

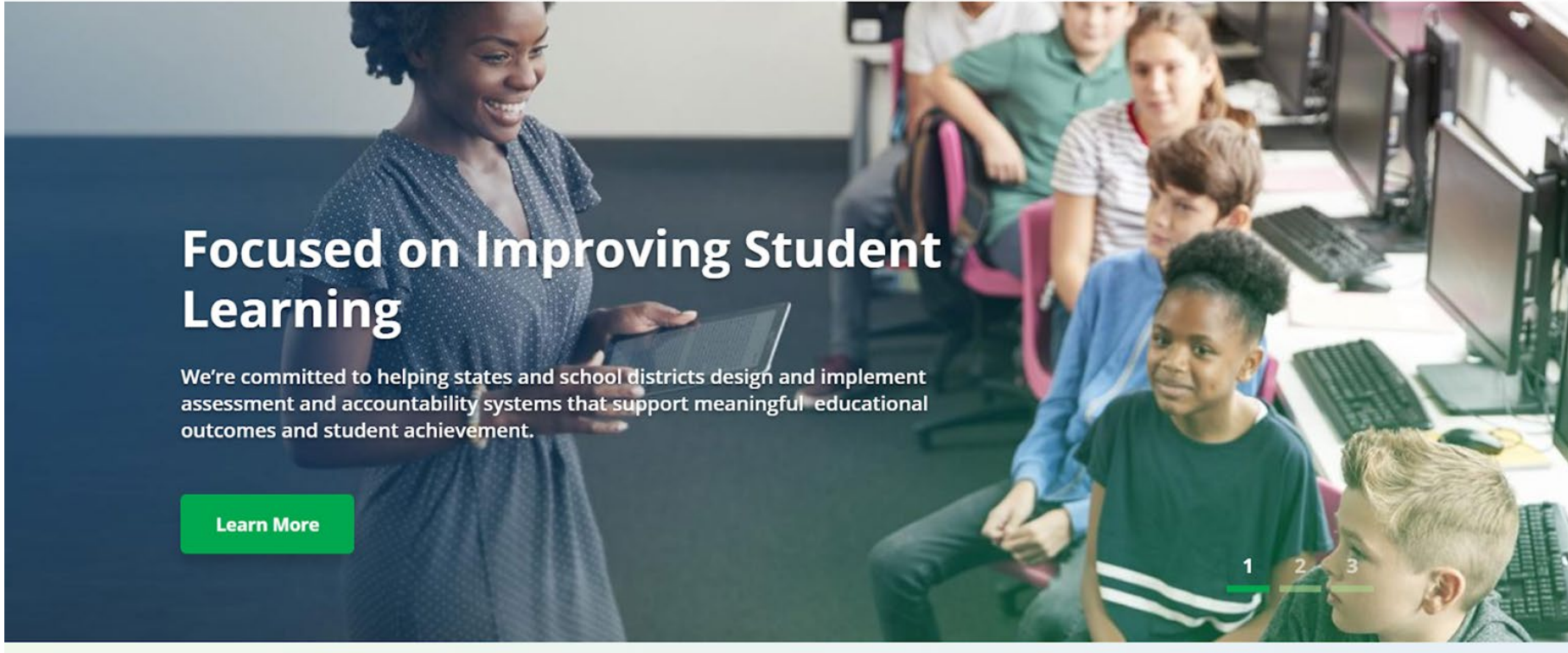


Implications for Summative Assessment

Michelle Boyer, Nathan Dadey, and Leslie Keng
Center for Assessment

Reidy Interactive Lecture Series, September 1, 2020





Focused on Improving Student Learning

We're committed to helping states and school districts design and implement assessment and accountability systems that support meaningful educational outcomes and student achievement.

[Learn More](#)

www.nciea.org

The [National Center for the Improvement of Educational Assessment, Inc.](http://www.nciea.org) (The Center for Assessment) is a Dover, NH based not-for-profit (501(c)(3)) corporation. Founded in September 1998, the Center's mission is to improve the educational achievement of students by promoting improved practices in educational assessment and accountability.

Overview

COVID-19 Response Resources

ESSA Accountability

Innovative Assessment and Accountability Systems

Comparability

The Center's COVID-19 Response Resources

State and district leaders are facing multiple concerns in response to widespread and potential long-term school closures due to the growing threat of COVID-19. The concerns are broad and consequential. Leaders are rightfully prioritizing the safety and welfare of students and the community. We have been inspired by the dedication and resourcefulness of leaders who are ensuring essential services, such as meals, are provided as well as facilitating innovative approaches to support remote learning.

Additionally, the school closures present substantial assessment and accountability implications and numerous challenges for state personnel. The Center for Assessment is poised to support our assessment and accountability colleagues around the country with technical, practical, and policy guidance and advice. We launched this page to help you efficiently find the resources you need during these uncertain times.

The resources are organized by the major categories of assessment and accountability and reflect the issues we anticipate state and district leaders will be dealing with over the next few weeks through next year. We hope you find these resources useful and if there is a question that you would like to see addressed, please [email us](#) or [tweet at us](#). We continue to wish you all the best in these uncertain times.

Featured Resources

- Restart & Recovery: Assessment Considerations for Fall 2020
- Classroom Assessment Learning Modules to Support Re-Entry Fall 2020 & Beyond
- Meeting the Moment: A Novel Format for RILS to Address Implications of the COVID-19 Pandemic

Accountability

- School Disruption Due to COVID-19A High-Level Overview of Likely Implications and Options for Assessment and Accountability
- School Disruption Due to COVID-19A High-Level Overview of Likely Implications and Options for Assessment and Accountability
- Considering Equity Within Accountability Systems in Response to Interruptions in Schooling: Making Accountability Systems Help
- The Outlook for ESSA School Accountability After COVID-19
- Dealing with Fallout from COVID-19 School Disruptions: What to do Next in Assessment and Accountability?
- Program Evaluations under COVID-19
- Rethinking School Accountability for the 2020-2021 School Year

Assessment

- We're All in This Together. Dealing Fairly with Assessment Contracts as Schools Cancel or Suspend Student Testing During the COVID-19 Crisis
- An Assessment Response to Anticipated Learning Gaps. Implications of School Closures on Assessment Needs
- In Search of Continuity of Student Learning After Extended School Closures
- Issues and Considerations that the COVID-19 Pandemic Presents for Measuring Student Growth
- Remote Learning Provides an Opportunity to Rethink Assessment (and Learning)
- Carpe Diem: Evolving Education After COVID-19
- Fall Educational Assessment: The Information You Need and How to Get It
- Summative State Assessments Can Wait!

www.nciea.org

 Current Initiatives

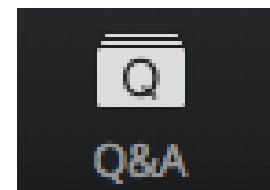
 COVID-19

Response

Resources

General Information & Zoom Protocols

- This webinar is being recorded and will be posted on the Center's RILS webpage: <https://www.nciea.org/events/rils-2020-implications-covid-19-pandemic-assessment-and-accountability>
- You can download this slide deck on the RILS webpage above
- **Introduce yourself in the chat**—your name and position (please make sure you've selected “all panelists *and attendees*”)
- Use Zoom's Q&A feature to ask questions at any time



Webinar Agenda

3:30 Welcome & Introductions

3:35 Technical Considerations Overview

Michelle Boyer, Nathan Dadey, and Leslie Keng, Center for Assessment

4:00 Panel Discussion – Moderated by Center Associates

Marc Julian, Senior Vice President – Psychometrics, DRC

Richard J. Patz, Distinguished Research Advisor, Berkeley Evaluation and Assessment Research Center, UC Berkeley

Ye Tong, Vice President – Psychometric and Research Services, Pearson

4:45 Moderated Q&A

5:00 Adjourn

Outline

1

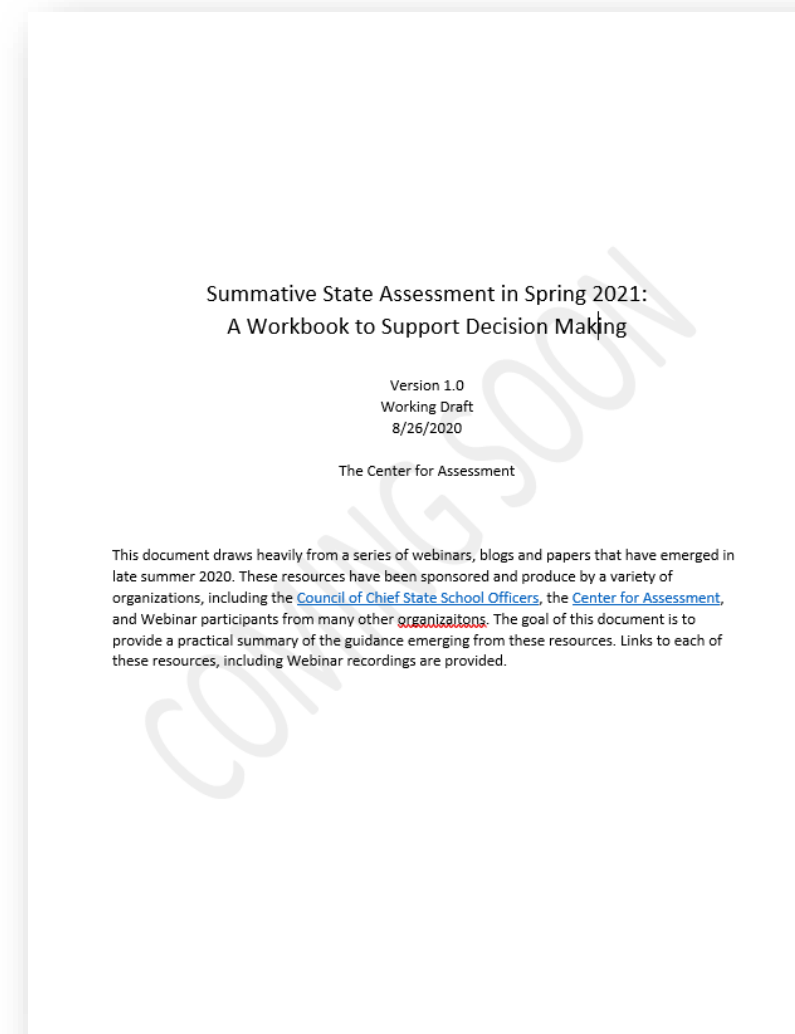
Overview of Technical Considerations

- Test Design
- Standard Setting
- Administration
- Field Testing
- Equating
- Score Interpretation & Use

2

Panel Discussion

- Greatest challenges in 2021
- Equating quality indicators
- Interpretation and use of scores



Technical Considerations Overview

Center for Assessment Associates

Center Speakers



Michelle Boyer

mboyer@nciea.org



Leslie Keng

lkeng@nciea.org



Nathan Dadey

ndadey@nciea.org

@NathanDadey

Introduction

- COVID-19 has led to disruption in schooling and suspension of testing in all states in spring 2020.
- The impact on schooling and testing in 2021 is still unclear, but differential impact by student groups is expected.
- There will be implications for various aspects of the annual development process of statewide summative assessments.
- States and their assessment vendors should develop a plan to address potential challenges in 2021. The planning should begin *as soon as possible*.

Goals and Assumptions

Goals

- Identify and address challenges to producing valid and reliable test scores in 2021 and beyond.
- Focus on useful approaches to controlling and evaluating equating accuracy under anticipated conditions.

Assumptions

- States will require summative test scores that meet professional standards for reliability, validity, and fairness.
- Those scores will need to be comparable to past and/or future scores.

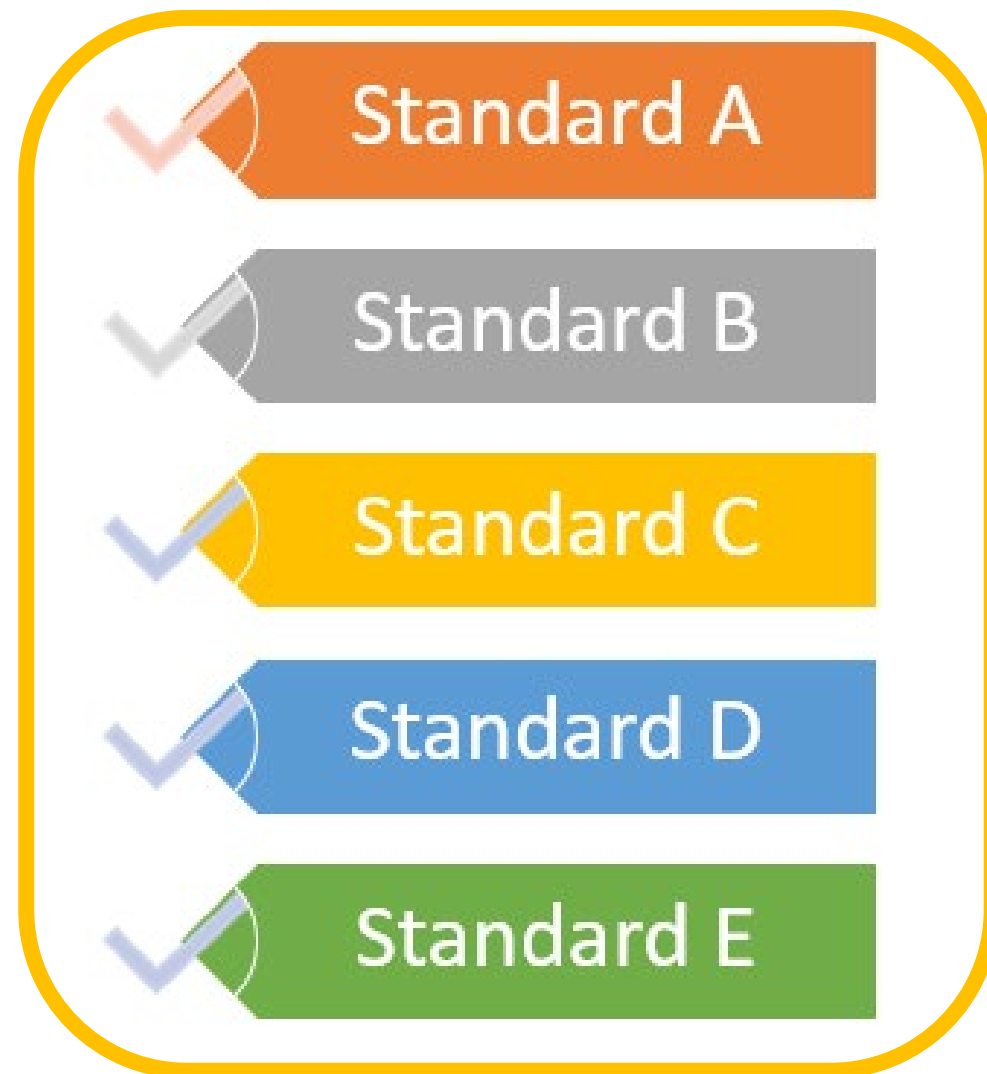
- 01 Test Design
- 02 Standards Setting
- 03 Administration
- 04 Field Testing
- 05 Equating
- 06 Interpretation and Use

- 01 Test Design
- 02 Standards Setting
- 03 Administration
- 04 Field Testing
- 05 Equating
- 06 Interpretation and Use

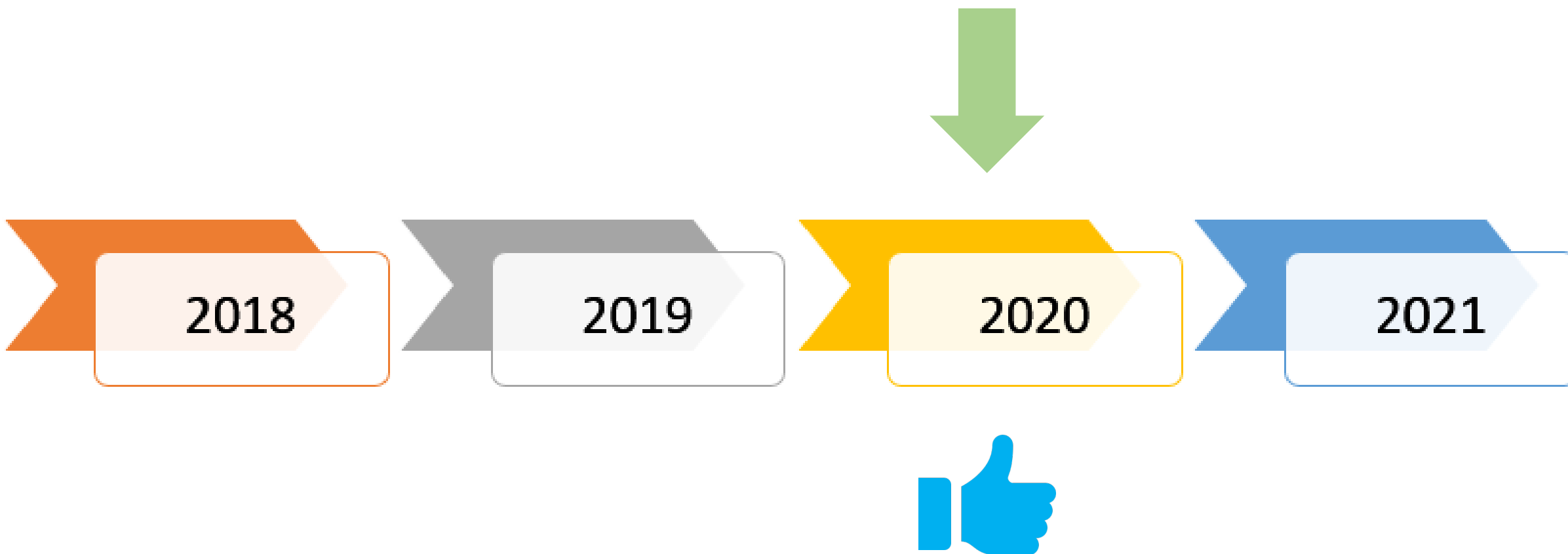
Test Design: OTL and Blueprints



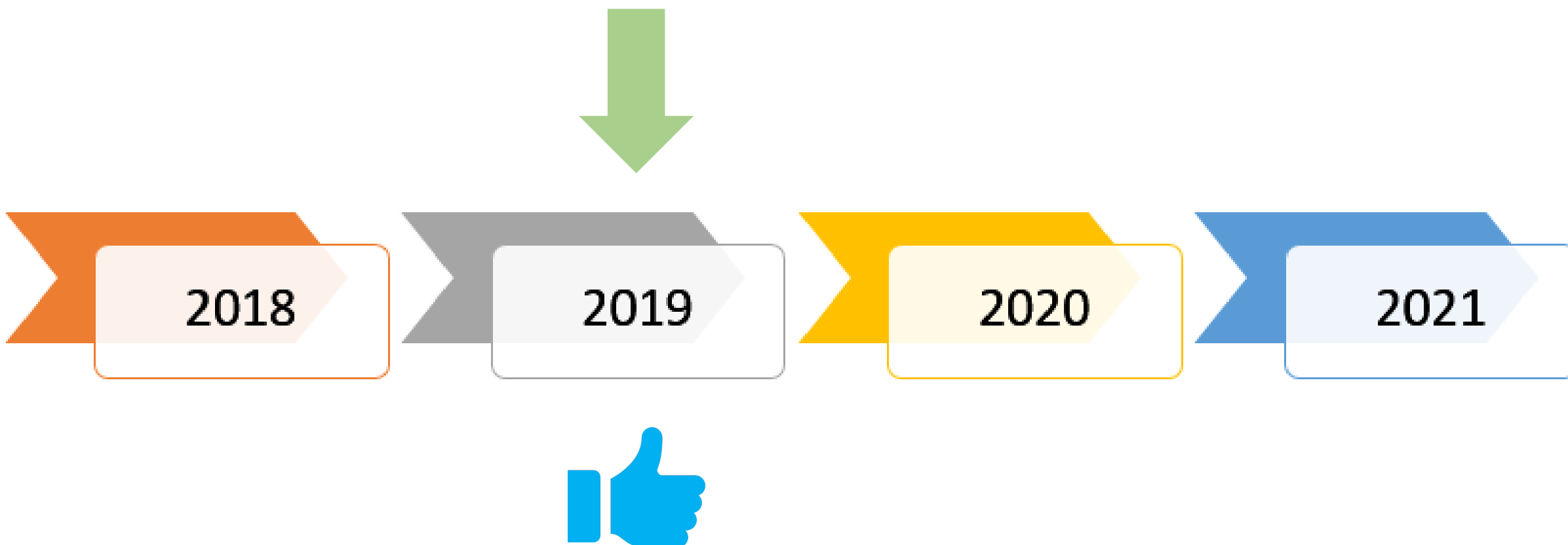
VS.



Test Design: Use of Previously Developed Tests



Test Design: Use of Previously Developed Tests



- 01 Test Design
- 02 Standards Setting**
- 03 Administration
- 04 Field Testing
- 05 Equating
- 06 Interpretation and Use

Standard Setting in 2021?



Questions and Issues to Consider

- Will as many students as previous years be able to achieve the highest levels of performance in 2021?
- Is it acceptable to exclude items from certain content strands in the standard-setting item sets or student profiles?
- If we assume overall performance will be depressed in 2021, what is the “real” level of performance we can expect in 2022 and beyond?
- If we know that COVID-19 disruptions affect students differentially, how should the standard-setting committee interpret differences in student group-level impact data based on 2021 performance?

If Standard Setting in 2021 is Needed...


- Consider a standard setting method that is less reliant on the ordering of items or persons to locate the cut scores.
- Present impact data as late as possible in the standard-setting process, e.g., after the second or third round of standard-setter judgments
- Establish criteria for reasonable impact data in subsequent administrations as the effects of learning loss gradually subside.

- 01 Test Design
- 02 Standards Setting
- 03 Administration**
- 04 Field Testing
- 05 Equating
- 06 Interpretation and Use

Administration

Instructional Context:

- Face-to-Face
- Hybrid
- Remote



Instructional contexts are mixed within schools and can fluctuate rapidly.

Administration

1. Face-to-Face
2. Remote:
 - Unproctored Internet-Based Testing
 - Proctored Internet-Based Testing



Akin to mode or
accommodation?

Considered in terms of: **logistics and safety, equity, security, and accessibility and accommodations.**

Key citations: [Keng, Boyer & Marion \(2020\)](#); [Camara \(2020\)](#); [Isbell & Kremmel \(2020\)](#); [Langenfeld \(2020\)](#); [Michel \(2020\)](#); [Steger, Schroeders & Gnambs \(2020\)](#)

Face-to-Face Testing

Logistics and Safety | Equity | Security | Accommodations

- Implementing social distancing and other safety measures
- Ensuring student and educators feel safe enough to test
- Recruiting proctors and test administrators
- Adjusting administration time and windows
- Providing remote testing options

Primarily from [Camara \(2020\)](#)

Online Remote Testing

Logistics and Safety | Equity | Security | Accommodations

- Scheduling assessments
- Providing support during the assessment
- Ensuring students have appropriate technology
- Ensuring students have sufficient familiarity with technology and online testing

Online Remote Testing

Logistics and Safety | **Equity** | Security | Accommodations

Are certain students or groups of students systematically disadvantaged by this type of administration?

In particular, do students have unequal access to:

- An appropriate device
- Internet connection
- Quiet space
- If needed, family support

Online Remote Testing

Logistics and Safety | Equity | **Security** | Accommodations

What safeguards will be in place to prevent testing improprieties?

- How will irregularities be defined, flagged, reported and handled?
- Will the test be proctored? If so, will:
 - Video proctoring be used?
 - Proctoring be conducted by a person, AI, or both?
 - Proctoring be live or based on a recording?

Some Potential Security Practices for Remote Administration

Logistics and Safety | Equity | **Security** | Accommodations

Administration

- Single testing time
- Narrow administration window
- Strict time limit
- Single access

Test Construction & Design

- Random item sequence
- Multiple forms
- Adaptive testing

Platform

- No changing answers after advancing
- Locked-down browser

From: [Langenfeld \(2020\)](#)

Remote Testing

Logistics and Safety | Equity | Security | **Accommodations**

How can we ensure that students have access to the full range of accommodations as in-person administrations?

Frequent and consistent communication on guidelines and procedures, as well as verification of implementation

→ To ensure that students who have been designated to receive accommodations receive their accommodations in ways that are intended.

Considering Potential Outcomes

Tested Students

- Census Testing
- Partial Testing with potentially unrepresentative data
 - Which can only be diagnosed in terms of collected data

Degree of Comparability

- What evidence do we have that scores obtained from face-to-face and remote testing are comparable? To what degree?

- 01 Test Design
- 02 Standards Setting
- 03 Administration
- 04 Field Testing**
- 05 Equating
- 06 Interpretation and Use

Field Testing

- Needed to Maintain Pool:
 - use higher tolerances for rejection and focus on revision
 - potentially informed by investigations of invariance of linking items
 - count on post-equating designs for these items
- Not Necessarily Needed:
 - replace field test items with additional equating items
 - remove to reduce testing time

- 01 Test Design
- 02 Standards Setting
- 03 Administration
- 04 Field Testing
- 05 Equating**
- 06 Interpretation and Use

Equating Foundation

Three features that influence the accuracy of an equating solution¹:

- Test content
- Conditions of measurement
- Examinee populations

Typically, standardized administrations and equating designs and procedures are used to control the influence of these features (to the extent possible), and we evaluate our solutions to check for any worrisome influence.

¹(Kolen, 2007)

Planning for Equating

Planning for 2021 summative assessments will need to consider all three.

Elements of planning:

1. Equating and Field Test Designs: implications for different years
2. Analysis and Acceptance Criteria (for status and for trend): specify the studies and analysis that will be require to establish adequacy of equating solutions and how how adequacy will be defined.
3. Plan for possible non-acceptance: Define alternate paths to addressing state and federal reporting and use requirements.

- 01 Test Design
- 02 Standards Setting
- 03 Administration
- 04 Field Testing
- 05 Equating
- 06 Interpretation and Use**

Considerations for Interpretation and Use

- Context differences will matter
- Gaps in access to high quality instruction will likely have consequences for our measures, and how we interpret their results.
- Opportunity-to-learn (OTL) data will be useful to contextualize score interpretations.
- Clear communication about results, and any limitations for interpretation, will be important to help stakeholders understand their meaning, and use them appropriately.

Panel Discussion

Introducing Our Distinguished Panelists!



Marc Julian

Senior Vice President –
Psychometrics
Data Recognition Corporation



Richard J. Patz

Distinguished Research Advisor
Berkeley Evaluation &
Assessment Research Center
University of California
Berkeley



Ye Tong

Vice President
Psychometric & Research
Services
Pearson



Some Initial Thoughts – Context Matters

- Take the opportunity to level set for each assessment – because each assessment administered in 2020/2021 has a different set of circumstances that will shape/guide/direct how to move forward.
- The ability to facilitate improved conditions for measurement from the 2020/2021 test administrations will be hampered by varying factors that will differ from client to client.
- Although the situation is unprecedented, we have a rigorous set of professional standards and a comprehensive set of psychometric tools that we will utilize to navigate the upcoming school year and beyond.



Reflections on Technical & Related Challenges to Summative Assessment in a Pandemic Context

- COVID-19 pandemic is damaging our students' educational progress
- Focus is needed to assess this damage and to remediate it
 - Identifying "the COVID effect" is not primarily an opportunity for research
 - Instead, a badly needed damage assessment, that may be preliminary (coarse), then refined
 - Required to catalyze and direct the required investment to restore student progress
- Stable summative assessment programs are the best available tools (initially) for quantifying the impact
 - Anything that undermines stability and comparability impairs their utility for this critical purpose
 - Multi-state consortia using common assessments (e.g., Smarter Balanced) are well positioned to support generalizable findings
 - Center for Assessment research publications address technical issues, generally support stability
- Other assessments can play a contributing role
 - An imperfect but stably administered assessment can provide useful information
 - Dropping in a new assessment for this purpose not likely to help
 - In time, NAEP and rigorous research should provide greater insight, more sensitive population measures
- Related Observations and Opinions
 - Instructionally-focused assessments should follow appropriately modified instructional strategies to restore student progress
 - Modifying educational standards in light of the pandemic seems inadvisable, but this is a policy question that technical measurement must follow (not lead)
 - Acknowledging limitations of any analogy: We don't lower safety standards when a hurricane damages a coastal highway
 - Re-directing investment, re-working accountability in the near-term, seem entirely appropriate, on path to restoration



2021 Summative Assessment

Ye Tong, Vice President, Psychometrics and Research, Pearson; President, NCME



- Viable, flexible and simple solution for the state
- Test design and blueprint
- Reuse of previous test forms
- Remote test administration considerations
- Test all students versus matrix sampling design
- Quicker reporting
- Types of information to collect and analyze



Panelist Question 1

Given the decisions you have had to make in the past about the appropriateness of equating designs and the acceptability of equating results, **what do you anticipate the biggest challenges** will be for states in using their results as intended?



Panelist Question 2

How can states **recognize if their equating results are not acceptable**? What are the kinds of red flags that signal that the results are not defensible or where heavy cautions on interpretations might be warranted?



Panelist Question 3

What would you anticipate that **scores be used for, and not used for in 2021?**



Moderated Q&A

What questions do you have for the panelists or presenters?

Top



Planning for 2021 Summative Assessment

- *Develop a 2021 research agenda, with a priori criteria for score quality related to specific intended uses.*
- *Think about student groupings--are there new groups and how, and what contextual data should be collected.*
- *Solicit technical advice early—and often.*
- *Document, document, document.*



Upcoming RILS Webinars

Day/Time	Topic	Strand
Sept 2 – 3:00-4:30pm	Outlook for Accountability	Accountability
Sept 16 – 1:00-2:30pm	Considerations for classroom assessment in a remote or hybrid context	Assessment in Support of Teaching & Learning

To Register: <https://www.nciea.org/events/rils-2020-implications-covid-19-pandemic-assessment-and-accountability>



www.nciea.org



This work is licensed under a
Creative Commons Attribution 4.0
International License. 50