

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.



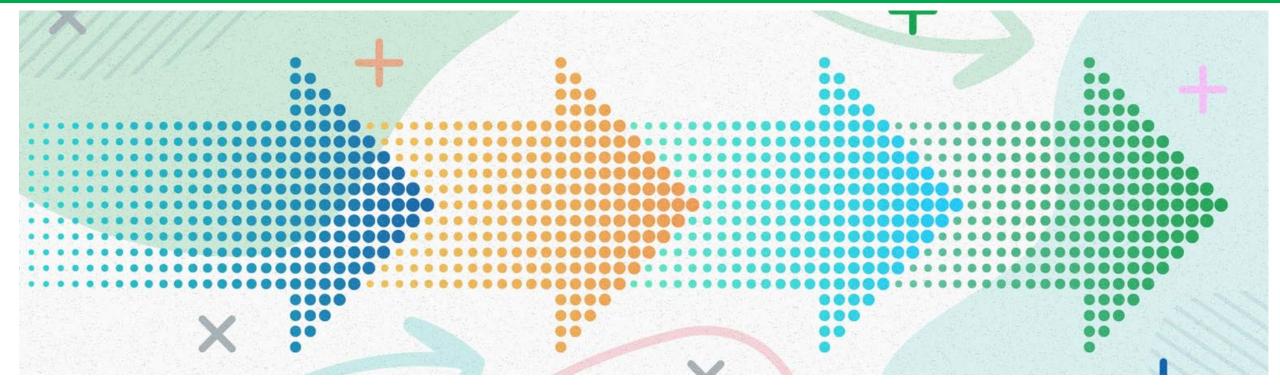
Tuesday, November 16, 1:00-2:30 PM ET

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



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

What will it take to make through-year assessment systems work to support students and educators?



Session 2: Claims, Designs, and Evidence

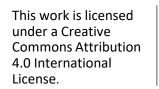
Through-Year Virtual Convening, November 15, 2021

Brian Gong, Nathan Dadey, Will Lorie & Scott Marion

The National Center for the Improvement of Educational Assessment









Invited Panelists







Leslie Nabors-Olah ETS

Courtney Bell WCER, UW

Karen Barton NWEA

1 Framing

Considering the logic and evidence of our theories of action and interpretive arguments

OutputUse Cases & Claims

- "Additional" use cases & claims
- Summative use cases & claims

Panelists Remarks
In depth considerations from invited participants

4 Question and Answer Session Facilitated Audience Interaction



1. Framing





A Definition of Through-Year Assessments

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
 - the production and use of a summative determination, and
 - at least one additional aim: instructional, monitoring, logistical

¹We suggest 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."



Location of Through-Year Assessment Challenges (ToA, IUA, Design, Evidence)

	Intended Use	
Type of Assessment	Summative	Instructional
State Summative	Widely worked on and researched; much technical support	
Interim		



Location of Through-Year Assessment Challenges (ToA, IUA, Design, Evidence)

	Intended Use	
Type of Assessment	Summative	Instructional
State Summative	Widely worked on and researched; much technical support	
Interim		Lots of examples of designs and uses; little formal ToA, IUA; less technical documentation



Location of Through-Year Assessment Challenges (ToA, IUA, Design, Evidence)

	Intended Use		
Type of Assessment	Summative	Instructional	
State Summative	Widely worked on and researched; much technical support	Very limited solutions so far (due to timing; limited subscores due to unidimensionality)	
Interim		Lots of examples of designs and uses; little formal ToA, IUA; less technical documentation	



Location of Through-Year Assessment Challenges (ToA, IUA, Design, Evidence) **Intended Use** Type of Instructional **Summative Assessment** Very limited solutions so far (due Widely worked on and State to timing; limited subscores due researched; much technical **Summative** to unidimensionality) support Most through-year assessment Lots of examples of designs and focus: uses; little formal ToA, IUA; less Interim How can traditional summative technical documentation

criteria be met? or

Can new criteria be developed?

Why?

What motivations underlie the shift to a through year assessment program?

Aims



How?

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

Theory of Action



What?

What is the set of assessments and corresponding assessment evidence?

Interpretive Argument

What evidence is needed?



Through-year programs are acting as interventions and should be evaluated as such.



Theory of Action

Interpretive Argument

Intended Design

- Inputs
- Actions
- Outcomes

- Claims
- Design Features

Supporting Evidence

- Are the actions implemented with fidelity¹?
- Do they lead to the intended outcomes¹?

 Are the claims about what students know and can do supported?

Building Out Connections

Theory of Action

- Are the actions implemented with fidelity?
- Do they lead to the intended outcomes?
- Is the assessment information provided sufficient for the given actions?
- Do the actions use the assessment information appropriately?

Interpretive Argument

- Are the claims
 about what
- students know and can do supported?



Use Cases & Claims

Theory of Action

Interpretive Argument

Use Case

Claim

School Identification & Support

Summative Claim

Instructional or Monitoring Use

Additional Claim(s)



Be as explicit as possible in the statement of the theory of action, especially in the statement of action mechanisms and intended effects, so as to allow for meaningful evaluation.

Bennett, Kane & Bridgeman, 2011, p. 46





2. Uses & Claims





2.A Instructional Claims & Uses



Use Cases & Claims

Theory of Action

Interpretive Argument

Use Case

Claim

Summative Use

Summative Claim

Instructional Use(s)

Instructional Claim(s)



Instructional Use(s)

Instructional Claim

What are the specific actions various actors are meant to take in reaction to the program and the information it produces?

In particular, how will these actions lead to better classroom teaching and learning?



Instructional Use(s)

Instructional Claim

- "Informing instruction" can include many possible actions, e.g.,: Abrams, McMillan and Wetzel (2015):
 - Modifications to whole class instruction
 - Working with students in small groups
 - Providing individualized support

- Even actions at this level of specificity are not detailed enough to support instruction.
- Ideally, these actions are informed by research on score use.

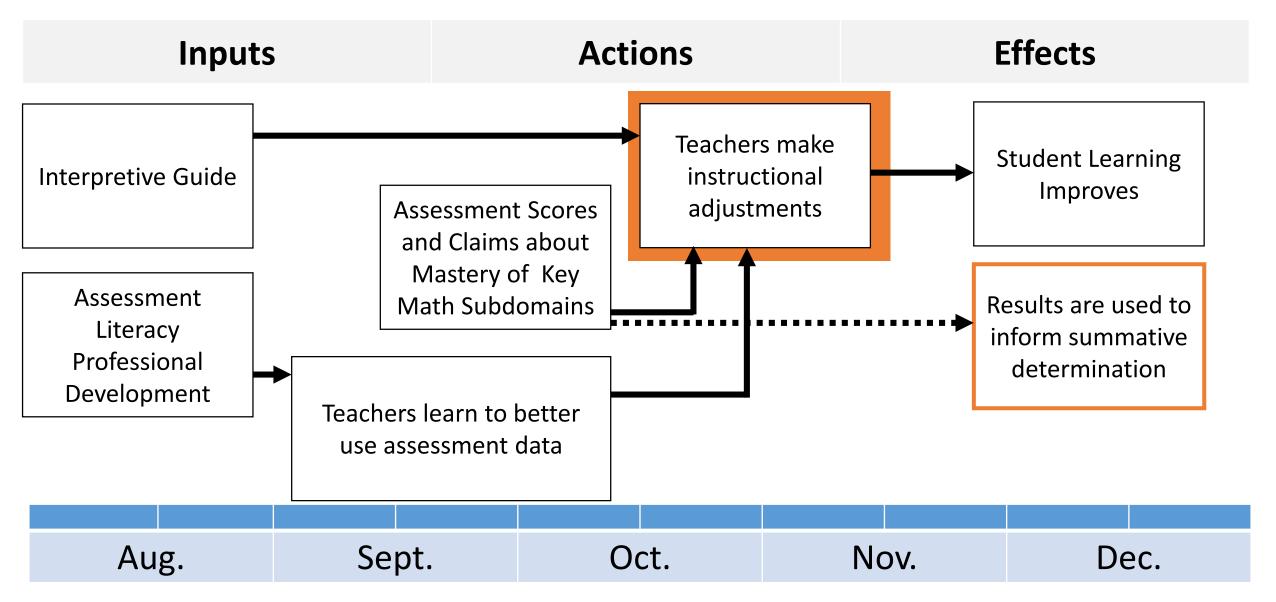
Instructional Use(s)

Instructional Claim

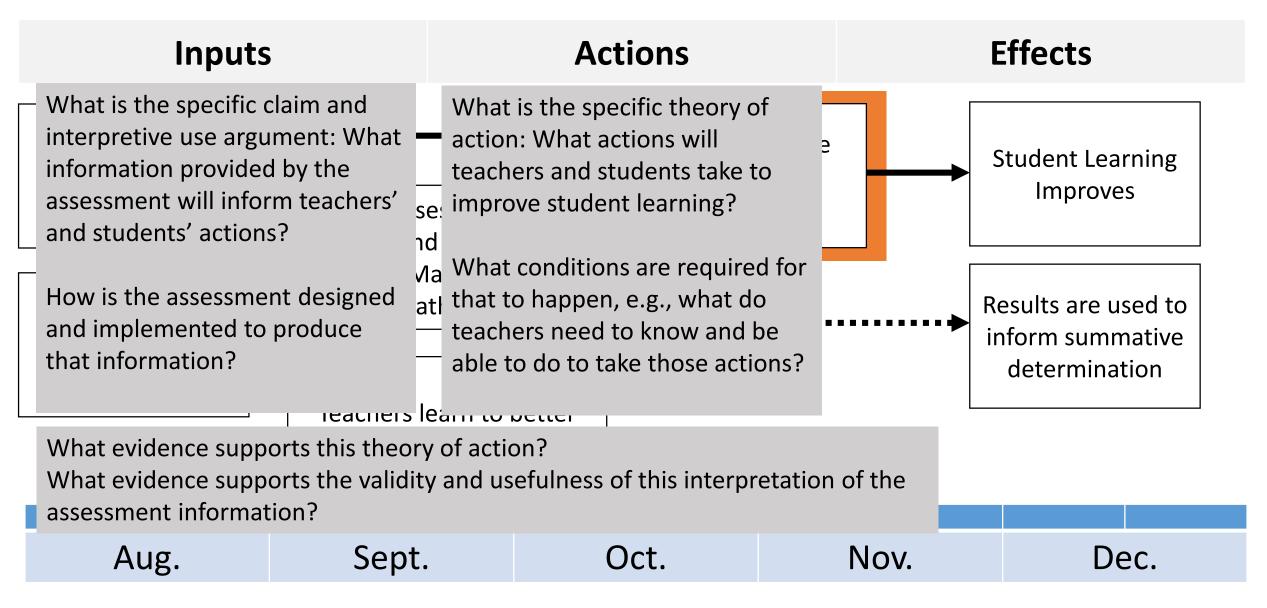
Let's consider an example theory of action for instructional use for a through-year assessment program (i.e., knowing the results also are intended to inform a summative determination) based on administration of an assessment in the fall semester.



Example High Level Theory of Action



Example High Level Theory of Action



Teachers make instructional adjustments

Supporting Detail

Teachers access reports in a timely fashion

the score reports
(mastery of math
subdomains)

Teachers triangulate assessment results with classroom assessments

Teachers implement supports at the individual and small group level

Teachers determine what kinds of instructional supports are needed

Student
misconceptions on
subdomain concepts
are addressed



Teachers determine what kinds of instructional supports are needed

Specific Assumptions

- Teachers have access to, and understand, a variety of instructional supports connected to the math subdomains
- Teachers can use assessment evidence to select amongst the various kinds of support
- The assessment evidence is needed to select amongst these kinds of support

Assessment provides needed information

Supporting Detail

Assessment Scores and Claims about Mastery of Key Math Subdomains

so that Student misconceptions on subdomain concepts are addressed Math subdomains are important for understanding current content and for learning subsequent content

claims of "mastery"
incorporate
appropriate depth,
breadth, cognitive
complexity,
generalizability, etc.

Claims of "mastery" lead to appropriate actions (e.g., go on)

Claims about "Non-Mastery" help teachers identify student misconceptions on subdomain concepts Scores are reliable and precise enough to support claims and actions

Scores are reported in way and time that support intended interpretation and use

Student results can be used in summative determination



Typical Constraints Posed by Summative Use

- Focuses on state content standards only
- Assessment instruments, administration conditions, times during the year, etc. are highly standardized; carefully designed accommodations and modifications provided to promote comparability and fairness
- Carefully controlled item presentation, mode, and test design to allow equating for comparable scores
- Scoring quality controlled for high accuracy and consistency to officially determined criteria
- Security: item exposure controlled; performance is attributable only to student without input from others or unfair advantage
- Underlying construct and scale usually fit with a unidimensional measurement model established at a certain point in time





2.B Summative Claim & Use



Use Cases & Claims

Theory of Action

Interpretive Argument

Use Case

Claim

Comparable Annual Determinations

Summative Claim

Instructional Use(s)

Instructional Claim(s)





Take Away Points

1

Values drive claims. Claims drive our choice of aggregation methods, not the other way around.

2

Designs that distribute content in multiple modules across the year require complex claims.

Summative Claim

To support annual determinations, we need a single summative score.

The creation of a single summative scores involves not only the application of an **aggregation method**¹,

but also consideration of values and corresponding claims.

¹Here we include both the application of a **measurement model** as well as additional post hoc steps like taking the maximum score.



School Identification & Support

Summative Claim

Restated, every aggregation method reflects specific values and supports specific claims.



School Identification & Support

Summative Claim

Value Judgement(s)

What value is placed on:

- Performance during the year?
- Performance at the end of the year?
- Changes in performance across the year?

Claims

What inference do we want to make about what students know and can do?, e.g.,:

- About "typical" student performance across the year?
- About student performance at the end of the year?

Score Creation

Implementation:

- Is the aggregation done within a measurement model, or in addition to a measurement model?
- How are the models, and thus time, addressed?



Theory on how learning occurs over time.

Summative Claim

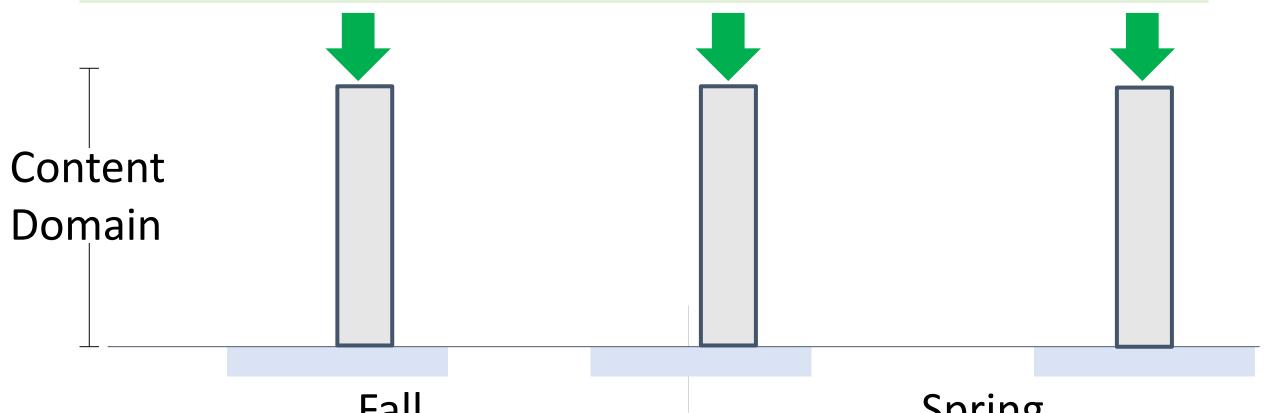
One way to understand this complexity is through comparisons to classroom grading.

E.g., is your class grade the sum of unit tests? A pass or fail based on a single end of year test?

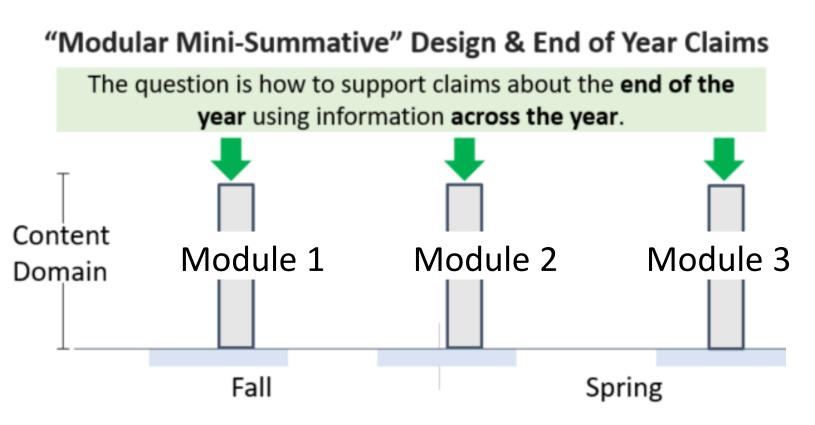
To start, let's consider claims based on the end of the year.

"Modular Mini-Summative" Design & End of Year Claims

The question is how to support claims about the end of the year using information across the year.



Methods that Support End of the Year Claims



- Use scores from modules 1 & 2 as priors for estimates of student scores on module 3
- Use an adaptive model that uses scores from modules 1 & 2 as starting points, but with full recovery

These approaches represent is our best guess about, and current understanding of, emerging designs. In a series of blogs Gong (2021a and 2021b) considers the tensions in supporting end of year claims.

Value Judgements & Claims: End of Year

These kinds of methods have been suggested, but not yet been implemented operationally.

Careful investigation of both summative score creation and learning & forgetting and opportunity to learn should accompany these approaches.



Take Away Points

1

Values drive claims. Claims drive our choice of aggregation methods, not the other way around.

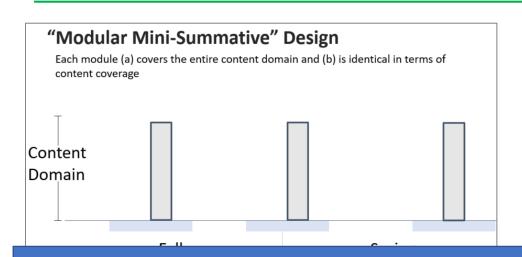
2

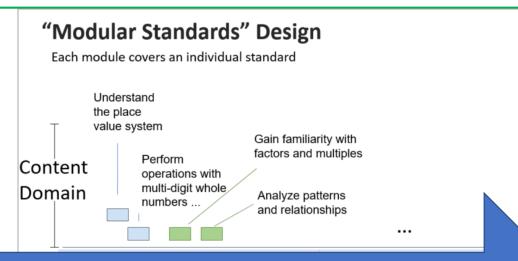
Designs that distribute content in multiple modules across the year require complex claims.

School Identification & Support

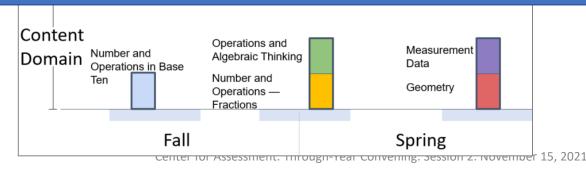
Summative Claim

Value Judgements & Claims: End of Year

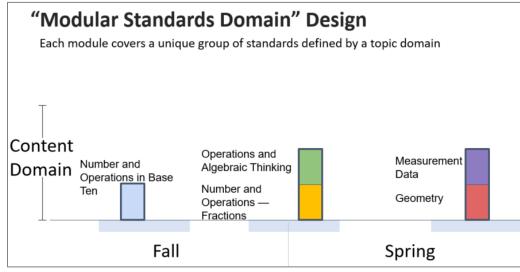


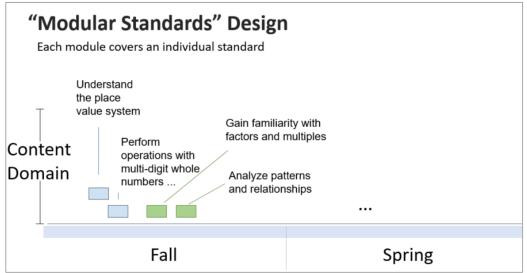


Increasing Complexity In Supporting End of Year Claims (due to the interaction of content and time)



Value Judgements & Claims: Across the Year





Here the interpretations are more complex:

- Performance on each content domain near the time of instruction?
- Number of mastered standards when "ready to test"?

This complexity arises because of the interaction of content and timing.

Value Judgements & Claims: Across Year

The interaction of content and time is a large problem of practice.

Additional work is needed to develop claims and supporting aggregation methods.



Take Away Points

1

Values drive claims. Claims drive our choice of aggregation methods, not the other way around.

2

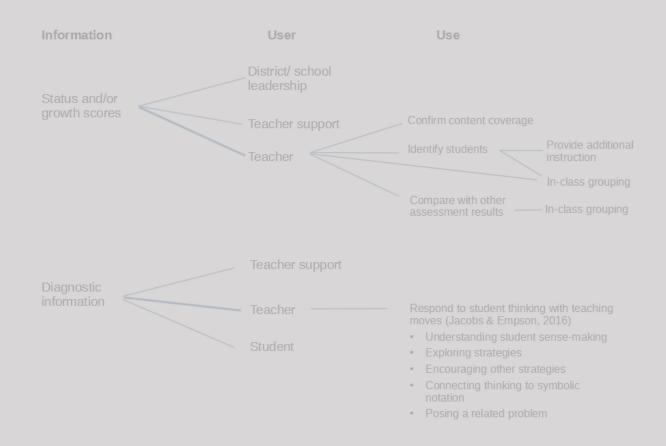
Designs that distribute content in multiple modules across the year require complex claims.



3. Participant Remarks



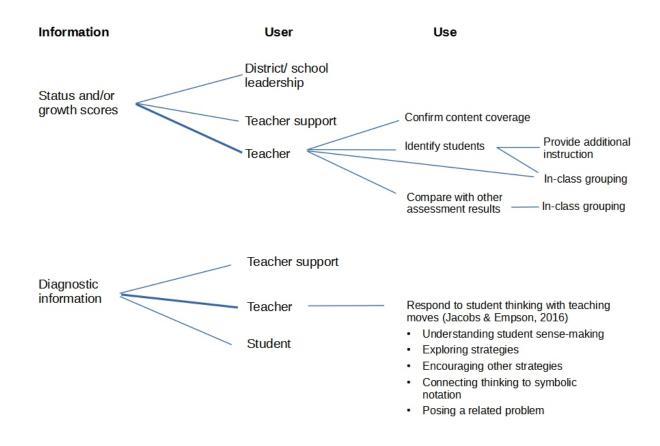
What is an "instructional use"?





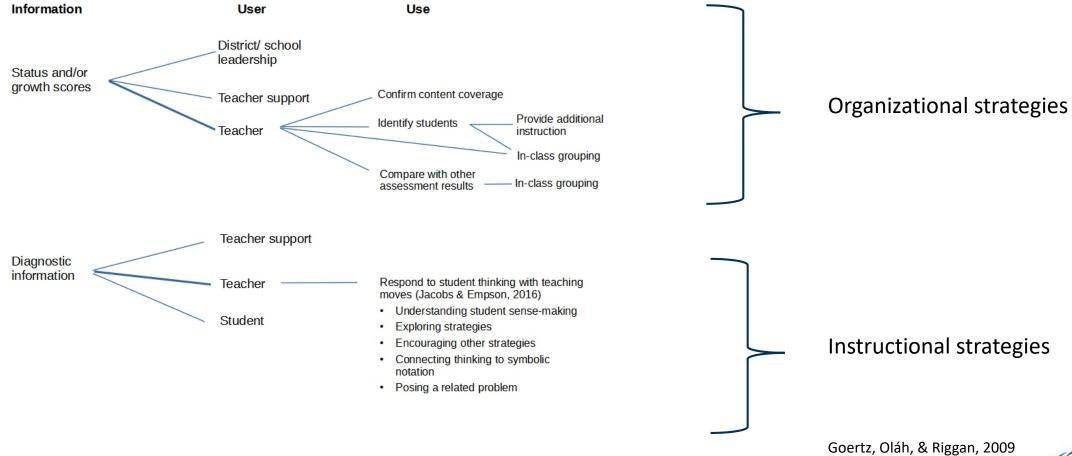
Leslie Nabors-Olah ETS

What is an "instructional use"?





What is an "instructional use"?







4. Facilitated Q&A



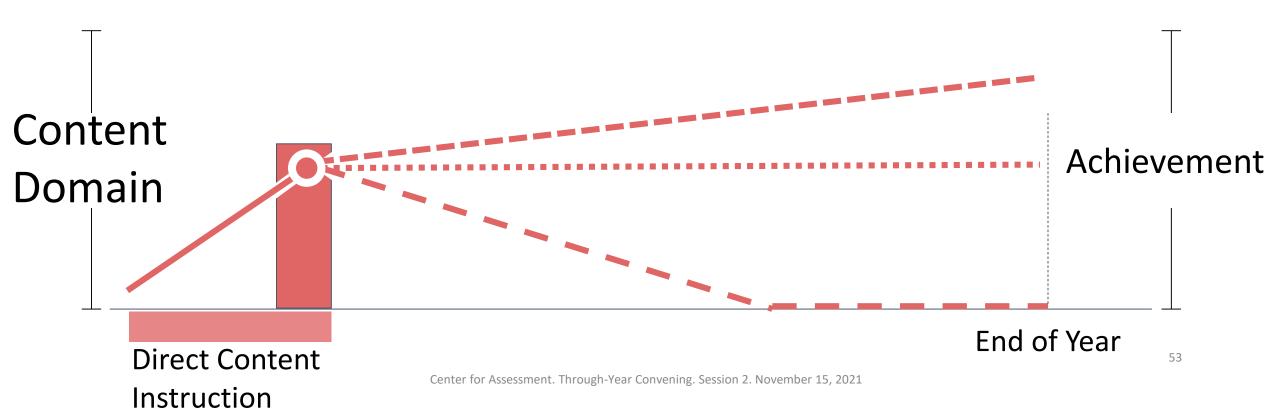


Supplemental Slides



Value Judgements & Claims: End of Year

Structuring the content & administration based on these assumptions



Five Most Frequent Instructional Uses Evident in the Research on Interim Assessments

	Use Description	k	Example
1.	Broad claim to modify or improve instruction	14	A total of 86% of teacher reported modifying instruction based on interim assessment results (Clune & White, 2008).
2.	Identify students to provide additional support	13	Results were used to identify students for supplemental instruction (e.g., software, working with volunteers, afterschool tutoring; Shepard et al [2011]).
3.	Identify what content to reteach	10	The administration hoped teachers would reteach with new strategies (Bulkley et al., 2010).
4.	Improve score on the summative assessment	10	Guide schoolwide improvement efforts to meet AYP (especially in low-performing schools; Bulkley et al., 2010).
5.	Differentiate instruction	9	Identifying students with similar patterns of performance on the assessment and using that to constructs groups to differentiate instruction (Blanc et al., 2010)

Note. Table created based on coding a total of 20 studies featuring interim assessments.



Full Consideration of Summative Score Creation



Use Cases & Claims

Theory of Action

Interpretive Argument

Use Case

Claim

School Identification & Support

Summative Claim

Instructional Use(s)

Instructional Claim(s)



To support annual determinations, we need a single summative score.

The creation of a single summative scores involves not only the application of an **aggregation method**¹,

but also consideration of values and corresponding claims.

¹Here we include both the application of a **measurement model** as well as additional post hoc steps like taking the maximum score.



School Identification & Support

Summative Claim

Restated, every aggregation method reflects specific values and supports specific claims.



School Identification & Support

Summative Claim

Value Judgement(s)

What value is placed on:

- Performance during the year?
- Performance at the end of the year?
- Changes in performance across the year?

Claims

What inference do we want to make about what students know and can do?, e.g.,:

- About "typical" student performance across the year?
- About student performance at the end of the year?

Score Creation

Implementation:

- Is the aggregation done within a measurement model, or in addition to a measurement model?
- How are the models, and thus time, addressed?



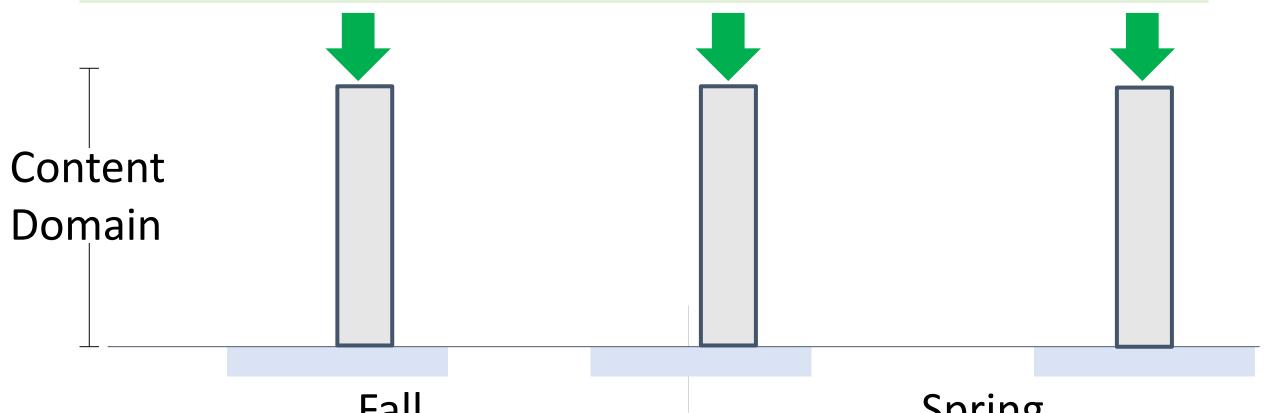
Theory on how learning occurs over time.

One way to understand this complexity is through comparisons to classroom grading.

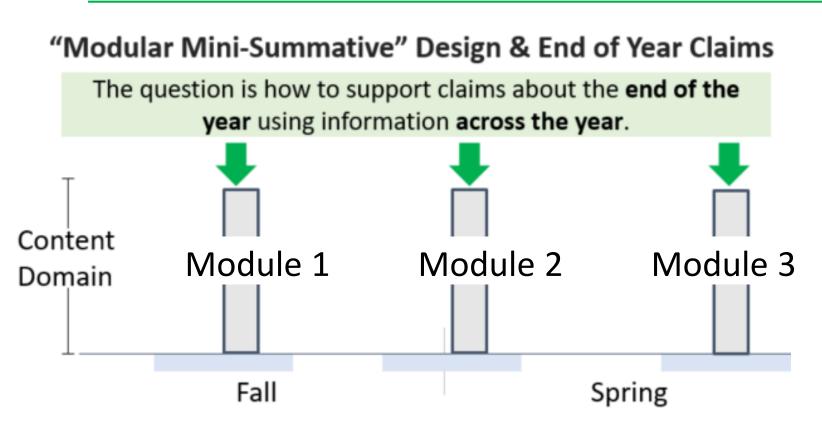
E.g., is your class grade the sum of unit tests? A pass or fail based on a single end of year test?

"Modular Mini-Summative" Design & End of Year Claims

The question is how to support claims about the end of the year using information across the year.



Methods that Support End of the Year Claims



- Use scores from modules 1 & 2 as priors for estimates of student scores
- Use an adaptive model that uses scores from modules 1 & 2 as starting points, but with full recovery

These approaches represent is our best guess about, and current understanding of, emerging designs. In a series of blogs Gong (2021a and 2021b) considers the tensions in supporting end of year claims.

Value Judgements & Claims: End of Year

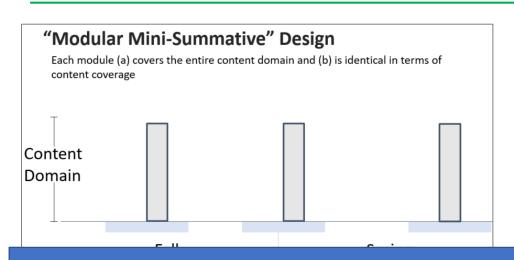
These kinds of methods have not yet been implemented operationally.

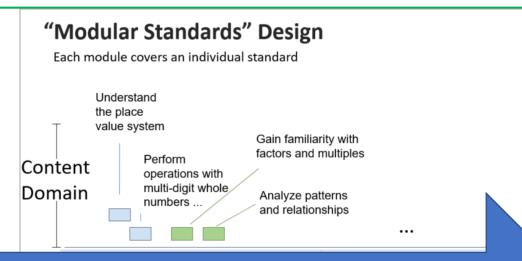
Careful investigation of both summative score creation and learning & forgetting and opportunity to learn should accompany these approaches.

School Identification & Support

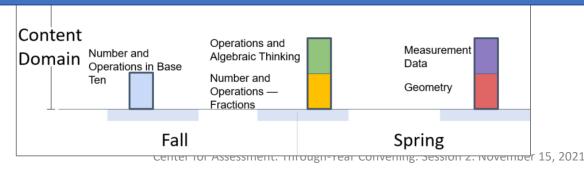
Summative Claim

Value Judgements & Claims: End of Year





Increasing Complexity In Supporting End of Year Claims (due to the interaction of content and time)



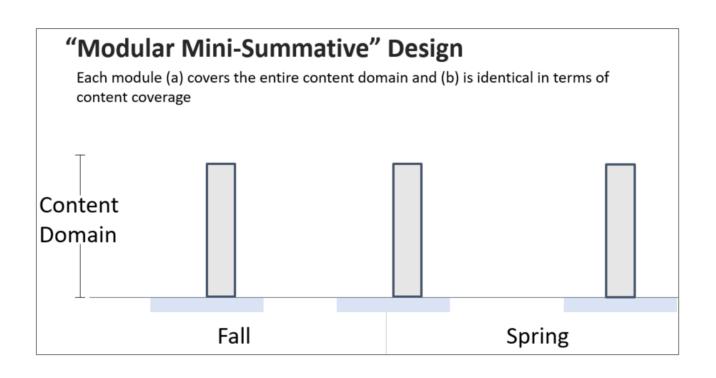
Value Judgements & Claims: Across the Year

- Alternatively, we could restructure the claims so that they are not rooted in the end of the year, e.g., a claim about:
 - "Typical" student performance across the year?
 - "Best" student performance across the year?
- Doing so means explicitly considering the value attached to each module and its associated portion of the content domain.
- Critically, states are currently constrained to report scores relative to the "full set" of grade-level standards.

School Identification & Support

Summative Claim

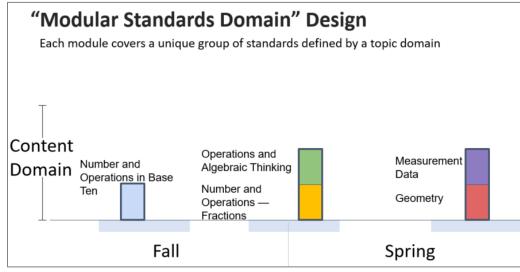
Value Judgements & Claims: Across the Year

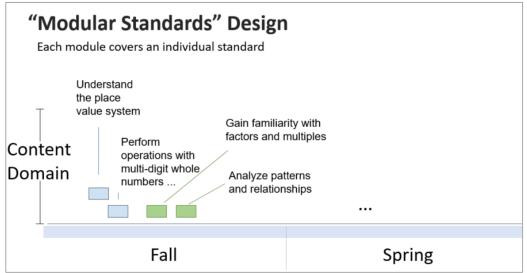


- Simple Sum
- Maximum Score (or above some cut point)
- Simple or Weighted Averages



Value Judgements & Claims: Across the Year





Here the interpretations are more complex:

- Performance on each content domain near the time of instruction?
- Number of mastered standards when "ready to test"?

This complexity arises because of the interaction of content and timing.

Value Judgements & Claims

Supporting a summative claim may follow this flow:

Making assumptions about, or empirically investigating, student learning & forgetting, opportunity to learn, or both

Structuring the content & administration based on these assumptions or findings

and, or,

Statistically adjusting for variation in student learning & forgetting and opportunity to learn

69



Open Questions for the Participants

- What evidence is needed to support the interpretive argument of the summative claim based on the through-year assessment program?
- Designing for multiple claims and uses entails various trade offs. What trade offs do we have to make to support a summative claim, and do any of those trade offs limit other uses?
 - How do we collect evidence to understand whether other claims and uses are being impacted?

