

## Through-Year Assessment Virtual Convening November 15-16, 2021

The National Center for the Improvement of Educational Assessment







#### Monday, November 15, 1:00-2:30 PM ET

Defining terms, considering aims, and diving into key design features.



#### Monday, November 15, 3:00-5:00 PM ET

Connecting use cases and claims, and the designs that support them, together to consider needed evidence.



#### Tuesday, November 16, 1:00-2:30 PM ET

In depth consideration of key big picture technical and logistical issues.



#### Tuesday, November 16, 3:00-5:00 PM ET

In depth consideration of key big picture technical and logistical issues.

### The Organizers













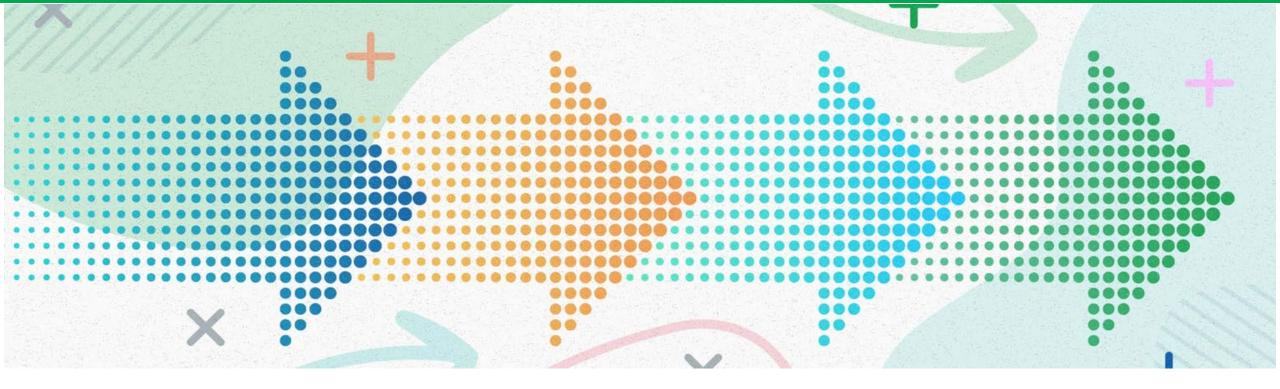
#### **Claims and Evidence**

- We are seeing a lot of strong claims and assumptions about what through-year designs can provide.
- This is not new—we've seen lots of overpromises associated with large-scale state tests for many years.
- This convening was designed to help us:
  - Develop some common language
  - Describe currently developing designs
  - Outline evidence necessary to support key claims
  - Provide insight on critical technical issues
  - Begin to outline a research and practice agenda



### A Brief Recap of Sessions 1-3

- **Session 1**—We offered a definition and described a few classes of through-year designs.
- Session 2—We took a deep dive into both summative and instructional claims associated through-year designs and described some of the evidence necessary to substantiate such claims.
- **Session 3**—We flexed our technical muscles and described several challenges that are either unique to or heightened with the use of through-year designs.
- All materials will be posted here by the end of the week:
  - https://www.nciea.org/events/claims-and-evidence-through-year-assessments-what-we-know-and-what-we-need-know



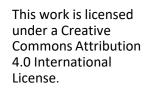
### Session 4: Threading the Needle

Through-Year Virtual Convening, November 16, 2021

Scott Marion, Will Lorié, Nathan Dadey, & Brian Gong
The National Center for the Improvement of Educational Assessment









#### **Our Orientation**

- Claims require evidence
- Stronger claims require stronger evidence
- We're pragmatists:
  - We rarely say, "No, that's impossible."
- We might eventually say, "nope, we don't see how this can work," but it will be after we have some decent evidence.
- This is our purpose for Session 4.







#### **Consequential Aims**

- As discussed in earlier sessions, through-year designs should be evaluated an intervention.
- Through-year designs are being proposed to address certain concerns with existing systems and to bring about positive changes (i.e., consequences).
- However, we don't limit our investigation to one side of the equation. We must search for and evaluate potential unintended negative consequences as a CORE part of our validity evaluation.

### The Role of Accountability

- Speaking of consequences, accountability constrains designs, but also carries many potential unintended consequences.
- The terrific panelists at our last RILS session this year spoke about this persuasively. Watch and listen <a href="here">here</a>.





#### **Panel Discussion**

 We're going to spend the rest of today's session engaged in panel discussions.

#### Part One:

• We asked each panelist to speak very briefly about they would design a system that can meaningfully address instructional or other high-priority aim, while meeting high-stakes accountability uses?

#### Part Two:

We will engage in an open discussion about several key issues.



## Our Brilliant Panelists!

### **Allison Timberlake**

Deputy Superintendent, Assessment and Accountability



### **Through-Year Assessment**

What do you think districts want and why can't the state do all the things districts want?

- Less testing or really, less high-stakes testing (accountability)
- High-stakes accountability less dependent on test scores
- Immediate feedback to adjust instruction
- Increased assessment literacy and buy-in, especially among teachers
- More local control



### **Through-Year Assessment**

What do you think districts want and why can't the state do all the things districts want?

- Serving multiple purposes impacts the test design
  - How can we build a test that meets these formative, instructional needs while ALSO meeting the high-stakes, accountability requirements?
  - Multiple purposes create implications (and tensions) for testing time, range of cognitive complexity assessed, range of item types used, score return times, etc.
- Challenges for a through-year assessment in an accountability paradigm
  - Through-year models increase accountability testing
  - When testing events "roll up" into the final score, they are no longer formative, and behavior will change
  - Administrative challenges
  - Loss of local control





Jeremy Heneger Director of Assessment

# Nebraska Student-Centered Assessment System (NSCAS) Growth



#### Lesson Learned...so far!

1. Conditions

2. Stakeholder Engagement

3. Design

#### **nwea** State Solutions

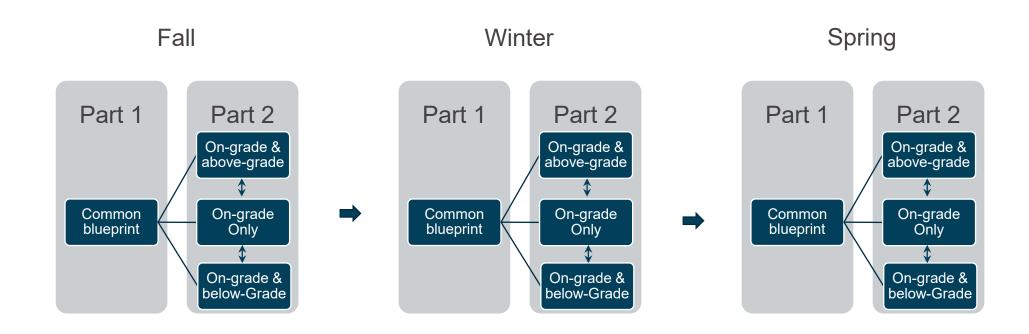
Threading the Needle with the Integrated Through-Year System

One coherent system to drive action

Garron Gianopulos, Learning and Assessment Engineer



### The Integrated Through-Year System of CATs



- Prior scores are used at start of new CAT.
- Prior scores are beneficial, but not required.
- Summative scores are updated in the spring--not aggregated.



### Innovative Technology

- Modern optimization engines can assemble tests to serve multiple purposes, provided that:
  - 1) item pools are developed and maintained well, and
  - 2) design trade-offs are acceptable to all decision makers
- + NSCAS Growth represents one instance of just such a trade-off:
  - We are giving up some summative score precision and accuracy in exchange for:
    - 1) better growth measures using the summative (blueprint) lens,
    - 2) student-centered information.



### Instructionally Embedded Assessments

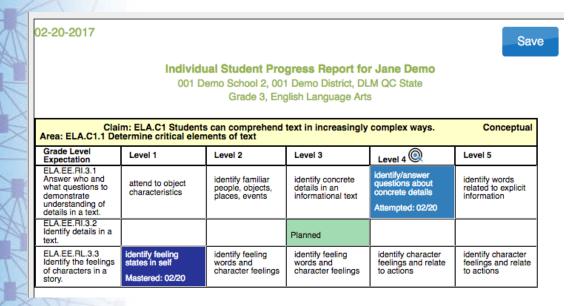
Meagan Karvonen

Director, ATLAS @ University of Kansas

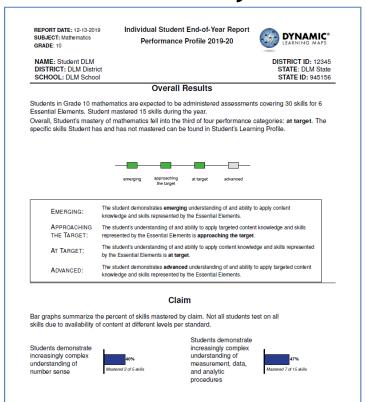


### Instructionally Embedded Assessments

### Results indicate what students know and can do adjacent to instruction

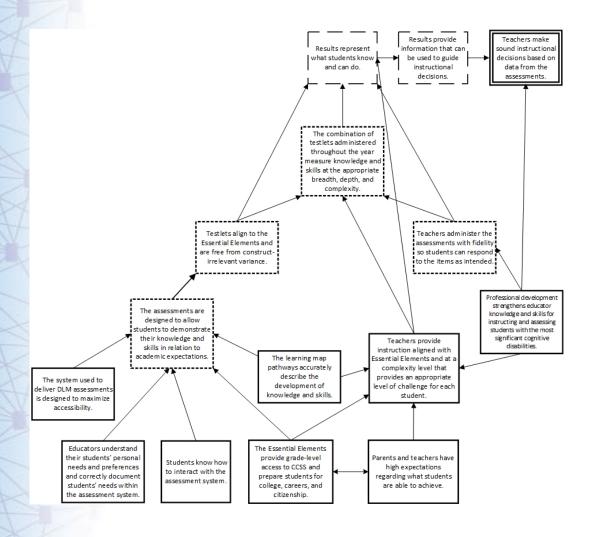


### Results indicate what students know and can do BY the end of the year





### A Good Theory of Action is Essential



#### Some extra claims:

- Teachers provide instruction aligned to the standards and that provides sufficient challenge
- Professional development strengthens teachers' knowledge and skills for instruction
- Teachers administer assessments with fidelity
- Results provide information that can be used to guide instruction
- Teachers make sound instructional decisions based on assessment data





### Approach (Will Lorié)

1. Where else is there a problem like this?

#### A problem where:

- There must be a summative determination, which
  - Depends at least in part on assessments before the end of the year
- These assessments have an additional purpose
- 2. How is the problem addressed in that context?
- 3. Can a TY solution be modeled on that example?



#### Summative classroom assessment

The appeal of through-year over end-of-year is based on an implicit analogy to summative classroom assessment.

- 1. Students are responsible for achieving learning goals, including intermediate goals
- 2. Students and teachers produce information w/r/t achievement of these goals
- 3. By processing this information, participants can bring about better outcomes on subsequent goals



#### How is the problem addressed in that context?

**Summative Classroom Assessment** Through Year Assessment

Intermediate goals Content division / scaffolding

Aggregation Aggregation





#### Can a Through-Year solution be modeled on Summative Classroom Assessment?

**Summative Classroom Assessment** 

**Through Year Assessment** 

**Curriculum ->** 

Intermediate goals

**Content division policy ->** 

Content division / scaffolding

Aggregation

Aggregation



#### Can a TY solution be modeled on SCA?

**Summative Classroom Assessment** 

**Through Year Assessment** 

**Curriculum ->** 

Intermediate goals

**Content division policy ->** 

Content division / scaffolding

**Grading policy ->** 

Aggregation

Aggregation policy ->

Aggregation



#### Can a TY solution be modeled on SCA?

**Summative Classroom Assessment** 

**Through Year Assessment** 

Curriculum ->

Intermediate goals

Content division policy ->

Content division / scaffolding

**Grading policy ->** 

Aggregation

Aggregation policy ->

Aggregation

Is there a content division and aggregation policy that can apply across variations in curriculum and instruction?





### **Proposal: Content Division Policy**

Two-module program: mid-year and end-of-year tests

- District or school selects any standards totaling 50% of the test domain.
- Mid-year is a domain sampling from these standards.
- End-of-year is a domain sampling from all standards.



### **Proposal: Aggregation Policy**

Two-module program: mid-year and end-of-year tests

- For any standard that was assessed on the mid-year and the end-of-year, take the higher of the two standard scores.
- The aggregate score is the blueprint-weighted sum of the standard scores.





### Why thread the needle this way?

- Content division policy follows <u>principle of local control</u>: Your curriculum, your choice.
  - 50% is an example. 0% is too little. 100% is too much.
- Aggregation policy incentivizes effort on each test.
- There should be no systematic relation between aggregate scores and content division choices; scores are comparable for typical accountability applications.
- Instructional utility of mid-year TY ~ instructional utility of CSA.



Ye Tong Vice-President, Psychometric Services Pearson Assessments

# Through Year Assessment Can Work, But Not Without Twists!

- 2 + 1 Through Year Strategy
- Alignment, scope and sequence
- Monitor within-grade and cross-grade growth
- Coherent and balanced assessment system (formative and instructional resources)



### **Brian Gong: Three Recommendations**

Design challenge: Through-year assessment design that meets high stakes accountability use

- Recommendation 1: Develop effective and valued separate balanced assessment elements first
  - Develop valued summative assessment uses, separate from accountability; make more valuable by incorporating contextual information; "right-size" administration burdens to benefits
  - Develop powerful instructional assessments and associated uses that demonstrably help increase student learning and school capacity and functioning (equitably, at scale)



#### **Brian Gong - 2**

- Recommendation 2: Address fundamental tension that external assessment and accountability are motivated in worst case by lack of trust in local assessment (even in state assessment without being monitored and approved by federal USED)—makes governance ineffectually distal (e.g., fed/state to school) and often distorts
  - Need to change the trust paradigm and/or what trustworthy assessment consists of, or will likely require high standardization criteria in through-year, which are fundamentally incompatible with many instructional assessment use cases
    - Possible design: As local entities demonstrate their results are "trustworthy" and/or "praiseworthy," then relax state requirements (and federal, e.g., Section 8401 could be better than through IADA's Section 1204)
    - Possible design: Develop a through-year assessment system at the district, school, or classroom level first, without state or federal constraints to address trust and move governance for instructional closer to the classroom



### **Brian Gong - 3**

- Recommendation 3: Continue to develop ways to address the most difficult conceptual and measurement challenges of through-year assessments, including:
  - Understand what claims reflect which values and views of learning and performance over time, especially the desired theories of action, such as a student not doing well and then, after instruction informed by assessment, doing better along a learning progression
    - Extend it to greater degrees of **flexibility**, e.g., different times, orders, evidence, learning targets for different students [extreme Rtl or competency-based learning would be test cases]
  - Develop and evaluate different methods to **aggregate evidence** with sound measurement properties, that match claims, theories of action, and values
  - Show what can be done with/without close ties to a specified curriculum, and to specified instruction



### **Nathan Dadey**



#### 1. Iterative

→Start small & be willing to walk away

#### 2. Realistic

- →Draw on past experiences and knowledge limitations
- 3. Evidence Based
  - →Gather evidence, especially disconfirming evidence
- 4. Evidence Based
  - →Success is in the details



#### **Additional Panelist Questions-1**

- 1. Much of the through-year push is coming from district leaders. What if we find that even if we give them this information, teachers don't know how to use the information to change instruction, or they don't know enough about instruction to change it productively? If so, who's responsibility is that?
- 2. What are any additional assessment literacy needs associated with through year designs compared to the normal challenges with increasing assessment literacy among key stakeholders?
- 3. If the through-year components test specific chunks of content, doesn't that have implications for local control of curriculum?



#### **Additional Panelist Questions-2**

- 4. What is the best way to advise and/or support states and other entities interested in pursuing a through year design? Can we make this work? Do we have to temper our claims?
- 5. As you all know, our field is thankfully wrestling with how to make more equitable assessments. Do you think there are any equity implications specific to a through-year designs?
- 6. If not through year designs, what are some other possible designs that might be able to meet some of the stated goals discussed in earlier sessions?
- 7. What would you do differently if we could loosen the reins on ESSA accountability?





### Lightening Round: One Minute Each



### **Closing Comments**

- All materials will be posted <u>here</u> shortly.
- Thank you to the organizers: Nathan, Will, and Brian!
- Thanks to our terrific collaborators: Chanda Johnson, Jeremy Heneger, Laine Bradshaw, Courtney Bell, Leslie Nabors-Olah, Karen Barton, Garron Gianopulos, Ye Tong, Meagan Karvonen, and Allison Timberlake—several of whom did double-duty
- We came together as a community to try to understand the opportunities and challenges associated with through-year assessment designs.
- We hope we've advanced your thinking—we know it advanced ours!
- We will continue share our thinking through blogs and papers and encourage you all to do the same!