

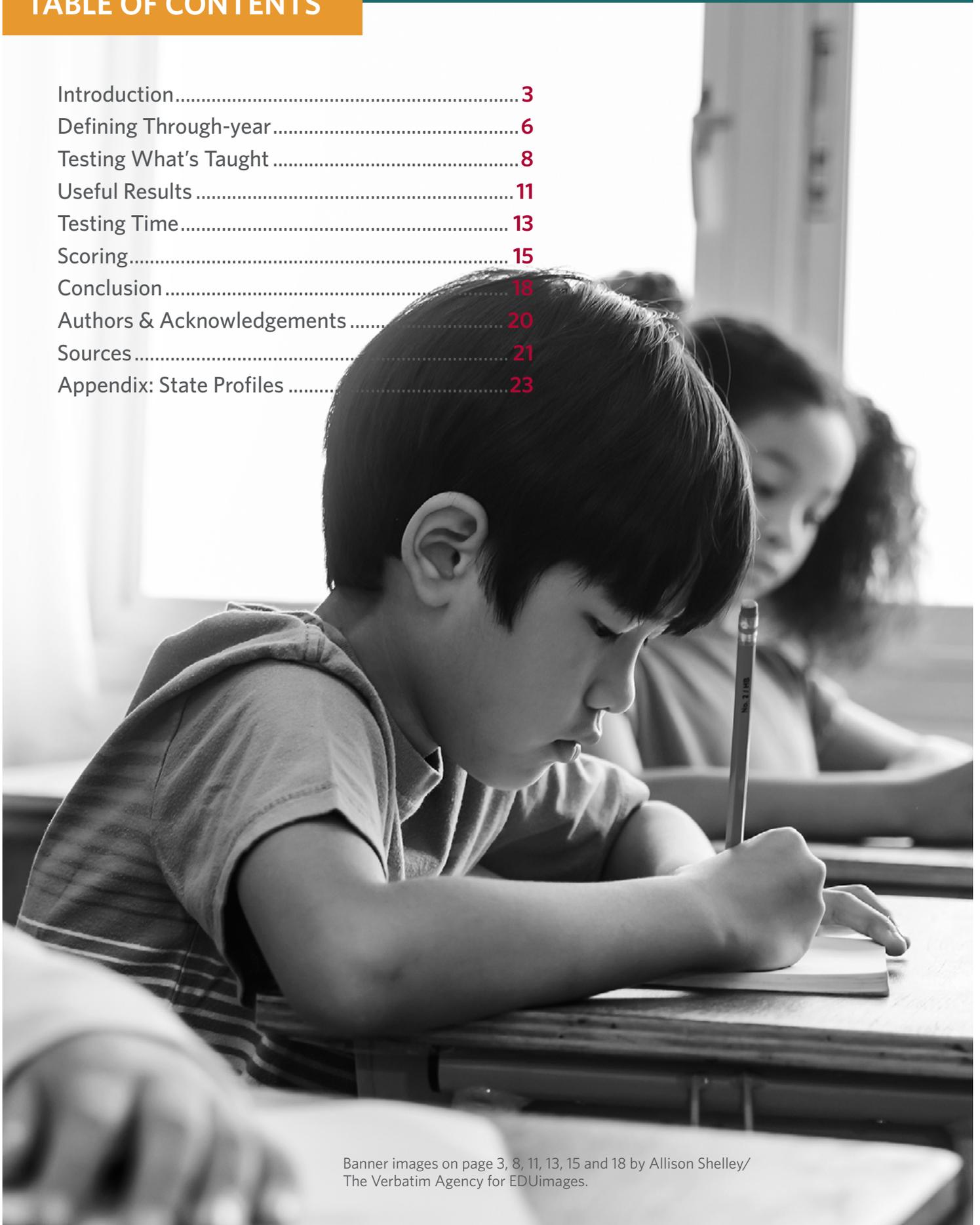


WHAT ARE THROUGH-YEAR ASSESSMENTS?

EXPLORING MULTIPLE APPROACHES
TO THROUGH-YEAR DESIGN

TABLE OF CONTENTS

| | |
|----------------------------------|-----------|
| Introduction..... | 3 |
| Defining Through-year..... | 6 |
| Testing What’s Taught | 8 |
| Useful Results | 11 |
| Testing Time..... | 13 |
| Scoring..... | 15 |
| Conclusion..... | 18 |
| Authors & Acknowledgements | 20 |
| Sources | 21 |
| Appendix: State Profiles | 23 |



Banner images on page 3, 8, 11, 13, 15 and 18 by Allison Shelley/
The Verbatim Agency for EDUimages.

INTRODUCTION



In recent years the number of states launching research and development initiatives on through-year assessments has multiplied. At their core, through-year assessments differ from states' traditional summative assessments by measuring student performance multiple times over the school year instead of during a single sitting close to the end of the year.

This change is driven by a shifting belief about what the purpose of state assessments should be—that state assessments can and should support instruction in addition to system evaluation. In other words, a belief that state tests can be both assessments OF learning and assessments FOR learning. States hope the administration of tests throughout the year¹ creates opportunities to address long-standing, legitimate concerns expressed by students, families and educators about traditional end-of-year summative assessments' inability to support teaching and learning because they are:

- + Disconnected from curriculum and instruction,
- + Provide results that do not inform instruction, and
- + Require undue time and resources

Education First believes these through-year assessment systems have the potential to be more equitable, focused and relevant for students, families and educators. In particular, we are interested in exploring the ways through-year models can strengthen the connection between assessment and instruction by timing assessments of learning immediately following relevant instruction or even aligning directly with curriculum. We refer to this as “testing what is taught, when it’s taught.”

¹Marion, S. (2021) identifies three common goals through-year assessments aim to address in alignment with these concerns.

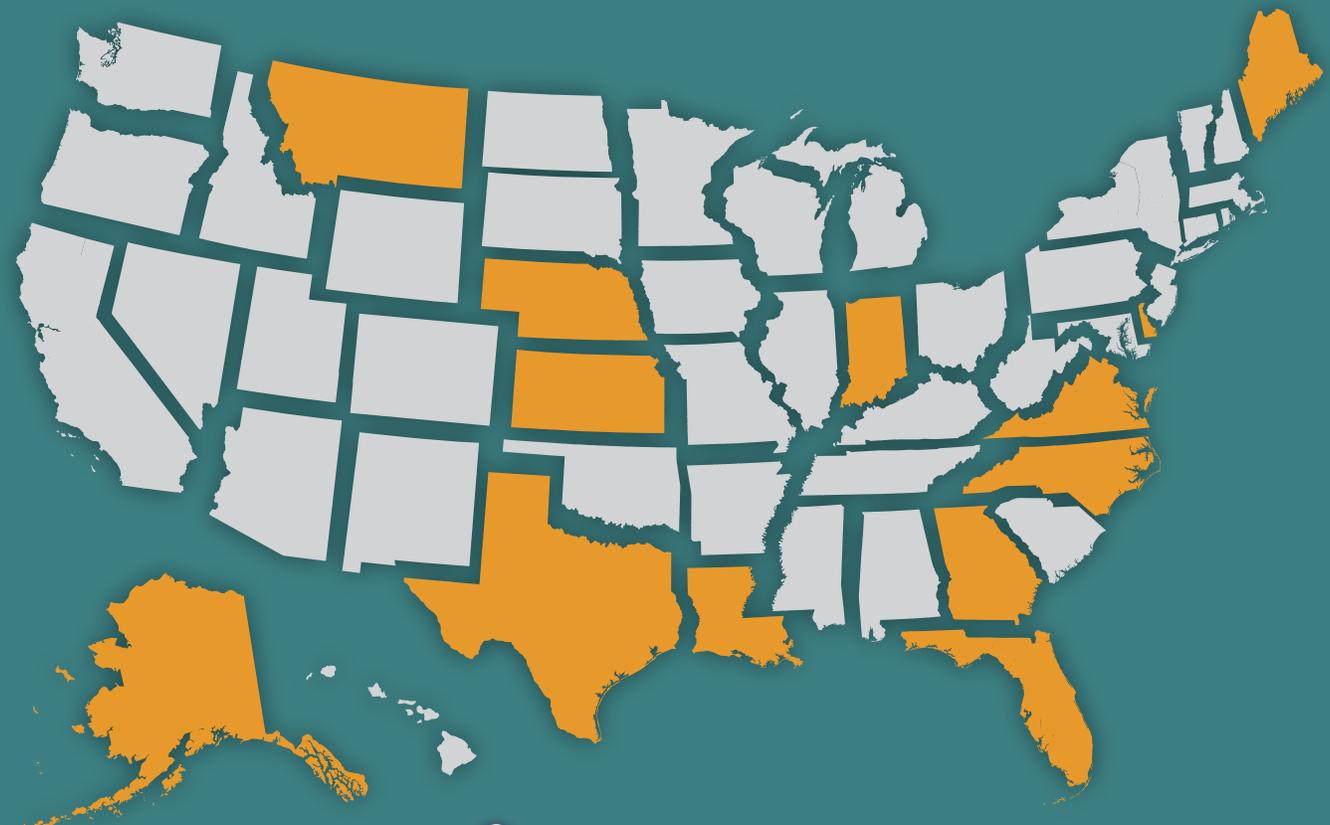
INTRODUCTION

In the fall of 2021, the Bill & Melinda Gates Foundation, the Walton Family Foundation and the Chan Zuckerberg Initiative funded Education First to 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. If the grant program is successful, by June 2023, the project will have seeded multiple new assessment designs and test prototypes available for development into full-scale operational systems. Preliminary results from an earlier pilot of a curriculum-embedded through-year model in Louisiana bolster our hope such models are more equitable and relevant for students, with NWEA **reporting** the model shrunk achievement gaps and increased student engagement compared to the traditional summative assessment.

States Exploring Through-year Assessments

At least 13 states are in varying stages of developing, piloting and implementing through-year assessment models. While there are similarities between state approaches, states are making unique design choices as well. **For detailed profiles of each state model, visit the [Appendix](#).**

STATES AND THEIR THROUGH-YEAR MODELS



 Click the + icon for more information

INTRODUCTION

USED has also signaled a continued interest in through-year assessments by providing both financial support and regulatory flexibility to multiple states. In terms of financial support, of the sixteen awards from the Competitive Grants for State Assessments made in 2020 and 2022, five support through-year assessment projects. In terms of regulatory flexibility, three of the four states with active waivers under the Innovative Assessment Demonstration Authority (IADA) are pursuing through-year assessments.

For this paper, we reviewed 18 assessment models being developed or prototyped in 13 states, each of which administer multiple assessments during the school year as part of their summative assessment system. In the **appendix**, we've included detailed profiles of each of these models in Alaska, Delaware, Florida, Georgia, Indiana, Kansas, Louisiana, Maine, Montana, Nebraska, North Carolina, Texas and Virginia.

Through exploring these models, this paper aims to build an understanding of the key differences in the characteristics, benefits and risks posed by the different through-year approaches states are pursuing and why states are or might consider these models. We've organized the paper around 5 key questions aimed at defining through-year assessments, exploring how states are designing through-year assessment models to tackle the three problems named above, and how states will use the models to meet federal accountability requirements.

- + What are through-year assessments?
- + How are states designing through-year assessments to test what is taught, when it's taught?
- + How are states designing through-year assessments to provide results that inform instruction?
- + How are states designing through-year assessments to change perceptions about the time and resources devoted to testing?
- + How are states producing single summative scores?

These questions will illuminate how states hope through-year assessments will create more coherent assessment systems better aligned to instruction, provide more frequent and meaningful information on student performance to students, families and educators throughout the year, and change perceptions of state testing—while retaining the capability for system leaders to disaggregate and consider results across classrooms, schools, school districts and states, without interfering in local curricular decisions.

DEFINING THROUGH-YEAR



What are Through-year Assessments?

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. This purpose of producing a summative score distinguishes through-year assessments from traditional interim or benchmark assessments which, while also administering multiple tests during the year, are not used in accountability systems.

Definitions of through-year assessments used in the field vary,² particularly in relation to which test administrations must count toward a single summative score. Some definitions set an expectation that data from all tests over the year should be combined to generate a summative score, others simply require all of the test administrations to be part of a single system—but are flexible or silent on whether every administration must contribute to the summative score. While all of the models we included in our review administer multiple assessments during the school year, many of them rely only on the final test to determine a summative score. We included these approaches in our review because they still represent a fundamental shift in a statewide assessment system away from a single end-of-year test. In addition, even for states currently choosing not to include every test administration in the calculation of a summative score, the structure lays the foundation for them to do so in the future.

By producing a summative score, through-year models could theoretically be used as part of a state's assessment and accountability system. For example, Louisiana used results from their Guidebooks 2.0 ELA through-year assessment operationally within their state system of school identification and support for the first time in 2021. While no fully operational through-year assessment systems have undergone federal peer review, several states are in the process of piloting through-year models they intend to use for federal accountability. The 2015 reauthorization of ESEA laid the groundwork for this potential by codifying into law an option for states to use "multiple statewide interim assessments that result in a single summative score that provides valid, reliable and transparent information on student achievement or growth."

² See Dadey, N. & Gong, B. (2017), Gianopoulos, G. (2019), and Dadey, N., Gong, B., Lorie, W., & Marion, S. (2021).

REGULATORY PATHWAYS

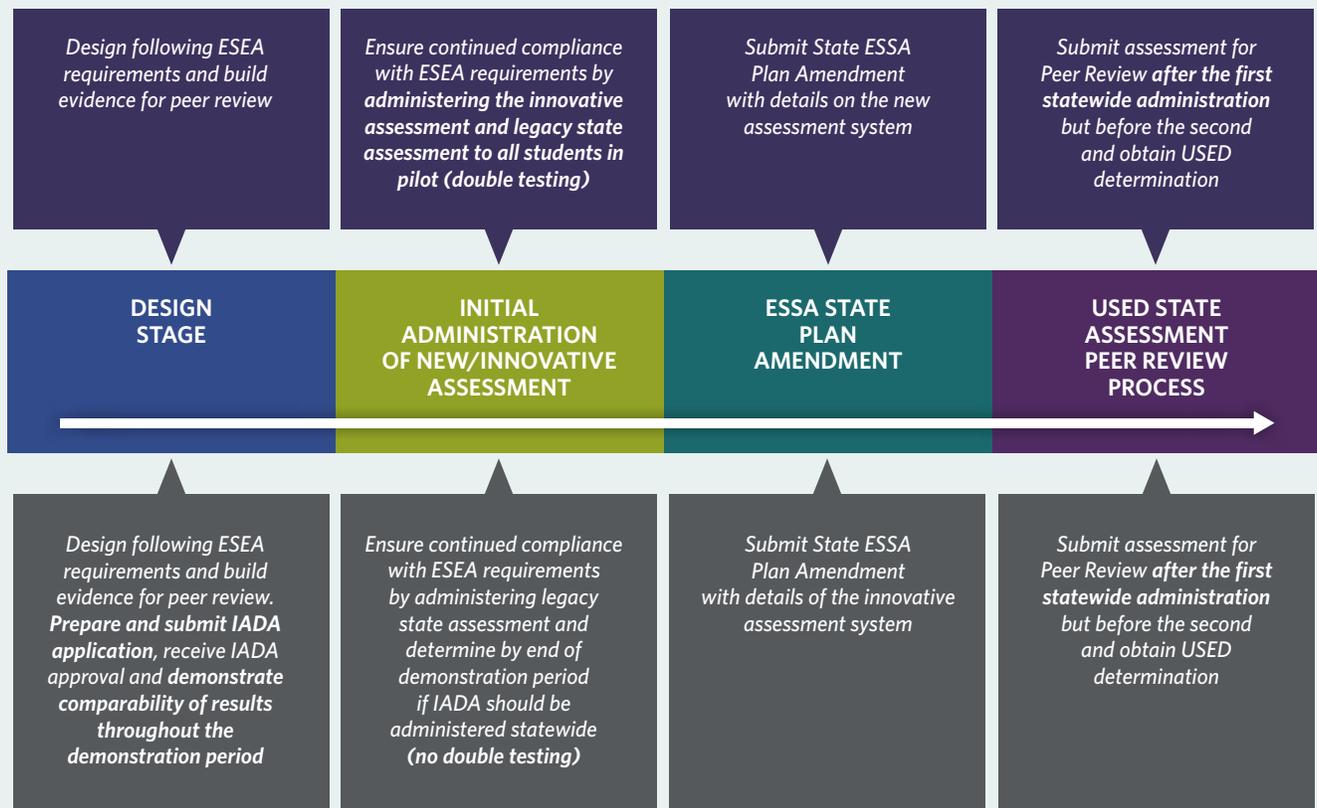
While interest in through-year assessments is growing, some states remain unsure of flexibility under ESSA regarding use of through-year assessments. ESSA created several new flexibilities to help states develop innovative approaches to assessment and reduce duplicative, unnecessary testing. Foresight Law & Policy and Education First reviewed two current avenues through which a state may develop a through-year assessment before submitting for peer review.

If a state continues administering its legacy assessment to all students while piloting an innovative assessment, a state may develop and pilot a through-year assessment that meets ESSA’s allowance for **multiple statewide interim assessments** without submitting an application or securing the Secretary of Education’s approval for a pilot.

States may also pursue an Innovative Assessment Demonstration Authority (IADA) waiver which grants special permission to qualifying SEAs to establish, operate and evaluate an innovative assessment system including competency-based assessments, performance-based assessments, etc. A state may be granted demonstration authority for up to five years. States must describe their approach to scaling the innovative assessment statewide by the end of the authorized demonstration period. Three states are currently testing through-year models using an IADA waiver (GA, LA, NC).

Each pathway has key differences in implementation. Both pathways require states to submit their assessment for USED’s State Assessment Peer Review Process, but Pathway 2 involves a few additional steps due to the demonstration period.

PATHWAY 1: MULTIPLE STATEWIDE INTERIM ASSESSMENTS



PATHWAY 2: IADA INNOVATIVE ASSESSMENT SYSTEM

TESTING WHAT'S TAUGHT



How are states designing through-year assessments to test what is taught, when it's taught?

All through-year models aim to create a more balanced, coherent assessment system. A handful of states are taking this innovation a step further with a goal of strengthening alignment between instruction and assessment by linking the content students are taught to state tests and measuring students' mastery of standards closer to when they are taught—testing what is taught, when it's taught. Three primary design choices illuminate trade-offs states are making in determining how closely to align through-year assessments with instruction and curriculum:

- + Whether each test administration measures all grade-level standards or a subset of standards.
- + Whether to align test administrations to *when* assessed standards are taught, such as through flexible administrations or alignment with scope and sequences.
- + Whether or not to connect assessments to curriculum.

As Nathan Dadey and Aneesha Badrinarayan layout out in their blog [In Search of the "Just Right" Connection Between Curriculum and Assessment](#), the results of these design choices can be conceptualized as a spectrum—ranging from assessments which remain unconnected to curriculum (like traditional state tests), to assessments which align with the pacing of content, to assessments which directly draw on the content found in specific curriculum.

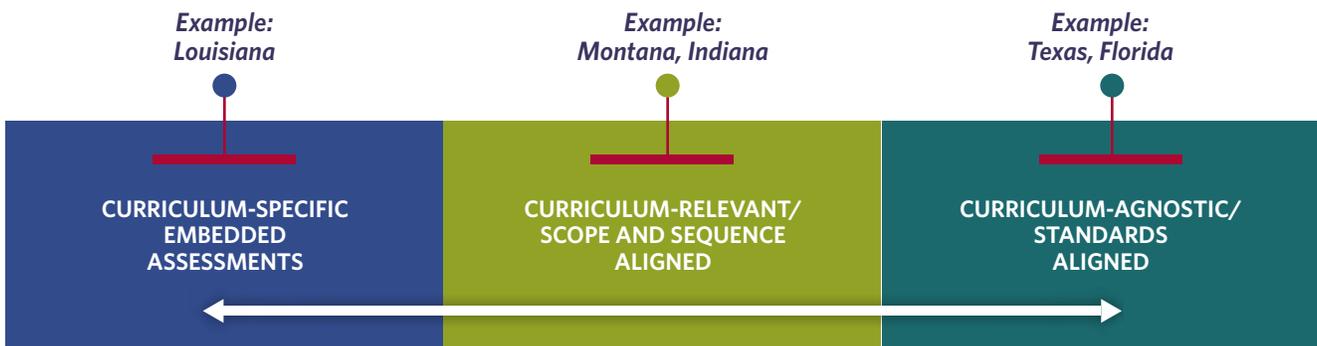


Figure 1. Demonstrates the spectrum of models states are exploring to connect through-year assessments to local curricula and scope and sequences.

Adapted from Dadey and Badrinarayan (2022)

TESTING WHAT'S TAUGHT

Assessing all-grade level standards or a subset of standards³

States with through-year assessments measuring all grade-level standards at each test administration, such as Texas, Florida and Virginia, are designing models to be applicable across schools, regardless of local curricular choices or instructional pacing. One trade-off with these models is students may spend time responding to assessment questions on standards they have yet to learn. In general, these states are utilizing computer adaptive tests which vary the questions students see based on their previous responses to help target the assessments toward the standards students have mastered. **By reporting item-level or subscore results back to educators quickly, valuable information could be gleaned about which students are behind, on, or above grade level earlier in the year.**

In contrast, states such as Louisiana and Montana's Instructionally Aligned Assessment System, North Carolina, and Georgia's Navy system are designing through-year assessments measuring only a subset of grade-level standards during each administration. These models are designed to provide more targeted measurements of student mastery and create the capability to assess student learning of specific standards immediately following relevant instruction. Educators may perceive these through-year assessments and the results they produce as more closely aligned to their instruction. The models, however, create logistical challenges states will have to solve — such as how to sync assessment administration with local instructional pacing.

Aligning test administrations to when assessed standards are taught

Some of the states measuring only a subset of grade-level standards at a time are leveraging this structure to try and sync the assessment of specific standards with the timing of instruction in those standards. **These states hope doing so will deepen the connection between assessment and instructional content while preserving local flexibility and agency around curricular choice and the pacing of instruction.** These states are taking a variety of approaches. For grades 3-8 in mathematics and English language arts (ELA), Montana is developing, testing, and scaling an Instructionally Aligned Assessment System of short "testlets" covering distinct sets of standards aligned with local scope-and-sequences and informed by subject learning progressions. These testlets will be administered 4-6 times throughout the year and will be aggregated to generate summative scores. Louisiana will also test the Instructionally Aligned Assessment System in grades 3-8 math. By aligning with scope-and-sequences, the intention of this model is to target the testing of relevant standards in sync with instruction. While early in its design, Indiana has also indicated their through-year model will allow for flexible administrators aligned with scope-and-sequence and North Carolina allows schools to administer its Check-ins 2.0 in any order. Georgia's Navy system is developing an even more fine-tuned system, composed of short, web-based tests, each of which assess an individual state standard. These tests can be flexibly administered by teachers on demand after instruction on the standard has occurred.

³ See Dadey, N. & Gong, B. (2017) for further discussion of this as a key distinction in through-year models.

TESTING WHAT'S TAUGHT

Connecting assessments to curriculum

So far, Louisiana is the only state attempting to align through-year assessments even more directly to instruction by connecting the tests to ELA curricula. While the models described in the last section build alignment between instruction and assessment by syncing with instructional pacing, Louisiana’s curriculum-connected through-year models deepen the alignment by featuring the same texts students studied as part of the curriculum. Louisiana hopes measuring student learning based on the exact content students were taught will promote deeper engagement with texts as well as be more equitable by reducing any disparities generated from students’ varying levels of background knowledge. To provide educators with some choice around the sequencing of instruction, Louisiana provides test forms covering different reading passages, so educators can select the test form that aligns with the reading passages they just taught.

Currently, Louisiana has developed through-year assessments connected with two high-quality English Language Arts curricula, Guidebooks 2.0 and Wit & Wisdom. However, these models can only be implemented in school districts that have adopted these specific curricula. To serve districts which have not adopted the Guidebooks 2.0 and Wit & Wisdom curricula, **Louisiana is now researching and developing a through-year assessment system designed to draw on common topics, texts, and domains found across high-quality ELA curricula used in the state.** The goal of this assessment is to balance the benefits of a curriculum-connected assessment with the capacity to be administered in any of the states’ school districts. All three of these assessment models will be administered three times throughout the year and will be aggregated to generate summative scores.

Table 1: Overview of states’ design choices related to testing what is taught, when it’s taught

| FEATURE | AK | DE | FL | GA ¹ | GA ² | IN | KS | LA ³ | ME | MT ⁴ | NE | NC | TX | VA |
|--|----|----|----|-----------------|-----------------|----|----|-----------------|----|-----------------|----|----|----|----|
| Assesses all grade-level standards each time | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ |
| Assesses a subset of standards each time | | ✓ | | ✓ | | | | | | ✓ | | ✓ | | |
| Syncs test with learning or scope and sequence | | ✓ | | ✓ | | ✓ | | ✓ | | ✓ | | ✓ | | |
| Curriculum- connected | | | | | | | | ✓ | | | | | | |

1: Georgia’s Navy System; 2: Georgia’s GMAP consortium; 3: Louisiana’s curriculum-connected assessments; 4: Louisiana is piloting IAAS in math

USEFUL RESULTS



How are states designing through-year assessments to provide results that inform instruction?

All the through-year assessment models we profiled share a stated purpose of providing data on student learning that local districts, schools and teachers can leverage to inform, guide and support instruction. Most fundamentally, states hope to achieve this by reporting assessment results multiple times a year, when the data is still timely and actionable. However, within this structure states still must make design choices which influence how useful the data is, including:

- + The type of items included on assessments
- + The design of reports

Variation in item type

As with all assessments, states must choose what types of test items to include on their assessments. This choice influences both the type of information assessment results can provide educators as well as how quickly the results can be provided. All the through-year models we reviewed include selected response items. At least 4 models are also including constructed response items in their test designs (Louisiana, Navy for writing, Texas for science and social studies and Montana's Instructionally Aligned Assessment System for ELA). These items support measuring higher-order thinking and assessing skills like writing and speaking where multiple choice items can not provide valid evidence of competency. However, since constructed response items are typically hand scored, their use can significantly increase the turn-around time for results. For example, while many of the models we reviewed provide immediate results, states using constructed response items reported turn-around times of 2-6 weeks. While this trade-off is not unique to through-year assessments, the choice is heightened by their purpose and structure:

- + Any delay in reporting results could work against the goal of having information in the hands of educators sooner than typical summative assessment.
- + Including constructed items on multiple test administrations each year may significantly increase the logistical burden on states for scoring responses.

USEFUL RESULTS

Given the benefits of constructed response items, states are looking for ways to include the items while limiting the impact of delayed result reporting, such as:

1. Using an automated scoring engine to reduce the scoring time of constructed responses. In their 2022 CGSA application, Louisiana indicated a plan to explore this approach. However, despite high degrees of agreement with human scorers, these engines often face critiques for relying on artificial intelligence in the scoring process.
2. Releasing scores to educators for all machine-scored items and following up with scores for constructed response items. Louisiana has indicated an intention to implement this approach as well.
3. Including constructed response items only on the final assessment, so scores from earlier administrations are not held up in scoring.
4. Including constructed response items in earlier administrations, but not including them in student scores until the final administration. This would allow teachers and students the opportunity to practice writing or speaking skills in an assessment setting.



The design of reports

At this stage we have little visibility into the design of reports. Some states have indicated they plan to provide more detailed data to educators on reports, such as item-level or subscore results, which could support educators in pinpointing where instructional adjustments are needed. As with all assessment reports, it will be critical that states provide both educators and families with accurate, easily understood and actionable reports along with guidance and support on utilizing results to inform future learning.

TESTING TIME



How are states designing through-year assessments to change perceptions about the time and resources devoted to testing?

Educators, families, and students' perspectives on through-year assessments are likely to be shaped by whether the time and resources devoted to testing feels worthwhile. For years, standardized tests have been criticized by educators and the public for taking too much time and being burdensome and disruptive in their administration. In promoting a shift to through-year assessments, some state leaders have emphasized a goal of reducing testing time.⁴

In practice, most through-year assessments do not appear to be reducing the total time students sit for testing over the course of the year. When we compared the total projected testing time of the through-year models we reviewed with the assessments they aim to replace, we found only Louisiana's curriculum-connected models are clearly reducing testing time. This reflects the fact that in comparison to current summative assessments, through-year models are administering a similar number, or more, test items over the course of the year in order to ensure each administration has enough items to report reliable scores.

While through-year assessments may not be reducing overall state testing time, states are hoping through-year assessments will change educators and families perspective on the value of that time in a number of ways:

- + By providing data on student performance at multiple points during the year, the time will be viewed as supporting instruction.
- + By spreading testing time across multiple, shorter administrations, assessment will be more integrated and less disruptive and burdensome. To support this, models like the Navy system and Montana and Louisiana's Instructionally Aligned Assessment System are utilizing tests which can be administered within a single class period.
- + By serving the same function as traditional interim or benchmark assessments, through-year assessments will reduce the need for locally administered exams thereby potentially reducing the total amount of time devoted to state and local testing each year.

⁴For example, see the [press release](#) announcing Florida's F.A.S.T. through-year model.

TESTING TIME



The extent to which through-year assessments will accomplish these goals, remains to be seen. As Nathan Dadey and Brian Gong (2017) note in **Using interim assessments in place of summative assessments? Consideration of an ESSA option**, schools and districts could choose to retain the use of traditional interim or benchmark assessments—causing an expansion rather than contraction in the overall amount of testing. Whether this happens will depend on local school systems evaluation of the utility of through-year assessments or the willingness of a state to dictate limitations to administering local interims. Additionally, current requirements for summative testing, such as high test security, also raise practical questions about whether states’ will truly be able to implement flexible testing windows and reduce the administrative burden of assessments or whether multiple administrations a year will actually increase the burden on the system, educators and students. Education First plans to explore how these implementation challenges are unfolding in future publications.

SCORING



How are states producing single summative scores?

As touched on earlier, states vary in their approach to generating single summative scores for use in federal accountability. Some states are choosing to use only the final administration for summative assessment, while others plan to aggregate across the year to create one final score. This choice is, at least in part, influenced by a state's design choices for which standards are assessed at each time point. It may also reflect a state's perspective on whether summative assessment should reflect what students know and can do at the end of the year or at any point during the year following instruction.⁵

Summative scores based on only a final test administration

States choosing to assess all standards at each time point could either aggregate scores from all administrations or only use the final administration to produce a summative score representing student mastery of all (or most) grade-level standards. Of the states we reviewed that assess all standards at each time, all but Louisiana's curriculum-connected models are currently planning to produce a summative score only from the end-of-year test. However, this choice may ultimately be influenced by what is learned during the piloting of models. Texas, for example, is exploring two potential cumulative scoring models with a guiding principle that earlier performance should be able to help but not hurt students' final scores.

While they do not contribute to a single summative score, several states are using multi-stage or multi-phase adaptive designs where performance on earlier administrations still informs what students see on the final summative test. In Texas' multi-stage adaptive design, each test administration is broken into two sections. In the first test in the fall, all students start with a router section. Depending on their performance, students are then routed to a second section that better fits their demonstrated ability. The next test administration, in the winter, then places each student in a section that best fits their demonstrated ability from the fall administration's second section. In this fashion, where a student starts on the final summative test is influenced by their previous performance. Nebraska and Georgia's MAP through-year assessments use a similar two section format that routes students mid-way through, but the design is also item-level adaptive and can return to earlier "phases." In North Carolina, for students who have completed at least two NC Check-Ins 2.0, the state uses a statistical model to route them to the appropriate adaptive item set on the flexible summative test.

⁵ See Gong, B. (2021) offering a detailed perspective on this tension.

Summative scores based on an aggregation of test administrations

States assessing only a subset of grade-level standards with each administration will likely need to combine across administrations to ensure that the single summative score represents the depth and breadth of the content standards, as required under federal peer review. For example, three models which assess only a subset of standards on each test — Delaware, Montana and Louisiana’s Instructionally Aligned Assessment System and Georgia’s Navvy system — have all indicated plans to calculate single summative scores by aggregating across administrations. Louisiana’s curriculum-connected models also aggregate across three administrations to produce a single summative score, despite assessing all standards on each administration.

There are benefits and potential drawbacks to aggregating scores across administrations, even in instances where states assess all standards. **Aggregation across time points can improve reliability of the final score and partially address regression to the mean that we see for higher and lower-achieving students.** Rather than relying on scores from a single snapshot in time that may be affected by random administration conditions (e.g. a fire drill, barking dog, etc.), averaging scores could reduce some of the impact of those disturbances.



At the same time, states that choose to average across all administrations will have to decide how to handle students who miss a test administration, such as chronically absent students or students who move throughout the year. For example, if a state uses three administrations during the year, some students who’ve missed a test will only have one or two scores when aggregating single summative scores. Unless states figure out another way to capture this data for students, this means students with fewer than three tests will have less precise final scores compared with those who completed the full test battery; and schools with higher proportions of these students may have less precise average scores used in accountability systems as well. Since certain populations of students are more likely to move during a year or miss an administration, such as migrant students and students living in poverty, how this challenge is addressed will have important equity implications.

SCORING

Growth

Through-year assessment models also influence the options states have for calculating growth. Currently, most states calculate growth using a statistical model to estimate the amount of student-level growth between the current and prior year(s) relative to similar students. One criticism of these models is that they often do not account for summer learning loss that may differentially impact students depending on their level of access to educational support and resources.

Through-year assessments provide an opportunity for states to calculate within-year growth, rather than across-year, by using the first administration as a baseline. This is a motivating purpose behind Virginia’s through-year model. How states can approach these growth calculations is influenced by whether the assessments are vertically scaled, a term for a set of design decisions that when taken together represent a continuum for a particular content area:

- + If the assessments are vertically scaled within-year, a simple gain score representing growth could be calculated by subtracting the fall score from the spring score.
- + If the assessments are not vertically scaled, states would need to use statistical models to estimate growth relative to similar students.

One key difference between these two is that a simple gain score would rely on just two scores (beginning of year and end of year), where a statistical growth model could be used more flexibly to include additional scores either from the winter benchmark or the previous year.

Table 2: Overview of states’ design choices related to producing single summative scores.

| FEATURE | AK | DE | FL | GA ¹ | GA ² | IN | KS | LA ³ | ME | MT ⁴ | NE | NC | TX | VA |
|--|----|----|----|-----------------|-----------------|----|----|-----------------|----|-----------------|----|----|----|----|
| Assesses all grade-level standards each time | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ |
| Assesses a subset of standards each time | | ✓ | | ✓ | | | | | | ✓ | | ✓ | | |
| Summative score based on final test | ✓ | | ✓ | | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ |
| Aggregates tests to create summative score | | ✓ | | ✓ | | | | ✓ | | ✓ | | | | |
| Multi-stage or Phase adaptive | | | | | ✓ | | | | | | ✓ | ✓ | ✓ | |

1: Georgia’s Navy System; 2: Georgia’s GMAP consortium; 3: Louisiana’s curriculum-connected assessments; 4: Louisiana is piloting IAAS in math

CONCLUSION



Conclusion

This paper explored trends within state approaches to through-year assessment design. We know leaders are trying to solve for concerns about alignment of assessment with instruction and curriculum, the limited utility and meaning of results to parents, students and teachers, and the perceived undue time and burden traditional assessments require. **New assessment systems will always require re-examining aspects of the current K12 accountability system**, which often begs the question of balance between assessment OF learning vs. assessment FOR learning and whether a single assessment system can achieve both goals. While we believe through-year assessments have the potential to achieve this balance and to build greater coherence between instruction and assessment, we share the concerns raised by colleagues in our field that replacing low-stake interims with only formative purposes with multiple higher-stake assessments linked to accountability could produce unintended consequences that negate the potential benefits of through-year models.⁶ The diversity of different approaches to through-year assessments across the country will aid in exploring whether these models have the intended impacts, what unintended consequences they create, and how to best design summative assessment systems that are more equitable, focused and relevant to students, families, educators and school leaders.

Looking Ahead

In March 2022, Education First convened an advisory group of State Education Agency (SEA) leaders to wrestle with these questions and deepen engagement and collaboration among states interested in exploring a through-year assessment model. Between March and June 2022, the group of 11 states engaged in four virtual convenings facilitated by Education First. These convenings offered opportunities for dialogue, knowledge building, and candid conversations about the nuanced design, implementation and technical challenges regarding implementing through-year assessment models.

These state leaders aimed to understand the variety of approaches states are researching and developing, as well as the reasoning behind different design choices. There are many technical, regulatory, design and implementation considerations that will continue to be explored in through-year assessment pilots across the country. For example, states' pilots will help answer technical questions related to aggregating summative scores and maintaining reliability, validity and

⁶ See Dadey, N. & Gong, B. (2017), Wallace (2022), and Marion, S. & Timberlake, A. (2022) for discussion on a number of other concerns or potential unintended consequences with through-year assessments.

CONCLUSION

comparability. We have also continued to convene assessment developers and partners to interrogate principles of equitable design, plan meaningful engagement with stakeholders and share learning together as they develop and prototype new through-year models.

Over the coming year, Education First and our partners will produce a series of publications and tools illuminating how states are addressing practical challenges and weighing trade-offs as they pilot their through-year assessment models. In consultation with state leaders and assessment experts, we've identified core questions at different stages of the research and development process:

- + How should states determine if a through-year assessment system aligns with their state goals?
- + What infrastructure and enabling conditions must be in place to implement a through-year assessment system and support transitioning to a new assessment system?
- + How can through-year assessment designs move toward coherence with what is being taught when many curricula are used within a state?
- + How are states addressing the logistical and operational implications of having multiple state assessment administrations?

We hope this research will facilitate shared learning across states and show how these design and implementation challenges can be addressed while improving the summative assessment experience for people most proximate to the problem—students, families, educators and school leaders. Further research will be needed to determine whether through-year assessments are meeting the promise of measuring students' learning based on what they've been taught and providing teachers with data to inform instruction while meeting state accountability goals. In particular, little quantitative research evaluating the impact on equity and student outcomes has been conducted so far; but we are encouraged that these questions are at the heart of many states' research and development efforts.



AUTHORS & ACKNOWLEDGEMENTS

Authors



Dave Powell
Education First
Senior Consultant



Senna Lamba
Education First
Associate



Khaled Ismail
Education First
Senior Consultant



Joshua Marland
President
Marland Strategy
Consulting

Acknowledgements

We want to thank the following partners for their time and expertise in reviewing and improving drafts of this publication:

- + **Aneesha Badrinarayan**, Learning Policy Institute
- + **Nathan Dadey**, Center for Assessment
- + **Garron Gianopulos**, NWEA
- + **Molly Minnick**, Walton Family Foundation
- + **Claire Voorhees**, Walton Family Foundation

We want to thank the leaders at State Education Agencies who reviewed their state's profile for accuracy and clarity:

- + **Dusty Shockley**, Delaware Department of Education
- + **Catherine Altmaier** and **Vince Verges**, Florida Department of Education
- + **Laine Bradshaw**, Navy Education
- + **Lynn Schemel**, Indiana Department of Education
- + **Beth Fultz**, Kansas State Department of Education
- + **Alissa Kilpatrick**, Louisiana Department of Education
- + **Janette Kirk**, Maine Department of Education
- + **Jeremy Heneger**, Nebraska Department of Education
- + **Shelby Armentrout** and **Tammy Howard**, North Carolina Department of Public Instruction
- + **Jordan Runge** and **Iris Tian**, Texas Education Agency
- + **Sarah Susbury**, Virginia Department of Education

The Through-Year Curriculum-Connected Assessment Grant Program founded in 2021 by the Bill & Melinda Gates Foundation, the Walton Family Foundation, the Chan Zuckerberg Initiative and Education First, was developed to support this work.

SOURCES

Alaska Department of Education and Early Development. (December 8, 2021). "Introduction to the Alaska System of Academic Readiness (AK STAR)." NWEA State Psychometrics.

https://education.alaska.gov/assessments/akstar/Introduction_to_AK_STAR.pdf

Alaska Department of Education and Early Development. (July 2021). "Alaska innovative assessment system, A connected approach: Bridging MAP Growth interim assessments and the annual summative."

https://education.alaska.gov/news/releases/2021/Alaska_Innovative_Assessment_System_A_Connected_Approach.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. (2017). Using interim assessments in place of summative assessments? Consideration of an ESSA option. Washington, DC: Council of Chief State School Officers.

https://ccsso.org/sites/default/files/2017-12/ASR_ESSA_Interim_Considerations-April.pdf

Dadey, N., Gong, B., Lorié, W., & Marion, S. (2021, Nov. 21). Defining Through-Year Assessment and the Motivations Behind Through-Year Programs. Center for Assessment.

<https://www.nciea.org/library/definitions-aims-and-use-cases/>

Delaware Department of Education. (2022). "Delaware Transitions to Innovative Through Assessment System for Social Studies by SY2024-2025."

<https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/111/SS%20Through%20Assessment%20System.pdf>

Florida Department of Education. (2022). "B.E.S.T. Assessments."

<https://www.fldoe.org/accountability/assessments/k-12-student-assessment/best/>

Gianopoulos, G. (2019). From Through-Course Summative to Adaptive Through-Year Models for Large-scale Assessment: A Literature Review. NWEA Research.

https://www.nwea.org/uploads/2021/06/From-Through-Course-Summative-to-Adaptive-Through-Year-Models-for-Large-scale-Assessment_NWEA_literatureReview.pdf

Gong, B. (2021, February 23). Why Has it Been So Difficult to Develop a Viable Through Year Assessment? Center for Assessment.

<https://www.nciea.org/blog/why-has-it-been-so-difficult-to-develop-a-viable-through-year-assessment/>

Kansas Assessment Program. (2022). "Introduction to the Kansas Assessment Program Predictive Interim Assessment."

<https://ksassessments.org/sites/default/files/scorm/introduction-to-the-predictive-interim-assessment/content/index.html#/>

Louisiana Department of Education. (2022). Materials from "Assessment Guidance: Innovative Assessment Program."

<https://www.louisianabelieves.com/resources/library/assessment-guidance>

Maine Department of Education. (2022). "Math & Reading: Transition to the Maine Through Year Assessment."

https://www.maine.gov/doe/Testing_Accountability/MECAS/NWEA

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

<https://www.nciea.org/blog/trying-to-serve-multiple-uses-with-through-year-assessments/>

Marion, S., and Timberlake, A. (2022, March 30). But Will It Work in Practice?

Practical Challenges to the Use of Through-Year Assessment for Accountability. Center for Assessment.

<https://www.nciea.org/blog/but-will-it-work-in-practice/>

Munyan-Penney, N. and Barone, C. (2020). Innovative Assessment Pilot: Louisiana's LEAP 2025 Humanities Assessment. Ed Reform Now.

<http://edreformnow.org/wp-content/uploads/2020/11/Louisiana-IADA-LEAP-2025-Humanities.pdf>

Munyan-Penney, N. and Barone, C. (2020). Innovative Assessment Pilot: Georgia's Georgia MAP and Navy Assessments. Ed Reform Now.

<http://edreformnow.org/wp-content/uploads/2020/11/Georgia-Innovative-Assessment-Pilot.pdf>

Navy Education. (2022). "Navy: Standard-by-Standard Diagnostic Assessment System."

<https://navyeducation.com/#navvy>

SOURCES

Nebraska Department of Education. (2022). "NSCAS Overview."
<https://www.education.ne.gov/assessment/nscas-system/>

Nebraska Department of Education. (2022). "NSCAS Growth."
<https://www.education.ne.gov/assessment/nscas-growth/>

North Carolina Department of Public Instruction. (2022). "North Carolina Personalized Assessment Tool."
<https://www.dpi.nc.gov/districts-schools/testing-and-school-accountability/state-tests/north-carolina-personalized-assessment-tool#additional-resources>

NWEA. (2021). Louisiana and NWEA: Creating innovative assessments to foster equity and deeper learning. NWEA State Solutions.
https://www.nwea.org/resource-center/white-paper/47609/Louisiana-IADA-Brief_NWEA_whitepaper-1.pdf/

Office of Elementary and Secondary Education. (2022). Competitive Grants for State Assessments Awards. Montana Office of Public Instruction: "Montana Alternative Student Testing Program."
https://oese.ed.gov/files/2022/08/MontanaCGSA2022application_Redacted.pdf

Office of Elementary and Secondary Education. (2022). Competitive Grants for State Assessments Awards. Louisiana Department of Education: "Testing What's Taught: Equity in Test Design."
https://oese.ed.gov/files/2022/08/LouisianaCGSA2022application1_Redacted.pdf

Office of Elementary and Secondary Education. (2019-2022). Materials under "2019 State Applications for IADA, Georgia."
<https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/iada/>

Office of Elementary and Secondary Education. (2018-2022). Materials under "2018 State Applications for IADA, Louisiana."
<https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/iada/>

Office of Elementary and Secondary Education. (2019-2022). Materials under "2019 State Applications for IADA, North Carolina."
<https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/iada/>

Office of Governor DeSantis. (2021, September 14). "Governor DeSantis Announces End of the High-Stakes FSA Testing to Become the First State in the Nation to Fully Transition to Progress Monitoring." News Release.
<https://www.flgov.com/2021/09/14/governor-desantis-announces-end-of-the-high-stakes-fsa-testing-to-become-the-first-state-in-the-nation-to-fully-transition-to-progress-monitoring/>

Preston, S. (June 28, 2022). "A case study in action: Delaware Social Studies Through Course Assessment. Learn how to change your assessment paradigm." Presentation at the Council of Chief State School Officers' National Conference on Student Assessment, Atlanta, GA.

Powers, S. (January 2022). "Georgia Innovative Assessment Pilot Program: Technical Assistance Annual Report." West Ed.
https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/Flexibility/GA_IAPP_Annual_Report_Yr%20-2_011222.pdf

Runge, J. (June 28, 2022). "Addressing the technical, logistical, and political challenges of through-year assessment" Presentation on "TTAP Overview- TYA" at the Council of Chief State School Officers' National Conference on Student Assessment, Atlanta, GA.

Schemel, L. Director of Assessment, Indiana Department of Education. Discussions with Sheila Briggs on June 22 and November 1, 2022.

Texas Education Agency. (September 16, 2022). "2022—2023 Texas Through-year Assessment Pilot (TTAP) FAQs."
<https://tea.texas.gov/sites/default/files/2022-TTAP-faq.pdf>

Uncomplicate Ed. (March 23, 2022). "Life in the F.A.S.T. Lane: Vendors Chosen for Florida's New Test."
<https://uncomplicateed.com/b-e-s-t-blog/f/life-in-the-fast-lane-vendors-chosen-for-floridas-new-test>

Virginia Department of Education. (2022). "Growth Assessments."
https://doe.virginia.gov/testing/test_administration/growth-assessment/index.shtml

Wallace, R. (2022, May 5). What Are Through-Year Assessments? Emerging Trends in Statewide Assessments. Data Quality Campaign.
<https://dataqualitycampaign.org/what-are-through-year-assessments/>

APPENDIX: STATE PROFILES

Appendix: State Profiles

Below we have included a profile for each of the through-year models we assessed with the goal of providing a snapshot of how each state has described the primary goals for their through-year assessment along with details on the test design, approach to summative scoring, implementation timeline and other features of the assessment system.

It should be noted this review does not capture all the through-year assessment models being researched and developed across the country or similar innovations exploring how to use multiple test administrations to create more balanced, coherent and instructionally useful assessment systems. For example, Dynamic Learning Maps’ instructionally-embedded alternative assessments utilize a through-year model. Illinois recently received a CGSA grant to develop a through-course model for high school Spanish language arts. Both Louisiana and Delaware are developing through-year assessments for science. Similarly, Nebraska has worked with a consortium of states to develop the Stackable, Instructionally-embedded, Portable Science (SIPS) Assessments project; a model which could be utilized as the foundation for a through-year assessment system. At a more localized level, Centerpoint is exploring how it could utilize curriculum-embedded interims it has created with a number of curriculum partners to create through-year assessments which produce a summative score across the multiple administrations. This approach would integrate testing into specific teaching and learning cycles aligned to the adopted curriculum.

Table 3: Comparative summary of states’ design choices and other assessment model features.

| FEATURE | AK | DE | FL | GA ¹ | GA ² | IN | KS | LA ³ | ME | MT ⁴ | NE | NC | TX | VA |
|--|----|----|----|-----------------|-----------------|----|----|-----------------|----|-----------------|----|----|----|----|
| Assesses all grade-level standards each time | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ |
| Assesses a subset of standards each time | | ✓ | | ✓ | | | | | | ✓ | | ✓ | | |
| Syncs test with learning or scope and sequence | | ✓ | | ✓ | | ✓ | | ✓ | | ✓ | | ✓ | | |
| Curriculum- connected | | | | | | | | ✓ | | | | | | |
| Summative score based on final test | ✓ | | ✓ | | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ |
| Aggregates tests to create summative score | | ✓ | | ✓ | | | | ✓ | | ✓ | | | | |
| Multi-stage or Phase adaptive | | | | | ✓ | | | | | | ✓ | ✓ | ✓ | |
| Item-level adaptive | ✓ | | ✓ | | ✓ | | | | ✓ | | ✓ | | | ✓ |
| Provides more regular data to educators and families | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

1: Georgia’s Navy System; 2: Georgia’s GMAP consortium; 3: Louisiana’s curriculum-connected assessments; 4: Louisiana is piloting IAAS in math

APPENDIX: STATE PROFILES

Alaska's Map Growth/Alaska System of Academic Readiness (AK STAR)

KEY FEATURES

GOAL

Alaska's MAP Growth Assessments aims to better reflect learning throughout the school year, increase coherence across interim and end-of-year tests, reduce testing time, and help teachers target instruction to students' individual needs. This innovative assessment system also increases opportunities for students and schools to show growth throughout the year.

DISTINGUISHING FEATURES:

- + Students will take two different item-level adaptive tests in the fall and winter for teachers to anticipate students' performance on the spring summative assessment, but these results will not impact where students begin the test on the spring summative.
- + Fall and winter interim assessments are connected to the spring end-of-year test for improved efficiency, cohesion and impact on teaching and learning, as they will provide teachers with immediate results from interim tests to improve instruction.

RELATION TO SUMMATIVE SCORE:

- + The 2 MAP Growth tests students take this year (2022-23) will provide a projected proficiency score, however, students' scores on the fall and winter MAP Growth will not determine where a student begins on the summative test.
- + The summative test given in the spring is an on-grade level adaptive test. The spring summative test will have two blueprints and will include a MAP Growth diagnostic as well as the traditional spring summative component.
- + Only the spring assessment is used for students' summative scores

IMPLEMENTATION TIMELINE

- + **First year of Implementation (2021-22 School Year):** Districts administered MAP Growth in the fall, winter and spring so that research could be done to support the production of RIT scores.
- + **Second year of Implementation (2022-23 School Year):** MAP Growth will only be administered in the fall and winter. Students will take the on-grade level adaptive summative in the spring. Next year, the state would like to have the summative adapt outside of grade level.

ASSESSMENT SYSTEM FEATURES

| | |
|---------------------------|--|
| Subjects and Grades | ELA & Math, 3-9 |
| Primary Vendor | NWEA |
| Number of Administrations | 2 (fall and winter) plus the spring summative test |
| Total Testing Time | 2 hours per test for 4 hours plus the spring summative |
| Results Turnaround Time | MAP Growth provides immediate results |
| Adaptability | Item-level adaptive |

APPENDIX: STATE PROFILES

Delaware's Social Studies Through-Course Assessment

KEY FEATURES

GOAL

Delaware's Social Studies through-year assessment aims to encourage a deeper level of learning by focusing instruction on fewer standards and providing actionable feedback.

DISTINGUISHING FEATURES:

- ⊕ Each of the three assessment administrations will focus on recently taught content and focus on fewer specific standards at each administration, so teachers receive actionable feedback to adjust instruction.

RELATION TO SUMMATIVE SCORE:

- ⊕ Students' scores will be based on aggregating the three test administrations.

IMPLEMENTATION TIMELINE

- ⊕ **First year of Implementation (2021-22 School Year):** Districts administered MAP Growth in the fall, winter and spring so that research could be done to support the production of RIT scores.
- ⊕ **Early 2020:** The Social Studies Coalition of Delaware voted to shift to through-year assessments in social studies in grades 4-8 (with a capstone assessment in Grade 11).
- ⊕ **2021-22 school year:** Interim field tests in 8th grade History.
- ⊕ **2022-23 school year:** Interim field tests in 4th grade Civics and History and 6th grade Geography
- ⊕ **2023-24 school year:** Interim field tests in 5th grade Economics and Geography and 7th grade Civics and Economics
- ⊕ **2024-2025 school year:** Delaware plans to implement fully and use this through-year assessment for state accountability.

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|---|
| Subjects and Grades | Social Studies, 4-8; at full implementation |
| Primary Vendor | Pearson |
| Number of Administrations | 3 (fall, winter, spring) |
| Total Testing Time | Average of ~30 minutes per testing administration (as per 2021-22 8th grade student data) for roughly 1.5 hours total |
| Results Turnaround Time | Students receive an immediate raw score Teachers receive scores on reporting platform within 1-2 days |
| Adaptability | Fixed forms |

APPENDIX: STATE PROFILES

Florida's Assessment of Student Thinking (FAST)

KEY FEATURES

GOAL

Florida's computer-adaptive coordinated screening and progress monitoring system, referred to as FAST, will help teachers track student progress in learning the B.E.S.T (Benchmarks for Excellent Student Thinking) standards in English Language Arts (ELA) Reading and Mathematics throughout the year to identify student learning gaps and target instruction, give parents and educators meaningful guidance to support their child's growth in real-time, and reduce overall testing time.

DISTINGUISHING FEATURES:

- + Florida will utilize the first two administrations of FAST (PM1 and PM2) to monitor student progress and provide educators and partners with real-time performance data. Each test will assess the full grade-level content based on the B.E.S.T. Standards. Following the baseline (PM1) and mid-year (PM2) administrations, the final administration (PM3) will provide a summary score that will accurately measure student mastery of the B.E.S.T. Standards.

RELATION TO SUMMATIVE SCORE:

- + While PM1 and PM2 are for informational purposes only, PM3 will be used for school accountability purposes (for grades 3-8 Mathematics and grades 3-10 ELA Reading ONLY) beginning in the 2023-24 school year.

IMPLEMENTATION TIMELINE

- + **2021-22 School Year:** Final administration of Florida's previous statewide, standardized assessment in ELA and Mathematics, the Florida Standards Assessments (FSA).
- + **2022-23 School Year:** First year of FAST assessments. Accountability will be paused for this year and summative scores will serve as an informational baseline only. However, school grades will be issued based on an equipercentile link back to 2021-2022 school grades.
- + **2023-24 School Year:** The third FAST assessment (PM3) will be used for accountability purposes (for grades 3-8 Mathematics and grades 3-10 ELA Reading ONLY).

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|--|
| Subjects and Grades | ELA Reading, VPK-grade 10; at full implementation Mathematics, VPK-grade 8; at full implementation |
| Primary Vendor | VPK-2: Renaissance Learning 3-10: Cambium Assessments, Inc. |
| Number of Administrations | 3 (beginning, middle, and end of school year) |
| Total Testing Time | The below information indicates the amount of time that students will be given to complete FAST assessments (ELA Reading and Mathematics) over the course of the school year. Some students may use less time, while others may need more. <ul style="list-style-type: none"> + VPK: 45 minutes (median) + K-1: 120 minutes (median) + 2: 127.5 minutes (median) + 3-5: 560 minutes + 6-8: 620 minutes + 9-10: 300 minutes |

Continues

Florida's Assessment of Student Thinking (FAST)

ASSESSMENT SYSTEM FEATURES

| | |
|---------------------------------------|--|
| <p>Results Turnaround Time</p> | <p>FAST VPK-Grade 2: Student-level results for all administrations of FAST are available to teachers immediately after testing.</p> <p>FAST Grades 3-10: Student-level results for PM1 and PM2 are available within one week of testing. Currently, PM3 results will be available by June 30th, but the Department hopes to report results by May 31st beginning with the 2023-24 school year.</p> |
| <p>Adaptability</p> | <p>All FAST assessments: item-level for mathematics; passage-based for ELA.</p> |

APPENDIX: STATE PROFILES

Georgia's MAP Assessment Partnership

KEY FEATURES

GOAL

The MAP Assessment Partnership's unified assessment system aims to reduce testing time, provide educators with instructional guidance, and challenge students to develop the higher order thinking skills they need to succeed in college and careers.

DISTINGUISHING FEATURES:

- + The test will blend NWEA MAP Growth benchmark assessments with items directly aligned to the Georgia Standard for Excellence (GSE) to create a unique design and through-year test experience.
- + The assessments are administered three times per year to measure student learning relative to grade-level expectations and adapt within, below, or above grade level based on the student's performance. Each assessment will provide timely data about students' achievement, show longitudinal academic growth within and across years, and maximize test efficiency for each student (e.g. if students demonstrate command of particular grade-level concepts in fall or winter, they do not necessarily need to be retested on them in spring).

RELATION TO SUMMATIVE SCORE:

- + The summative proficiency score will be based on the final test administration. Fall and winter test scores will be used as priors on future adaptive tests to improve score estimation. In other words, the fall test tells the winter where to begin and the winter test informs where the spring begins.

IMPLEMENTATION TIMELINE

- + **2021-22 School Year:** Completed field test
- + **2022-23 School Year:** First through-year administration for comparability evidence
- + **Fall 2023:** Statistical comparability evidence provided
- + **2023-24 School Year:** Comparability evidence to be reviewed and approved
- + **2024-25 School Year:** First operational administration in lieu of current Georgia Milestones

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|---|
| Subjects and Grades | ELA & Math, 3-8; at full implementation Science, 5 & 8; at full implementation |
| Primary Vendor | NWEA |
| Number of Administrations | 3 (fall, winter spring) |
| Total Testing Time | Roughly 90 minutes per test for about 4.5 hour total |
| Results Turnaround Time | Results are available 1-2 hours after testing administration |
| Adaptability | Multi-phase, item-level adaptive |

Note: At the time of publication, Georgia's MAP Assessment Partnership was on hold.

APPENDIX: STATE PROFILES

Georgia: Putnam County Consortium's Navy System

KEY FEATURES

GOAL

The Putnam County Consortium is partnering with Navy Education to pilot an innovative, through-year assessment system (the Navy system) which aims to support the teaching and learning of the Georgia Standards of Excellence by increasing the quality of data collected and therefore providing instructionally-relevant feedback for schools.

DISTINGUISHING FEATURES:

- + Navy is a standards-level assessment system that pinpoints student competencies of individual state standards, using short, web-based assessments providing actionable, real-time feedback to support personalized instruction.
- + Educators can flexibly administer on-demand assessments to fit their instructional schedule, that will provide actionable and reliable diagnostic information, creating a learning-focused assessment system. Each standard includes approximately 8 questions and students can be retested on each standard up to two additional times after the initial administration.

RELATION TO SUMMATIVE SCORE:

Several ideas are being considered for determining how the final summative score should be calculated. Some of the ideas that Navy has considered include:

- + Calculating the percentage of standards/competencies reached as the summative score (will not require a year-end summative assessment)
- + Using a weighted percentage of standards learned

IMPLEMENTATION TIMELINE

- + **2021-22 School Year:** Gather additional field testing data and comparability evidence for grades 3-8 ELA and math
- + **2022-23 School Year:** Gather additional field testing data and comparability evidence for HS ELA and math
- + **Spring 2023:** Statistical comparability evidence provided and reviewed
- + **2022-2023 School Year:** Comparability evidence to be reviewed and approved for grades 3-8 math and ELA
- + **2023-2024 School Year:** Upon approval, first operational administration in lieu of current Georgia Milestones

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|--|
| Subjects and Grades | ELA and Math, 3-8 and HS; at full implementation Writing, 3-8 and HS; at full implementation |
| Primary Vendor | Navy Education (owned by Pearson) |
| Number of Administrations | Administered on demand throughout the school year (up to individual teachers) |
| Total Testing Time | Varies. Each district utilizes Navy as a classroom assessment system and gives different numbers of assessments, based on students' pacing and needs. |
| Results Turnaround Time | Navy Math and Reading Competency Checks: Individual student level and aggregate reports provided in real-time Navy Writing Checks: Results provided within 2 weeks due to extended responses |
| Adaptability | The Navy system is designed to be customized to the standards each student is learning. Based on the content students' are learning, teachers assign assessments for a standard or any combination of standards on an as-needed basis. Within these administrations, Navy uses a dynamic assessment generation that is non-adaptive. |

APPENDIX: STATE PROFILES

Indiana's Smarter Balanced Interims

KEY FEATURES

GOAL

Indiana's interim assessments will provide more actionable reports for teachers, make interim assessments more meaningful and aligned, and tie assessments to instruction. Professional learning will be provided throughout the implementation with modules and workshops on the topics of assessment literacy, test administration, reporting and application for instruction.

DISTINGUISHING FEATURES:

Indiana is currently creating and designing three interims for mathematics and English/language arts in grades 3-8 that are flexible and based on curriculum scope and sequence, along with a shortened summative, that will provide educators with more detailed reports.

- + Schools and teachers will be able to administer the tests in whichever order best aligns with their curriculum.
- + Indiana is considering creating two forms for each interim which schools will be able to use flexibly, such as by: 1) administering one form to all students at the conclusion of the instruction, 2) administering both forms to all students—one form as a pre-test prior to instruction and the other form as a post-test after instruction, or 3) administering one form to all students and using the second form for a subset of students after the implementation of interventions for students in need.

RELATION TO SUMMATIVE SCORE:

- + Information in the interims will be used to create a claim score which is not a portion of the final summative score. The summative will be the scale score. The claim score will inform reporting categories to inform educators, students, and families on what students know and can do and next steps.

IMPLEMENTATION TIMELINE

- + **2024-25 School Year:** Implement interims with 20% of the state
- + **2025-26 School Year:** Implement interims and shortened summative statewide

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|--|
| Subjects and Grades | ELA and Math, grades 3-8; at full implementation |
| Primary Vendor | Cambium Assessments, Inc. |
| Number of Administrations | 3 interims, 1 summative |
| Total Testing Time | TBD, blueprint development with educator committees in 2022-2023 |
| Results Turnaround Time | Immediate for interims, within 12 business days for summative |
| Adaptability | Initial interims will be fixed, goal to move to adaptive interims by 2027-2028 |

APPENDIX: STATE PROFILES

Kansas' Predictive Interim Assessment & Interim Mini Tests

KEY FEATURES

GOAL

Kansas' long-term goal is to use assessments to help teachers make data-driven instructional decisions to help students learn better. These interim assessments allow educators to evaluate students' knowledge and skills relative to academic goals and can inform decisions at the school or district level.

DISTINGUISHING FEATURES:

- ⊕ Along with the state summative test, Kansas is implementing Predictive Interim Assessments, an optional assessment given three times a year in ELA and math that evaluates where students are in their learning and whether they are on track to perform well on future summative assessments.
- ⊕ The predictive interim uses clusters of standards that are also assessed on the Kansas Summative. The three administrations together cover all standards and provide teachers with information about how their students are doing relative to state curricular standards. Students are exposed to the same content and types of questions that they'll see on the required EOY Kansas Summative assessment.
- ⊕ Interim mini-tests are an optional assessment available year-round in ELA and math that provides ongoing feedback during instruction to aid teachers in adjusting instruction.

RELATION TO SUMMATIVE SCORE:

- ⊕ These predictive assessments predict students' scores on the summative. The state summative assessment is still given once per year in the spring to evaluate student learning for the year.

IMPLEMENTATION TIMELINE

These assessments are currently implemented but are optional.

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|---|
| Subjects and Grades | Math and ELA, 3-12 |
| Primary Vendor | State created |
| Number of Administrations | 3 times (Fall, winter, spring) Predictive interim assessments are available during three two-week testing windows during the school year |
| Total Testing Time | Each assessment has between 17 and 25 questions Roughly 45 - 60 minutes per session for a total of roughly 3 hours |
| Results Turnaround Time | No information provided |
| Adaptability | N/A |

Louisiana’s Curriculum-connected ELA Assessments

KEY FEATURES

GOAL

The goal of these through-year assessments is to engage students in deeper learning and make assessments more equitable and relevant to instruction. The Louisiana Department of Education established a content-rich curriculum through its ELA Guidebooks, where students build their background knowledge through units centered around general themes and texts. Rather than assessing specific skills, such as summarizing passages, Louisiana is building these new assessments based on passages from the content-rich curriculum to be more responsive to instruction and promote deeper engagement with texts. This is also more equitable as it reduces any disparities generated from students’ varying levels of background knowledge.

DISTINGUISHING FEATURES:

ELA Guidebooks 2.0 and ELA Wit & Wisdom:

- + Test blueprints are curriculum based with both ‘hot’ and ‘warm’ reads built around a specific content unit with curriculum-specific anchor texts. (*hot reads are based on texts included in the curriculum, warm reads are related to the same subject as the ‘hot read’ but new to the student, giving students the ability to show how they can extend knowledge obtained through the unit of instruction).
- + In the fall and winter, districts choose between 2 test forms to align with taught content (with each form measuring the same standards). All students take the same form during the Spring administration, which also includes an end-of-year essay.

Curriculum-relevant ELA (also called Crawfish):

- + To accommodate districts who do not use Guidebooks 2.0 or Wit & Wisdom, Louisiana is creating a curriculum-relevant through-year assessment that is less dependent on any specific curricula.

RELATION TO SUMMATIVE SCORE:

- + The Guidebooks 2.0, Wit & Wisdom and Crawfish models all plan to produce a summative score by aggregating results from all three test administrations. They will explore several approaches to generating summative scores but intend to weigh each administration equally.

IMPLEMENTATION TIMELINE

- + **2022-23 School Year:** All: 6,7,8 operational. Grade 5 field tests in Guidebooks and Wit and Wisdom
- + **Crawfish:** Will begin prototyping in winter/spring 2022

ASSESSMENT SYSTEM FEATURES

| | |
|---------------------------|--|
| Subjects and Grades | All: ELA, 3-8; at full implementation |
| Primary Vendor | All: NWEA |
| Number of Administrations | All: 3 administrations(fall, winter, spring) |
| Total Testing Time | All: Unit models are 8-11 items and should take about 45-60 minutes to complete |
| Results Turnaround Time | All: Intend to have scores come within 2-weeks of the test administration. Written portions are hand scored and will come 1 month-6 weeks later. |
| Adaptability | Non-adaptive |

APPENDIX: STATE PROFILES

Maine's Through Year Assessment

KEY FEATURES

GOAL

Maine's Through Year Assessment Program in reading and math provides a flexible assessment system with timely and actionable data to support educators and students, and aims to reduce lost instructional time.

DISTINGUISHING FEATURES:

- ⊕ Students are assessed two to three times per year for performance monitoring, and teachers receive interactive reports and tools designed to provide timely and actionable information.
- ⊕ The Maine Through Year Assessment will produce two scores in the spring for each student: a summative, criterion-referenced score based on the student's performance according to grade-level state standards, and a norm-referenced RIT score useful for measuring a student's growth over time.

RELATION TO SUMMATIVE SCORE:

The summative score is determined by a student's performance on the summative portion of the spring assessment, which can adapt within one-grade level above and below the assessed grade. In addition, an RIT score is determined by all questions on the spring assessment, including both the summative portion and a diagnostic portion from NWEA's MAP Growth item bank which can adapt across multiple grade levels. The required fall and optional winter assessment administrations will also produce RIT scores so that growth can be measured not only between academic years but also throughout the year.

IMPLEMENTATION TIMELINE

- ⊕ **2021-22 School Year:** Began implementing NWEA MAP Growth as an interim assessment in reading and mathematics.
- ⊕ **2022-23 School Year:** Transitioning to Maine Through Year Assessment model for math and reading, with NWEA MAP Growth in the fall and winter and the first Maine Through Year Assessment administration in the spring.
- ⊕ **2023 - onward:** Full implementation of the Maine Through Year Assessment model.

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|--|
| Subjects and Grades | Reading and Math, 3-8 and second year of high school; at full implementation |
| Primary Vendor | NWEA |
| Number of Administrations | 2 required (fall and spring), third test is optional (winter) |
| Total Testing Time | Approximately 60-minute sessions for each content area administration during fall and spring administration windows. |
| Results Turnaround Time | Due to standard setting in the summer of 2023, student results from the spring 2023 assessment will be delayed. For future assessment administrations, however, we expect SAUs to receive results within 24 to 48 hours. |
| Adaptability | Computer Adaptive |

APPENDIX: STATE PROFILES

Montana's ELA Instructionally Aligned Assessment System (referred to as MAST)

KEY FEATURES

GOAL

Montana's Instructionally Aligned Assessment System for ELA is designed to offer a coherent, continuous and useful balanced assessment system to better align assessment more closely with learning, and provide meaningful, actionable information on learning to students, parents and educators.

DISTINGUISHING FEATURES:

- ⊕ The Instructionally Aligned Assessment System uses short "testlets" that cover on-grade learning standards and are designed to be aligned with local curriculum scope and sequence.
 - The testlets are designed to be flexibly administered across several sessions throughout the school year as students learn, offering greater coherence with the taught curriculum and providing more timely and actionable assessment results.
 - ELA testlets will be designated as beginning-of-year, middle-of-year, or end-of-year, and the testlets will be differentiated based on the complexity of passages and the skills being assessed.
- ⊕ The testlets follow a genre-based design. For each testlet, students will read two passages and answer a series of questions designed to measure key reading skills. According to New Meridian, the test will use cognitive diagnostic models (CDMs) to report meaningful student profiles that convey specific skills students have and have not yet mastered at a grain size level that can support action.

RELATION TO SUMMATIVE SCORE:

- ⊕ No decision has been made on how a summative score will be calculated but Montana intends to aggregate results from all testlets to produce a summative score.

IMPLEMENTATION TIMELINE

- ⊕ **2022-23 School Year:** Will pilot the Instructionally Aligned Assessment System in grades 5 and 7
- ⊕ **2025-26 School Year:** Expected to be fully operational

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|---|
| Subjects and Grades | ELA, 3-8; at full implementation |
| Primary Vendor | New Meridian |
| Number of Administrations | Teachers/districts can choose to administer testlets individually or in batches (for a total of 8 ELA testlets) |
| Total Testing Time | 40 minutes per testing session for no more than four hours of total testing time over the academic year |
| Results Turnaround Time | Most items will be machine-scorable to ensure timely reporting of results, however, according to New Meridian, there will be some constructed response questions scored by an external vendor |
| Adaptability | Non-adaptive in the pilot phase, but may explore adaptability in the future |

APPENDIX: STATE PROFILES

Montana and Louisiana’s Instructionally Aligned Assessment System for Math

KEY FEATURES

GOAL

The Instructionally Aligned Assessment System for math (referenced to as MAST in Montana) is designed to offer a coherent, continuous and useful balanced assessment system to better meet the goals of assessing for learning.

DISTINGUISHING FEATURES:

- ⊕ The math testlets are organized around 12 “strands” that represent grade-level content, as defined in the CCSS, with 10-13 items per testlet. Within each strand, the items will be designed across several levels of cognitive depth that reflect increasing depth of understanding of content.
- ⊕ Schools and districts can administer the testlets in any order, to fit their local scope and sequence.

RELATION TO SUMMATIVE SCORE:

- ⊕ No decision has been made on how a summative score will be calculated but the intention is math testlets could be aggregated to produce a summative score.

IMPLEMENTATION TIMELINE

- ⊕ **2022-23 School Year:** Will pilot the Instructionally Aligned Assessment System in grades 5 and 7
- ⊕ **2025-26 School Year:** Expected to be fully operational

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|--|
| Subjects and Grades | ELA, 3-8; at full implementation |
| Primary Vendor | New Meridian |
| Number of Administrations | Teachers/districts can choose to administer tests one by one (for a total of 12 administrations) or in batches of 2 or 3 throughout the year |
| Total Testing Time | Students should be able to complete 3 testlets in 40 minutes, total testing time should be about 3 hours |
| Results Turnaround Time | All test items will be machine scored |
| Adaptability | Leverages adaptive testing |

APPENDIX: STATE PROFILES

Nebraska’s Student-Centered Assessment System (NSCAS) Growth

KEY FEATURES

GOAL

The NSCAS Growth will produce information to help support learning throughout the year while also providing summative data at year’s end, streamlining assessment and prioritizing information that helps educators, parents and students drive learning.

DISTINGUISHING FEATURES:

- + The test uses a multi-phase, item-level adaptive design that will have two phases.
 - In the first phase of the test (about 27 questions), which is more diagnostic, the test will adapt on grade level to determine the student’s ability level.
 - The second phase of the test (about 13 questions) will provide a more tailored blueprint that may adapt outside of grade-level and include more challenging questions.
- + According to the Nebraska Department of Education, “each test picks up where the student left off on the last test.” For example, if a student already demonstrated grade-level proficiency (On Track) in the winter, the spring test will pick up where the student left off, and continue to adapt into the College & Career Ready level within the standards (and up into the next grade level if student performance merits it).
 - This design provides multiple chances for students to demonstrate proficiency while also eliminating the retesting of students on concepts they have already mastered
- + The test blueprint for each test is aligned to state standards and the state’s general summative blueprint. If a student gets a question they haven’t learned yet and answers it incorrectly, they may see that question again on the next administration of the test.

RELATION TO SUMMATIVE SCORE:

- + The spring assessment produces the summative proficiency scores, which reflect a student’s grade level performance at year’s end. Student’s proficiency on the fall and winter administrations influences where the adaptive summative starts. NDE states they “will use the highest score for public reporting and accountability from the year as long as the student meets the summative blueprint on the administration in order to minimize the impact of a single “bad” test experience in the spring. We will also report growth data from fall to spring once the fall and spring assessments are required.”

IMPLEMENTATION TIMELINE

- + **2020-21 School Year-Planning and Design:** Developed item-bank and carried out linking study
- + **2021-23 School Year-Preliminary Transition:** Piloted NSCAS Growth in the winter, carried out a study on the stability of through year scores to RIT scores
- + **2023-24 School Year-Operational Transition:** Will make a full transition to the through-year model

ASSESSMENT SYSTEM FEATURES

| | |
|---------------------|---|
| Subjects and Grades | ELA and Math, 3-8; at full implementation |
| Primary Vendor | NWEA |

Continues

APPENDIX: STATE PROFILES

Nebraska's Student-Centered Assessment System (NSCAS) Growth

| ASSESSMENT SYSTEM FEATURES | |
|----------------------------------|---|
| Number of Administrations | 3 (fall, winter, spring) Fall and winter are optional for 2022-2023 school year, fall and spring will be required for 2023-2024 |
| Total Testing Time | 90 minutes per test for roughly 4.5 hours total |
| Results Turnaround Time | Scores come in 24 hours. An RIT score is produced within 72 hours of completing the test. |
| Adaptability | Multi-phase, item-level adaptive |

APPENDIX: STATE PROFILES

North Carolina’s Personalized Assessment Tool (NCPAT)— Check-Ins 2.0 + Flexible Summative

KEY FEATURES

GOAL

NCPAT provides teachers, students and parents with usable information about student performance on grade-level content standards throughout the year. The NC Check-Ins 2.0 are designed to provide immediate and detailed feedback on student performance to inform classroom instruction for individual students’ needs, and provide reliable estimates to inform a student’s summative assessment experience.

DISTINGUISHING FEATURES:

- ⊕ The NCPAT includes three interims and a flexible summative assessment aligned to North Carolina grade level content standards for reading and math.
- ⊕ The Check-Ins 2.0 test is delivered online but is not adaptive. The flexible summative is the only test that is adaptive (multi-stage).
- ⊕ Check-Ins 2.0 tests are aligned to NC content standards. Schools/districts can choose to administer the tests in any order.
 - Teachers can only give Check-Ins 2.0 that are on grade level.

RELATION TO SUMMATIVE SCORE:

- ⊕ The Flexible Summative will have the same types of questions as the current EOG test. Both the flexible summative and the current EOG will have the same statistical equated scale score. The scores reported from the flexible summative will be comparable to that of the current EOG and reported using the current achievement level standards.
 - Data from students who have completed at least two NC Check-Ins 2.0 by April 1st will be used in a statistical model to provide information to route them to the appropriate adaptive set for the flexible summative. Only students with at least two NC Check-Ins 2.0 will be able to take full advantage of the measurement enhancements of the flexible summatives.

IMPLEMENTATION TIMELINE

- ⊕ **2019-20 School Year:** Pilot began.
- ⊕ **2022-23 School Year:** As the pilot enters its fourth year, North Carolina plans to pilot both NC Check-Ins 2.0 and the flexible summative in grades 4 and 7 reading and mathematics at selected pilot schools. A linking scale will be confirmed.
- ⊕ **2023-24 School Year:** Grades 3 and 6 will be available.

ASSESSMENT SYSTEM FEATURES

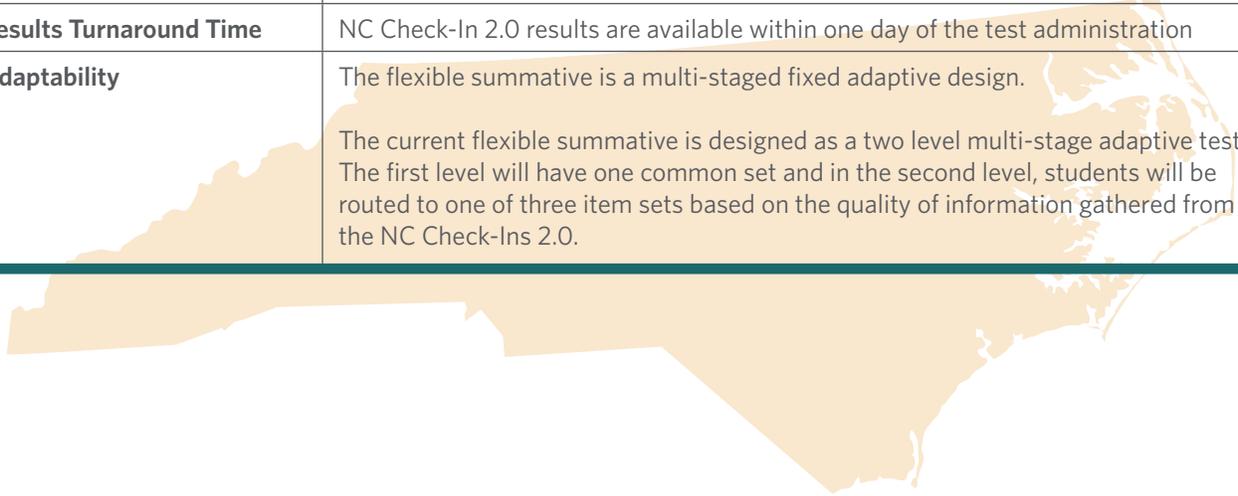
| | |
|----------------------------------|---|
| Subjects and Grades | Reading and Mathematics, 3-8; at full implementation |
| Primary Vendor | State develops its own assessments in partnership with Technical Outreach for Public Schools (TOPS) at North Carolina State University |
| Number of Administrations | 3 (+1 flexible summative assessment). The NC Check-Ins 2.0 are optional and it is up to schools and districts to choose if they plan to administer all 3. The state however recommends that if schools want to experience the full benefits of the NCPAT system, they should plan to administer at least two NC Check-Ins 2.0 by April 1st. |

Continues

APPENDIX: STATE PROFILES

North Carolina’s Personalized Assessment Tool (NCPAT)— Check-Ins 2.0 + Flexible Summative

| ASSESSMENT SYSTEM FEATURES | |
|--------------------------------|--|
| Total Testing Time | Test administrations for NC Check-Ins 2.0 is less than 90 minutes |
| Results Turnaround Time | NC Check-In 2.0 results are available within one day of the test administration |
| Adaptability | <p>The flexible summative is a multi-staged fixed adaptive design.</p> <p>The current flexible summative is designed as a two level multi-stage adaptive test. The first level will have one common set and in the second level, students will be routed to one of three item sets based on the quality of information gathered from the NC Check-Ins 2.0.</p> |



APPENDIX: STATE PROFILES

Texas' Through-Year Assessment Pilot (TTAP)

KEY FEATURES

GOAL

The goal of the pilot is to create a through-year assessment system that provides more timely and frequent feedback to inform instruction, offers multiple opportunities for students to show what they've learned and allows for in-year growth information, as well as minimize disruption to instructional time.

DISTINGUISHING FEATURES:

- + TTAP will be administered three times a year (and replace local benchmarks) and be full scope for every testing opportunity (covering entire curricular standards proportionately to the STAAR (State of Texas Assessments of Academic Readiness) design) to allow districts to preserve their local scope and sequence.
- + The test is a computer based multi-stage adaptive test. Each of the three administrations is broken up into two sections with a routing form and a second half. All students start the test with the medium level form. Depending on performance in the first half, students are either routed to a low, medium or high form. The next test administration picks up where the student left off.
 - By better aligning the questions that students see with their true abilities, TTAP will shorten the amount of questions needed to still get the same level of reliability.

RELATION TO SUMMATIVE SCORE:

- + The method for calculating the summative score is still to be determined. TTAP is exploring two potential cumulative scoring options where earlier performance can help but not hurt students' final scores.
 - Option 1: If the student scores the strongest in the third testing opportunity, their final score is used as the cumulative score for the year and their prior testing performance does not impact their cumulative score.
 - Option 2: If a student scores stronger in prior testing opportunities, compared to the last test, a weighted average formula is used to calculate the cumulative score. Their prior testing performance would help their cumulative score.

IMPLEMENTATION TIMELINE

- + **2022-23 School Year:** Pilot Launch
- + **2023-2026:** Pilot continues for three more school years
- + **2026:** Final pilot report to the legislature — earliest possible decision to potentially replace current state summative with TTAP model. TTAP will not be used for accountability during the pilot.

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|--|
| Subjects and Grades | Piloting in Science, 5; Math, 6 & 7; Social Studies, 8 |
| Primary Vendor | Pearson, Cambium |
| Number of Administrations | 3 (fall, winter, spring) |
| Total Testing Time | The first two test administrations have around 20 items each and the third test administration has around 30 items. |
| Results Turnaround Time | For Tests 1 & 2 score reports will be available the Monday after the one-week testing window. Scores for social studies and science will have a 2-week turnaround as the constructed response sections have to be hand scored. |
| Adaptability | Multi-stage adaptive |

APPENDIX: STATE PROFILES

Virginia's Growth Assessments

KEY FEATURES

GOAL

According to HB 2027, the purpose of the Growth Assessments is to provide measures of individual student growth over the course of the school year and for educators to use student growth data from these assessments to improve reading and math instruction throughout the year.

DISTINGUISHING FEATURES:

- + The test blueprint will correspond with the Standards of Learning for the student's grade-level. The test will give the student mostly on-grade level questions.
- + The test is item-level adaptive and can give students some questions above or below grade-level based on student responses.
- + Item-level reports are provided for each test attempt to include a brief item descriptor, the associated Standard of Learning being assessed, item difficulty range (Low, Medium, High), and answered correctly or incorrectly.

RELATION TO SUMMATIVE SCORE:

- + The Growth Assessments will not replace the current summative test at this time.

IMPLEMENTATION TIMELINE

- + **2021-22 School Year:** State legislation required the implementation of growth assessments beginning in the fall
- + **2022-23 School Year and beyond:** State legislation required the implementation of growth assessments in both fall and mid-year/winter

ASSESSMENT SYSTEM FEATURES

| | |
|----------------------------------|---|
| Subjects and Grades | Reading and Math, 3-8; at full implementation |
| Primary Vendor | Pearson |
| Number of Administrations | 2 (fall, winter), followed by the Standards of Learning test in the spring |
| Total Testing Time | Growth assessments are untimed, but each administration is estimated to take about half as long as the spring summative Standards of Learning test. |
| Results Turnaround Time | For online tests, results are available within 24 hours. For paper tests, results are available within 96 hours after receipt for scoring. |
| Adaptability | Item-level |

