



# ***The Role (?) of Interim Assessments in Balanced Systems of Assessment***

Scott Marion

National Center for the Improvement of Educational Assessment

***Reidy Interactive Lecture Series (Session #1)***

***Portsmouth, NH***

September 26, 2019

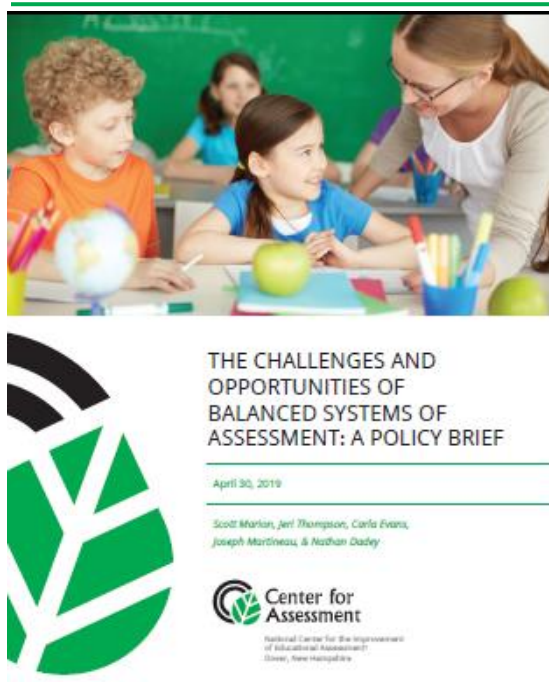


# Overview

---

- Introduction to balanced assessment systems
- Components and levels of balanced assessments systems
- The role of interim assessment

# A Tricky Balance



- We've been thinking and writing about balanced assessment systems for many years
- We've partnered with several states and districts to try to improve the balance of their systems
- Some have claimed we might be tilting at windmills



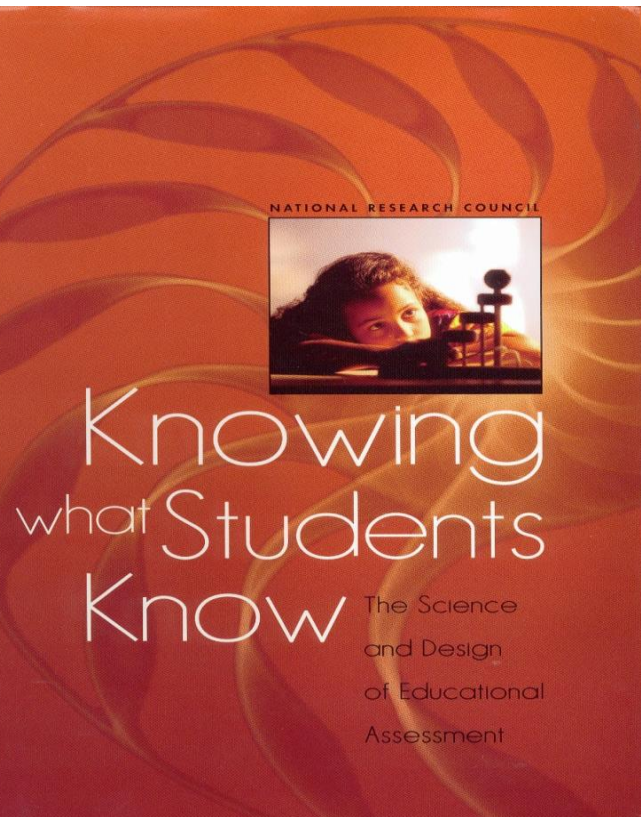
# Overview of Balanced Assessment Systems

---



# A Call for Balanced Assessment Systems

Assessments *at all levels*—from *classroom to state*—will work together in a system that is *comprehensive, coherent, and continuous*. In such a system, assessments would provide a variety of evidence to support educational decision making. Assessment at all levels would be linked back to the same underlying model of student learning and would provide indications of student growth over time (NRC, 2001, p. 9).



## Balanced Assessment Systems to Serve Multiple Purposes

Assessment systems designed to serve multiple purposes require thoughtful planning about which data will be privileged at each level (Chattergoon & Marion, 2016). For example:

- ✓ **standardized** vs. **dynamic/flexible**
- ✓ **uniform** vs. **variable** dates
- ✓ **independent** vs. **scaffolded** performance
- ✓ **delayed** vs. **immediate** feedback
- ✓ **stringent** vs. **less stringent** technical requirements



Accountability/Monitoring ← → Instructional

# Properties of Balanced Assessment Systems (NRC, 2001)

---

- 1. Comprehensiveness** – “a range of measurement approaches should be used to provide a variety of evidence to support educational decision-making”
- 2. Coherence** – “the conceptual base or models of student learning underlying the various external classroom assessments within a system should be compatible”
- 3. Continuity** – “assessments should measure student progress over time”

## Two additional criteria (Chattergoon & Marion, 2016)

---

4. **Efficiency** means getting the most out of assessment resources and eliminating redundant, unused, and untimely assessments.
5. **Utility** is a key criterion for assessment system quality should be the degree to which the system provides the information necessary to support the intended aims
  - Follows from a well-articulated theory of action that specifies the various intended outcomes for the system and the processes and mechanisms by which these outcomes will be realized.



# Coherence

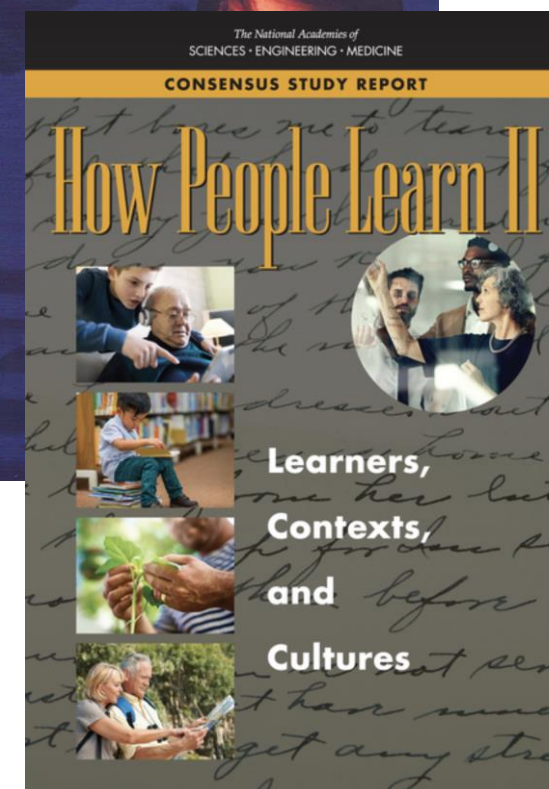
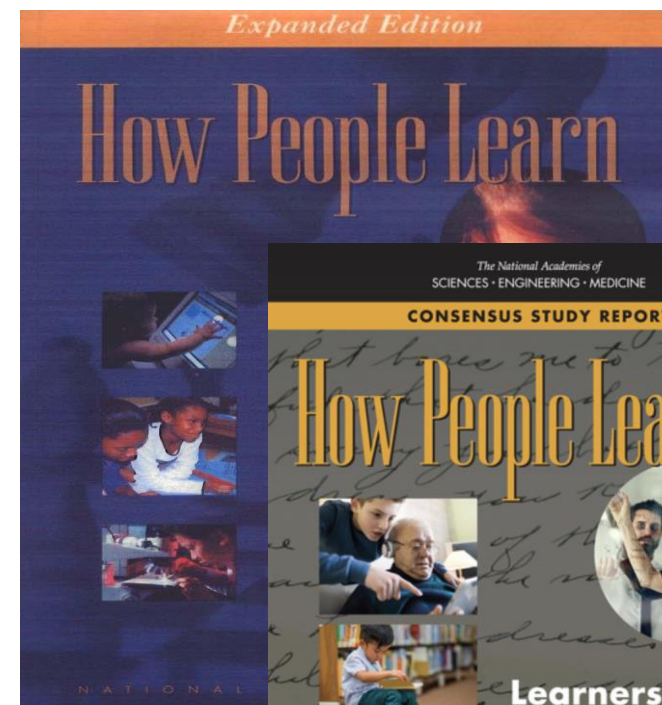
---

- Vertical **Coherence** – conceptual base or models of student learning underlying the various **external** and **classroom** assessments within a system should be compatible
- Horizontal **Coherence** – alignment among curriculum, instruction, and assessment along a common set of learning goals

# Vertical Coherence - Not Just Any Model of Learning

Assessments and assessment systems must be based on **research-based models of learning**.

Adherence to outdated, naïve, and/or implicit notions of learning are an **impediment** to assessment literacy and assessment reform.



# Horizontal Coherence

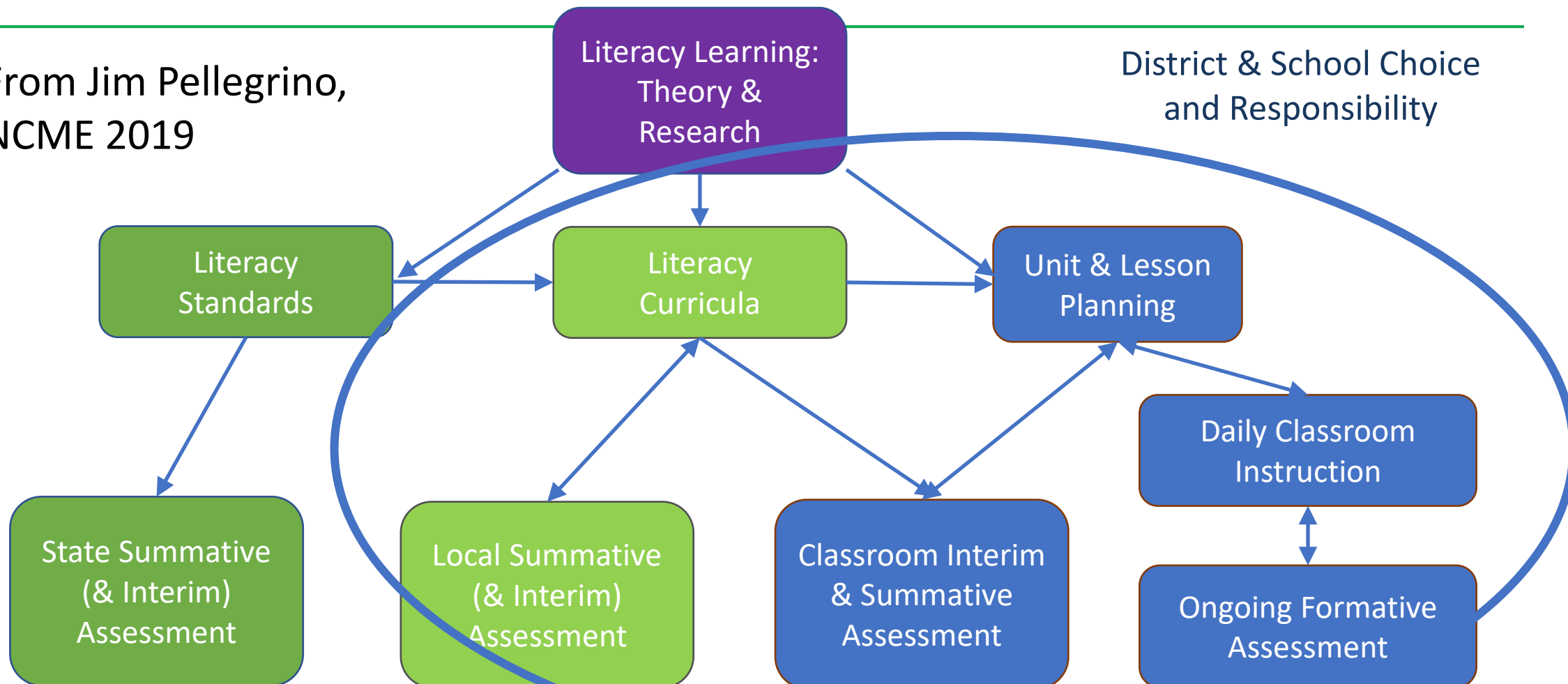
---

- Vertical Coherence – conceptual base or models of student learning underlying the various external and classroom assessments within a system should be compatible
- Horizontal Coherence – alignment among curriculum, instruction, and assessment along a common set of learning goals
- When this is done well, it is a considerable benefit to low-performing schools

# A “Simple” Model for a Coherent and Coordinated System of Curriculum, Instruction & Literacy Assessments

From Jim Pellegrino,  
NCME 2019

District & School Choice  
and Responsibility



# Layers and Components of Balanced Assessment Systems

---



# Multiple Layers of Assessment Systems

**“Minicosm”:  
District-Level Assessments**

**A key locus of control**



**Microcosm:  
Classroom-Level Assessments**



**Macrocosm:  
State System of Assessments**



See: Chattergoon, 2016; Marion, 2018; Shepard, et al., 2018

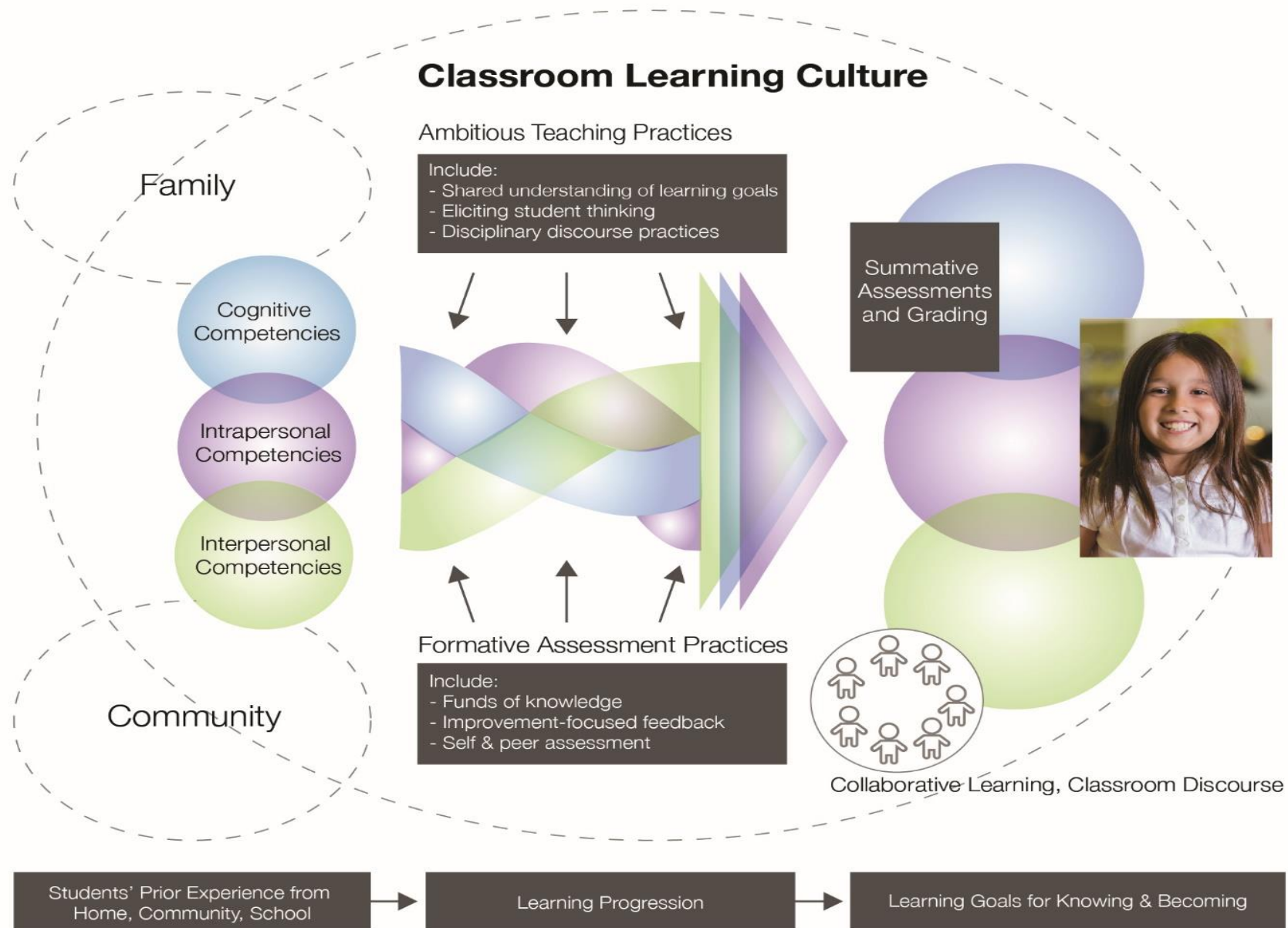
# Components of Balanced Assessment Systems

---

- Quick small group activity at your table (5 minutes total)
- Very quickly choose to focus on either a state, district, or classroom assessment system (1 minute)
- See if you can come to agreement at your table regarding the **critical** components of your assessment system, such as:
  - State summative
  - State interim
  - Unit-based assessments
  - Weekly quizzes
  - Classroom summative (e.g., grades)
  - District interim
  - Formative assessment (instruction)
  - Projects/exhibitions
- Note: these are only a few possible components



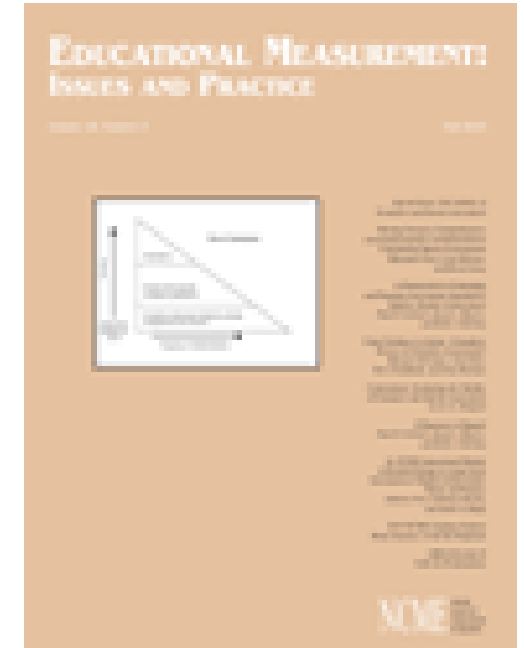
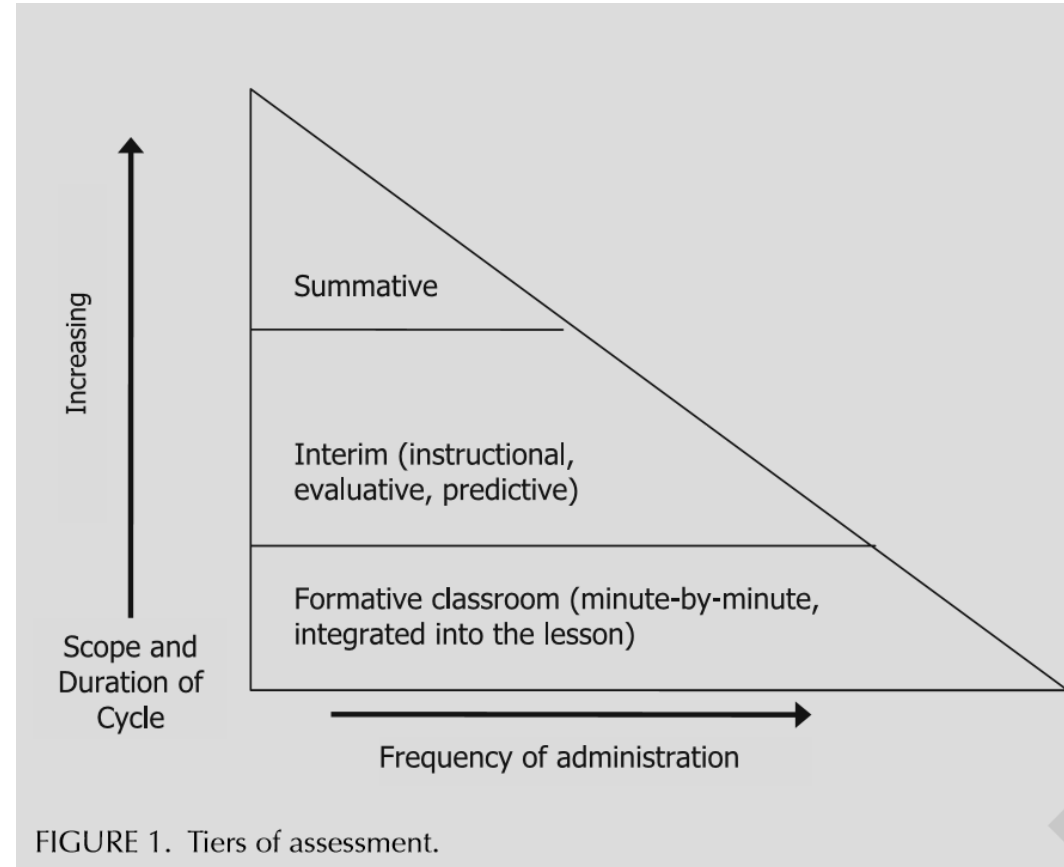
# Here's Lorrie Shepard's (2019) answer for classrooms...





# Perie, Marion, & Gong, 2009

This well-publicized figure might have led many to believe balanced assessment system designers needed to pick assessments from each of these three levels...**That's not what we meant!**



# The Role of Interim Assessments?

---

- We've seen a massive proliferation of interim assessments in the past 20 years
- What role, if any, should commercial interim assessments play in balanced assessment systems?
- Can they enhance the quality and balance of assessment systems? How?
- Can they threaten the balance and utility of such systems? How?



# Purposes and Uses

---

- In 2009, we outlined three broad categories of uses for interim assessments:
  - Instructional
  - Evaluative
  - Predictive
- We learned early on these categories were too broad to guide use whether as part of an assessment system or not
  - Brian’s terrific paper for this conference illustrates the importance of specifying what we mean by “instructional”
- The rest of the presentations this morning will dig into this need for specificity to help frame selection, evaluation and use



# One Lay of the Land: Preliminary Findings from Two High Level Scans

Nathan Dadey & Calvary Diggs

*The National Center for the Improvement of Educational Assessment*

September 26<sup>th</sup>, 2019

Reidy Interactive Lecture Series

Portsmouth, NH



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

The landscape of interim assessments has changed substantially over the last 20 years

due, in part to,

▶ Policy

▶ Technology

▶ Scale

The goal of this work is to share **two high level scans** that hopefully help us get a better understand that landscape.

# 1 **State Supported Interim Assessment Options**

Summary of interim assessments supported by state department of educations, as reported on department websites.

---

# 2 **Review of Available Literature on Interim Assessments**

A review of published literature dealing with the use of interim assessments.

# Summary of State Supported Interims

---





# State Supported Interim Assessments

---

- Increasingly, states are supporting schools and districts by providing access to or guidance on interim assessments.
  - We think this is reflective of a shift towards systems of assessments. That is, many states are now considering how assessments they can influence can work together.
- Given this, goal of this work is to understand what types of assessments are being supported by examining all **50 state department of education websites**.

Note: a previous version of this work was shared at the 2018 NCME Classroom Assessment Conference and the 2018 Reidy Interactive Lecture Series.

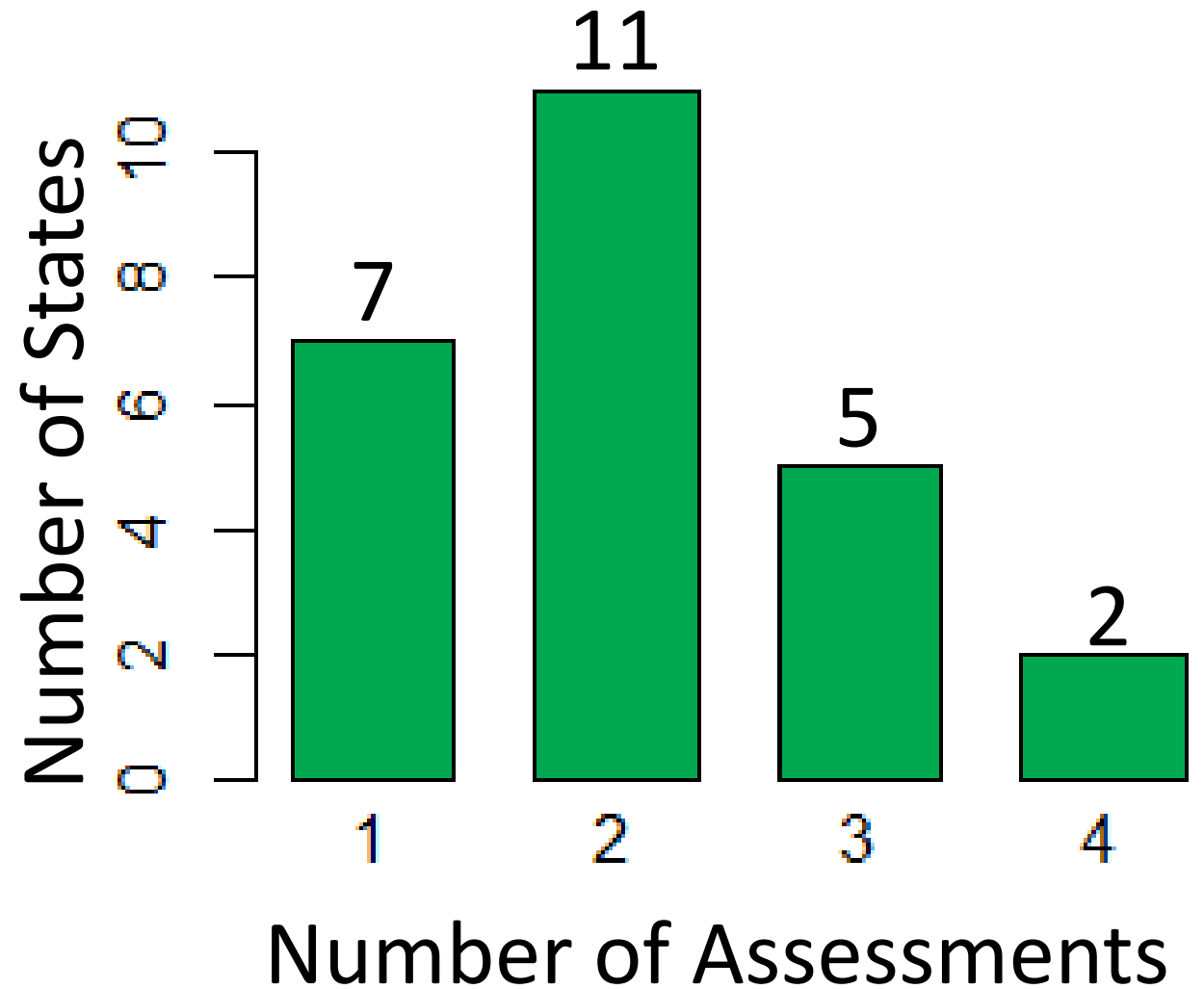


# 25

state websites had  
information on  
34 unique interim  
assessments.

Notes: The definition of interim is used quite broadly here (e.g., included early literacy screeners). Each lists of approved assessments are counted as a single assessment in the above counts. At least five states had materials that addressed the idea of systems of assessments: [LA](#), [MI](#), [RI](#), [VT](#) and [WY](#).

The number of assessments per state ranged from 1 to 4, with a median of 2.



Most states (22)  
provided assessments  
**directly to districts,**  
often through online  
platforms.

The remaining 3  
states provided lists of  
approved assessments:

- Early Literacy: [CO](#), [MI](#) & [SC](#)
- K-12 Math & ELA: [SC](#)

Distinguishing further among the assessments  
involves considering the  
interaction of **content** and **time**.

I.e., How, if at all, is the content domain divided up  
and assessed over time?

Note: the paper *Matching Instructional Uses with Interim Assessment Designs* (Gong, 2019) dives deeply into this issue and focuses on the claims. Gong ([2010](#)) also explores this issue, as does Dadey & Gong ([2017](#)).

Multiple terms are floating around in the literature to address these types of design choices.

*Mini-Interim*

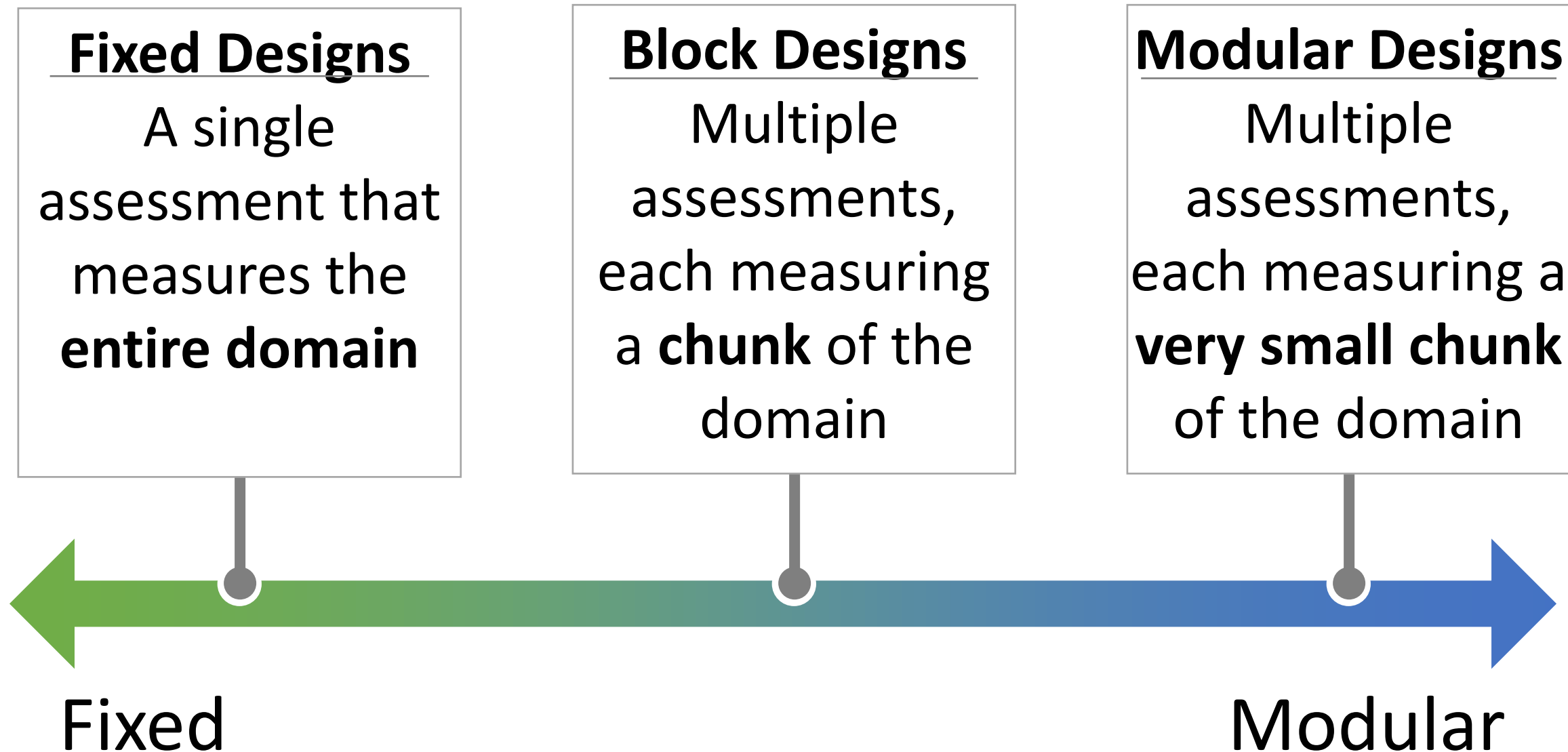
*Modular*

*Different Blueprint*

*Mini-Summative*

*Micro-Modular*

*Same Blueprint*



	Number of		
Design	States	Assessments	Unique Assessments
Fixed	19	23	14
Block	13	15	6
Modular	8	11	11

Notes: List of approved assessments are excluded in the above counts.



# Literature Review of Interim Assessment Use

---



“School districts across the country have [rapidly] adopted interim and benchmark assessments... Thus, it is especially important that well-conceived, empirical studies of the effects of such programs be carried out.”

(Shepard, [2010](#))

Our suspicion is that research on the uses of interim assessments has been emerging.

The goal of this project is to provide a **review** of this body of literature, with an emphasis on how the assessments are **used**.

# Literature Review

4,000



141



20

articles eligible for inclusion after cleaning, based on searches of key terms & assessment names within ASP and ERIC (which limits the review).

articles eligible for coding, after applying [inclusion/exclusion rules](#), within additional articles included with certainty.

articles coded thus far.

# Studies by Year

(N = 141)

Number of Studies

30

20

10

0

2001

2002

2003

2004

2005

2006

2007

2008

2009

2010

2011

2012

2013

2014

2015

2016

2017

2018

Year

# Focusing on Use for the Coded Studies

---

We coded each discrete use within each of the 20 studies as:

1. **Instructional, Evaluative or Predictive**<sup>1</sup> following Perie, Marion, & Gong ([2009](#)), and then
2. Created **additional codes with each category**, to better describe the specific ways in which the assessments are use following a grounded approach (following Hook, [2015](#)).

<sup>1</sup>Notes: In cases where predictions are used in a clearly instructional or evaluate manner, we have retained them within the predictive category.

# Uses by Study (n=20)

20

Studies contained at least one  
**instructional** use.

16

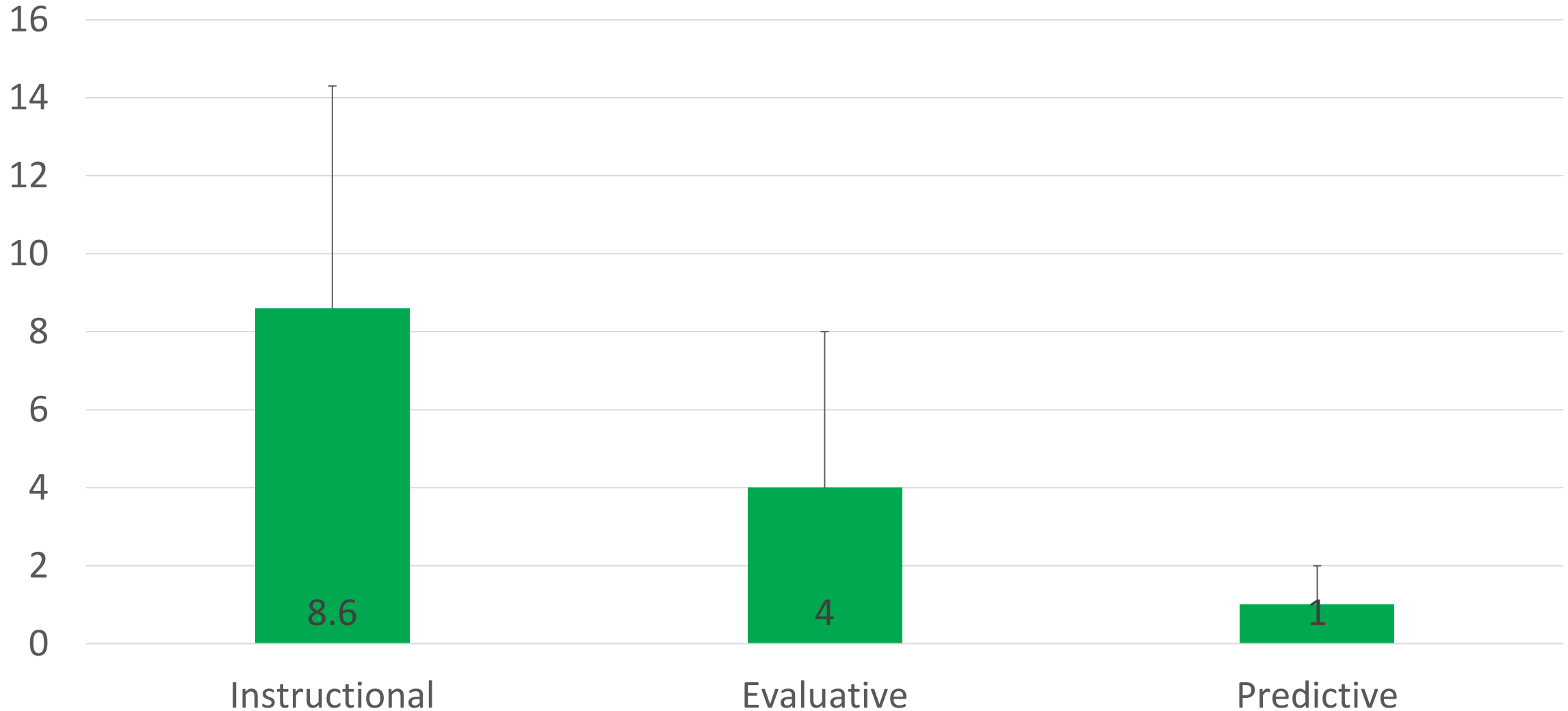
Studies contained at least one  
**evaluative** use.

10

Studies contained at least one  
**predictive** use.

# Number of Coded Discrete Uses by Study

Average Number of discrete uses within study





# Top Five Discrete Instructional Uses

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

	Use Description	<i>k</i>	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.

# Preliminary Conclusions

---

- Interim assessments are used in a multitude of ways.
- Description of assessment use is widespread, examination of efficacy is not.
- Much of the way in which use is describe is unspecific, meaning evaluation is difficult even when it is conducted as we have little idea about *how* the interim assessments are used.



# Setting the Stage

Erika Landl & Juan D'Brot

*The National Center for the Improvement of Educational Assessment*

*Reidy Interactive Lecture Series (Introduction)*

***Portsmouth, NH***

September 26-27, 2019



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

# Goal of RILS

---

- The purpose of RILS is to engage in discussions around a problem of practice
- Goal of RILS 2019: Leverage the experience of experts and practitioners to address the selection, use and evaluation of interim assessments within the context of a balanced assessment system
  - Engage in activities that revolve around a set of tools to help district leaders engage in a thoughtful process focused on identifying and prioritizing assessment needs and defining the types of assessment necessary to support those needs

# Interim Assessments: The Challenge

---

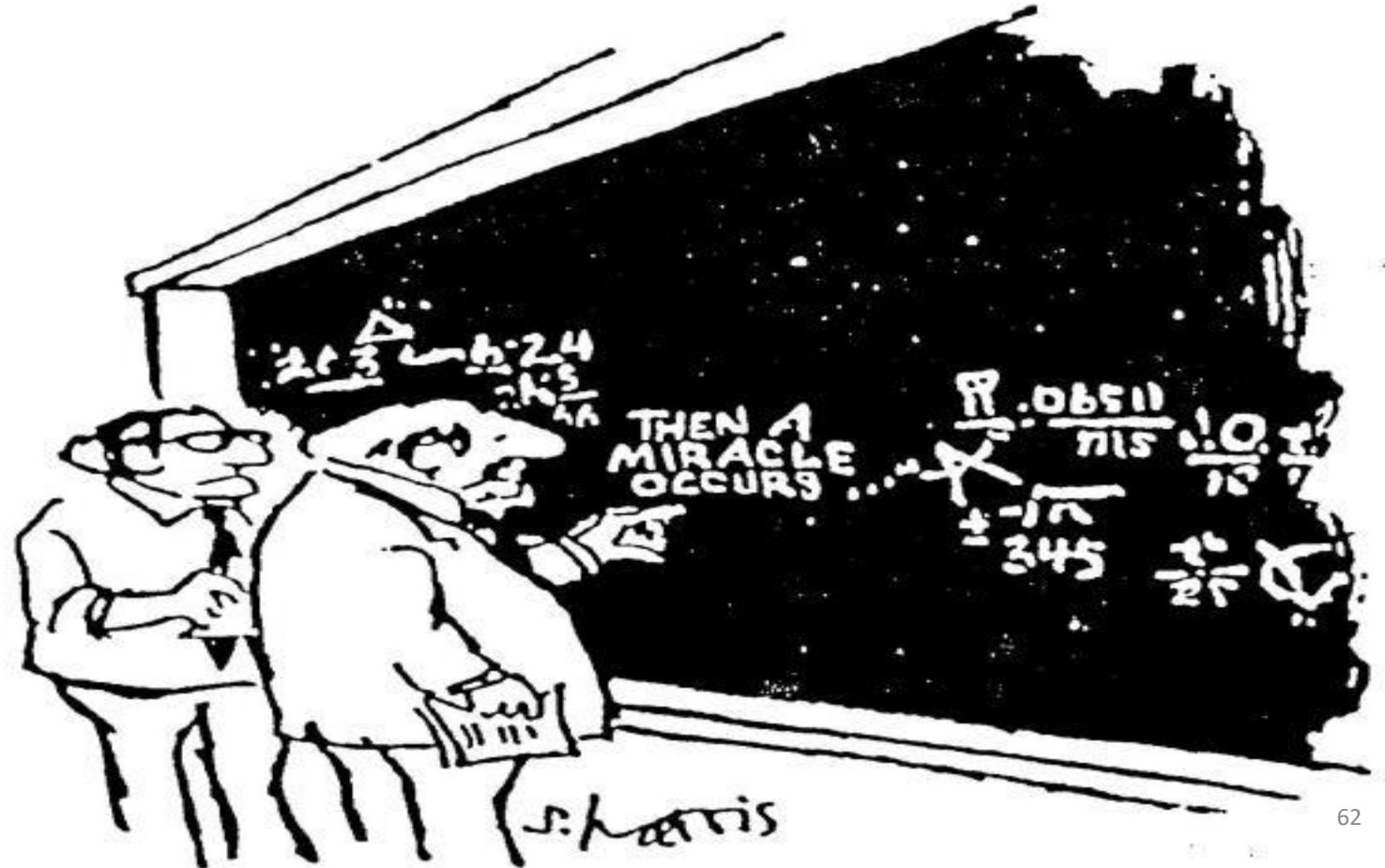
- Many barriers that impede the appropriate selection and use of interim assessments (Martineau, 2018)
  - Influence of politics, policy and political boundaries (hierarchy) on decisions
  - Influence of commercialization and proliferation of assessments
  - Lack of attention to curriculum and learning in the design of systems
  - Lack of assessment literacy at multiple levels of the system
- This is not unique to interim assessments

A common problem at state and/or district levels is that assessment components are not conceptually coherent.

They often conflict and as a consequence their use doesn't lead to the desired outcomes of educational improvement.

It is essential to make **EXPLICIT** one's assumptions and "theory of action" for the system of assessments.

*"I think you should be more explicit here in Step Two."*



# Interim Assessments: Mitigating Barriers

---

- Mitigate barriers by having a strong theory of action
- Theory of action must be in part based on how assessment information can support a vision of teaching and learning
- Consider what is needed and how it will provide for the outcomes desired
- Again, this is not unique to interim assessments *but* there are few requirements or guidelines that drive interim use

# Goals of the Toolkit

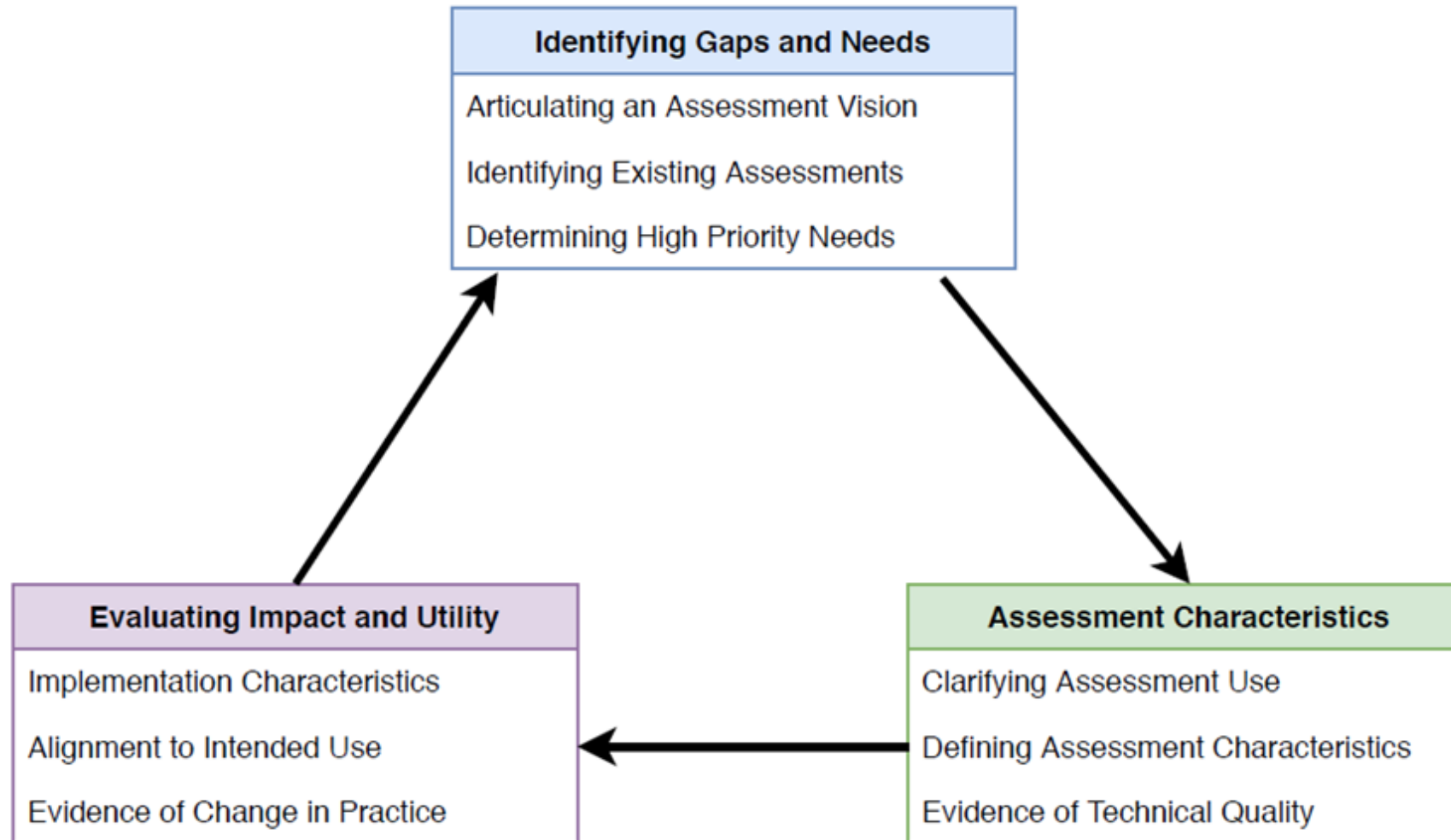
---

- To support a thoughtful and systematic *approach* to the selection and evaluation of an assessment *aligned to a clear theory of action*
- *Promote collaboration* among educators and leaders who select and implement interim assessments
- Articulate the factors that *influence assessment characteristics and features*
- **Not to advocate for or condemn any specific type of assessment product or tool**



# What does it do? Summary of the Toolkit

---



# Ok...So What are We Trying to Accomplish?

---

- Desired end product is a multilevel system of assessments
  - Assessment tools are designed to serve an intended purpose, use and user
- The levels are articulated and conceptually coherent
  - They share the same underlying concept of what the targets of learning are at a given grade level and what the evidence of attainment should be.
  - They provide information at a “grain size” and on the “time scale” appropriate for translation into action.

# Defining Interim

Summative <-----Interim-----> Formative

Purpose/Use

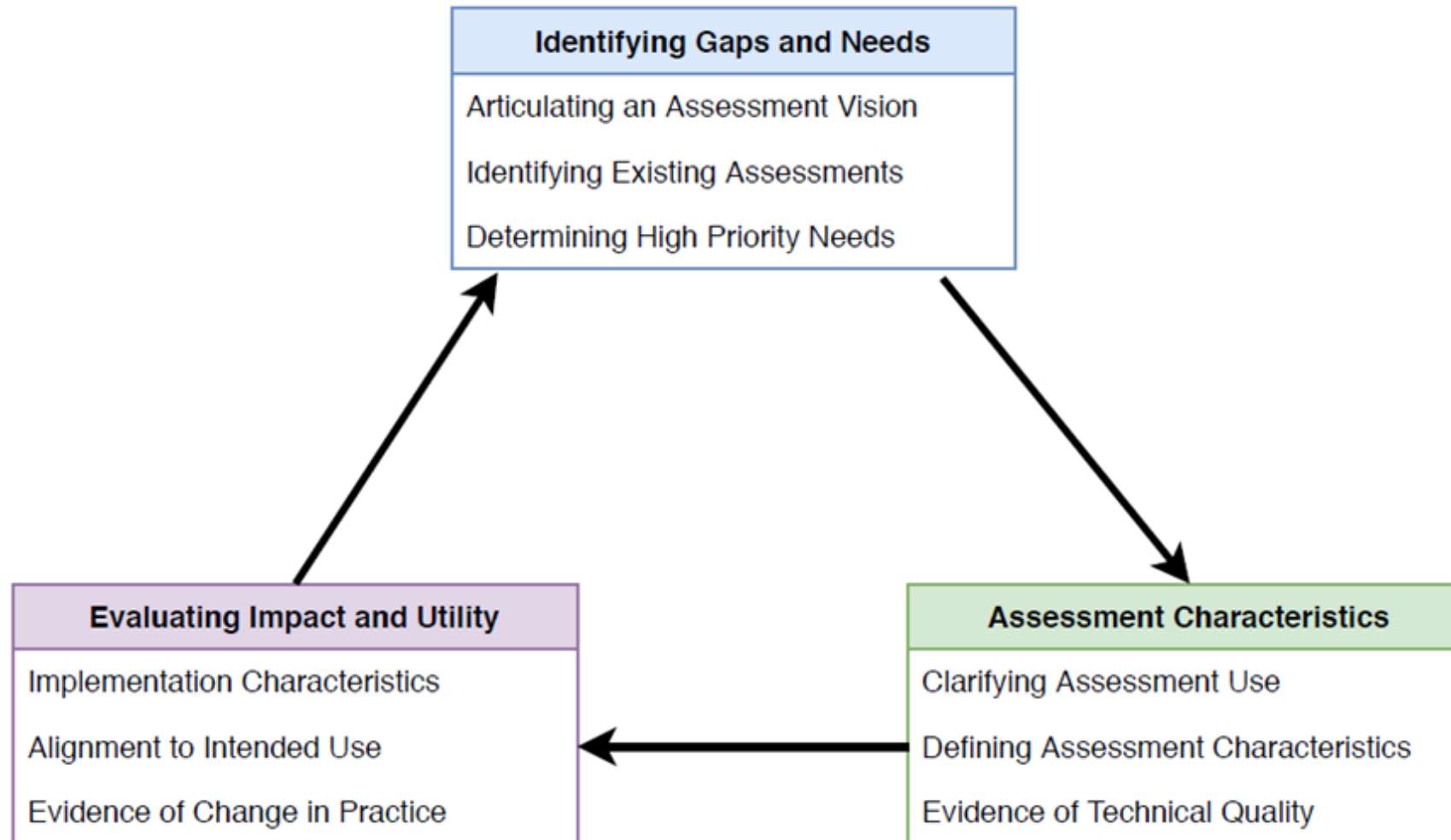
Accountability,  
Program Evaluation,  
Promotion/Retention;  
Grading

Elicit and use evidence  
of student learning to  
improve instruction.

		Tier/Type of Assessment		
		Summative	Interim	Formative
Owner	Teacher	<ul style="list-style-type: none"> <li>• <b>Final</b> exams, projects, and papers</li> </ul>	<ul style="list-style-type: none"> <li>• Graded quizzes and homework</li> <li>• Unit projects, papers, and exams</li> <li>• Mid-term exams</li> <li>• Marking period exams</li> </ul>	<ul style="list-style-type: none"> <li>• Strategically planned check-ins</li> <li>• Homework used to provide feedback and revision before grading</li> </ul>
	District	<ul style="list-style-type: none"> <li>• <b>Common</b> final exams, projects, and papers</li> <li>• Common assessments for testing out of a course/credit</li> <li>• Common graduation assessments</li> </ul>	<ul style="list-style-type: none"> <li>• Common unit exams, mid-terms, and marking period exams</li> <li>• Common quarterly assessments</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Not applicable</b></li> </ul>
	State	<ul style="list-style-type: none"> <li>• Annual state tests</li> <li>• End of course state tests</li> </ul>	<ul style="list-style-type: none"> <li>• State-provided within-year common assessments</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Not applicable</b></li> </ul>

# Recall: Summary of the Toolkit

---



# Steps to Support Identify, Implement, and Evaluate

---

1. Specify a vision of teaching and learning, including valuable student knowledge, skills, and dispositions;
2. Articulate how assessment information that can support this vision should be used;
3. Determine how existing assessment information is being used;
4. Identify the gap between use and high priority assessment information needs;
5. Determine key assessment design, administration, and reporting characteristics that align to intended uses; and
6. Engage in an evaluation of the impact and utility of these assessments

# Phase 1: Identifying Gaps and Needs

---

## Phase 1 of the Toolkit Includes 4 Sections

1. Establishing or overall vision for teaching and learning
2. Articulating your vision for assessments
3. Understanding the assessment lay of the land
4. Identifying your highest priority needs based on how assessments are used and how they should be used

# Phase 2: Assessment Characteristics and Quality

---

## Phase 2 of the Toolkit Includes 3 Sections

1. Clarifying how you intend to use interim assessments
2. Defining the assessment characteristics that align to intended use
3. Identifying and evaluating evidence of technical quality

# Phase 3: Evaluating Impact and Utility

---

Phase 3 of the Toolkit is not yet Developed, but it should Evaluate whether Assessments

1. Are aligned to your theory of action
2. Are used as intended
3. Lead to changes in behavior as expected



# Applicability: Evaluating Interim Assessments

Scenario	State Responsibility	District Role
State provides common, voluntary interim with specified windows	Intended purpose and use of interim assessment results; administrative requirements to support intended use at the state level; limitations of information, etc.	Evaluate how interim assessment results supports district goals/vision so information can be used appropriately. Determine need for additional assessments based on perceived gaps.
State provides modular interim assessments with on-demand administration	Intended purpose and use of interim assessment results. Why the assessment was selected and how it meets state goals, etc. Factors to consider when evaluating when to administer the assessment, and how to use/interpret results.	Evaluate whether/how the interim assessment aligns with the district’s goals/vision. Determine whether the assessment will add value given the district’s data needs or whether a different assessment should be considered (in lieu of or in conjunction with the state test).
District has been using an interim assessment	Broad characteristic and features that should be considered/prioritized as part of selection process to support coherence with state assessment and vision (e.g., alignment to content standards, etc.).	Evaluate whether the current interim assessment supports district goals/vision, can be used as needed, and is having the intended impact.
District looking to select an interim assessment		Evaluate the type of interim needed (formal/non-formal); Identify an interim assessment that aligns with the district’s goals/vision and meets its data needs.

# Current Constraints

---

1. Length
2. Complexity of Concepts
  - a) Although developed to be user friendly, there are still concerns that many of the concepts are not things that district leaders could address without the support of an assessment expert.
3. Guidance Around the Need for/Appropriateness of Formal Tool
  - a) The tool is really focused more on helping districts think about and prepare for the evaluation/selection/development of a formal instrument.
  - b) It may be the case that the high priority needs/uses defined at the end of Phase 1 are better served by an informal tool. Currently there is nothing in place to help them independently make that distinction.
4. Guidelines for Evaluating Different Options
  - a) This helps identify desired assessment characteristics and highlight high priority evidence in light of a high priority need/use(s), but does not provide guidance for selecting among different options. There is not a scoring sheet that helps districts make a final decision.
  - b) It is presumptive for us to think that we could divine all possible contexts for “their best assessment.”

# Structure of the Meeting

---

- We will be hosting discussions that highlight issues to consider when developing or selecting assessment tools
- Panels will provide contextually-based perspectives and how they've addressed this problem of practice
- We will be sequentially thinking about
  - Vision and Theory of Action
  - Assessment Characteristics aligned to a Theory of Action
  - Evaluating Impact and Utility

# Small Group Discussions

---

# Small Group Discussion

---

- We will be asking you to engage in small group discussions around three questions focused on
  - Information needs
  - Assessment characteristics
  - Evaluation of impact and utility
- Activity:
  - Please identify a note taker for your table
  - Individually, jot down responses to the questions on the next slide
  - As a group, share your thoughts and identify differences and similarities
  - See if you can come up with any agreed upon recommended practices
  - Be prepared to share (briefly!) the highlight of your discussion

# Small Group Discussion - Key Questions

---

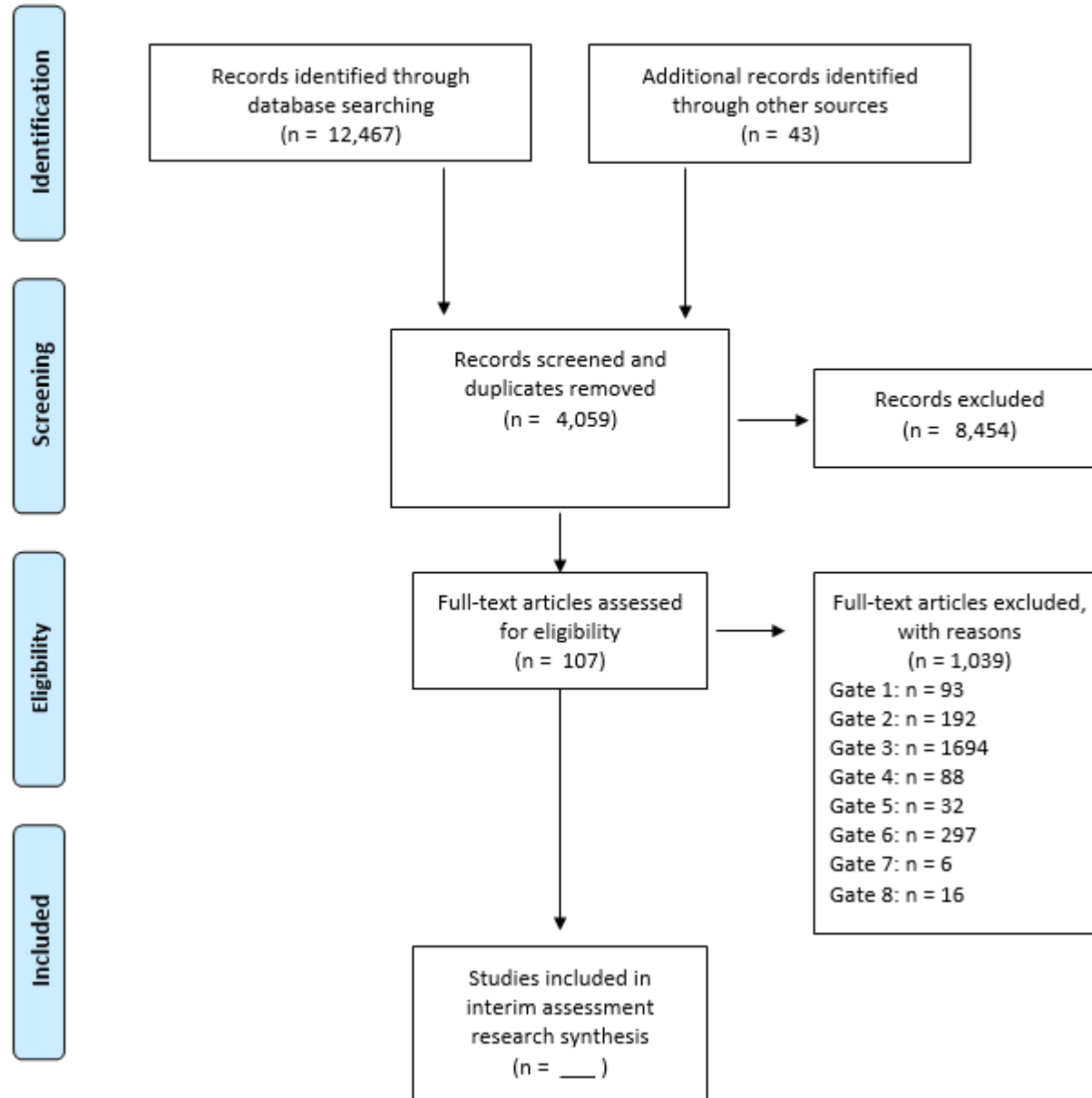
- 1. Evaluation of Information Needs:** What procedures are currently in place (or should be in place) at a school or district to identify high priority needs for assessment information?
- 2. Specification of Assessment Characteristics:** What procedures do you use (or should you use) to identify and select assessments that will meet your needs? How do you evaluate the fidelity and appropriateness of particular assessment design?
- 3. Evaluation of Impact and Utility:** What practices do you have (or should you have) in place to evaluate whether current interim assessments are having the impact intended?

# Supplemental Slides

---



# PRISMA 2009 Flow Diagram





## Inclusion Criteria

1. English language, conducted in the US, published since 2000
2. Conducted with a population from K-12 settings
3. An assessment was a key component of the study and functioned as an IV
4. The assessment was academic in nature
5. The assessment was administered by school or affiliated staff for school use (i.e., not administered by parents, clinics, outside consultants)
6. The assessment can be described as an “interim assessment”
  - i. multiple measurements points or used mid-way through the term
  - ii. assessment data were used for a purpose that may be broadly described as evaluative, instructional, or predictive
  - iii. assessment data can be (or were) aggregated to represent the entire grade, school, district, or state
  - iv. if not aggregated, data would still be relevant to a purpose (evaluative, instructional, predictive) at a higher level (e.g., school, district, state)
7. The interim assessment is/was commercially available or was developed by a school or system(s)
8. The study evaluates an interim assessment using a traditional experimental or quasi-experimental design, observational methods, or recollections.

Exclusion (The inclusion criteria were designed to identify studies and assessments that fall under our current typology of an, “interim assessment.” In the current conceptualization, interim is used to denote assessments that are not clearly used for formative or summative purposes. However, there are various ways that each of these concepts could be defined, which, in turn, would provide a different approach to the inclusion and exclusion process. This study aims to provide a wide, scoping search of the literature. Exclusion criteria for both summative and formative assessments were defined as follows,

Focal assessment is clearly used for formative purposes (i.e., teacher administered, informal, conducted frequently, and data can not be aggregated and meaningfully generalized beyond the individual classroom)

Focal assessment is clearly used for summative purposes (i.e., an end of the year state test)

The following criteria were set to allow studies who used interim assessments for traditionally summative or formative purposes, thereby allowing the study to explore the range of their utility for a variety of applications.

# Method

---

- Literature Review
  - Rapid review (Grant et al., 2009)
  - Four means of identification
- Inclusion Criteria
  - “Interim assessment”
  - Additional Gate
- Coding
  - General characteristics
  - Discrete uses

