

A Pilot Year in Review

WHAT HAVE WE LEARNED ABOUT
THROUGH-YEAR ASSESSMENTS?

About this publication

A Pilot Year in Review: What have we learned about through-year assessments?

This publication interrogates key takeaways from through-year assessment pilots administered during the 2022-2023 school year. We explore key design decisions, enabling conditions and implications for future research and practice. This publication is part of a series published through Education First's Through-year Curriculum-Connected Assessment Grant Program.

Acknowledgements:

In 2021, Education First began work with assessment developers and state education agencies on researching and developing a new generation of through-year solutions connecting what is taught with what is tested by aligning assessments with scope and sequences or curriculum. This work is supported by the Bill & Melinda Gates Foundation, the Walton Family Foundation and the Chan Zuckerberg Initiative. Learn more about the next generation of curriculum-connected through-year assessments [here](#).

Authors



Dave Powell
*Senior Consultant
Education First*



Emma Fortier
*Associate
Education First*



Senna Lamba
*Associate
Education First*



Khaled Ismail
*Principal
Education First*

Table of Contents



01. Executive Summary
02. Why Through-Year Assessment Models?
03. Lessons Learned
04. Exploring the Models
05. Where Do We Go Next?
06. Appendix

EXECUTIVE SUMMARY



Introduction & Background





*While state summative assessments serve an important role in our education system, **they have the potential to improve through various innovations***

Education First believes students, educators, families and state leaders need more equitable, focused and relevant assessments that strengthen the connection between assessment and instruction and better align what is tested with what is taught

Since 2019, we've led grant and coaching programs to advance innovations in assessment, reporting and accountability



Grants have supported a range of innovations, including **through-year assessments, computer-adaptive assessments, comprehensive graduate portfolios, whole-child measures and equity indicators.**

The curriculum-connected through-year assessment grant program has focused on:

Incentivizing R&D to support states to focus on curriculum-connected through-year models

Building connections and buy-in among federal advocates and policymakers for change, including supporting CGSA grant writing

Facilitating a community of practice among grantees, innovative states and developers pursuing through-year assessment

Sharing our learnings and thought leadership with the field

The models we invested in must address the needs of stakeholders and advocates, and meet the following criteria:



Disaggregate data for essential student populations and provide data across schools and districts.



Eventually scale statewide or organization-wide *(if shown to be successful)*



Be able to **integrate within the state accountability system in the future**, even if that integration requires policy change



Many states are exploring through-year assessments to address some long-standing, legitimate concerns about traditional end-of-year summative assessments

Stakeholders (including students, families and educators) often see traditional end-of-year summative assessments as:

Lacking utility to teaching and learning

Providing untimely results that do not inform instruction

Requiring a large footprint on the overall system

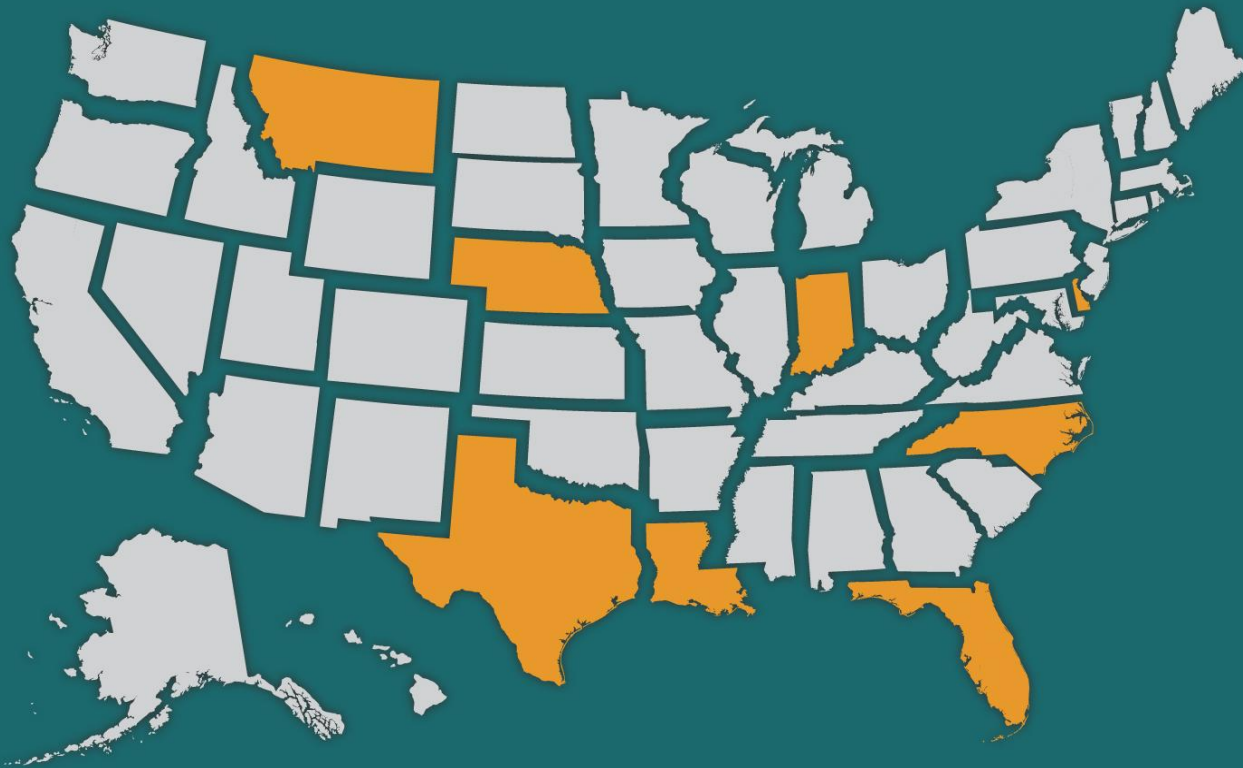
(in terms of the resources needed, time for preparation and administration)

Misaligned to what and when students are taught and their curriculum

Read more about the reasons for the growing interest in through-year assessment models in Education First's publication, "[What are Through-year Assessments?](#)"



This report synthesizes learnings from 8 states who tested different versions of through-year assessment models in the 2022-2023 school year

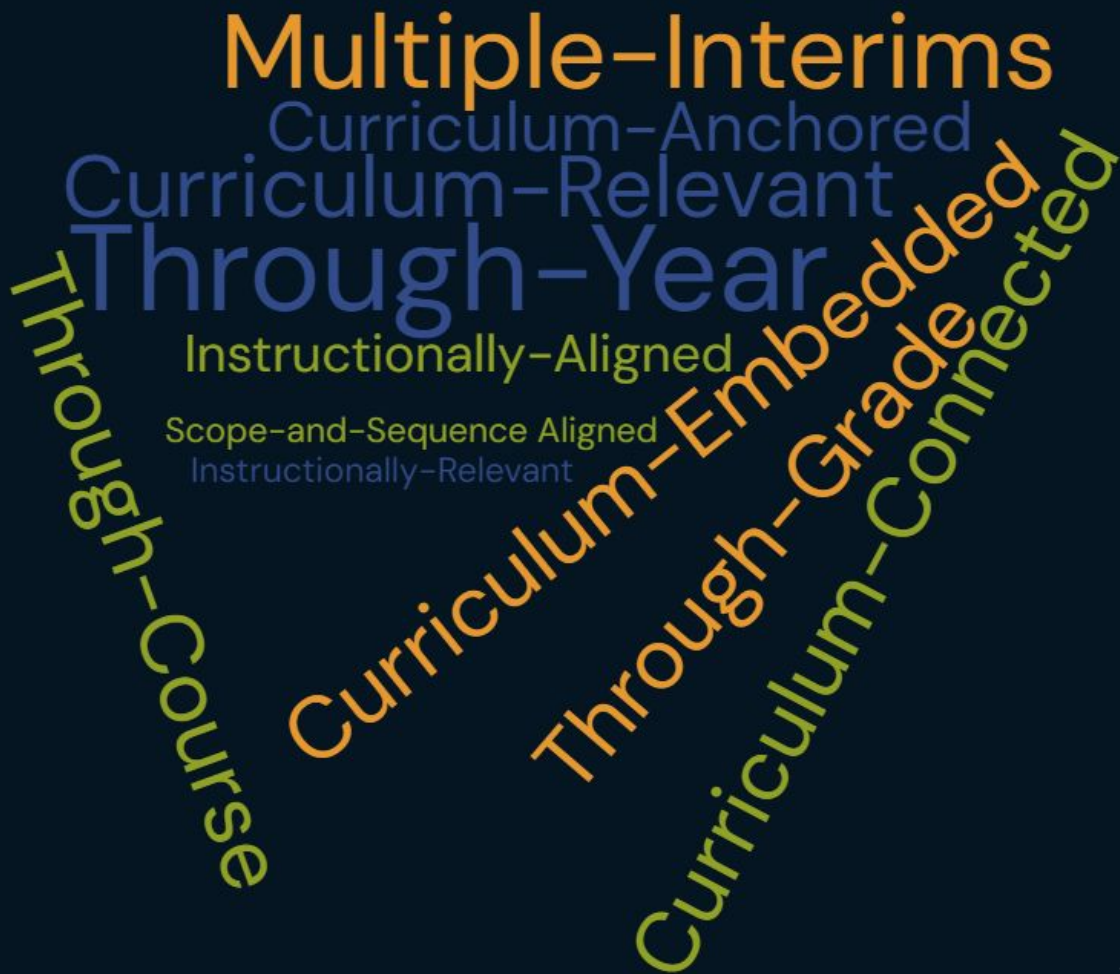


We also incorporated learnings from CenterPoint’s Through-Year Illustrative Mathematics (IM)-Aligned Interims study with district partners in Maryland and Wisconsin. In total, 13 states were currently exploring, developing, or testing through-year models in the 2022-23 school year.

Through-year assessment designs differ, and the field has varying levels of alignment on definitions.

These definitions are continuing to evolve as the field develops.

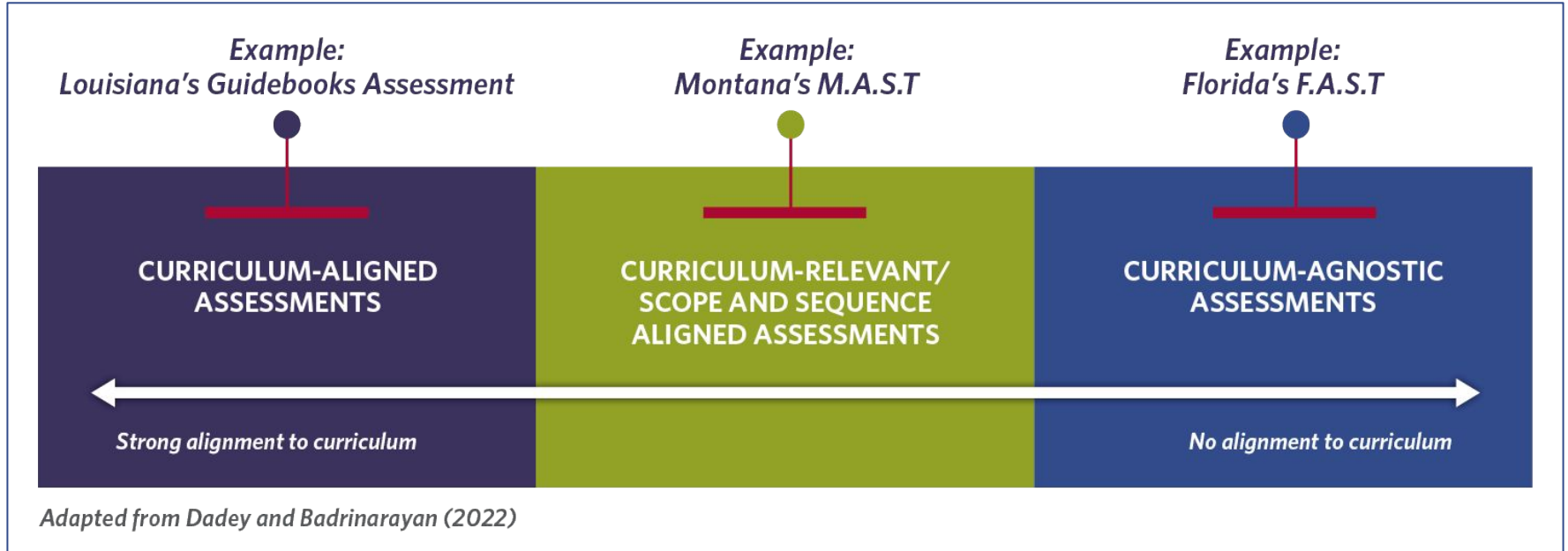
Our thinking on definitions has also evolved as we've continued to learn about through-year models.





All through-year assessment models involve multiple administrations, but they vary in the degree to which they connect to curriculum and instruction

We've seen a spectrum of models and approaches in the states we studied.





In this report, we use the following definitions that build on our prior thinking and the work of others in this field:

Through-Year Assessment Models

Through-year assessment models administer multiple tests throughout the school year as part of an assessment system designed to produce a single summative score meeting federal and state accountability requirements. Through-year assessment models are also referred to as “through-course” by some states.

Curriculum-Agnostic Approach

Through-year assessment models that **test the entire content domain (or grade-level standards) throughout the year at each testing administration**, and do not try to align content tested to curriculum.

Curriculum-Aligned Approach*

Through-year assessment models that directly **draw on the content found in specific curriculum**. This model is also referred to as “curriculum-specific” or “curriculum-embedded.”

Curriculum-Relevant Approach

Through-year assessment models that can be **flexibly aligned with multiple curricula, a scope and sequence or pacing of content**. This approach is also referred to as “scope and sequence aligned”, “instructionally relevant” or “instructionally aligned.”

*In previous publications, we referred to this as curriculum-specific embedded

With these definitions in mind, this publication explores the following research questions



Primary Research Questions:

What are the lessons learned from a group of states who piloted through-year models in the 2022-2023 school year?

What are key implications and recommendations for scaling the models?

What are the outstanding questions, needs and considerations for the future of innovations in assessment?

Across the states we reviewed for this publication,
over
2.5 million
students tested using a through-year assessment during the 2022-2023 school year.

We answered these research questions through stakeholder engagement, literature reviews and interviews



Methodology:

1

Synthesis of stakeholder engagement findings, surveys, prototyping and piloting reports from three assessment developers and two states

2

Literature and artifact review

3










Interviews and focus groups with 20 state leaders, assessment developers and field leaders

Key Findings & Recommendations



We examined the following models and discussed lessons learning in-depth with the leaders implementing them



	Delaware (DE)	Through-Course Assessment
	Florida (FL)	Florida Assessment of Student Thinking (FAST)
	Indiana (IN)	Through-Year Assessment (ILEARN (State Summative) Redesign)
	Louisiana (LA)	Guidebooks and Wit & Wisdom — CrawFish Model
	Montana (MT)	Montana Alternative Student Testing Pilot Program (MAST)
	Nebraska (NE)	Student-Centered Assessment System (NSCAS)
	North Carolina (NC)	North Carolina Personalized Assessment Tool (NCPAT)
	Texas (TX)	Texas Through-Year Assessment Pilot (TTAP)
	Districts in MD and WI	CenterPoint's Through-Year Illustrative Mathematics-Aligned Interims

Each of the states vary in how they approach the design of their through-year assessment, the degree to which they connect to curriculum and the goals they aim to achieve



Feature	LA*	LA**	MT	DE	FL	IN	NE	NC	TX
Each administration assesses the depth and breadth of grade-level standards	+	+			+		+	+ <i>ELA</i>	+
Each administration assesses a subset of standards			+	+		+		+ <i>Math</i>	
Curriculum-aligned	+								
Curriculum-relevant		+	+	+		+			

*Guidebooks and Wit & Wisdom **CrawFish Model

Note: CenterPoint's Through-Year Illustrative Mathematics-Aligned Interims studied in districts is curriculum-aligned and assess the depth and breadth of standards.

Each state's approach to through-year assessment design depends on the problem(s) they are attempting to address, their underlying beliefs and their state's context



These solutions are not one-size-fits-all. States have come up with a myriad of solutions based on...

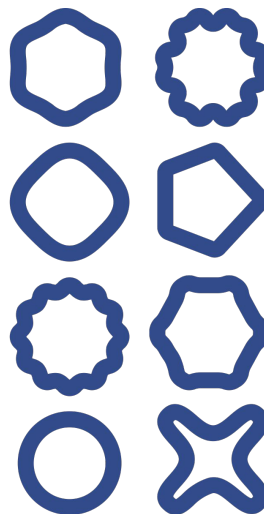
1 The problems they intend to solve for

2 Underlying beliefs and assumptions about teaching and learning

(Including how and when students should acquire, retain and demonstrate knowledge)

3 Individual state contexts

(local control of curriculum, adoption of HQIM)



In this publication, we studied a range of through-year approaches and will share what we've learned about the considerations and trade-offs.

The field is nascent in understanding the degree to which different approaches may solve for different problems and more research is needed to demonstrate impact



There is currently limited data on the impact of different approaches on student outcomes or educator experience with implementation.

For example:

- To what extent might curriculum-aligned through-year assessments reduce inequity caused by disparities in background knowledge?
- How does a scope and sequence aligned or curriculum-relevant model provide flexibility for teachers and support students based on the science of learning and development? What might be different in ELA compared to math?





In addition to looking across the 8 states in this publication, we take a deep dive into three assessment developers' and their state and district partners' through-year model designs

Developer	Partner	Model
NWEA	Louisiana	CrawFish Model: curriculum-relevant, through-year assessments for ELA that can work with two HQIM (Guidebooks and Wit & Wisdom)
New Meridian	Louisiana, Montana*	MasteryGuide Assessments: curriculum-relevant, through-year ELA and math testlet designs
CenterPoint	Two urban districts	Illustrative Mathematics (IM)-Aligned Interims: curriculum-aligned, through-year math assessments aligned to the Illustrative Mathematics scope and sequence

*Montana refers to this as MAST

States and developers we interviewed are grappling with both their aspirations for through-year assessments and scaling the model



1

Our partner states and assessment developers are making a bet that if you test throughout the year, align those tests to what is taught and provide timely reporting—student learning and outcomes will improve.

Research supports the claim that providing timely feedback, making course corrections and increasing the coherence between instruction, curriculum and assessment bolsters student learning.

2

State leaders identified three key enabling conditions critical for any state considering a transition to a through-year assessment:

- Strong partnerships and external support
- Coherent and expert internal capacity
- Diligent planning and communication

3

States are confronted with the ways current assessment systems, behaviors and incentives are oriented to support traditional end-of-year summative assessment models.

A transition to through-year models would require fundamental shifts to build buy-in, supportive infrastructure and ensure implementation fidelity.

A great deal of professional learning and stakeholder engagement is required to ensure buy-in and fidelity of implementation



Stakeholder Engagement

States and developers are continuing to engage stakeholders, including educators, students and families, at multiple points in time and in multiple ways, throughout their pilot and implementation.



Professional Learning

While states have provided professional learning on understanding what through-year assessments are, meeting districts, educators, families and students' desire for support in interpreting results to inform instruction is an emerging focus.

As through-year models scale and integrate into accountability systems, we recommend states and districts consider reducing duplicative testing and aligning intended purposes with the tests used



States, their partners and districts must consider the ways that a through-year summative system should be situated within a balanced assessment system.

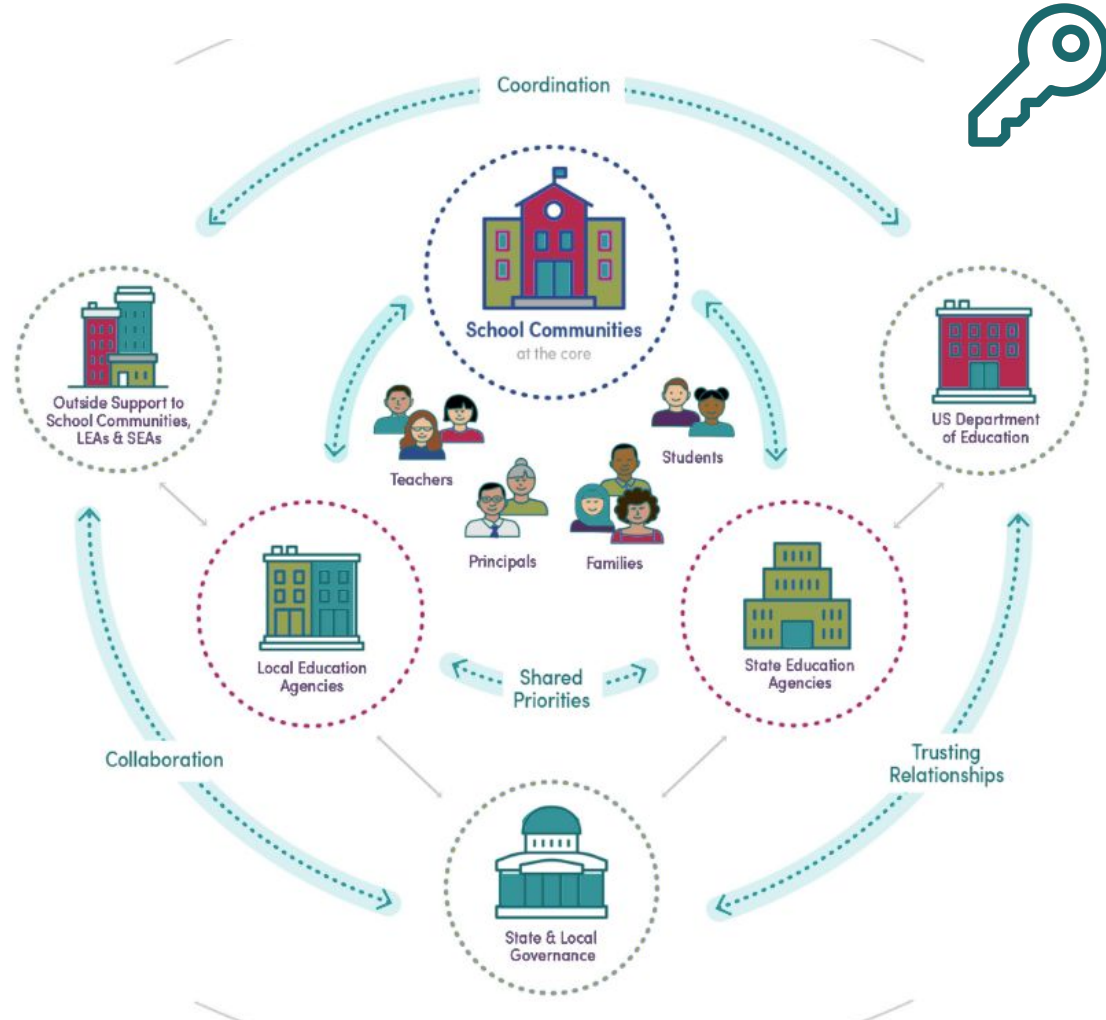
Depending on the design and approach that a state is taking with their through-year model, **what, if any, additional benchmark or interim assessments are needed at the district level?**

If the data provided from the through-year system yield actionable and timely results that support the same purposes, **can districts reduce the amount of overall testing that students and teachers are experiencing?**

What supports will schools, districts and educators need **to use the data from through-year models to bolster instruction?**

States and their partners must also focus on **clear, coherent and systematic implementation** in a way that **builds and deepens buy-in of stakeholders**.

Iterating on the **test design, utility, reporting and information with key stakeholders** including teachers, parents and students can **ensure buy-in through the change process**.



Overall, the field still has a lot to learn about the impact of and process of designing and implementing through-year assessments



It is too early to tell the degree to which each model or approach will improve student outcomes. **Each state needs to define the problems they are solving for and develop the model that best meets their local contexts.**

For states interested in:

Testing what students are taught closer to the point of instruction

Providing timely results that support teachers to use the data from the tests to inform instruction

Creating balance within an assessment system and reducing the overall footprint of testing over the course of the year

The research and lessons learned from states described in the following sections support aligning through-year assessments to curriculum and/or scope and sequence.

WHY THROUGH-YEAR ASSESSMENT MODELS?

THE POTENTIAL BENEFITS, TENSIONS AND ONGOING ISSUES



The Research Supporting Through-Year Assessments



Education First believes that if states test throughout the year, align those tests to what is taught and provide timely reporting—student learning will improve



If you...

- + Test throughout the year
- + Align the content of the test to what students have recently learned
- + Provide reports in a timely manner

Then you can...

- + Address some of the disparities in background knowledge
- + Provide more frequent and timely feedback to students and instructors
- + Create space for course corrections
- + Support teachers in planning instruction & scaffolding material
- + Measure the acquisition of knowledge more effectively

Which will...

- + Improve student experience and outcomes
- + Make assessments more equitable
- + Improve coherence between instruction, curriculum and assessments



**Improve
student
learning**

The research supports our hypothesis: Through-year assessments that connect more closely to what students are taught have the potential to improve student learning



By providing timely feedback to students

By creating space for course corrections

By creating greater coherence between instruction, curriculum and assessments

By increasing the acquisition and retention of knowledge

By decreasing the role of background knowledge in student performance

This includes curriculum-aligned and curriculum-relevant through-year assessments

Student learning improves when students are given timely and relevant feedback



What the research says

Students learn best when they receive timely and relevant feedback. Research shows that timeliness of feedback is critical to its effectiveness in order for it to resonate and impact a students' next task.

Claim

If the reports provided by curriculum-aligned and curriculum-relevant through-year assessments provide timely formalized feedback for students and teachers, then student learning will improve.

Student learning improves when teachers make course corrections based on their needs



What the research says

Course corrections are when an instructor uses the information gained from an assessment to inform and change their instruction. This could mean providing acceleration or reshuffling a scope and sequence. These kinds of behaviors can support student learning by increasing the personalization of the student experience.

Instruction that adapts to individual students to help them access grade level instruction can improve student learning.

Claim

Curriculum-aligned and curriculum-relevant through-year assessments provide more frequent and timely reports of student performance. If teachers use these reports to scaffold instruction for students to access grade level instruction, student learning will improve.

Student learning improves with increased coherence between curriculum, instruction and assessments



What the research says

When assessments are integrated into coherent systems that include high-quality curriculum and rigorous instruction, and are moving towards a united goal of improving student outcomes, student learning can improve as a result. Coherence with curriculum often more sharply connects with student learning than standards because it is tied to specific content.

Claim

If curriculum-aligned and curriculum-relevant through-year assessments increase coherence between instruction, curriculum and assessments, then student learning will improve.

Student learning—and in particular, the acquisition and retention of knowledge—improves with increased opportunities to retrieve information



What the research says

The act of retrieving information helps improve student learning by increasing the acquisition and the retention of knowledge. An assessment requires students to retrieve information.

Claim

If curriculum-aligned and curriculum-relevant through-year assessments increase the number of times student must retrieve information they've learned, then the acquisition and retention of knowledge will increase, and student learning will improve.

Through-year assessments that connect to what students are taught can potentially reduce the impact of disparities in background knowledge on student performance



What the research says

Background knowledge plays a significant role in reading comprehension. Students do not have equitable access to opportunities to develop background knowledge and often are penalized for this when confronted with cold reads that are included in traditional end-of-year summative assessments.

Claim

If curriculum-aligned and curriculum-relevant through-year assessments include material students have seen before, which can decrease the impact of inequitable access to building background knowledge on a student's performance, then student learning will improve (and become more equitable).

Cold reads: texts that students have not read before and are typically unrelated to material taught in class

Warm reads: texts students have not read in class but are typically related to the information and knowledge they encountered in class

Hot reads: texts students have read in class

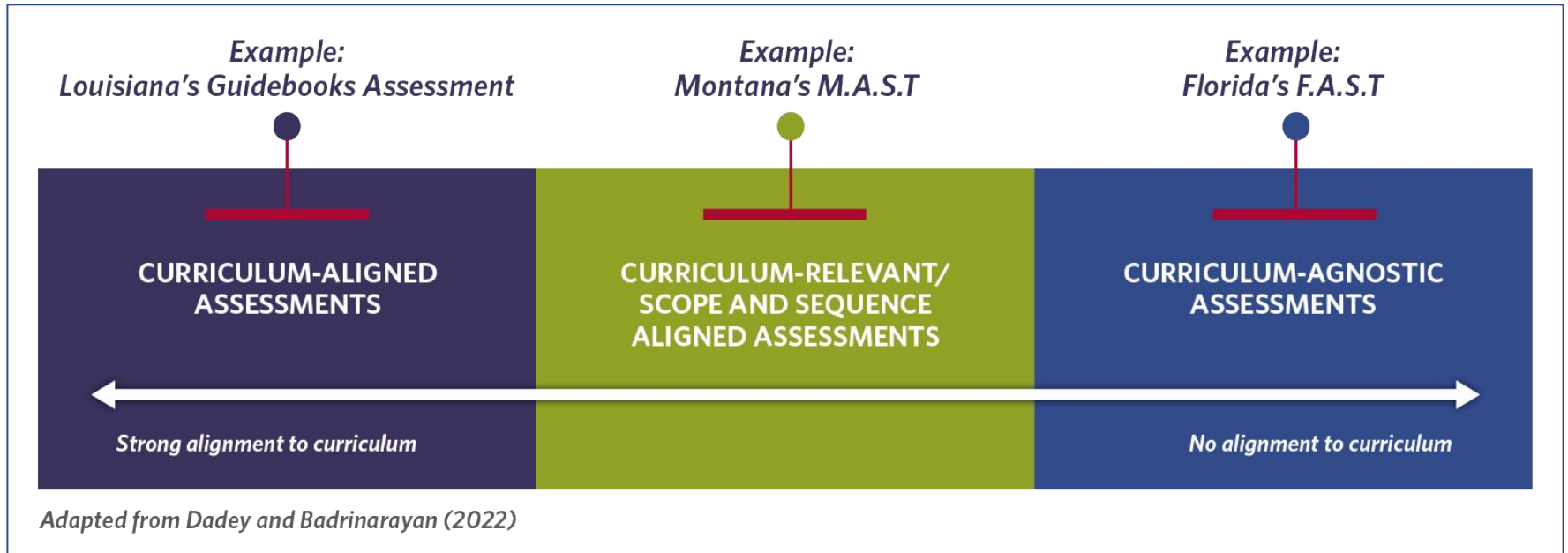
Tensions and Trade-Offs



Different through-year assessment approaches leverage the research differently and connect to instruction in varying ways



Through-year Approaches



Within these three approaches, there are still some tensions and trade-offs that states and assessment developers are working through



Alignment to Curriculum and/or Scope and Sequence:

What is the balance between local control and state assessments?



Timing:

When are students expected to demonstrate mastery of the content they've learned?



Scoring:

Should states measure the retention of knowledge or the acquisition of knowledge?

As through-year assessments become more tightly tied to a curriculum or scope & sequence, both the potential benefits and the logistical challenges may increase



Assessments **more tightly tied** to a curriculum or scope & sequence



The likelihood of **students being tested on material they have learned** in a timely manner

The **logistical challenges** of ensuring students have been taught material before they're tested on it

Curriculum-relevant and curriculum-aligned approaches have potential benefits depending on the level of flexibility they provide and how closely they are aligned with curriculum



Curriculum-relevant approaches

- The ability to transfer the assessments to different curricula which may make scaling statewide potentially more feasible

- Students will be tested on what they have been taught
- Students will be tested in a timely manner
- Instructors will be able to receive actionable data for instructional use

Curriculum-aligned approaches

- The ability to test students on specific material/texts they've been taught
- The ability to minimize the impact of previous background knowledge on test performance

On the other hand, both approaches also surface tensions related to implementation and local control depending on how closely they are aligned with curriculum



Curriculum-relevant approaches

- Districts need flexibility in the timing of when assessments are administered in order to meet their curriculum needs
- If this flexibility is not reached, then students may be tested on material they have not learned yet

- HQIM needs to be followed with fidelity, otherwise students will be tested on materials they have not learned yet

Curriculum-aligned approaches

- Data is needed on what curriculum instructors are using and the fidelity with which instructors are following the curriculum
- Assessments are not easily transferable to other districts not using the same curriculum

States vary on whether they test students on all standards during each administration or a subset of standards



Each administration tests students on *the depth and breadth* of the state standards

Each administration tests students on *a subset* of the state standards

Potential Benefits

The model can be used in all schools within a state, regardless of local curriculum or instructional pacing

Students will be tested on material they have recently been taught

Tensions

Students may be tested on standards that haven't been taught yet

State testing is meant to measure mastery by the end of the year rather than proficiency at the time of testing

States (of the eight identified)

Louisiana, Florida, Nebraska, North Carolina (ELA), Texas

Montana, Delaware, Indiana, North Carolina (Math)

States developing through-year models have different perspectives on whether data should be used to inform instruction, predict and/or contribute to summative scores



Purpose of the Data from Assessments

To inform instruction

**To predict how students
will perform at the end
of the year**

**To contribute to
summative scores for
accountability**

Note that these purposes are not mutually exclusive

States' varying perspectives also drive their design decisions



State Goal

Design Decision

- + Measure acquisition of knowledge and use scores to inform instruction.
- + Students are given multiple opportunities to demonstrate their mastery of knowledge



- + Use a blueprint that is aligned to what students have been taught immediately before administration
- + Use aggregate scores from multiple administrations to calculate summative



For example:
Montana

- + Measure the retention of knowledge and use scores to predict how students will perform on the final summative test
- + Students are only given one opportunity to demonstrate their mastery of knowledge



- + Use a blueprint that incorporates the full scope of the standards during each administration
- + Use end of year administration for calculating summative



For example:
Nebraska

How a state calculates their summative score also underlines fundamental beliefs about what we expect of students



Some believe...

Students should demonstrate mastery of standards after they've been taught the content

Acquisition of knowledge and growth throughout the year is what we care about

Accountability should drive improvement throughout the year

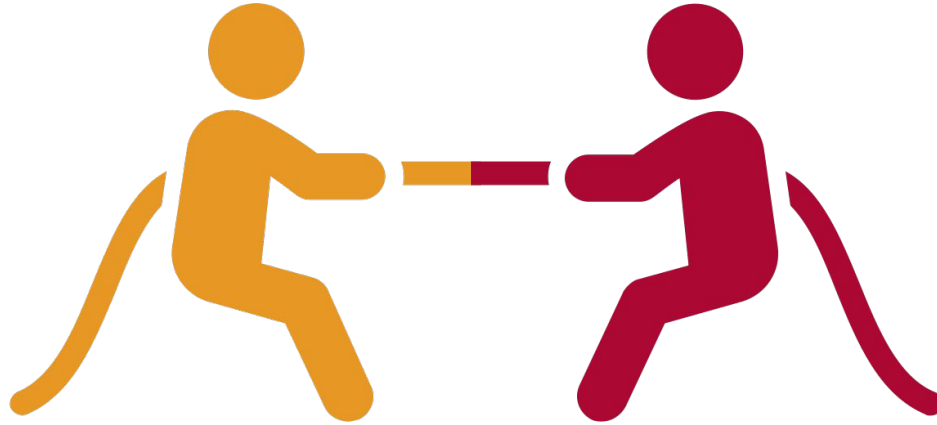
States, members of measurement community, instructors, caregivers, etc

Others believe...

Students should demonstrate mastery of standards at the end of the year

Retention of knowledge is what we care about

Accountability plus interims might distort implementation



Overall, many states are still figuring out how to calculate their summative score and there is a lot of variation among them

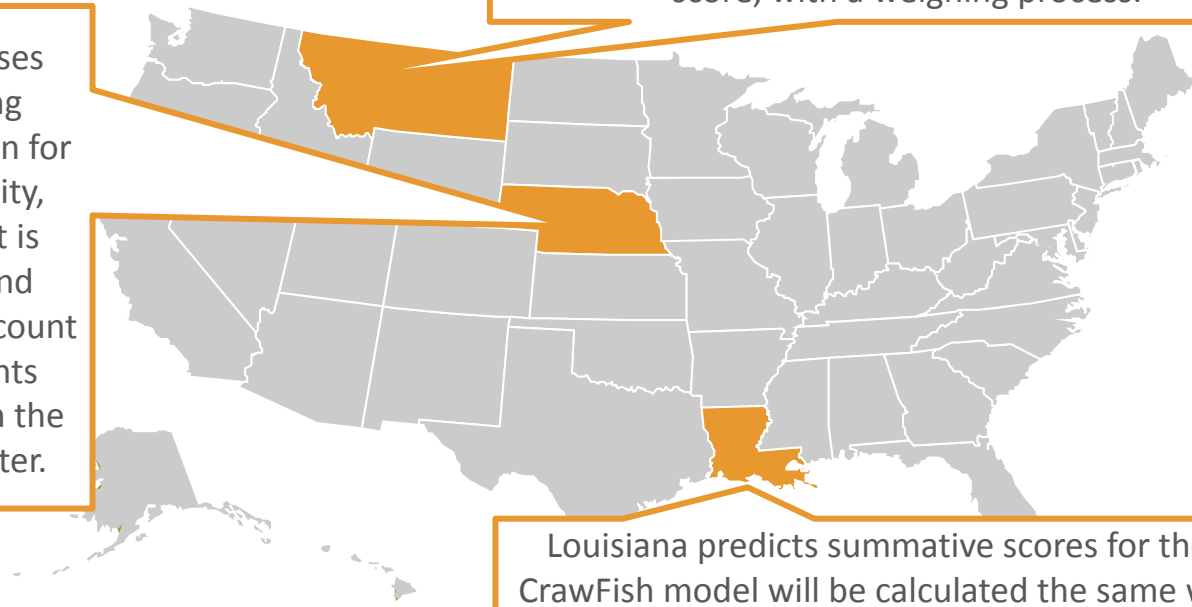


Montana eventually plans to aggregate mini-scale scores from each testlet into an overall summative score, with a weighing process.

Nebraska uses their spring administration for accountability, but the test is adaptive and takes into account how students performed in the fall and winter.

Across the states we reviewed, there is no agreement on how to calculate scores and varying opinions on the implications of those different choices.

Louisiana predicts summative scores for the CrawFish model will be calculated the same way as their curriculum-aligned models—pooling data to estimate scale scores.





While some of these tensions and considerations are unique to through-year models, many would be true for any approach that deviates from a traditional summative test

The summative testing infrastructure in the United States' education system is built around end-of-year assessments.

Changing this system will require more effort and reexamining of fundamental aspects of what is measured, for what purpose and how the data is used.

There are key enabling conditions and lessons learned from states testing these approaches that are worth learning from.

LESSONS LEARNED

WHAT HAVE WE LEARNED ABOUT THE INFRASTRUCTURE, CONDITIONS, RESOURCES AND CAPACITY REQUIRED TO PILOT THROUGH-YEAR ASSESSMENTS?



Piloting through-year assessment models requires certain enabling conditions to be in place, strategic stakeholder engagement and professional learning



Enabling Conditions

What infrastructure and enabling conditions must be in place to pilot and implement a through-year assessment system?

What resources and level of capacity is required to pilot and implement a through-year assessment system?



Stakeholder Engagement

How have states and assessment developers engaged stakeholders and people proximate to the problem? What forms of stakeholder engagement were more effective?

What are stakeholders' views on through-year assessments so far?



Professional Learning

How have states and assessment developers address professional learning for educators?

What have been the biggest challenges in addressing professional learning?

We will cover these three infrastructure components in detail on the following slides

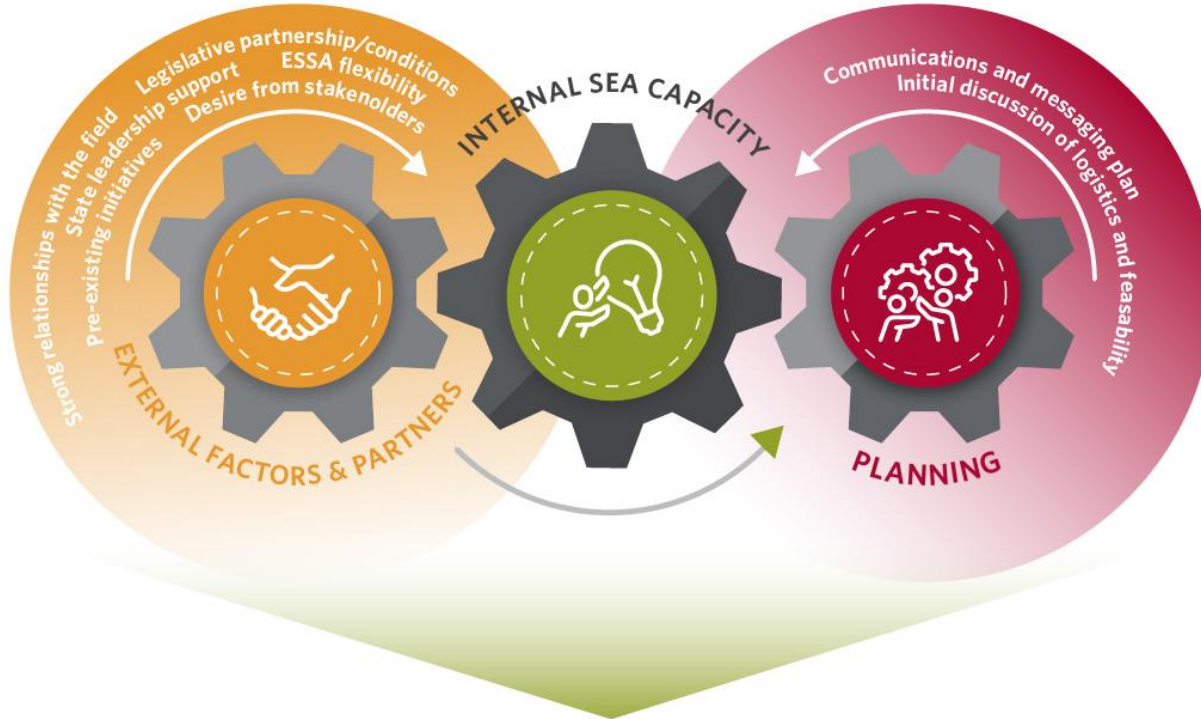
Enabling Conditions





The enabling conditions for through-year assessments fall into three main areas: external factors & partners; internal State Education Agency (SEA) capacity; and planning

Articulating a vision for a future state



Many of the states and assessment developers we spoke to identified these factors as key to their successful pilot and implementation.

We believe these enabling conditions are essential for all states attempting to pilot through-year assessments.

States and developers identified six external factors and partnerships as key enabling conditions that allowed their through-year assessment model to get off the ground



External factors and partnerships identified as key enabling conditions

- 1 ESSA flexibility
- 2 Desire from stakeholders
- 3 Strong relationships with the field
- 4 Legislative partnerships/conditions
- 5 State leadership support
- 6 Pre-existing initiatives

We will describe each one in more depth on the following slides...

ESSA provided states a path for considering through-year assessments as a possibility



1

ESSA flexibility

ESSA created a new academic assessment option for states—the allowance for multiple statewide interim assessments—along with a new flexibility for developing innovative approaches to assessment, the Innovative Assessment Demonstration Authority (IADA). Additionally, if a state continues administering its legacy assessment to all students while piloting an innovative assessment, a state may develop and pilot a new assessment, such as a through-year assessment. For several of the states we talked to, this new option and development pathways were key in allowing them to explore the potential of through-year assessments.

“It came down to ESSA and the flexibility it gave states in designing... we don’t have to have the same test form as the last decade... this set the stage for looking at what can we do?”

- Director of Statewide Assessments, Nebraska Department of Education

“When ESSA was reauthorized, it promoted through-year assessment systems so that helped.”

- Education Associate, Social Studies Assessment, Delaware Department of Education

Read more about the regulatory pathways for through-year assessments in Education First’s publication [“What are Through-Year Assessments?”](#)

Most states considered through-year assessments as a response to stakeholders' frustration with the current summative assessment system and desire for a new test



2

Desire from stakeholders

In some states, stakeholders, including educators and families, expressed the desire for a change in assessments, namely a **need for more timely test results**. A number of states assembled task forces and commissions to originally gather stakeholder perspectives' and the recommendations from many of these groups was a through-year assessment model.

In Texas, the legislature put together a commission on how to improve assessment and accountability and ***“one of the recommendations was to look at replacing the summative with an integrated, formative through-year model.”***

- Department of Assessment and Reporting
Associate Commission, Texas Education Agency

“We heard loud and clear that everyone wanted data in the hands of teachers to help adjust instruction.”

- Section Chief, Test
Development, North
Carolina Department of
Public Instruction

“We heard from parents and teachers over and over again that we don't get results back in a timely fashion.”

- Assistant Deputy Commissioner,
Division of Accountability, Research,
and Measurements, Florida
Department of Education

Collaboration and strong relationships with partners in the field seems to be a necessary enabling condition prior to piloting through-year assessments



3

Strong relationships with the field

State leaders and assessment developers we spoke to discussed collaboration in different ways, including collaboration across state government branches, partnerships with districts, community members and teacher associations, and a partnership with a strong assessment developer.



North Carolina

The state agency's partnership with NC State University has been a key enabling condition. The state develops assessments in close partnership with NC State University, and works collaboratively to design and develop test items and forms.



Delaware

Delaware has a strong social studies coalition with representatives from all districts. **State level leaders' relationships with this coalition allowed them to partner together to roll out a social studies through-year assessment.**

"We didn't move forward until the SS coalition was on board... their leadership and membership were really important."

- Education Associate, Social Studies Assessment, Delaware Department of Education

A strong assessment developer is one of the key partnerships for successful implementation



In Nebraska, having a strong assessment developer was key in building capacity.

The state looked for a developer that had the capacity to handle a state level innovative assessment.

“It came down to which vendor, what the vendor has in their background as far as test design, what it is they can scale up to at the state level. The fact that they [NWEA] already had a national imprint was very important.”

- Director of Statewide Assessments,
Nebraska Department of Education

For some states, guidance or mandates from the state legislature is what made through-year assessments possible in the first place



4

Legislative partnerships/conditions

For five states that we spoke with, legislative conditions played a key role in the transition to through-year assessments.



Indiana

In Indiana, House Enrolled Act 1251 in 2011 legislated the **DOE to streamline and prioritize standards**, and to realign assessments to these prioritized standards.



Texas

The **Texas Legislature passed House Bill 3906 in 2019**, which **mandated changes to the annual standardized tests and directed the SEA to create and pilot an integrated through-year assessment** that would monitor student progress, inform classroom instruction and potentially replace summative assessments. The bill came with a lump sum of annual funds which allowed TEA to pilot their through-year assessment.



Florida

In Florida, having the governor's backing was key. **Governor DeSantis signed Senate Bill 1048 into law in March 2022**, officially **replacing the Florida Standards Assessment with progress monitoring to measure students' growth**.



For other states, support from state leadership is what made through-year assessments possible

5

State leadership support

State leadership support has been a key enabling condition for both assessment developers and SEAs.



Montana and Louisiana



- + **New Meridian described how collaboration with Montana and Louisiana was essential to the success of their pilot, and the importance of the capacity and commitment of both the state leader and staff.** The Montana Office of Public Instruction's dedication to the success of their innovative through-year assessment led them to request a Field Test Flexibility Waiver from USED. Multiple members of Congress and the Montana Governor wrote letters of support for the waiver, and USED approved the waiver in August 2023.
- + NWEA also shared how collaboration with state leaders in Louisiana has been key to their work.

"The leader at the state has to be ready to take on the work of selling their vision. They have to be ready to convince not only their field, but often their own team of the benefits of a new through-year design."

- Senior Vice President, State Partnerships, New Meridian

Pre-existing initiatives can be an especially helpful enabling condition



6

Pre-existing initiatives

In Louisiana, the state built on their previous work investing in HQIM infrastructure to design their Guidebooks through-year assessment and pilot their Innovative Assessment Program (IAP). This made their transition to building the CrawFish through-year model easier.

- + Develops the **ELA Guidebooks 2.0 curriculum**.
- + **Reviews and rates curriculum** for quality.
- + **Incentivizes schools** to adopt high-quality materials.
- + **Vets professional learning** for alignment to high-quality curriculum.

- + Applies for **Innovative Assessment Demonstration Authority (IADA)** to **build a through-year assessment aligned to Guidebooks 2.0**, and later Wit & Wisdom.
- + IADA helped with initial through-year assessment work, the **Innovative Assessment Program**
- + The **IADA application forced a lot of this early development**, and that opportunity crystallised the theory of action.

In addition, the existing infrastructure of the **Innovative Assessment Program** helped significantly with **buy-in and roll-out** as Louisiana already had “curriculum-aligned” assessments that made the transition to the CrawFish model easier. The new through-year model had the same platform, technical supports, manuals and communication.

Internal State Education Agency (SEA) capacity is key for implementing through-year assessments, both at the start of piloting and throughout the process



Strong theory of action and vision
Staffing capacity and knowledge
Internal coordination

Internal SEA capacity factors identified as key enabling conditions

1

Strong theory of action and vision

2

State level personnel and capacity

3

Staff knowledge and expertise

4

Internal coordination

We will describe each one in more depth on the following slides...

A strong theory of action and a vision for the *why* behind the shift to through-year assessments at the SEA level is critical



1

Strong theory of action and vision



*“States have to have a **vision for assessment that needs to prevail through leadership changes.**”* - State Assessment Leader

- + Almost every state and assessment developer we spoke with emphasized the **importance of having a theory of action for the through-year assessment, with clear goals**
- + The theory of action for through-year assessments should fit into the state’s overall vision for assessment

*“You have to have a theory of action. You have to be very clear about what your goal is - **processes have to be in place and have to be well thought out.**”*

- Director of Statewide Assessment, Nebraska

NWEA also recommended that states and assessment developers **“develop a strong theory of action from the start and reflect on it often.”** As they’ve progressed in their work, they’ve had to make numerous decisions and pivot quickly. The theory of action acts as a roadmap to guide their decisions.



"When we talk about implementation [of through-year assessments], we have to consider how implementation may span statewide administration changes and the implications of that. This [implementing through-year assessments] requires buy-in across leaderships."

- State Assessment Leader



State level capacity needs include additional personnel and training to implement through-year assessments



2

State level personnel and capacity

Piloting and implementing through-year assessments involves switching from a system everyone is familiar with to a new way of thinking and operating, as well as dealing with the logistical challenges that come from administering a test multiple times a year. States described how they had to adjust both personnel and ramp up technology capacity.

Developers described how implementing a through-year assessment system is like starting up a new assessment program which requires consistent capacity.

“You need consistent staff who are owning what is needed... There’s a strong vision from the top, which helps, but you need staff.”

- Vice President, Product Strategy, New Meridian

Consultants supporting states highlighted the need to grow state department capacities, especially during transitions between assessment models.

“Coming from a statewide summative assessment, the psychometric and content teams are not big, only a couple people. That model is not going to be enough long term for through-year assessments. It’s a hard reality.”

–Senior Associate, Center for Assessment

We spotlight how two states approached this on the following slides...

Delaware addressed capacity through creating a new position at the SEA level



For their social studies through-year assessment, Delaware created a new position to oversee the curriculum, instruction and professional learning side. This allowed another state leader position to move over to focusing on assessment full time.

“You need somebody on both ends of this. If the intent is to improve instruction, you need someone overseeing instructional resources.”

- Education Associate, Social Studies Assessment, Delaware Department of Education

Florida addressed capacity issues through adding staff and ramping up their technology capacity



Regarding district capacity, Florida shared that **districts were also able to adapt quickly to the required additional technology capacity and ramp up their capacity.**

However, Florida's Assistant Deputy Commissioner also noted that collecting and dealing with student data three times a year, as opposed to one, did stretch some districts' capacity.

*"We did add two FTEs because of the additional grades... **The workload certainly exploded.** For example, our reporting team usually only has to worry about reporting tests once a year and now it's 3 times a year - that's not a small undertaking - the ability to manage test administration, testing system."*

- Assistant Deputy Commissioners, Division of Accountability, Research, and Measurements, Florida Department of Education

*"Another piece was the technology capacity. **We needed to have a pretty robust ramp up to have the devices and do computer based testing in additional grades.**"*

- Assistant Deputy Commissioner, Division of Accountability, Research, and Measurements, Florida Department of Education

While feasibility and logistics can be a challenge for states' capacity, a few states felt that the logistical strain may be less on districts



It is too early to estimate the full impact of through-year assessments on district capacity, but states and assessment developers pointed out that most districts already have structures and systems in place for administering multiple interim assessments a year.

*“We had a lot of people say multiple administrations through the year - everyone will hate it! We haven’t heard these reactions. **Districts and schools know what to do.** It hasn’t been as hair-on-fire to do multiple assessments as people thought it might be.”*

- Senior Vice President, State Partnerships, New Meridian

In Louisiana, NWEA shared they haven't heard concerns from the state about district capacity during their collaboration piloting through-year assessments—but wondered whether districts will consider replacing their current interim assessments with the new state through-year model in the future.

And in certain states, through-year assessments fill a gap for districts that do not have strong interim assessment infrastructures



For some districts, particularly smaller or rural districts, **through-year assessments provide interim assessments they may otherwise not have.**

North Carolina

North Carolina also shared that the NC Check-Ins have received lots of support, especially from small districts. **While large districts have resources to buy benchmark assessments, smaller districts appreciated having these available from the state.**

Texas

In Texas, many of the participating districts were rural districts. The fact that the pilot through-year assessment is aligned to state standards is another incentive for these districts.

For these districts, *“it can take a lot of effort or money for them to develop their own local benchmarks and for that reason they were open to doing a pilot put together by the state.”*

- Strategy and Operations Manager, TEA

CenterPoint partnered directly with districts and found that districts need consistent capacity and want guidance around data use and interpretation



Many districts have the infrastructure in place to transition to a through-year system, but consistent capacity and support with data use are key.

Consistent capacity at the district level is important.

“Some districts may experience changes in leadership or have new staff members for the current year, making it challenging to establish stable collaborations.”

- CenterPoint

Districts want more insight into the data and guidance on interpreting results to inform instruction

CenterPoint found that **districts will engage when they meet them where they are and help them to understand the data they receive.**

Besides additional personnel, having staff with deep knowledge and expertise in assessments is a crucial factor



3

Staff with deep knowledge and expertise

Staff with deep knowledge and expertise in assessments is important in different ways

They may have prior experience with changing assessment systems

Through-year assessments require strong technical knowledge

*“States that don’t have a robust and knowledgeable staff, they’re going to struggle. We have a large assessment unit and most of us have been around for a number of years... We work with our national experts, our other colleagues in other states. **That breadth and depth of knowledge and connections throughout the country made it helpful.**”*

- Assistant Deputy Commissioners, Division of Accountability, Research, and Measurements, Florida Department of Education

*“On the SEA side, we needed **to have someone who knew psychometrics** and who could talk with the vendor about what we need from the state’s accountability piece - we need to ensure it’s a valid test.”*

- Director of Statewide Assessments, Nebraska

*“The amount of items that are written that have to be reviewed, **the amount of embedding that needs to be done** on the current end of grade test to develop a bank to develop all these different assessments while maintaining the current assessments... **it has to be strategic.**”*

- Section Chief, Test Development, North Carolina Department of Public Instruction

Strong internal coordination and coherence between the curriculum & instruction team and the assessment team within an SEA can also be a determinant of success



4

Internal coordination

One of the main goals of through-year assessments is to drive and improve instruction. To achieve this fully, it helps if the curriculum/instructional office and assessment office within the SEA communicate and coordinate, especially for states using curriculum- connected and/or scope and sequence aligned models.

*“The curriculum office and assessment office need time and space together to do the work... **There’s a need for participation and buy-in from multiple arms of the state agency.**”*

- Director, State Innovative Solutions, NWEA

For states considering through-year models that connect more closely to curriculum, thinking through the communication between the curriculum and assessment offices is key. In Louisiana, for their ELA Guidebooks curriculum-aligned model, because ELA Guidebooks is created by the state, the curriculum and assessment offices are in close contact. Both offices understand the curriculum roadmap and the through-year assessment outline.

State Spotlight: Delaware is aiming for strong ties between its instructional and assessments offices



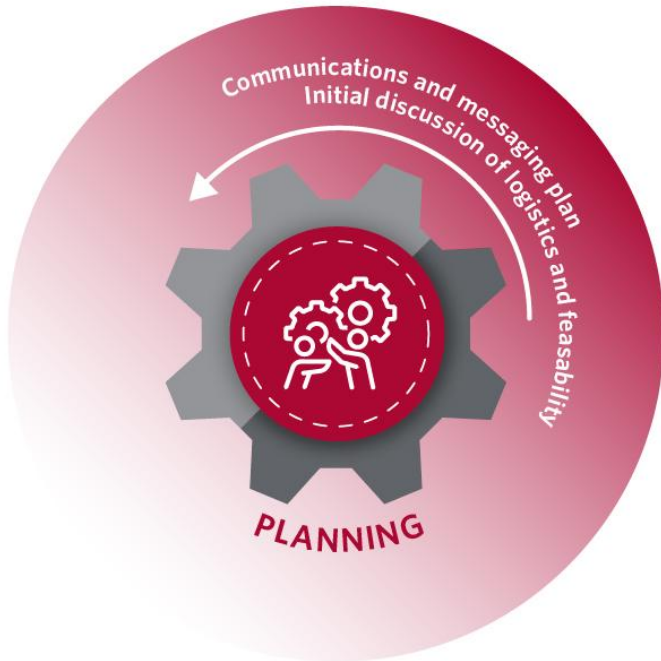
- + With the shift to a social studies through-year assessment, there was a huge push funded by the department to create **instructional resources aligned to the through-year assessments**.
- + The lessons, called Model Lessons, outline the standards on each assessment and include a **planning guide paced to the assessment administrations**.
- + The **through-year assessment development happens at the same time as the instructional resource development** which gives more credibility to the whole system.
- + As mentioned earlier, Delaware created a **new position for curriculum instruction and professional learning** to accompany the shift to through-year assessments, and the two offices coordinate.



*"You can't really think of it as just assessment - **it's a teaching and learning system with assessment at the end to help you see how it's gone** - that's what resonates with folks."*

- Education Associate, Social Studies Assessment, Delaware Department of Education

States have found that a strong communications and logistics plan is a critical enabling condition for articulating a vision for a future state



Planning factors identified as key enabling conditions

1

Initial discussion of logistics and feasibility

2

Communications and messaging plan

We will describe each one in more depth on the following slides...

Initial discussions about logistics, feasibility and the time & commitment needed for through-year assessments can help to level set expectations



1

Initial discussion of logistics and feasibility

It is crucial that SEA staff, district partners and assessment developers understand what implementing a through-year assessment actually means, and the time and effort that is involved. It is also key that states think through the logistics of coordinating with districts, test design, scoring and reporting.

“Level setting about the time scale. These things take longer to stand up, but are potentially more impactful to states and students.”

–Senior Associate, Center for Assessment

Communication must clearly message the purpose of a state's through-year assessment and expectations for how it should be used



2

Communications and messaging plan

States have experienced challenges setting expectations with educators about the fundamental purpose of their through-year assessment and what the assessment does and doesn't do.

States must be clear on if the assessment is meant to be diagnostic, formative, a benchmark, summative—or a combination of these purposes.

"Teachers would want to see [the through-year assessment] as a fully formative tool, but it more so serves the role of a replacement to locally adopted benchmarks that takes a snapshot of student performance throughout the year."

— Strategy and Operations Manager, Texas Education Agency

"One of biggest challenges is what it does and what it doesn't do—what are the purposes and how it's being designed. Some (educators) thought it would roll everything up and it's not. Our colleagues do a really good job of emphasizing the formative aspect."

— Senior Director of Accountability and Testing, North Carolina Department of Public Instruction

Strong communication plans are essential for setting expectations and establishing purpose



States and assessment developers shared best practices for communication plans

- + Communicate constantly, consistently and cyclically
- + Reach all stakeholders at every phase of design and implementation
- + Gather input at every phase
- + Coordinate communication centrally from the SEAs
- + Partner closely with assessment developers and platform providers to develop and deliver communications and technical support
- + Build on existing systems to reach stakeholders where they are expect

Communication within districts presents challenges

While Delaware's social studies coalition was an essential partner, the SEA found members didn't always fully articulate what was happening to other district leadership, exposing assumptions about intra-communication and requiring additional layers of communication.

Celebrate schools and educators piloting assessments

New Meridian made banners for piloting schools and certificates for teachers to honor their role and build buy-in to the vision.

"As pilot participants, helping to shape the next assessment, we know this is not easy."

– Senior Vice President, New Meridian

Stakeholder Engagement



All assessment developers mentioned the importance of stakeholder engagement throughout their through-year pilot process



*“You need to engage with stakeholders and PPP (people proximate to the problem) from the start. **It absolutely needs to happen.**”*

- Director, State Innovative Solutions

*“We place significant value on the **feedback of our stakeholders**, particularly those deeply involved in the teaching and learning process.”*

- CenterPoint

*“Stakeholder engagement and diverse voices from PPP will remain **key drivers of the development process.**”*

- New Meridian

States and assessment developers each defined “stakeholder” differently, although most included students, caregivers and educators



Students

Caregivers

Educators

States and assessment developers also engaged:

- + Principals
- + Counselors
- + Superintendents
- + Administrators
- + Test coordinators
- + Experts in education measurement
- + Accessibility advisory groups
- + Assessment advisory groups

States and assessment developers found that engaging stakeholders in different ways and at different times served a variety of purposes



Types of stakeholder engagement:

- + Focus groups
- + Cognitive labs
- + Empathy interviews
- + Surveys
- + Committees
- + Rapid testing
- + Interviews
- + In-person workshops
- + Quarterly feedback sessions

At the Outset

"TEA didn't have idea of what this should've looked like from very beginning - we very much relied on our stakeholders to inform our work and help us understand what folks are looking for when it comes to TYA" - Strategy and Operations Manager, TEA

At Specific Decision Points

Louisiana and NWEA used feedback from instructors to inform the design of their score reports

At the End of the Pilot

Montana's feedback sessions at the end of their pilot year showed immense improvement in teacher opinions about TYA

Each of the states and assessment developers engaged with stakeholders in their own way, and each plan to continue that engagement



CenterPoint	Louisiana and NWEA	Montana and New Meridian
Educator surveys	Educator focus groups on test design sketches	A week-long, in-person item development workshop with educators
Empathy interviews with students	A student survey on reporting completed by over 13,000 students	Student surveys as part of the second and fourth administrations
<i>To come:</i> Continued engagement with PPP (people proximate to the problem)	<i>To come:</i> Caregiver engagement on reports	<i>To come:</i> Cognitive labs with students

Stakeholder feedback directly impacted the design of their through-year pilots



These are three examples:

Montana and New Meridian

Empathy interviews with educators and school leaders



The creation of culturally responsive questions for indigenous students in Montana

Louisiana and NWEA

Educator focus groups on test design sketches



The funnel design being chosen for Louisiana's CrawFish design

Montana and New Meridian

Educator focus groups



The continued development of the configurator tool & the reporting suite reporting

Despite the value of stakeholder feedback, assessment developers found structural barriers to be in the way of engaging with students and parents



Student Engagement

Gathering student feedback is challenging due to privacy and legal concerns and the legality of those interactions varies by state.

However, surveys were an effective tool for gathering significant input from large numbers of students.

Parent & Student Engagement

Recruiting parents and students for more direct interactions—such as empathy interviews—was a persistent challenge—and developers are still figuring out how to do this more effectively

*“Recruitment efforts by the states, social media marketing, direct compensation, and contact with parent/student groups yielded **minimal leads for students and parents that are willing to talk about their experiences with assessments.**”*

- New Meridian

Overall, stakeholder engagement shows an excitement for through-year assessments and a dissatisfaction with the current assessment system



Excitement for Through-Year Assessments

*“Curriculum-aligned assessments are **more representative of what students are actually learning**, more helpful tools for teachers, they could use the results – [assessments] would be less discouraging.”*

- District educator who worked with CenterPoint

*“The focus [groups] were **generally excited about and interested in the testlet system.**”*

- Focus group on Louisiana’s math pilot with New Meridian

*“The recent Student Survey from Spring 2022 uncovered that a large percentage of the students **preferred the Innovative Assessment to the regular LEAP 2025 assessment (66% to 34%).***

*For those students who preferred the IAP, many called out that **the preference was related to the fact that the test was administered after the unit of study (37%).**”*

- Louisiana’s Innovative Assessment Program, data shared by NWEA

Overall, stakeholder engagement shows an excitement for through-year assessments and a dissatisfaction with the current assessment system



Dissatisfaction with the Current Assessment System

For example, New Meridian gathered feedback from instructors about their current end-of-year assessment systems:

- “1. **Too much time** is dedicated to formal assessment in the classroom.*
- 2. Assessments given throughout the year and at the end of the year are **not well aligned to classroom instruction.***
- 3. Assessments **cause stress** for students.*
- 4. **Educators want to see measures of growth** just as much as they want to see proficiency.”*

- New Meridian

Professional Learning



“One instructional coach told us he’s finally able to have real conversations with teachers about what is happening in their classroom.”

- Education Associate, Social Studies Assessment, Delaware



States and developers provided professional learning on through-year assessments in two main areas



1

Developing districts', educators' and staff's understanding of the assessment (the design, structure, administration guidance, the testing platform)

2

Supporting educators to use the data from through-year assessments to inform teaching and learning

Both areas of professional learning are important to the success of through-year assessments, and states are in different phases of development and success with each area.

1

States have approached professional learning on understanding through-year assessments in different ways



Webinars

A number of states have conducted webinars on the what and how of their through-year assessment



Online modules

Some states have created independent online modules and video trainings for teachers on the test design and reports



Contracting with assessment developers

Some states contract with assessment developers to handle professional development and train educators and distinct staff

Supporting educators to actually use the data from through-year assessments to inform instruction is an evolving area



Most states are still figuring out how to address this type of professional learning. Some states are trying to support educators' use of data by producing detailed test reports.



Delaware has tried to create a reporting system that gets at the information and grain-size teachers need. **The test reports break down results by summary, by standard and individual student reports** so that teachers can dive into the data. The instructional resources from the state aim to actually help teachers do something about the results.



Nebraska has revamped their test report to **communicate more information on proficiency and create actionable reports with detailed data.**



Montana has plans to develop and launch a trainer-led, multi-session, in-person PD on data-driven instruction and develop a teacher community of practice for learning from each other.

States face a number of challenges, both logistical and strategic, in implementing professional learning around through-year assessments



Lack of capacity

- + **Many states face capacity issues to create and implement professional development** and often have to contract out with partners or other organizations.
- + In some states, the professional development piece is expected to be handled by the instructional/curriculum side of the state agency.

Level of fidelity in districts

- + **District implementation and capacity can pose a challenge as well.**
- + Districts have different levels of fidelity in training for teachers, particularly when it comes to the data and interpretation side of professional learning.

Finding time for professional learning

- + Teacher and substitute shortages make finding time for in-person professional learning more challenging
- + Many schools struggle to find time for educator professional learning communities to discuss data

Ensuring all teachers access professional learning

- + **Ensuring that professional learning and training reaches every teacher is a challenge.** In North Carolina, despite multiple avenues of professional learning, an end of year survey revealed that a handful of teachers had not received training.

In sum, we have learned that having the right enabling conditions, engaging stakeholders continuously and establishing a professional learning infrastructure are key



Enabling Conditions

States, along with their assessment developers, must think through certain external factors and partnerships, ensure they have strong internal SEA capacity, and carefully consider their communication and logistics plans.



Stakeholder Engagement

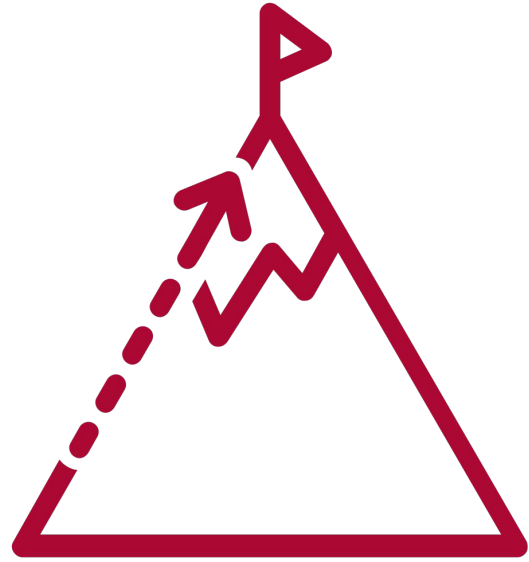
States and developers must engage stakeholders, including educators, students and families, at multiple points in time and in multiple ways, throughout the pilot and implementation.



Professional Learning

While states have provided professional learning on understanding what through-year assessments are, meeting districts, educators, families and students' desire for support in interpreting results to inform instruction is an emerging focus.

Common Challenges



Across the models piloted in our grant and other states, there were a few common challenges and areas for further exploration in through-year test design



Testing students on content they haven't been taught yet



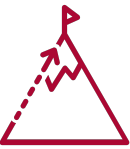
Knowing which curricula districts are using



Negotiating issues of local control



In some states' through-year pilots, students were tested on content and skills they had not been taught yet, which required additional communication to stakeholders



In their initial pilot, due to logistical considerations, **New Meridian was not able to let districts and schools choose the order of testlets.** In math particularly, many students were likely tested on material they hadn't learned yet.

For states who are doing **full-scope through-year models, the first administration presents material teachers have not covered yet.** States such as Florida have tried to address this by messaging that this is a byproduct of the nature of the design and that results will likely be lower in the first administration, but messaging has been a challenge.

In North Carolina, the SEA tried to group the standards in a way that made sense instructionally and surveyed educators on the order they typically teach and group standards. However, **due to local control of curriculum, the order the state through-year groups standards may be different from a locality.**

Delaware is one state that has addressed this challenge head on



In Delaware, **state leaders worked with the social studies coalition for the 8th grade history assessment to decide which time period each assessment covered.** That agreement allowed the through-year assessments to actually cover content students had learned.

For states considering curriculum-aligned or curriculum-relevant models, determining the curricula districts use has posed a challenge



There are two parts to this challenge:

1

Many states do not have a systematic way of collecting data on which curricula districts use.

2

Districts and individual educators implement curriculum with different levels of fidelity—and states have little visibility on this variation.

Example: As Louisiana attempted to launch their CrawFish model, they came upon this challenge. The state initially considered five curricula to incorporate into their CrawFish model, but **struggled to collect data and understand the landscape of curriculum use in the state**. It also wasn't clear if the curricula were being implemented “at the level of fidelity required to scale a curriculum-aligned or curriculum-relevant assessment.”

States intending to pursue a through-year assessment related to curriculum should consider implementing an annual systematized data collection of curriculum use in the state.

For states considering curriculum-aligned or curriculum-relevant models, negotiating local control of curriculum is an ongoing issue



Our education system in this country is designed on the premise that districts have local control over their curricular choices.

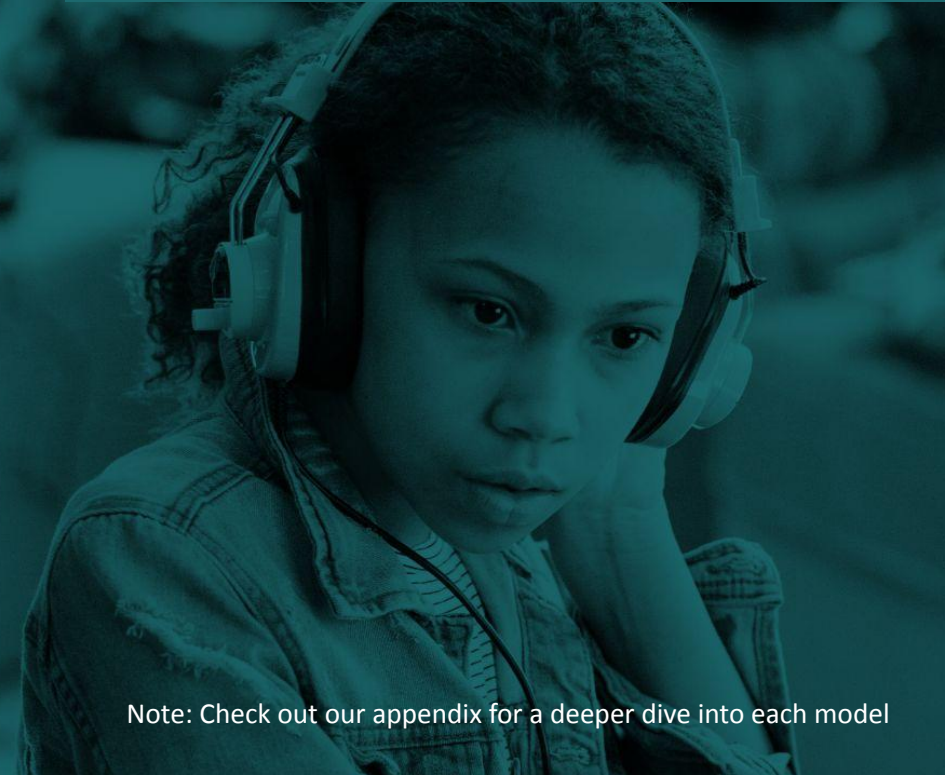
For states trying to design a through-year assessment model that is curriculum-aligned or curriculum-relevant, this poses a challenge. **They must figure out a way for the through-year assessment model design to work with any curriculum used by districts in that state.**



Louisiana is the state in the country best positioned to implement a curriculum-connected through-year assessment, as at least 85% of the state uses two curriculum. However, even in Louisiana, they face challenges in creating a model that is applicable for all districts.

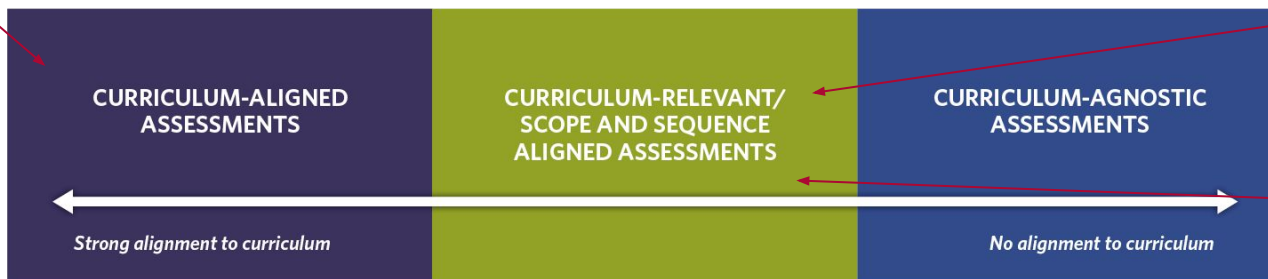
EXPLORING THE MODELS

HOW HAVE STATES AND ASSESSMENT DEVELOPERS APPROACHED THE CHALLENGE OF ALIGNING THROUGH-YEAR ASSESSMENTS TO CURRICULUM AND/OR SCOPE AND SEQUENCES?



Note: Check out our appendix for a deeper dive into each model

For the past year, we funded three assessment developers to pursue curriculum-connected, through-year assessments



Adapted from Dadey and Badrinarayan (2022)

In this section, we provide a deep dive into these three models and cover the following questions:

1



What is the developer trying to accomplish?

2



How is the model accomplishing this?

3



What did the developer learn so far and how is that influencing their plans moving forward?

Through-Year:
CenterPoint's
Curriculum-Aligned
Interim
Assessments



CenterPoint's Curriculum-Aligned Interim Assessments are based on the premise that students can effectively demonstrate their knowledge through assessments that are closely tied to instruction

Overview

CenterPoint is conducting research on the **feasibility of a curriculum-aligned through-year model aligned to Illustrative Math (IM)**

Model

Curriculum-Aligned

Goals

CenterPoint aims to answer the following research questions in this study:

- Do the CenterPoint IM Interim assessments measure similar learning outcomes as summative assessments?
- Do the CenterPoint IM Interim assessments predict students' performance on summative assessments?

Partners

CenterPoint collaborated with two K-12 urban districts located in distinct regions of the U.S. collectively serving around 160,000 students. Both districts serve diverse student populations.

What problem is this model solving?

CenterPoint believes that students can effectively demonstrate their knowledge through assessments that are closely tied to curriculum and instruction

CenterPoint aims to create a curriculum-aligned through-year assessment tied to the Illustrative Mathematics curriculum



Model Overview

- CenterPoint chose the Illustrative Mathematics (IM) curriculum for math assessments because it is comprehensive and rated green on Ed Reports. CenterPoint's IM-aligned interim assessments have received certification from the IM organization.
- The assessments are designed to be administered three times a year (fall, winter, spring) to provide a comprehensive picture of students' progress within the curriculum.
- The current study has thus far focused its efforts on the middle school grades.
- The ultimate goal is to broaden the scope of the study and its reach to include multiple states to include CenterPoint's interims in kindergarten through high school These assessments include:
 - 3-4 interims per grade level
 - Each interim takes approximately 50 minutes
 - Each interim includes a variety of item types and cognitive complexity levels



CenterPoint conducted a study to determine the feasibility of generating a summative score using their IM-aligned interims



Study Design

This study aimed to determine the concurrent validity between the two assessments (to what degree they measure the same outcome).

- They analyzed results from two district's CenterPoint IM interim assessments against those districts' summative results from 2021-2022 school year using an IRT (item response theory) model.
- This included placing the results from both the summative and IM assessments on a common scale to determine the degree to which the results are correlated.
- The objective was to analyze the “thetas,” or student abilities, in their performance for both the interim and summative assessments

Findings

The initial findings demonstrated varying levels of correlation between CenterPoint's Illustrative Mathematics Curriculum-Aligned Interim Assessments and the summative assessments of the two urban districts. The analysis **“offer initial evidence that the CenterPoint IM interim assessments have the potential to yield similar summative outcomes.”**

Dr. Hong Jiao, Professor at the University of Maryland, College Park collaborated with CenterPoint on the psychometric analyses of this study

Moving forward, CenterPoint has plans to increase the scope of their study and further explore growth measures and scale scores



CenterPoint's Future Research Plans are to:

Conduct a similar study involving additional testing grades for mathematics

Extend this study to other CenterPoint Curriculum-Aligned Interim Assessments

Explore the possibility of using standard setting to create a scale score similar to summative assessments, allowing for a **more detailed comparison between interim and summative assessments**

Investigate the potential for **measuring growth over time**

Consider the practicality of linking items between interim and summative assessments to **establish a meaningful connection between content and skills**





Louisiana hopes assessing students on content related to what they were taught will promote deeper engagement with texts and reduce disparities generated from students' varying levels of background knowledge

Overview

In partnership with Louisiana, NWEA created the CrawFish model so that the state's through-year assessments for ELA could **work for two HQIM: Guidebooks and Wit & Wisdom**

Model

Curriculum-Relevant

Goals

Louisiana and NWEA aims to accomplish the following goals through their work:

1. Provide the same level of, or better, information to educators (instructionally-oriented and practical, actionable information)
2. Incentivize deep engagement in the material and texts throughout the year

Partners

Louisiana Department of Education (LDOE): 23 districts; John Hopkins University, Center for Assessment, Odell Education

What problem is this model solving?

Create a curriculum-relevant, through-year assessment that aligns to all of the possible curricula within the state and provides useful information and results tied to curricula.



The CrawFish Funnel honors background knowledge while connecting to different curriculum

1 Unique, unit-based “Knowledge” sections

Hot read: Students see different texts here based on the curriculum they are using (e.g. from Guidebooks). All texts are thematically related.

Example: A text about the Civil Rights movement that students have already seen

2 A common thematically related warm read text and item set in the “Application” section

Warm read: All students see the same text here. It is thematically related to the text from section #1.

Example: A new text on the Civil Rights movement that students have not seen before

3 A common writing prompt in the “Synthesis” section

All students see the same prompt here. It is thematically related to the texts from sections #1 and #2.

Example: A question is posed about the Civil Rights movement

The scoring design will match what is currently used in the Innovative Assessment Program:
Data is pooled across units and scale scores are estimated using an IRT model



Louisiana plans to administer a full-year pilot during the 23-24 school year and expand the curricula included in the CrawFish Funnel

In the 2023-2024 school year, the CrawFish Funnel pilot will be expanded to:

- A full-year pilot with three windows of administration
- Continue with Grade 5
- Connect to additional HQIM

The goals of the expanded pilot are to:

Recruit additional schools and additional curricula for the CrawFish funnel

Continue to refine report prototypes

Continue understanding how assessment data can be used to inform instruction

**MasteryGuide
Assessments**





New Meridian's flexibly administered testlets aim to provide actionable data to inform instruction throughout the year

Overview

New Meridian has created an assessment system of **short, modular “testlets”** in ELA and math **designed to be flexibly administered** across several administrations and **align with local scope and sequences**

Model

Scope and sequence aligned

Goals

- To create a ***coherent, continuous and useful assessment to better meet the goals of assessment for learning*** by administering frequent, mini-assessments aligned to local curriculum that provide **actionable data** to inform instruction throughout the year
- To replace traditional EOY statewide assessments and interims

Partners

Louisiana Department of Education (LDOE)

- Grades 5 & 7, math

Montana's Office of Public Instruction (OPI)*

- Grades 5 & 7, math & ELA

What problem is this model solving?

New Meridian's user research found the following problems with the current assessment system:

- Traditional statewide assessments return data too late to provide instructional value
- Interim assessments don't align to content taught and pull away from a coherent curriculum plan



New Meridian is using this pilot database to inform the initial development of a summative scoring model

The model would generate a **mini-scale score based on a student's performance on an individual testlet**, and then **aggregate all the mini-scale scores into an overall summative score**.



The model will **weigh scores** based on the timing of the testlets' administration and their content.

New Meridian plans to develop summative and predictive scoring models to help Montana's OPI incorporate testlets into the state accountability system by the 2024-2025 school year.



New Meridian learned that users want expanded reports and actionable data to inform instruction

Concerns in initial pilot



Changes moving forward

User Experience



Users were generally positive about the functionality and usability of the platform. **Montana educators liked the student reports they received but requested expansions to the reporting system** including more details about the testlets and a classroom view of the data. States and district administrators requested reports at the school, district and state levels.



New Meridian is taking this feedback into account for the expanded pilot next year.

The Importance of Actionable Data



In this pilot, the testlets' order was predetermined and did not match some schools' scope and sequences. As a result, **students may have taken a math testlet assessing material they had not yet learned. This reinforced the importance of reporting actionable data in a timely manner**, as teachers do not want to waste time on assessments that provide information they cannot use.



The ability to tailor the administration of testlets to the local scope and sequence in a school system is important. This has influenced New Meridian's plans going forward, as we will describe in a following slide.



Inconsistency around test timing and the difficulty of test items had led New Meridian to make revisions

Concerns in initial pilot

Changes moving forward

Testlet Length



The amount of time estimated for a student to start and finish a test doesn't include all the administration time that is involved (e.g. logging in). **The testlets took longer than expected for some and the messaging around timing was inconsistent.** While New Meridian estimated that math testlets would take 10-15 minutes, OPI messaged to schools that it would take no more than 10 minutes, which led to inaccurate time expectations on teachers' end.

New Meridian is working to rapidly revise forms and plans on generating a timing estimate that includes non-assessment time.

Test Item Difficulty



Diagnostic analysis following each administration showed that most students were placed into certain skills profiles: **either showing they mastered every skill or no skills.** As a result, New Meridian recalibrated the assessment after the pilot closed, with more items in ELA and math. In ELA, New Meridian also adjusted the difficulty of the texts as student performance was closely tied to passage complexity. In math, New Meridian did a deep dive review of current test items to adjust the difficulty of items across testlets.

The math content team is updating guidelines for item writers to decrease the number of difficult items while still including enough of a range of difficulty to provide actionable feedback.



Moving forward, New Meridian has plans for an expanded pilot in both Montana and Louisiana

In the 2023-2024 school year, New Meridian plans to expand the MasteryGuide Assessment pilot to



- Grades 3, 5, 6 and 7
- Math



- Grades 3 - 8
- Math & ELA

The goals of the expanded pilot are to:

Continue to develop the configurator tool to help districts align the ELA and math testlets to their unique scope and sequence

Improve usability based on user experience data

Continue to compare MasteryGuide's results with current summative assessments

Create additional items (in grades 3-8) to prepare for an operational field test in the 2024-2025 school year

Continue to develop expanded reporting tools

WHERE DO WE GO NEXT?

IMPLICATIONS AND RECOMMENDATIONS
FOR THE FIELD



There is a compelling evidence base that supports connecting what is taught to what is tested



Research supports the hypothesis:
Through-year assessments that connect more closely to what students are taught have the potential to improve student learning.

*This includes through-year models that are directly aligned with a specific curriculum (**curriculum-aligned**) or that can be flexibly aligned with multiple curricula, a scope and sequence or pacing of learning (**curriculum-relevant**).*



These models have the **potential to solve for concerns with summative assessment** related to:

- utility for instruction
- disparity in background knowledge
- incoherence between assessment, curriculum and instruction.

There are tensions and trade-offs with different approaches, but through-year assessments have the potential to *reduce* overall testing



While the assessments may require more time over the course of the year...



...it may provide an opportunity for local education agencies to **limit or supplant their interim, benchmark and diagnostic assessment systems** and reduce overall testing during the year...



...while also **providing more timely and useful data** back to educators.

States and assessment developers, like the work of **NWEA and New Meridian in Louisiana and Montana**, are piloting this approach.

State leaders shared recommendations for others considering piloting through-year assessments

Start with and communicate with your stakeholders

Consider assessment developer capacity and alignment

Make sure you fully understand what through-year assessment models entail

Be clear on your theory of action and goals

Ensure you have buy-in from relevant stakeholders

Create a coalition

The Center for Assessment outlines [Ten Key Considerations](#) for states considering through-year assessment

Check out Education First's [Through-Year Assessment Toolkit](#) to explore tools and resources to navigate the change

Words of wisdom from a state leader on starting on a path towards through-year assessments...

“Is the purpose you have identified and the use you’ve identified—is it worth getting there? Are you creating purpose that actually brings value to the experiences of teachers and students?... Make sure you have listened to the folks that count, the ones who will have to live with this system.”

Director of Statewide Assessment, Nebraska

As through-year models scale and integrate into accountability systems, we recommend states and districts consider reducing duplicative testing and aligning intended purposes with the tests used

States, their partners and districts must consider the ways that a through-year summative system should be situated within a balanced assessment system.

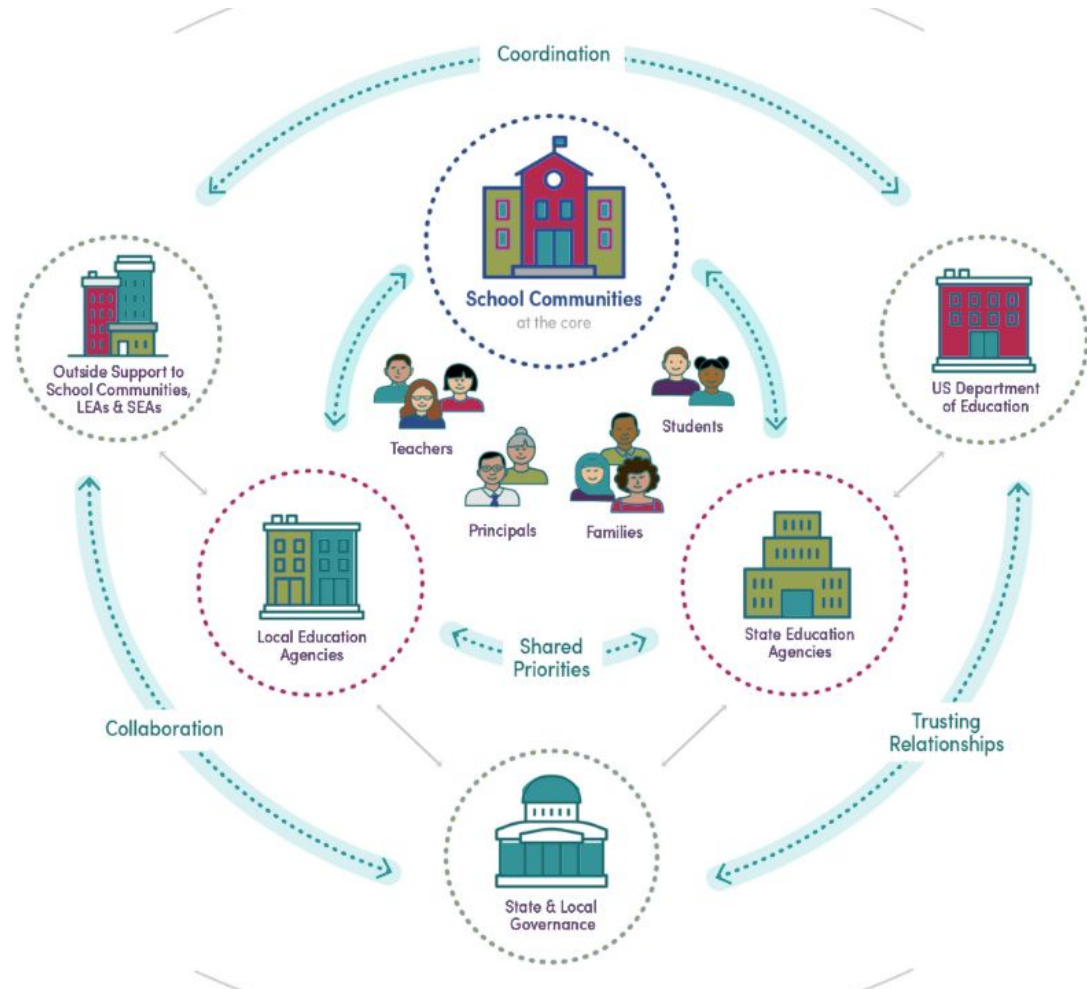
Depending on the design and approach that a state is taking with their through-year model, **what, if any, additional benchmark or interim assessments are needed at the district level?**

If the data provided from the through-year system yield actionable and timely results that support the same purposes, **can districts reduce the amount of overall testing that students and teachers are experiencing?**

What supports will schools, districts and educators need **to use the data from through-year models to bolster instruction?**

States and their partners must also focus on **clear, coherent and systematic implementation** in a way that **builds and deepens buy-in of stakeholders**.

Iterating on the **test design, utility, reporting and information with key stakeholders** including teachers, parents and students can **ensure buy-in through the change process**.



Overall, the field still has a lot to learn about the impact of and process of designing and implementing through-year assessments

It is too early to tell the degree to which each model or approach will improve student outcomes. **Each state needs to define the problems they are solving for and develop the model that best meets their local contexts.**

For states interested in:

Testing what students are taught closer to the point of instruction

Providing timely results that support teachers to use the data from the tests to inform instruction

Creating balance within an assessment system and reducing the overall footprint of testing over the course of the year

The research and lessons learned from states described in this report support aligning through-year assessments to curriculum and/or scope and sequence.

No state has yet undergone federal peer review to operationalize their through-year assessment system

In partnership with Foresight Law & Policy, we conducted research to **explore how the peer review process is set up to accommodate innovations in assessment.**

- + Our findings indicate that **there are opportunities for state leaders to pursue more innovative approaches within the current requirements**, but **ambiguities in these requirements undermine confidence** that innovative approaches will pass peer review.
- + The Peer Review Guide **does not address the multiple assessments approach** allowed by ESEA, nor provide examples of evidence for such an approach.

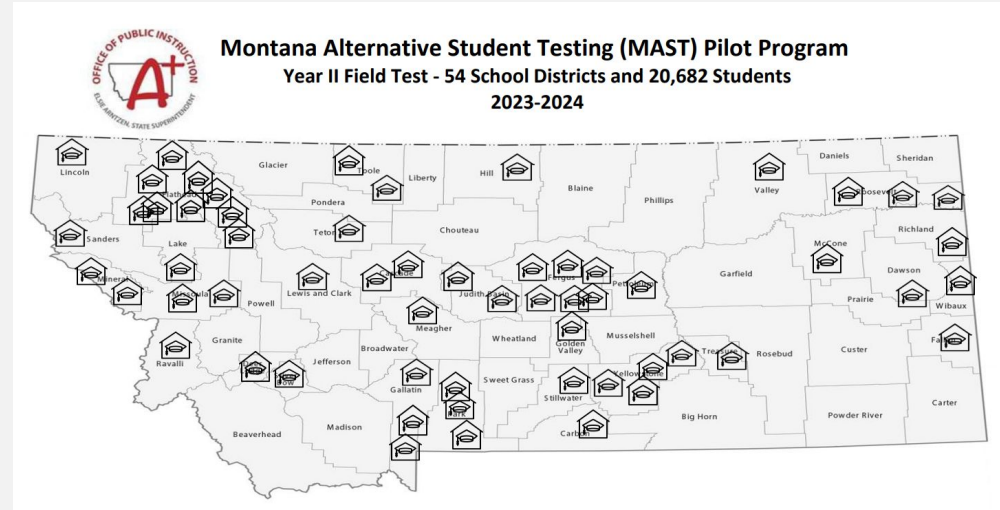


We submitted five recommendations to USED to better position peer review for innovative assessments

- 1 Elevate and clearly signal the path states can take for innovative approaches by communicating what is possible under currently law
- 2 Update the Guide, including updating or changing the examples of evidence and clarifying ambiguities
- 3 Increase opportunities for engagement between Department staff and state leaders
- 4 Recruit, select and assign peer reviewers who are experts on proposed innovative assessment systems
- 5 Integrate tutorials on innovative assessment systems into current training process

However, Montana submitted and were approved for a Field-Testing Flexibility Waiver from the U.S. Department of Education

- Montana's OPI submitted the waiver in May 2023 and was approved in August 2023
- **The waiver ensures that students, teachers and district leaders participating in the Montana Alternative Student Testing Pilot Program (MAST) are not overburdened with double testing during the 2023 - 2024 school year**
- Montana received public support from a number of education-based advocacy groups
- The Department approved Montana's waiver because of how the through-year assessment is **“expressly designed to provide educators with more frequent and timely feedback on their instruction”** and they **“determined that this waiver will advance student academic achievement.”**



Read the full letter from the Department of Education [here](#)

We suggest additional research and future efforts focus on the following key questions

- What types of professional learning would support educators in using the data from through-year assessments to drive instruction?
- What behaviors might be incentivized as a result of the shift to incorporating accountability into testing throughout the year?
- What changes in the federal peer review process, guidance and examples of evidence can support states transitioning to through-year assessments?
- What types of supports will families, students and policymakers need to understand, interpret and use new types of data and reports throughout the year?

Ed First plans to continue partnering with states implementing through-year assessments and supporting policy changes needed

In our next phase of work, we plan to:

Facilitate a community of practice of states implementing through-year assessment models

Deepen understanding of trade-offs, tensions and promises of new summative models

Host a convening for state leaders, assessment developers and experts in the field on through-year assessments

Continue sharing our learnings and thought leadership with the field

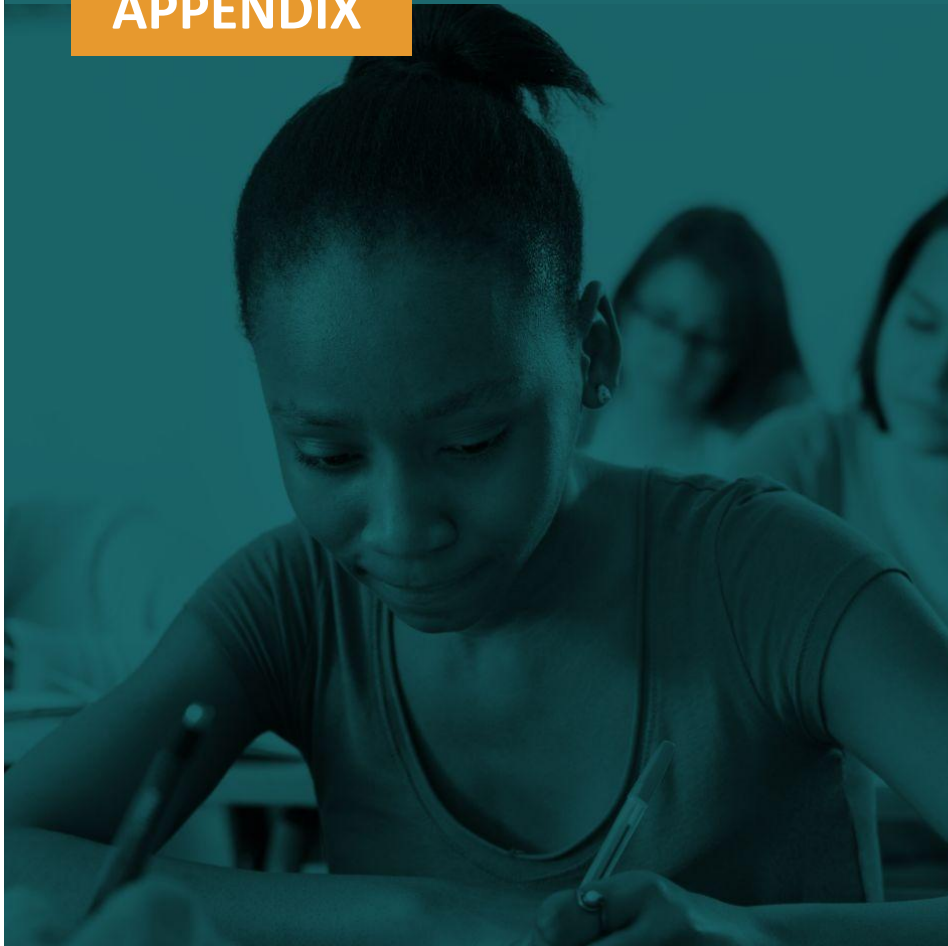
Acknowledgements

We want to thank the following individuals for their time and expertise and sharing their learnings with us:

- Elizabeth Gilbert, Jim Mirabelli, Laura Slover and Melissa Peplinski, CenterPoint Education
- The Center for Assessment team
- Dusty Shockley and Michael Feldman, Delaware Department of Education
- Sheila Briggs, Amanda Kim, Fatima Omar and Jennifer Vranek, Education First
- Vince Verges, Florida Department of Education
- Lynn Schemel, Indiana Department of Education
- David Steiner, Johns Hopkins University
- Thomas Lambert, Louisiana Department of Education
- Krystal Smith and Sharyl Allen, Montana's Office of Public Instruction
- Trudy Clark, Nebraska Department of Education
- Ashley Eden, Kristopher John and Chris Gentile, New Meridian
- Tammy Howard and K. Maxey-Moore, North Carolina Department of Public Instruction
- Leslie Mugan, NWEA
- Elizabeth Davis, Odell Education
- Iris Tian, Jamie Kwan and Jose Rios, Texas Education Agency



APPENDIX



Sources



Sources



Cromley, J. & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Education Psychology*, 99(2), 311-325. doi:10.1037/0022-0663.99.2.311\

Dadey, N., Evans, C. M., & Lorie, W. (2023). Through-year assessment: Ten key considerations. The National Center for the Improvement of Educational Assessment.
<https://www.nciea.org/wp-content/uploads/2023/03/Ten-Key-Considerations-Through-Year-Assessment-Report-March-2023-F.pdf>

Dadey, N., and Badrinarayan, A. (2022, April 21). In Search of the “Just Right” Connection Between Curriculum and Assessment: Considering Options Between Curriculum-Specific and Curriculum-Agnostic State Assessment. Center for Assessment. <https://www.nciea.org/blog/in-search-of-the-just-right-connection-between-curriculum-and-assessment/>

Dadey, N., & Gong, B. (2023). An introduction to considerations for through-year assessment programs: purposes, design, development, evaluation [Research report]. Retrieved from Smarter Balanced website:
<https://portal.smarterbalanced.org/library/en/2023-sb-consideration-of-technical-issues.pdf>

Davis, M. H., & Karunathilake, I. (2004). The adaptive curriculum. *Medical teacher*, 26(6), 501–503.
<https://doi.org/10.1080/01421590412331285388>

Sources



Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2018). What makes for effective feedback: Staff and Student Perspectives. *Assessment & Evaluation in Higher Education*, 44(1), 25–36. <https://doi.org/10.1080/02602938.2018.1467877>

Education First. (2022). What are through-year assessments?: Exploring multiple approaches to through-year design. Retrieved from <https://www.education-first.com/wp-content/uploads/2023/01/What-are-Through-year-Assessments-1.pdf>

Karaman, P. (2021). The effect of formative assessment practices on student learning: A meta-analysis study. *International Journal of Assessment Tools in Education*, 8(4), 801–817. <https://doi.org/10.21449/ijate.870300>

Krupat, E., & Dienstag, J. L. (2009). Commentary: Assessment is an educational tool. *Academic medicine : journal of the Association of American Medical Colleges*, 84(5), 548–550. <https://doi.org/10.1097/ACM.0b013e31819f7fb9>

Marion, S. (2021, September 29). Trying to Serve Multiple Uses with Through Year Assessments. Center for Assessment.

Roediger, H. L. (2014). How tests make us smarter. *The New York Times*.

Sources



The National Academies Press. (2001). Implications of the New Foundations for Assessment Design. In *Knowing What Students Know: The Science and Design of Educational Assessment* (pp. 177–220). chapter.

Shepard, L. A., Penuel, W. R., & Davidson, K. L. (2017). Design principles for new systems of assessment. *Phi Delta Kappan*, 98(6), 47–52. <https://doi.org/10.1177/0031721717696478>

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021) The Role of Background Knowledge in Reading Comprehension: A Critical Review, *Reading Psychology*, 42:3, 214-240, DOI: 10.1080/02702711.2021.1888348

Wiggins, G. (2012, September). Seven keys to effective feedback. *Educational Leadership*, 70(1), 10-16. Retrieved August 22, 2023, from <http://www.ascd.org/publications/educational-leadership/sept12/vol70/num...>

Zhou, Y., Ye, X., & Liu, Y. (2022). The influence of personalized learning intervention system on student learning a study of junior middle school. *Interact. Technol. Smart Educ.*, 19, 441-459.

Through-Year Curriculum-Connected Assessment Grant Program Theory of Action



We set out at the beginning of the project with the following theory of action



If we

Assemble an advisory group of state and district leaders to work with assessment and curriculum vendors to engage PPP and begin research and design efforts for through-year prototypes

Undertake a research agenda in partnership with technical experts, researchers, thought leaders and PPP to share answers to critical questions needed for implementation of through year models

And...

Rapidly test through-year prototypes and methods that are explicitly designed to generate a single summative score for each student

We set out at the beginning of the project with the following theory of action



Then...

Policy and advocacy organizations will have new information to improve policy and conditions to enable use of new assessments in accountability systems

Assessment vendors can develop new scope and sequence and curriculum-connected solutions that can be administered throughout the year

States and districts will have more choices in the marketplace

Educators, students and parents will derive more value from the assessments

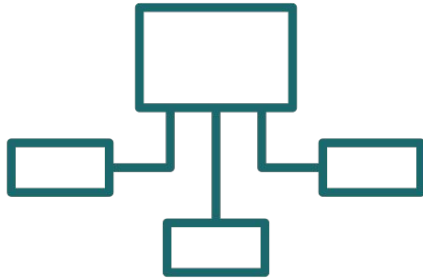
Which...

Can address PPP's concerns about utility of current assessments, inadequate score reporting and difficulty using data to inform instruction

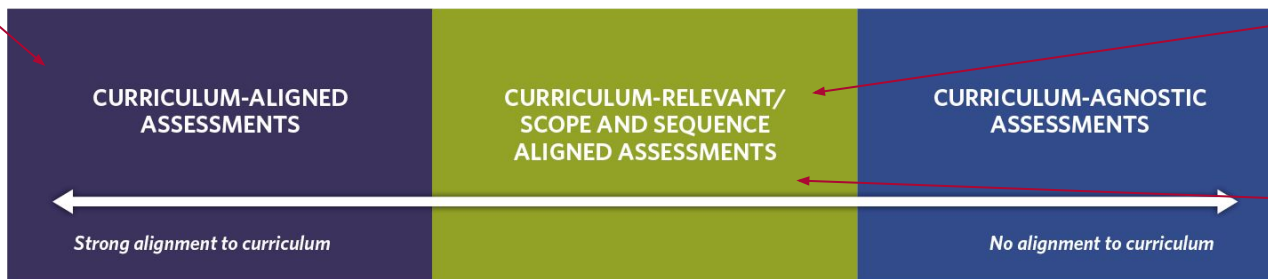
Which will contribute to...

Policymakers and the public having more accurate information by which to compare and assess student achievement across schools and districts, and inform future authorization of EESA

Exploring the Models in Depth



For the past year, we funded three assessment developers to pursue curriculum-connected, through-year assessments



Adapted from Dadey and Badrinarayan (2022)

In this section, we provide a deep dive into these three models and cover the following questions:

1



What is the developer trying to accomplish?

2



How is the model accomplishing this?

3



What did the developer learn so far and how is that influencing their plans moving forward?

Through-Year:
CenterPoint's
Curriculum-Aligned
Interim
Assessments



CenterPoint's Curriculum-Aligned Interim Assessments are based on the premise that students can effectively demonstrate their knowledge through assessments that are closely tied to instruction

Overview

CenterPoint is conducting research on the **feasibility of a curriculum-aligned through-year model aligned to Illustrative Math (IM)**

Model

Curriculum-Aligned

Goals

CenterPoint aims to answer the following research questions in this study:

- Do the CenterPoint IM Interim assessments measure similar learning outcomes as summative assessments?
- Do the CenterPoint IM Interim assessments predict students' performance on summative assessments?

Partners

CenterPoint collaborated with two K-12 urban districts located in distinct regions of the U.S. collectively serving around 160,000 students. Both districts serve diverse student populations.

What problem is this model solving?

CenterPoint believes that students can effectively demonstrate their knowledge through assessments that are closely tied to curriculum and instruction



CenterPoint's study addresses the challenge of potential misalignment between summative assessments and instruction in the classroom

The Challenge

There is a **potential misalignment between summative assessments and a curriculum's scope and sequence**, instruction and the measurement of those learnings by interim assessments. This exemplifies the **tension between when and how often students show mastery**. Educators try hard to navigate the tension by supporting and valuing each of the assessments in their suite. Yet, the substantial emphasis on the single end-of-year measure remains a hurdle.



The Solution

By analyzing the CenterPoint Illustrative Mathematics Interim assessments alongside summative assessments, CenterPoint observed commonalities in intended measures and outcomes, spanning standards, domains and results. This enables them to assess predictability for the summatives and **investigate the potential of the interim assessments to become the primary indicator of student learning**. Assisting educators in understanding these interrelationships, coupled with guidance in navigating this understanding, will help educators effectively assist their students.

CenterPoint aims to create a curriculum-aligned through-year assessment tied to the Illustrative Mathematics curriculum



Model Overview

- CenterPoint chose the Illustrative Mathematics (IM) curriculum for math assessments because it is comprehensive and rated green on Ed Reports. CenterPoint's IM-aligned interim assessments have received certification from the IM organization.
- The assessments are designed to be administered three times a year (fall, winter, spring) to provide a comprehensive picture of students' progress within the curriculum.
- The current study has thus far focused its efforts on the middle school grades.
- The ultimate goal is to broaden the scope of the study and its reach to include multiple states to include CenterPoint's interims in kindergarten through high school These assessments include:
 - 3-4 interims per grade level
 - Each interim takes approximately 50 minutes
 - Each interim includes a variety of item types and cognitive complexity levels



CenterPoint conducted a study to determine the feasibility of generating a summative score using their IM-aligned interims



Study Design

This study aimed to determine the concurrent validity between the two assessments (to what degree they measure the same outcome).

- They analyzed results from two district's CenterPoint IM interim assessments against those districts' summative results from 2021-2022 school year using an IRT (item response theory) model.
- This included placing the results from both the summative and IM assessments on a common scale to determine the degree to which the results are correlated.
- The objective was to analyze the “thetas,” or student abilities, in their performance for both the interim and summative assessments

Findings

The initial findings demonstrated varying levels of correlation between CenterPoint's Illustrative Mathematics Curriculum-Aligned Interim Assessments and the summative assessments of the two urban districts. The analysis **“offer initial evidence that the CenterPoint IM interim assessments have the potential to yield similar summative outcomes.”**

Dr. Hong Jiao, Professor at the University of Maryland, College Park collaborated with CenterPoint on the psychometric analyses of this study



As CenterPoint moves forward, they are considering how to better meet students' and educators' needs in their design choices

In response to feedback from people most proximate to student learning, CenterPoint is considering some design changes in their assessments. Some of the potential design changes include:



Providing students **more opportunities to demonstrate their modeling and reasoning skills**, further assessing the breadth of the curriculum

Offering more **choice in the types of topics or context related to mathematics**

Building in **additional opportunities for culture and community**

Developing **additional professional learning opportunities to support educators and students**

Moving forward, CenterPoint has plans to increase the scope of their study and further explore growth measures and scale scores



CenterPoint's Future Research Plans are to:

Conduct a similar study involving additional testing grades for mathematics

Extend this study to other CenterPoint Curriculum-Aligned Interim Assessments

Explore the possibility of using standard setting to create a scale score similar to summative assessments, allowing for a **more detailed comparison between interim and summative assessments**

Investigate the potential for **measuring growth over time**

Consider the practicality of linking items between interim and summative assessments to **establish a meaningful connection between content and skills**





Louisiana hopes assessing students on content related to what they were taught will promote deeper engagement with texts and reduce disparities generated from students' varying levels of background knowledge

Overview

In partnership with Louisiana, NWEA created the CrawFish model so that the state's through-year assessments for ELA could **work for two HQIM: Guidebooks and Wit & Wisdom**

Model

Curriculum-Relevant

Goals

Louisiana and NWEA aims to accomplish the following goals through their work:

1. Provide the same level of, or better, information to educators (instructionally-oriented and practical, actionable information)
2. Incentivize deep engagement in the material and texts throughout the year

Partners

Louisiana Department of Education (LDOE): 23 districts; John Hopkins University, Center for Assessment, Odell Education

What problem is this model solving?

Create a curriculum-relevant, through-year assessment that aligns to all of the possible curricula within the state and provides useful information and results tied to curricula.



The CrawFish Model seeks to find a balance between curriculum-aligned and curriculum-agnostic through-year assessments in order to scale the solution statewide

Curriculum-Aligned

Louisiana created **curriculum-aligned** ELA through-year assessments for Guidebooks and Wit & Wisdom



The Challenge

Curriculum adoption varies across the state, so **curriculum-aligned models can not be used statewide**. It would be too expensive and impractical to develop curriculum-aligned models for all high-quality ELA curriculum.

- Many school systems adopt **more than one high-quality curriculum**
- Student movement between districts during the year can mean exposure to multiple curriculum.



Envisioned Solution

Develop an **ELA curriculum-relevant assessment** by drawing on common domains, topics, and texts found across curricula that could be used with any ELA curriculum—allowing Louisiana to **scale statewide** while **still incentivizing deep engagement with texts and reducing reliance on background knowledge**.

- Will complement or replace current curriculum-aligned assessments.

Louisiana planned to use “hot reads” of common texts across curriculum and “warm reads” of texts on related topics to achieve their goals



Hot reads allow students to engage with texts on the assessment which they have already studied in class, reducing the disparities generated by students' varying levels of background knowledge

Warm reads allow students to apply the common background knowledge on a topic generated from hot reads to a new text on the same topic

Hot reads: texts students have read in class

Warm reads: texts students have not read in class but are topically related to the information and knowledge they encountered in class



However, a curriculum analysis revealed that there were few common texts and topics across the five most commonly used curricula in Louisiana from which to build the CrawFish Model

From the curriculum analysis...

- **No texts were shared across all five curricula.**
 - While some texts were shared between two or three curricula at Grades 3 and 4, there was no overlap in Grade 5 for all five curricula.
- **Some topics were shared across the curricula**, such as marine animals and emotions, but overwhelmingly there were minimal shared topics.
- **Very little was shared across the curricula at sub-topic levels.**

This analysis was conducted by Dr. David Steiner and his team at the Institute for Education Policy of John Hopkins University.



As they couldn't use common texts across curricula, NWEA created two prototypes to find a statewide assessment model that honors background knowledge in a different way

A solution needs to:

- Align to all of the possible curricula within the state
- Provide useful information and results

CrawFish Funnel

The assessment builds background knowledge *based on shared themes across curricula*

Build Together Knowledge Model

The assessment builds background knowledge *within the test and draws on texts and learning from other subjects, such as social studies and science*



Louisiana and NWEA chose the CrawFish Funnel because educators supported how it retains the commitment to using background knowledge established in ELA classrooms

A solution needs to:

- Align to all of the possible curricula within the state
- Provide useful information and results



CrawFish Funnel

- The assessment uses background knowledge from different texts found in individual curriculum around a shared theme
- Educator feedback raised challenges with connecting an ELA assessment to background knowledge built in other subjects, as the Build Together Knowledge Model proposed.



The CrawFish Funnel honors background knowledge while connecting to different curriculum

1 Unique, unit-based “Knowledge” sections

Hot read: Students see different texts here based on the curriculum they are using (e.g. from Guidebooks). All texts are thematically related.

Example: A text about the Civil Rights movement that students have already seen

2 A common thematically related warm read text and item set in the “Application” section

Warm read: All students see the same text here. It is thematically related to the text from section #1.

Example: A new text on the Civil Rights movement that students have not seen before

3 A common writing prompt in the “Synthesis” section

All students see the same prompt here. It is thematically related to the texts from sections #1 and #2.

Example: A question is posed about the Civil Rights movement

The scoring design will match what is currently used in the Innovative Assessment Program:
Data is pooled across units and scale scores are estimated using an IRT model



Despite roadblocks, Louisiana and NWEA prototyped one administration of the CrawFish model with almost 400 students as part of its broader ELA through-year pilots



Table 2: Sampling Design

Administration	Total Actual Count	Form A #	Form A %	Form B #	Form B %	Form C #	Form C %
WW	1,413	608	43	608	43	198	14
Guidebooks	2,422	1,114	46	1,114	46	194	8

“Form C” represents the CrawFish model prototype



Louisiana plans to administer a full-year pilot during the 23-24 school year and expand the curricula included in the CrawFish Funnel

In the 2023-2024 school year, the CrawFish Funnel pilot will be expanded to:

- A full-year pilot with three windows of administration
- Continue with Grade 5
- Connect to additional HQIM

The goals of the expanded pilot are to:

Recruit additional schools and additional curricula for the CrawFish funnel

Continue to refine report prototypes

Continue understanding how assessment data can be used to inform instruction

**MasteryGuide
Assessments**





New Meridian's flexibly administered testlets aim to provide actionable data to inform instruction throughout the year

Overview

New Meridian has created an assessment system of **short, modular “testlets”** in ELA and math **designed to be flexibly administered** across several administrations and **align with local scope and sequences**

Model

Scope and sequence aligned

Goals

- To create a ***coherent, continuous and useful assessment to better meet the goals of assessment for learning*** by administering frequent, mini-assessments aligned to local curriculum that provide **actionable data** to inform instruction throughout the year
- To replace traditional EOY statewide assessments and interims

Partners

Louisiana Department of Education (LDOE)

- Grades 5 & 7, math

Montana's Office of Public Instruction (OPI)*

- Grades 5 & 7, math & ELA

What problem is this model solving?

New Meridian's user research found the following problems with the current assessment system:

- Traditional statewide assessments return data too late to provide instructional value
- Interim assessments don't align to content taught and pull away from a coherent curriculum plan



New Meridian’s math testlets are organized by standard clusters and designed to align to local scope and sequence

Model Features

- The math model includes **12 individual testlets aligned to instructionally coherent clusters of standards**. The current design for the 2023 - 2024 year includes 14 testlets.
- Each testlets primarily assesses a single content strand. New Meridian designed items to assess different levels of cognitive depth within each strand.
- The math testlets are designed to be **spread across multiple administration windows**, and the number of testlets administered per window depends on districts’ local scope and sequence.

Initial Pilot

The math testlets assess content that could be taught in a variety of different sequences. **In this prototype, the order of administration was predetermined due to technical and logistical issues.**

This pilot only had machine-scored test items.



Future Plans

In future piloting, **educators will be able to choose from a bank of testlets and administer them in a schedule that works for their instructional timing.**

Future piloting will include **at least one constructed response math performance task** in a fifth administration, and the final design will include several performance task-based testlets.



The math testlets are designed to align to high-quality curricula and can be administered flexibly

- The math testlets each assess a single content strand and assess content that could *“reasonably be taught in a variety of different sequences.”* (New Meridian)
- Districts and schools will be able to choose the order of testlet administration to align with the order of content in their scope and sequence

Example: Aligning with Eureka

Module 1 of the Eureka 5th grade math scope and sequence focuses on place value and decimal fractions. Schools following this scope and sequence could administer the **Place Values- Power of 10** and **Place Values- Represent, Compare & Round** testlets as their first test administration at the end of this module.



Future Plans

The first unit of the Illustrative Math 5th grade scope and sequence focuses on Finding Volume. Schools following this scope and sequence could administer the **Units of Measurement** and **2 Dimensional Shapes** testlets as their first administration at the end of this module.



New Meridian's ELA testlets feature four administrations with increasingly complex text



Model Features

- ELA testlets assess **standards-based reading comprehension skills with increasingly complex passages and tasks.**
- In this pilot, there were **four sets of paired testlets given in four designated administration windows** - each testlet includes 1-2 passages
- The **testlets increase in textual complexity and in the skills assessed as the year progresses.**

Initial Pilot

In this pilot, all ELA test items were machine-scored.

New Meridian worked with the states to identify culturally relevant passages. In Montana, New Meridian included texts by Indigenous and rural authors.







Future Plans

Future piloting will include three sets of paired testlets administered in at least three separate windows, plus a writing-task based testlet

Future piloting will include **the development of multiple testlets with different textual complexities** for each administration to provide more flexibility for educators.



The ELA administrations increase in complexity of text and analysis throughout the year, following a developmental cognitive theory of acquisition of reading skills¹

Administration 1	Administration 2	Administration 3*	Administration 4
<ul style="list-style-type: none">▪ Includes readily accessible texts with explicit ideas▪ Evaluating single text elements 	<ul style="list-style-type: none">▪ Includes moderately complex texts that require light inferences▪ Some synthesis of texts 	<ul style="list-style-type: none">▪ Repeat of passages in Administration 1▪ Purpose is to measure progress 	<ul style="list-style-type: none">▪ Includes moderately to highly complex texts and items that require inferencing▪ Synthesizing texts 

*This describes the design for the small-scale pilot in the 2022-2023 school year. In the next pilot phase, the testlet in administration 3 will be a writing-task testlet



New Meridian is using this pilot database to inform the initial development of a summative scoring model

The model would generate a **mini-scale score based on a student's performance on an individual testlet**, and then **aggregate all the mini-scale scores into an overall summative score**.



The model will **weigh scores** based on the timing of the testlets' administration and their content.

New Meridian plans to develop summative and predictive scoring models to help Montana's OPI incorporate testlets into the state accountability system by the 2024-2025 school year.



New Meridian learned that users want expanded reports and actionable data to inform instruction

Concerns in initial pilot



Changes moving forward

User Experience



Users were generally positive about the functionality and usability of the platform. **Montana educators liked the student reports they received but requested expansions to the reporting system** including more details about the testlets and a classroom view of the data. States and district administrators requested reports at the school, district and state levels.



New Meridian is taking this feedback into account for the expanded pilot next year.

The Importance of Actionable Data



In this pilot, the testlets' order was predetermined and did not match some schools' scope and sequences. As a result, **students may have taken a math testlet assessing material they had not yet learned. This reinforced the importance of reporting actionable data in a timely manner**, as teachers do not want to waste time on assessments that provide information they cannot use.



The ability to tailor the administration of testlets to the local scope and sequence in a school system is important. This has influenced New Meridian's plans going forward, as we will describe in a following slide.



Inconsistency around test timing and the difficulty of test items had led New Meridian to make revisions

Concerns in initial pilot

Changes moving forward

Testlet Length



The amount of time estimated for a student to start and finish a test doesn't include all the administration time that is involved (e.g. logging in). **The testlets took longer than expected for some and the messaging around timing was inconsistent.** While New Meridian estimated that math testlets would take 10-15 minutes, OPI messaged to schools that it would take no more than 10 minutes, which led to inaccurate time expectations on teachers' end.

New Meridian is working to rapidly revise forms and plans on generating a timing estimate that includes non-assessment time.

Test Item Difficulty



Diagnostic analysis following each administration showed that most students were placed into certain skills profiles: **either showing they mastered every skill or no skills.** As a result, New Meridian recalibrated the assessment after the pilot closed, with more items in ELA and math. In ELA, New Meridian also adjusted the difficulty of the texts as student performance was closely tied to passage complexity. In math, New Meridian did a deep dive review of current test items to adjust the difficulty of items across testlets.

The math content team is updating guidelines for item writers to decrease the number of difficult items while still including enough of a range of difficulty to provide actionable feedback.



To address local scope and sequence, New Meridian plans to allow districts to select the order in which they administer testlets

In the next pilot, New Meridian plans to include a **configuration tool** that will allow schools/districts to schedule testlets to fit their local scope and sequence. **With this tool, schools and districts could choose the order of testlet administration.**

- The tool will **allow schools to enter the curriculum they use and would make a recommendation for the order of testlet administration**
 - The curriculum entered would need to be an identifiable enacted curriculum
- New Meridian is currently collecting curriculum data
- Focus groups with math and ELA educators to gather initial feedback on the prototype configurator tool helped determine what information educators want to see when scheduling testlets and provided insight into the reports educators need to see





Moving forward, New Meridian has plans for an expanded pilot in both Montana and Louisiana

In the 2023-2024 school year, New Meridian plans to expand the MasteryGuide Assessment pilot to



- Grades 3, 5, 6 and 7
- Math



- Grades 3 - 8
- Math & ELA

The goals of the expanded pilot are to:

Continue to develop the configurator tool to help districts align the ELA and math testlets to their unique scope and sequence

Improve usability based on user experience data

Continue to compare MasteryGuide's results with current summative assessments

Create additional items (in grades 3-8) to prepare for an operational field test in the 2024-2025 school year

Continue to develop expanded reporting tools

THANK YOU!

