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

The National Center for the Improvement of Educational Assessment
Zoom protocols, yeah, we know you know

• Please ensure your name is visible as a participant
• Please introduce yourselves via chat
• We will be monitoring the chat and Q & A for questions and comments throughout the webinar

...(not quite old hat)
The Convening

• There is a rapidly accelerating interest in through-year assessments, even though such designs have been allowed at least since the passage of ESSA.

• We are interested in the technical, policy, and practical implications of these approaches.

• Hence, this convening...
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
What will it take to make through-year assessment systems work to support students and educators?
The Organizers

Will Lorié  Nathan Dadey  Brian Gong  Scott Marion
States Developing or Considering a Through-Year Model

• Alaska
• Florida
• Georgia
• Kansas
• Louisiana
• Maine
• Montana
• Nebraska
• North Carolina
• Texas
• Virginia

Plus organizations like DLM
Strong Assumptions Require Strong Evidence

• Our goals
  ▪ Develop some common language
  ▪ Describe currently developing designs
  ▪ Outline evidence necessary to support key claims
  ▪ Provide insight on critical technical issues associated with through-year designs
  ▪ Begin to outline a research and practice agenda

• We are not advocating or criticizing any particular through-year program
We Need to Stay Humble

• This convening is drawing together researchers and practitioners in the spirit of collaboration

• We have done our best to situate the presentations in extant research and emerging design, but we have likely missed key pieces of work

• We hope that through collaboration, we all emerge with a better sense of both possibilities and pitfalls
Session 1: Definition, Aims, and Use Cases
Through-Year Assessment Virtual Convening, November 15, 2021

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

1. Definitions and Aims
   Defining through-year assessment programs and the motivations for these programs

2. From Aims to Design
   Moving from aims to program theory and assessment design

3. Invited Presentations
   Presentations from invited participants on emerging through-year programs

4. Question and Answer
   Facilitated Audience Interaction
Take Away Points

1. Be specific about why a through-year program is being pursued. A through-year program only serve a limited number of purposes, or aims, well.

2. Be specific on what a through-year program is meant to accomplish and how it will be accomplished.

3. The design of the through-year program should be informed by, and inform, the theory of action.
1. Definitions & Aims

Defining through-year assessment and the motivations behind through year programs.
One Definition

Through-course summative assessment means an assessment system component or set of assessment system components that is administered periodically during the academic year. A student’s results from through-course summative assessments must be combined to produce the student’s total summative assessment score for that academic year.

An Expanded Definition

The defining characteristics of a through-year\(^1\) assessment program are that it is:

- Administered through **multiple distinct administrations** across a school year, and
- It is meant to support both (a) the production and use of a **summative determination**, and (b) **one additional aim**.

We introduce “aims” to capture a core motivation behind through-year designs: to accomplish “something else” while also creating summative determination.

---

\(^1\)We suggest that the term “through-year” assessment be used as a general term to describe these models, regardless whether the time period is a “year,” “semester,” or “course.”
Through-year assessment system has multiple distinct administrations meant to support both (a) the production and use of a summative determination, and (b) one additional aim.

• A through-year assessment may be a type of “balanced assessment system” since through-year assessments often incorporate interim and summative assessments.

Balanced assessment system has multiple types of assessments (e.g., summative, interim, formative) that serve distinct purposes – the interim and formative are not necessarily combined to yield a summative determination.

• A balanced assessment system may incorporate much more than just interim and summative assessments.
On “Aims”

The reasons why stakeholders want a through-year program are likely diverse, complex and often unstated or understated.

These reasons can be grouped into three categories of “aims”:

- Logistical
- Monitoring
- Instructional

Likely there are multiple aims, however, some will take precedence in design. Likely, one aim, or a very small set will be served well, and conflation of aims by various stakeholders will invariably lead to problems.
Logistical aims are typically oriented on the “footprint” statewide assessment, e.g.:

• Reduce overall assessment footprint (e.g., replace separate district/interim and state assessments with one system)
• Better fit into school calendars
• Reduce test anxiety (perhaps by providing multiple opportunities to “pass”)

Potential Unintended Consequences
• District assessments retained, leading to a larger footprint
• Each administration leads to disruption to school calendars
• Each administration causes anxiety
Logistical

Example: Reduced Testing Time

• Many through-year assessment designs will not result in less testing time than just a single summative assessment
  ▪ If content coverage and reliability are maintained

• Some hope for a through-year assessment design where multiple interim assessments replace the end of year summative assessment
Administrators are looking to understand where attention is most needed, e.g., grades, classrooms, schools, often to:

→ Determine where additional investigation is needed as part of a process of continuous improvement, which may involve:
  - Site visits
  - Direct coaching
  - Development of improvement plans

Potential Unintended Consequences or Limitations
- Over reliance on through-year results, leading to inappropriate conclusions
- Data used punitively
• How a through-year assessment program’s design and results support the instructional practice of educators.

• “Informing instruction” can include many possible actions\(^1\), e.g.,:
  - Modifications to whole class instruction
  - Working with students in small groups
  - Providing individualized support

\(^1\)In the context of interim assessment, see Dadey & Diggs (2019) and Abrams, McMillan, & Wetzel (2015) for possible categories of instructional uses.
• A natural inclination is to seek instructional relevance whenever a score report is provided.

• Through-year programs have more administrations and thus reports (even if they follow same template).

• Bottom line: instructional uses should be addressed through the theory of action, and subsequent design, of a through-year program.
  - Even if the intended aims are not instructional, the program must find ways to properly direct use of the assessment results, e.g., anticipate and mitigate unintended negative consequences.
"... a conceptualization of formative assessment as finer-grained or more frequent evaluation subverts its significance by occluding the meaning, nature, and the promise of formative assessment practice." (Heritage, 2010, p. 12).
In some cases, already existing assessments will be used to create a through-year assessment program.

Careful consideration of aims is still needed and should include taking both current and proposed uses of assessments.
Be specific about why a through-year program is being pursued. A through-year program only serve a limited number of purposes, or aims, well.

Be specific on what a through-year program is meant to accomplish and how it will be accomplished.

The design of the through-year program should be informed by, and inform, the theory of action.
2. From Aims to Design

Moving from aims to program theory and assessment design
2.A Theory of Action

Using theory of action to connect aims to assessment design
Why?
What *motivations* underlie the shift to a through-year assessment program?

How?
What is the through-year *program*? What are the intended outcomes and *how will they be achieved*?

Theory of Action
This framing is built on the distinctions proposed by Bennett et al. 2011, who characterize the same ideas in terms of a theory of action and measurement argument. These ideas also echo throughout literature on validity theory, e.g., Kane (2006, p. 53) speaks of a semantic interpretation and a decision inference. Cronbach and Messick also discuss these ideas in detail.
How? What is the through-year program? What are the intended outcomes and how will they be achieved?

Theory of Action

This framing helps us define:

- What the effects, or intended results, are.
- What actions are meant to lead to the intended effects.
- What inputs are needed to prompt and support the actions.

For each and every aim.

There are a number of ways to visualize and model theory of action, including Logic Models (e.g., Frechtling, 2007, W.K. Kellog Foundation, 1998) or Driver Diagrams (e.g., Bennett & Provost, 2015). A more complex logic model framework might include outputs, short- & long-term outcomes and impact.
How? What is the through-year program? What are the intended outcomes and how will they be achieved?

Theory of Action

Which helps us define and understand:

- **What** do we need to know about what students know and can do?
- **When** do we need to know it?, and
- What are students, teachers, leaders, and others supposed to do about it?

Interpretive Argument
Within a through-year model, assessments are designed to support multiple claims about what students know and can do.

- **Summative Claims**
  - A claim meant to support use in a state’s system of school identification and support (i.e., an annual determination), based on the collection of evidence from across the assessments
  - “The student is proficient in the state’s mathematics content and skills standards at the end of the year”

- **Additional Claims – E.g., Instructional or Administrative**
  - “The student does not yet understand place value in early Fall”

**What?**

What is the set of **assessments** and corresponding assessment evidence?

**Interpretive Argument**

Supported **claims** (at multiple points in time) & corresponding design
Example: Theory of Action Relating Instruction to Assessment Design

**Inputs**

- Assessment information (claim): “Student’s score indicates Student knows X at level Y”
- Teacher Interpretive Judgment: “Student needs help on these aspects of X”
- Teacher Instructional Repertoire: “I could do A, B, or C to help Student learn the identified aspects of X, but should not do D”

**Actions**

- Teacher does B instructional/learning actions with Student

**Effects**

- Student learning improves on targeted aspects of X

Teacher Contextual Knowledge:
- Overall goals, how student learned in past, upcoming curriculum, other students’ needs, etc.

### Example: Theory of Action Relating Instruction to Assessment Design

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Actions</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment information (claim): “Student’s score indicates Student knows X at level Y”</td>
<td>Teacher does B instructional/learning actions with Student</td>
<td>Student learning improves on targeted aspects of X</td>
</tr>
<tr>
<td>Teacher Interpretive Judgment: “Student needs help on these aspects of X”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher Instructional Repertoire: “I could do A, B, or C to help Student learn the identified aspects of X, but should not do D”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Teacher Contextual Knowledge:** Overall goals, how student learned in past, upcoming curriculum, other students’ needs, etc.

**Assessment design:** The assessment must produce this specific type of assessment information/claim at the time when needed to support this specific type of instruction (action/use).
Developing a **theory of action** is a complex and messy endeavor, and there are no shortcuts, only tradeoffs.

Developing **claims** are supportive of, and supported by, a theory of action is also a complex and messy endeavor, and there are no shortcuts, only tradeoffs.
Be specific about why a through-year program is being pursued. A through-year program only serve a limited number of purposes, or aims, well.

Be specific on what a through-year program is meant to accomplish and how it will accomplish it.

The design of the through-year program should be informed by, and inform, the theory of action.
2.B Assessment Design

Considering a key design feature – content & administration
Some Key Design Features

1. Content & Administration
2. Connections to Curriculum & Learning Theory
3. Standardization and Security
4. Reporting & Reporting Metrics
5. Aggregation Methods

See also Dadey & Gong, 2017; Gianopulos, 2019 for additional framing around the design of through-year assessment programs.
Content & Administration

- There are numerous ways content might be organized across multiple assessments
  - The key driver is the intended uses
  - Restated, content organization is solely determined by additional uses and claims

- Claims and evidence are deeply tied to the timing of the content assessed
Across assessments, what is the:

- **Number and timing** of assessments
- **Content distribution** across assessments
- **Grain-size** of the content divisions
- **Articulation** between the divisions
- **Flexibility** of the administration of the content divisions

**Administration**

- **Windows**
  - Defined vs. Open
- **Order**
  - Fixed vs. Flexible
- **Control**
  - Who decides what is administered and when
A Hopefully Helpful Heuristic

Typical End of Year Summative Testing

Module

Content Domain

Fall Spring

Administration Window
So now let’s consider some possible variations on content and administration that mirror emerging designs.
“Modular Mini-Summative” Design

Each module (a) covers the entire content domain and (b) is identical in terms of content coverage.
An alternative design is divide up the content domain.

The design question then becomes how to do so.
"Modular Standards Domain" Design

Each module covers a unique group of standards defined by a topic domain.

- **Fall**
  - Number and Operations in Base Ten
  - Operations and Algebraic Thinking
  - Number and Operations — Fractions

- **Spring**
  - Measurement Data
  - Geometry
"Modular Standards" Design
Each module covers an individual standard

- Understand the place value system
- Perform operations with multi-digit whole numbers ...
- Gain familiarity with factors and multiples
- Analyze patterns and relationships

Content Domain

- Fall
- Spring
“Modular Standards” Design

Each module covers an individual standard

- Understand the place value system
- Perform operations with multi-digit whole numbers...
- Gain familiarity with factors and multiples
- Analyze patterns and relationships
- ...

Content Domain

Fall  |  Spring
Some Tensions

• Learning does not occur in discrete chunks that align to standards, rather the knowledge, skills and abilities that underlie a progression of learning map across multiple standards across multiple domains

• Standards are not the only way to structure the content domain across time, e.g.,:
  - Complexity or sophistication, perhaps along a learning progression
  - Degree of scaffolding of content
  - Genre of text (ELA), bundle of performance expectations (Science)
Connections to Other Design Features

• The way in which content and administration is structured alone does not define a through-year assessment program

• A specific design for content and administration doesn’t dictate decisions about other design features
  ▪ E.g., various approaches to summative score creation can be paired with a given content and administration design
  ▪ However, some decisions about content and administration tend to pair better with specific decisions about other features (e.g., the mini-summative design and reporting “growth”)

Center for Assessment. Through-Year Convening. Session 1. November 15, 2021
The design of content and administration is essential in supporting and informing claims, both instructional and summative.

The challenge of through-year program is how to make those work well together.
1. Be specific about why a through-year program is being pursued. A through-year program only serve a limited number of purposes, or aims, well.

2. Be specific on what a through-year program is meant to accomplish and how it will accomplish it.

3. The design of the through-year program should be informed by, and inform, the theory of action.
In Summary
Take Away Points

1. Be specific about why a through-year program is being pursued. A through-year program only serves a limited number of purposes, or aims, well.

2. Be specific on what a through-year program is meant to accomplish and how it will be accomplished.

3. The design of the through-year program should be informed by, and inform, the theory of action.
Considering “Success”

• The usefulness of a through-year assessment program lies in:
  ▪ How well specified the theory of action is, and how well the theory of action works in practice
  ▪ How well the design serves the theory of action

• The implementation of a through-year model intentionally introduces an additional use, or uses, onto the statewide summative assessment
  ▪ Doing so means that the success of the through-year model is much more contextually based than state summative assessment programs
3. Deep Dive with Invited Participants
Chanda Johnson
Deputy Assistant Superintendent of Academic Content, Office of Teaching and Learning
Louisiana Department of Education

Jeremy Heneger
Director of Statewide Assessment
Nebraska Department of Education

Laine P. Bradshaw
Founder and CEO
Navvy Education, LLC
Assistant Professor
Educational Psychology Department
University of Georgia
Participant Presentation and Discussion

• What are the key problems and issues you are trying to solve?
• How does your theory of action help solve these problems or issues?
• What assessment information does your through-year design provide, and how is that information used within that theory of action?
4. Q&A
Supplemental Slides
An additional in-depth consideration of content
An Alternative Look at Content
Content for instructional assessment purposes

• There are a near infinite number of ways content might be organized to serve instructional assessment purposes
  ▪ Many instructional purposes
  ▪ Several ways to approach each instructional purpose
    • Curriculum
    • Instruction
    • Assessment

• Claims and evidence are deeply tied to content assessed
  ▪ And the timing is especially important for summative claims
Examples of content selection and organization

• When to assess?

What is the instructional use for an assessment at this point in time?
What is the claim that needs to be supported at this point in time?
What evidence is need to support the claim at this point in time?
What differences (in use, claim, evidence) are there for different points in time, if any?
Examples of content selection and organization

• When to assess; how many times during the year?

What is the instructional use for an assessment at this point in time?
What is the claim that needs to be supported at this point in time?
What evidence is needed to support the claim at this point in time?
What differences (in use, claim, evidence) are there for different points in time, if any?
  • Why would you assess three times before the end of the year, rather than two times? Four times?
  What would you do differently?
Examples of content selection and organization

• What content assessed, when; relation to instructional order?

Assess only the content that has most recently been taught?
Examples of content selection and organization

• What content assessed, when; relation to instructional order?

Assess every time all the content that is included in the summative claim?
Examples of content selection and organization

• What content assessed, when; relation to instructional order/time?

What if specific content/skills are taught across the year, or come back at different times or in different combinations?
Examples of content selection and organization

- What content assessed, when; relation to instructional order; what grain-size?
Examples of content selection and organization

- What content assessed, when; relation to instructional order?

What about rework or retesting to demonstrate learning within the year prior to the end of year?
What content should be retested, if any?
Examples of content selection and organization

• What content assessed, when; does order matter?

How flexible/standardized can/must the assessments be: different content (individual teachers choose individual content to match their curricula; different times; different orders; different for students within classrooms, different for individual students, etc.?)

Who decides?
Several other important content aspects

• Amount of scaffolding
• Degree of sophistication, “deeper learning”
• Generalization, application
• Independence
• Facility, automaticity
Alternative Content Designs
“Modular Mini-Summative +” Design

Each module (a) covers the entire content domain and (b) is identical in terms of content coverage, with the exception that each module contains additional items on a specific chunk of the domain.

Gianopulos, 2019 proposes this design.
“Modular Overlapping Domain” Design

Each module covers a partially overlapping group of standards defined by a topic domain.