (NC)</i> <i>Thomas Lambert (LA)</i> <i>Sam Ribnick (MA)</i>
	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!