THROUGH-YEAR ASSESSMENT: TEN KEY CONSIDERATIONS

March 2023

Nathan Dadey, Carla M. Evans & Will Lorié
Center for Assessment

Center for Assessment | 25th ANNIVERSARY

National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

The National Center for the Improvement of Educational Assessment, Inc. (the Center for Assessment) is a New Hampshire-based not-for-profit (501(c)(3)) corporation. Founded in September 1998, the Center's mission is to improve student learning by partnering with educational leaders to advance effective practices and policies in support of high-quality assessment and accountability systems. The Center for Assessment does this by providing services directly to states, school districts, and partner organizations to support state and district assessment and accountability systems.

To mark 25 years of advising states and school districts, the Center for Assessment is issuing a series of papers on important, timely aspects of assessment and accountability. This paper is one installment in that series.

ACKNOWLEDGMENTS:

We acknowledge the support and feedback from our colleagues at the Center for Assessment, especially Catherine Gewertz, Erika Landl, and Scott Marion. Any errors and omissions are our own.

SUGGESTED CITATION:

Dadey, N., Evans, C. M., & Lorié, W. (2023). Through-year assessment: Ten key considerations. The National Center for the Improvement of Educational Assessment.

THROUGH-YEAR ASSESSMENT: TEN KEY CONSIDERATIONS

TABLE OF CONTENTS

THROUGH-YEAR ASSESSMENT: TEN KEY CONSIDERATIONS

INTRODUCTION

Through-year assessment is enjoying a surge of attention. More than a dozen states are designing, piloting, or implementing versions of this assessment model to replace their current end-of-year summative test for accountability. As an organization that advises states on assessment and accountability, we understand the complexities of these kinds of assessment models. Through-year programs may provide a number of benefits, but these potential benefits must be clearly articulated and evaluated in light of their required trade-offs. We urge states to proceed with caution.

This paper outlines ten interrelated considerations for states exploring through-year designs for their federally mandated state assessment program. We use the broad definition[1] of a through-year assessment program created in 2021 by our colleagues Nathan Dadey and Brian Gong:

A through-year assessment *program* is:

1. Administered in multiple distinct sessions during a school year, and
2. Intended to support (a) the production and use of a summative determination, and (b) one additional aim.[2]

Before we detail important considerations about through-year assessment, let's review briefly what has led to the current interest in these programs.

State leaders, policymakers, educators, and families have been unhappy with key aspects of federally mandated standardized assessments since they were first required by the 1994 version of the Elementary and Secondary Education Act (ESEA), the main federal law that governs K-12 education. Two issues frequently top the list of complaints about these state assessments: 1) they take too much time, and 2) the results arrive too late in the year, with too little useful information, to help teachers adjust instruction.

These issues—and the desire for assessment information that can help education leaders monitor and support schools—are the biggest drivers of the current surge in interest in through-year assessment.

This isn't the first such surge in interest, only the latest. The U.S. Department of Education's 2010 "Race to the Top" grant program invited assessment designs that based students' annual proficiency determinations on tests given throughout the year, rather than just once, at year's end.[3] The

[1] The field is still debating whether the "through-year" label should be restricted to programs that base annual proficiency determinations on both within-year and end-of-year results. The definition we use in this paper is sufficiently broad to include programs that use only end-of-year results in annual determinations.

[2] Dadey, N., Gong, B., Lorié, W., & Marion, S. (2021). Through-year assessment virtual convening, Session 1: Definition, aims, and use cases. National Center for the Improvement of Educational Assessment. https://www.nciea.org/library/definitions-aims-and-use-cases/

[3] U.S. Department of Education. (2010). *Catalog of Federal Domestic Assistance (CFDA) Numbers:* 84.395B (Comprehensive Assessment Systems grants) and 84.395C (High School Course Assessment Programs grants). 75 FR 18171. Washington, DC. https://www.federalregister.gov/d/2010-8176

Partnership for Assessment of Readiness for College and Careers (PARCC), one of the two state consortia that won federal grants to design those tests, proposed a model that featured multiple test sessions throughout the year. PARCC dropped that model in its final design, citing concerns about cost, testing time and local control of curriculum.[4]

The 2015 version of the ESEA, the Every Student Succeeds Act, permitted states to use "multiple interim statewide assessments" to create annual student proficiency determinations, further establishing through-year as an option for federally mandated standardized assessments.[5] The U.S. Department of Education has also encouraged states to explore the idea through its Innovative Assessment Demonstration Authority and Competitive Grants for State Assessments programs.[6,7] Currently, 13 states, including some we advise, are exploring through-year assessment programs.[8]

In our 25 years of working with states, we've seen clearly that every assessment program requires trade-offs. Because they pose many conceptual, technical, and practical challenges, through-year programs will require even harder choices among competing priorities and goals than typical programs do. Our organization detailed many of these challenges and trade-offs in our 2021 through-year assessment virtual convening.[9] The evidence-based research on through-year assessment programs has not deepened substantively since then.

This paper is divided into three sections, each exploring a cluster of important considerations about through-year assessment: (1) their conceptualization and design; (2) their relationship to curriculum and instruction; and (3) their logistical and technical challenges. In each section, we provide practical recommendations for states exploring through-year assessment programs. We end with concluding thoughts and an appendix that organizes the through-year assessment programs states are currently using—or considering—into common models. Finally, we note that each program will likely uncover new and unique challenges. The considerations we present should not serve as a comprehensive checklist.

We hope this paper finds its way into the hands of state superintendents, department of education leaders, board of education members, and state legislators, to help inform their thinking if they are considering through-year assessment programs for accountability. We urge states to proceed slowly and to carefully interrogate the trade-offs inherent in any through-year assessment program, bearing in mind their state's vision and goals.

We urge states to proceed slowly and to carefully interrogate the trade-offs inherent in any through-year assessment program.

[4] Gewertz, C. (2011, July 7). State consortium scales back common-test design. Education Week. https://www.edweek.org/teaching-learning/state-consortium-scales-back-common-test-design/2011/07

[5] Dadey, N. (2018, August 24). When it comes to getting summative information from interim assessments, you can't have your cake and eat it too. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/assessment/when-it-comes-getting-summative-information-interim-assessments-you-cant-have-your

[6] Marion, S. & Evans, C. (2021, August 20). Following their lead: Conversations with innovative assessment demonstration authority state leaders. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/following-their-lead/

[7] Munyan-Penney, N. & Barone, C. (2020). Innovative assessment pilot: Louisiana's LEAP 2025 humanities assessment. Education Reform Now. http://edreformnow.org/wp-content/uploads/2020/11/Louisiana-IADA-LEAP-2025-Humanities.pdf

[8] EducationFirst. (2022). What are through-year assessments? https://www.education-first.com/download/12842/?tmstv=1673362241

[9] National Center for the Improvement of Educational Assessment. (2021). Through-year assessment virtual convening. https://www.nciea.org/library/?_resource_events=convening

CONCEPTUALIZATION AND DESIGN

Consideration 1: Research on through-year assessment is limited.

With a few notable exceptions, through-year assessment programs are not operational. They're just now being explored and developed, so research on these programs is limited.

Until the research base is more robust, those interested in through-year assessment programs cannot fully substantiate the many claims that have been made about their potential value, such as their ability to generate instructionally useful information. The field also lacks definite solutions to the many technical, logistical, and policy challenges that arise as a state moves to a through-year assessment model.

Until the research base is more robust, those interested in through-year assessment programs cannot fully substantiate the many claims that have been made about their potential value.

Most of the currently available literature, including this paper, offers guidance and considerations on various aspects of design, implementation, and evaluation. But it does not examine implementation or provide evidence of how effectively through-year assessment programs achieve the various goals their supporters have set for them.

We are eager to see the body of research develop through the work that more than a dozen states, with their various partner organizations, including the Center for Assessment, are doing. There is still a lot to learn about how—and whether—these assessment programs can achieve their goals, and what, if any, negative unintended consequences may emerge.

Consideration 2: No through-year program can solve all assessment problems.

States that want to replace their current end-of-year assessments with a through-year program are typically trying to solve one or more problems. They may want to reduce the amount of time testing takes up at the end of the school year, for example, or make test results more useful to classroom teachers. Perhaps they're hoping to give students multiple opportunities to show what they know and can do, or they want to get data more frequently to help leaders monitor programs and provide student support services or interventions.

Building a through-year program that addresses these kinds of problems requires acknowledging a difficult truth: Like any assessment program, through-year assessment programs cannot solve every problem. It might be able to address just a few—and perhaps only one—well. Prioritizing the most important problems for a through-year model requires trade-offs.

Like any assessment program, through-year assessment programs cannot solve every problem.

A through-year program that is meant to reduce overall testing time, for example, likely cannot also provide more instructionally useful information to teachers. Providing the kinds of fine-grained information helpful to teachers requires more time, often in the form of more assessment, not less.

Federal accountability also constrains the ways stakeholders identify and prioritize the problems they want their through-year system to address.[10] The current federal technical requirements for state summative assessments lead to assessments that are highly standardized, generally value breadth over depth, are easily and reliably scored, and focus on whether students have developed proficiency on the state's academic standards.[11] Many of the problems that advocates would like to see through-year solve, however, would require pushing back against these kinds of requirements.

Consider a hypothetical through-year assessment program that produces student proficiency data to meet federal accountability requirements, but is also expected to generate information that's helpful in guiding classroom instruction. One approach to producing instructionally useful information would be to investigate what teachers need to know about their students throughout the year and design assessments to reflect that.

Given the curricular and instructional diversity in a typical state, however, the list of what teachers need to know would likely suggest many distinct assessments, in various formats, each relating in a different way to the state's academic content standards. Accountability, however, requires assessments to be comparable across all students. The surest way to achieve comparability is standardization: to have common items or tasks that, taken together, cover the breadth and depth of the state's academic content standards. These kinds of tensions must be navigated for a through-year assessment program to address its problems and meet its goals.

Can a through-year program aiming for instructional utility find the right trade-offs between the kinds of information that may help educators and the need for comparability and standardization? Perhaps the list of what teachers need to know could be translated into a manageable set of assessments that can be flexibly administered. But, if so, would those tests support sufficient levels of comparability? These questions, and others, remain unanswered. Careful research on high-quality implementation is needed.

In the meantime, states considering through-year assessment must prioritize the problems they want to address, clearly articulate their goals, and formulate a plausible approach to meeting them.

States considering through-year assessment must prioritize the problems they want to address, clearly articulate their goals, and formulate a plausible approach to meeting them.

Agreeing on the problem(s) a through-year program is meant to solve, or the goals it should achieve, requires substantial stakeholder engagement. Stakeholders are likely to disagree, so reaching a shared understanding on the priorities of the program could be challenging. Even with clear priorities and a limited number of test purposes, states will still have to navigate the tensions inherent in addressing those problems while still meeting, or redefining, the technical requirements for statewide accountability assessments.

[10] Gianopulos, 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

[11] Badrinarayan, A. & Dadey, N. (2023). In press.

Consideration 3: Each through-year assessment design requires trade-offs.

Through-year assessment designs vary;[12,13] we have categorized them into a number of possible models (currently, states have only adopted a limited subset of these models; see appendix). To choose a design wisely, states must make clusters of interconnected decisions. Ideally, all these decisions should be grounded in a theory of action that clearly articulates what the through-year assessment program is meant to do and how it will do it.[14,15]

Design choices that states make under one theory of action will often mean they can't embrace an alternative theory of action. If a through-year program's theory of action holds that teachers need to know how their students are doing on material taught in a recent unit, for example, then the program should, at a minimum, allow relevant content to be assessed after each unit and, ideally, connect to the curriculum in rich ways.

If, instead, the theory of action holds that teachers need to know how their students would perform if they took the end-of-year test at mid-year, and how their results compare to others across the state, then the test should be more curriculum-agnostic. The assessment program designed for the first theory of action cannot provide the information for the second, and vice-versa. Thus, each design path requires trade-offs in the information it provides.[16]

Here's a sample of just some of the decisions states will face, and the trade-offs they might need to make, based on their chosen theories of action:

[12] Gianopulos, G. (2019).

[13] National Center for the Improvement of Educational Assessment. (2021).

[14] Badrinarayan, A. & Dadey, N. (2023).

[15] Bennett, R. E., Kane, M., & Bridgeman, B. (2011). Theory of action and validity argument in the context of through-course summative assessment. https://www.ets.org/Media/Research/pdf/TCSA_Symposium_Final_Paper_Bennett_Kane_Bridgeman.pdf

[16] Marion, S. F. (2023). Validity for through-year assessments. In R. J. Tierney, F. Rizvi, & K. Erkican (Eds.), *International Encyclopedia of Education* (Vol. 13, pp. 212–219). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.09071-0

EXAMPLE DECISIONS	TRADE-OFF(S)
Whether and how the through-year assessment program connects to curriculum.	Curricula vary widely across schools and districts in any given state.[17] Connecting through-year assessments to curricula, in the hope of providing instructionally relevant information, requires navigating the complexities of differences in curriculum implementations across the state. Through-year designs that allow for greater connection to curricula and instruction, such as those that afford a great deal of flexibility, pose challenges to comparability and standardization.
What content to test, including whether each administration should assess the depth and breadth of the state's academic content standards, or just a subset of these standards, and if so, whether standards will be revisited across administrations.	Greater local control, in the form of flexibility, allows choice on which standards are assessed, and when. But flexibility is in tension with the standardization required for accountability uses. The technical and logistical demands of flexibility are not insignificant, and will require close partnerships between states and local entities, which adds burden and requires capacity.
When to test, and whether there will be any flexibility in the timing or order of administrations.	
How often to test, and how many administrations there will be during the year.	The more test administrations states choose, the more often they can get information that's useful for program monitoring. But more testing sessions means more time dedicated to testing and less to instruction.
The weightings, business rules, or other procedures by which each administration contributes to a student's annual proficiency determination.	Some states may put all (or most) of the weight of a student's annual determination on the final through-year assessment to signal the importance of end-of-year performance, to mitigate the impact of missing data, or to address concerns with opportunity to learn. Doing so, however, may result in perceptions that earlier administrations do not "count" and are therefore less important.

All these decisions affect how well a through-year assessment program can solve the original problems that motivated the move away from a single end-of-year state test. States should carefully articulate the likely trade-offs before deciding to work toward a through-year assessment program.

[17] Dadey, N. & Badrinarayan, A. (2022, April 21). In search of the "just right" connection between curriculum and assessment. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/in-search-of-the-just-right-connection-between-curriculum-and-assessment/

Consideration 4: The needs of students with disabilities and English learners must be considered in the earliest stages of design.

Every state assessment program—through-year or not—must provide all students with a fair opportunity to show all they know and can do, without having to overcome factors irrelevant to what is being tested. All student groups, including English language learners and students with disabilities, should be able to participate in through-year assessment programs in ways that yield valid inferences about what they know and can do.

One alternate assessment consortium, Dynamic Learning Maps, has developed through-year assessments for students with the most significant cognitive disabilities.[18] But there has not yet been corresponding work on the ways general through-year assessment programs can support—and affect—students who need accommodations and support. Emerging lessons from work on alternate assessments, as well as from interim assessment, can be helpful.

In a recent report, the National Center on Educational Outcomes[19] offered recommendations about how state departments of education can ensure that all students participate in interim assessments. The organization divided its concerns and recommendations into five areas: data use, interpretation and reporting; participation, accessibility, technical issues, and the role of academic standards.

At a minimum, states should use universal design guidelines,[20] their current test design guidelines, and the recommendations on interim assessments as they consider the needs of students with disabilities and English language learners in the earliest stages of designing or implementing a through-year assessment program.

Additionally, states should consider all students in the earliest stages of design. They should interrogate how information collected throughout the year might disadvantage students who typically need more time and opportunities to learn, or who have more variability in the scope and sequencing of their learning.

In those interrogations, states could focus on factors that may have different impacts for various student groups, such as English learners and students with disabilities. Students from these groups often have patterns or progressions of learning that differ from the general population of students, as well as from one another. These differences have implications for through-year assessment, including whether and how to include score information from earlier in the year in annual proficiency determinations. Other issues—such as diminished instructional time and disruptions to learning because of frequent testing—may also be more pronounced for these groups of students.

A through-year assessment program that works well for most students may not work well for specific student groups.

[18] The Dynamic Learning Maps consortium has offered an "instructionally embedded" model since 2016 (see Clark & Karvonen, 2021). Clark, A. K., & Karvonen, M. (2021). Instructionally embedded assessment: Theory of action for an innovative system. *Frontiers in Education, 6,* 401. https://www.frontiersin.org/articles/10.3389/feduc.2021.724938/full

[19] Lazarus, S. S., Hinkle, A. R., Liu, K. K., Thurlow, M. L., & Ressa, V. A. (2021). Using interim assessments to appropriately measure what students with disabilities know and can do: Advisory panel takeaways and NCEO recommendations (NCEO Report 427). National Center on Educational Outcomes. https://nceo.umn.edu/docs/OnlinePubs/NCEOReport427.pdf

[20] Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). Universal design applied to large scale assessments (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. https://nceo.info/Resources/publications/OnlinePubs/Synthesis44.html

Ultimately, a through-year assessment program that works well for most students may not work well for specific student groups.

Summary Recommendations for Conceptualization and Design

- **Go slowly.**
  - Consider leveraging the experiences of states that are already designing through-year models. States that move forward early should rely heavily on rapid prototyping, stakeholders' input, and small-scale pilots to continuously test claims and gather evidence about technical quality, usefulness, and consequences.
  - Consider extending procurement timelines to allow the state, and its vendors, time to think through, pilot, and address the complex issues involved.
- **Think long-term.**
  - Any move to through-year should reflect a state's long-term planning and vision. In some cases, it may be worth temporarily extending the current assessment system to allow for more planning time.
- **Develop a theory of action with stakeholder feedback.**
  - Choose one problem—or cluster of related problems—as a target (in addition to the basic purpose of measuring student proficiency).
  - Recognize the necessary trade-offs and choose a design that's best suited to the one or two purposes the state has prioritized.
  - Clearly articulate a theory of action for the proposed through-year program that's based on the state's vision, goals, and policy environment. States should take the time needed to define and articulate the problems they want the through-year system to solve.
  - Engage deeply with students, families, teachers and policy leaders. Draw on all possible resources to determine how the through-year program will solve the chosen problems.
- **Design with intention for all students.**
  - Design for, and then examine the system in terms of, the experiences of student groups such as English language learners and students with disabilities. A shift to testing throughout the year will have different implications for these groups than for the general population.
- **Consider the entire assessment ecosystem.**
  - Through-year programs add at least one purpose to the state summative assessment program, essentially asking the program to do more. States should carefully consider whether the state summative assessment program can and should do more, or whether the purposes are better served outside of that program.
  - In some cases, states might address their added purpose, or purposes, by developing assessments and resources *outside* of the summative state assessment program—ideally a balanced, coherent *system* of assessments. That would mean having separate assessments that work together, instead of a single assessment that tries to serve multiple purposes.

States should carefully consider whether the state summative assessment program can and should do more.

CURRICULUM AND INSTRUCTION

Consideration 5: Through-year assessment programs could reduce local control over curriculum.

Through-year assessment programs require states to decide how and when they will assess students. These decisions affect how teachers, schools, and district administrators handle instruction.

Typically, statewide assessments test students on the depth and breadth of the state's content standards at the end of the year, which allows districts to structure the timing and pacing of instruction as they wish, as long as they cover the most important content by the time students are tested. If states test students multiple times per year, however, districts must consider *when* they teach content. Currently, such decisions vary widely from district to district, reflecting curriculum variation.[21]

If states test students multiple times per year, however, districts must consider when they teach content.

Some states exploring through-year assessment are addressing this issue by testing the full depth and breadth of state content standards in each test administration. With this approach, each administration mirrors the others. This is similar to what districts and schools are already doing when they administer interim assessments that cover a full set of grade-level standards in the fall, winter, and spring.[22] Other states have chosen to assess a subset of the standards in each test administration (See appendix for details on states using each of these two approaches).

The theoretical benefit of the first option is that instruction would not need to be tethered to the testing schedule, allowing teachers to design the scope, pace, and sequence of their instruction as they wish. It might help teachers track students' growth across the series of test sessions, and help identify students who consistently need extra support. In practice, however, this option may encourage schools to cover more state standards earlier in the year than prescribed by their curriculum, or to cover content quickly ahead of a test, and then go back and dive more deeply. This model also increases the chance that students will be tested on topics they haven't studied yet, particularly in mathematics.

The theoretical benefit of the other approach—covering only a subset of the standards in each test administration—is that each test can focus on what has just been taught. But variation in curriculum means that some scopes and sequences would align better than others with the assessed content. Some educators might adjust their instruction to match the assessments, which reduces their discretion and agency, while others may accept that misalignment, which could influence their students' test performance.

Through-year assessment programs that give schools and districts more control over the timing of test administrations, such as with modular assessments or "testlets" that can be arranged around instruction, might help districts retain greater control over curriculum. But those designs greatly increase logistical complexity. They also could encourage teachers to move away from deeper learning goals, which ask students to integrate across standards in sophisticated ways.

[21] Dadey, N. & Badrinarayan, A. (2022).

[22] Typically, commercial interims do not assess a state's academic content standards; they assess the standards of a related framework (such as the Common Core State Standards) with a reasonable degree of alignment to the state's content standards.

Consideration 6: Using state tests—through-year or not—to directly inform instruction requires radical changes.

State summative assessments are designed to provide annual determinations of student proficiency that, when combined with other information, serve as the basis for identifying schools in need of additional support and intervention in federally mandated accountability systems. Tests designed for this purpose influence classroom instruction *indirectly*, by triggering those supports and interventions and by enabling state officials to monitor school quality and student success across years. State summative assessments also signal what is valued, in terms of performance on the state standards. These mechanisms for informing instruction happen over the course of years, and don't affect students' instruction in the year they took the test.

Most people who want tests that inform instruction picture a different process, one that occurs in shorter time cycles and affects the instruction of the students who took the test. In such a scenario, a teacher would receive results immediately after the test is given, review those results, and adjust instructional plans for the next few lessons.

An assessment program that informs instruction in this more *direct* way must have certain characteristics that are absent from typical state accountability testing programs. Each test should be closely connected to the instruction that immediately precedes each test administration, for example. It should be relevant to upcoming curricular units, and it should communicate results in a way that enables teachers to plan future instruction for their current students.

Building in this kind of direct instructional utility is a big ask for a state testing program that must also sample the breadth and depth of a state's content standards. Building in any one of these characteristics is a challenge. Here, we illustrate how the third—communicating results in a way that enables teachers to plan future instruction—extends state assessment well beyond its current boundaries.

A test that is directly useful to instruction must be designed to capture, and report, qualitative information on how well students are learning, and what they need, in the context of a lesson or unit. This is information that makes student thinking visible, and serves as evidence of what needs to change in teaching or learning, for which students, and why. Current state tests do not provide this kind of information.[23]

For a state test to directly inform instruction, then, it must change in an important way: It must produce reports that speak to the quality of student work. Those reports may need to illustrate the ways in which students have met or fallen short of the learning goals of prior units. They may need to show actual student work,[24] possibly annotated by test scorers or instructional partners. Those reports will need to address practical instructional implications at levels

For a state test to directly inform instruction, then, it must change in an important way: It must produce reports that speak to the quality of student work.

[23] Evans, C. M. (2022, June 22). Can a state test inform instruction?: The typical, the improbable, and the possibilities in the messy middle. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/can-a-state-test-inform-instruction/

[24] Marion, S. (2019, March 21). Five essential features of assessment for learning. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/five-essential-features-of-assessment-for-learning/

deeper and more specific than the boilerplate language often found on state assessment reports. State assessment would be drawn into curriculum and instruction, territory it has long avoided.

Creating assessments that produce fine-grained accounts of how students are learning the enacted curriculum—information teachers can use directly to inform instructional strategies—requires more than a revision of current state assessments. It requires designing entirely new assessments and fundamentally shifting the way curriculum, instruction, and state assessments interact.

This shift, in turn, alters the relationship between states and local districts. A state must take on additional roles in student learning that were once reserved for local education agencies. When a through-year assessment program is meant to directly inform teaching and learning, the state becomes responsible for providing that information. For this to happen, state agency staff must deeply understand teaching and learning and how they play out in multiple contexts across the state. Filling these roles demands greater capacity from states, their vendors, and other state partners.

Consideration 7: Through-year programs may compound existing tensions between accountability and instruction.

State assessment is a high-stakes proposition for schools. Their performance is rated publicly through test-based accountability systems. Those ratings can spark public criticism of schools' performance. They can also have farther-reaching implications, such as influencing property values, since ratings are often used as selling points for homes and neighborhoods.

The literature is full of cases documenting the ways accountability pressure can have a negative impact on instruction.[25] Teachers often feel they should focus heavily on what is tested. Doing so can narrow the curriculum, resulting in a far heavier focus on English language arts and mathematics than on social studies, science, art, music, and physical education. Teachers may also shift curriculum content to focus on basic skills, or they may spend too much time on test-preparation activities.

Given these widely noted negative effects of current state tests, states considering a through-year assessment model must ask two important questions: 1) Will adding more state assessment during the year compound these problems? 2) Are there specific design or mitigation strategies that can be used to make sure this doesn't happen?

States might be able to mitigate a through-year system's potential negative effects on instruction by deciding not to use the results from within-year administrations—the fall and winter sessions, for instance, in a system that features fall, winter and spring administrations—when calculating students' summative proficiency determinations. This would eliminate the stakes attached to the earlier test administrations, and might reduce the pressure teachers feel to mimic in their classrooms the assessment practices and content emphases on the state tests.

Alternatively, states might choose to design a through-year assessment program based on rich performance tasks or other models of good classroom assessment, which would shape instruction in different ways.

[25] Faxon-Mills, S., Hamilton, L. S., Rudnick, M., & Stecher, B. (2013). New assessments, better instruction? Designing assessment systems to promote instructional improvement. Washington, DC: RAND Education. https://www.rand.org/pubs/research_reports/RR354.html

Whatever pathway they choose, states should be proactive on the issue of tests' instructional impact, and use their theory of action to logically connect the design of their through-year assessment program to the types of instructional actions they would like to support.

Summary Recommendations for Curriculum and Instruction

- **Define, create and share expectations with stakeholders.**
  - Communicate with stakeholders clearly about how through-year would inform instruction (broadly speaking) and how it would not.
  - Carefully manage expectations that through-year assessments will directly inform instruction. Families and educators will likely seek instructional utility, even if the goals of the program don't include it.
- **Define what it means to inform instruction.**
  - If the through-year assessment program is meant to inform instruction, clearly define the processes, actions or activities that will inform instruction, and who will be performing them.
  - Critically examine arguments about the specific ways the through-year assessment program can and cannot inform instruction.
- **Understand and address the varied scopes, sequences and pacings of curriculum and instruction.**
  - Analyze and articulate the ways local control will be affected by decisions about the interaction of assessed content with curriculum scope and sequence.
  - Review curriculum materials to evaluate the implications for the timing and content coverage of each test, or rely on national work that does so as a starting point.[26]

[26] Cole, S. K. & Swanson, C. (2022). Content progressions & clustering across instructional materials: Viability for supporting the design of a through-year assessment model. Smarter Balanced. https://portal.smarterbalanced.org/library/en/sb_content-progressions_through-year-assessment.pdf

LOGISTICAL AND TECHNICAL

Consideration 8: Through-year assessment programs likely require more time, money, and resources than current state testing programs.

Moving to a through-year assessment program will demand more of state and local agencies in both design and implementation.

States will have to design, administer, monitor, and report on multiple tests over the course of the year—a heftier responsibility than with only one summative test per year. This will almost certainly require states to assign additional personnel, ask their vendors to supply this added administrative capacity, or both.

Almost every aspect of statewide assessment is complicated by a through-year model: development and maintenance of the item pool, form assembly, administration, scoring, scaling, equating and reporting. These complications are more pronounced in some through-year assessment models than in others, but all require greater complexity than do typical statewide assessment programs. Below we examine just two areas—administration and missing data—to illustrate the complexities involved in through-year assessment programs.

In a through-year program, testing imposes greater demands not only on the state, but on schools and districts.[27] Local administrators will have to spend more time on—and potentially assign more personnel to—testing, since it will occur several times a year, with all the complexities inherent in giving standardized tests for accountability.

In a through-year program, testing imposes greater demands not only on the state, but on schools and districts.

They'll have to train current staff to administer these new tests, for example, and divert those staffers from their normal duties several times a year. Schools and districts will also need to reconfigure school schedules to accommodate multiple test sessions. Reporting and rostering records will need to be updated, and staff assigned to additional reporting data-validation cycles.

States and their local districts using through-year assessments will also have to spend more time on security and compensating for missing data. Student mobility and absences translate into missing data. In typical end-of-year testing programs, make-up days capture most, if not all, absent students. Introducing multiple make-up days to the school year would almost certainly disrupt school operations.

Other ways to address missing data come with their own challenges. At one extreme, a program could base students' proficiency calculations only on the scores from the year's final test administration, so schools would need makeup scores from only that session. But this solution could create an incentive for students to skip school or otherwise opt out of testing on within-year testing days, since those don't "count."

At the other extreme, if schools offer make-ups for every test day absence, they could easily find themselves continuously testing, especially at large schools with high levels of chronic absenteeism. Whatever solution they choose, schools and districts are likely to find that variation in the number of

[27] New Teacher Center. (2022). Through-year assessments: Practical considerations for LEA implementation. Smarter Balanced. https://newteachercenter.org/resources/through-year-assessments-2022/

test sessions students complete will be a source of measurement error, and possibly bias, since student absenteeism correlates with test performance.

Consideration 9: Through-year programs will reduce overall testing time only if districts use them in place of their own interim tests.

Some states are drawn to through-year assessment programs because they think it will reduce the overall amount of time students spend taking tests. For that to happen, state and local assessment practices must dovetail in important ways.

No through-year assessment program to date appears to reduce the amount of time students spend on state-required tests. In fact, compared to current statewide summative assessments, they actually increase the amount of time students spend on those tests.

Imagining any significant reduction in overall testing time requires recognizing that state-required tests are just part of the assessment picture.

Many districts require their own interim tests several times a year, on top of school and classroom assessments. One study found that students actually spend more time taking district-required tests than they do on state-required assessments.[28]

If local districts drop their own interim assessments and use their state's new through-year assessment program instead, then total testing time may be reduced. Such a system might also afford greater coherence, since its parts would be designed to work together. Currently, local interim tests aren't necessarily aligned to state content standards, and don't use the state test's reporting scale or achievement level descriptions.

If a district opts to maintain its own interim assessments in addition to administering the state's mandated through-year assessment program for accountability, overall testing time will increase. In addition, testing time would increase for districts that don't currently require interim tests in a shift to a through-year assessment program, since their students would be moving from one required test to two or more.

If a district opts to maintain its own interim assessments in addition to administering the state's mandated through-year assessment program for accountability, overall testing time will increase.

Consideration 10: Making claims about student proficiency is more challenging with a through-year assessment program than with current state testing.

State academic content standards are end-of-year expectations. Accordingly, state tests assess what students know and can do at the end of the year, and the claims we make—what we say about what students know and can do—is based on the end of the year. Through-year assessment programs break this chain of logic by assessing what students know and can do not only at the end of the year, but at multiple points during the year.

[28] Lazarin, M. (2014). Testing Overload in America's Schools. Center for American Progress. https://eric.ed.gov/?id=ED561097

The shift to through-year assessment calls into question the summative claims that can be made about student proficiency. If students are learning throughout the year, then the earlier administrations can't capture what they'll know and be able to do later in the year. By the end of the year, students may have forgotten, or better mastered, the content that was already tested. Importantly, between the earlier administrations and the end of the year, students may have covered content that would have enabled them to score better if they'd known it during those earlier test sessions.[29,30,31]

When students haven't yet experienced the full year's instructional time, states must explicitly recognize that assessment information gathered earlier in the year provides limited evidence for end-of-year student proficiency, or they must revise that claim. Through-year assessment requires us to consider what is most important: what students know and can do at the end of the year, or what students know and can do at various points during the year. To be clear, either approach is reasonable. Each approach, however, reflects different priorities about student learning.

Consider this issue by comparing the creation of annual student determinations to grading. Should a student's grade be based on an end-of-year final test alone, or on the average of three end-of-semester tests and the end-of-year final? Traditional statewide summative assessment generally functions like the former, a grade based only on an end-of-year test. Through-year assessment, however, potentially allows for other options, like an average across tests.

The averaging option isn't the only one for through-year assessment programs; there are a number of possible approaches to the creation of annual determinations. The key is to clarify the state's values about *when* students demonstrate proficiency on *which* standards. This clarification helps define the claims the state wishes to make about students.

This conceptual shift has technical implications. States will have to decide how to aggregate results across test administrations to produce required annual determinations of student proficiency. How the results are aggregated—and whether and how weights are applied—will affect the range of claims states can make about student performance.[32,33,34]

The shift in claims also has practical implications. An annual proficiency determination is typically compensatory: Higher performance in one part of the end-of-year test, such as number sense and operations in math, can offset lower performance in another. In through-year assessment programs that divide content and assign it to specific test sessions, it's unclear how this kind of compensatory relationship would work. This becomes particularly challenging if the goal is to support a claim about students' *end-of-year* proficiency. Alternatively, states may seek to redefine the claim they will make.

[29] Dadey, N. & Gong, B. (2017).

[30] Gong, B. (2021a, March 3). Could two through year assessment designs provide both summative and instructional information? An exploration of feasibility for a through year assessment system. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/could-two-through-year-assessment-designs-provide-both-summative-and-instructional-information/

[31] Gong, B. (2021b, February 23). Why has it been so difficult to develop a viable through year assessment? National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/state-testing/why-has-it-been-so-difficult-develop-viable-through-year-assessment

[32] Dadey, N. & Gong, B. (2017).

[33] Lorié, W. (2021, May 5). Educating Pablo: A play about through course assessment systems. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/educating-pablo-a-play-about-through-course-assessment-systems/

[34] Wise, L. L. (2011). Picking up the pieces: Aggregating results from through-course assessments. Center for K-12 Assessment & Performance Management at ETS. https://www.ets.org/Media/Research/pdf/TCSA_Symposium_Final_Paper_Wise.pdf

Student groups that tend to improve more than others between earlier and later test administrations could be at a disadvantage in a model that divides up content and assigns it to specific test sessions, assuming all administrations of a through-year test "count." Solving this technical challenge is not trivial. States will need to work closely with their technical advisors and vendors to decide on claims, align aggregation models appropriately, and evaluate disproportionate impact on student groups. Finding out what students, parents, teachers and education leaders value in a test's claims, and taking that into account in design, must also be considered.

Some states are getting around these thorny claims-related issues by excluding within-year results from their annual proficiency determinations (See appendix). Students' proficiency determinations are based only on their end-of-year assessment. This is the most common model among states currently designing or piloting through-year assessment programs (See Model 5, appendix).

Summary Recommendations for Logistical and Technical

- **Analyze the necessary investments and potential returns.**
  - Evaluate the return on investment, both for the state education agency and local districts. Interrogate the logistics, including:
  - How much money, time, and effort will the through-year program require over its lifespan? Can the state provide them over the long term?
  - How many additional full-time equivalent personnel will be needed?
  - How many test days will through-year add to the current state test schedule? Is this additional time reasonable? Is it legal? (Some states cap testing time.)
  - How will the required testing and make-up days differ across schools of different sizes and levels of chronic absenteeism?
- **Examine local interim assessment practice.**
  - States should ask their districts whether they can (or should) use a state-provided through-year in place of their local assessments.Through-year cannot reduce overall testing time if districts that use interim tests keep using them.
- **Define the summative claim.**
  - Consider the kinds of claims about student performance the through-year program aims to support. Can the claim be clearly articulated? Is it an end-of-year claim, or an alternative conditional on timing and content?
  - Articulate the parameters that would need to be set around timing, content, and aggregation to support the intended claim. Are those parameters acceptable to stakeholders?
  - Consult with technical experts and stakeholders on the implications of various ways of handling score aggregation and what that means for student absences on testing days.

CONCLUDING THOUGHTS

Implementing a through-year assessment program is rife with challenges, and states are facing those challenges on a daunting landscape: little is known about how well these programs work, and under what conditions.

Many people connected to education, from state legislatures and education department leaders to assessment vendors and advocacy groups, are advancing arguments about the value and promise of through-year programs for accountability. Those visions may ultimately be validated by research and practice.

But right now, there isn't enough research to show that through-year programs will resolve the two most frequently cited problems with state assessments: That they take too much time, and don't directly support instruction. As we've shown here, through-year programs also pose many significant technical, practical and policy challenges.

Until there is sufficient evidence that through-year programs will do more good than harm, we urge states to proceed slowly and deliberately. If through-year is to become a force in the positive evolution of assessment, it will need a strong foundation, and states can play important roles in building it.

We urge states exploring through-year to begin by specifying their vision, goals, and theory of action, and by asking: What problem(s) are we trying to solve, and how, exactly, would a through-year assessment program solve them? States must interrogate all claims about what through-year systems can reasonably do, and how they would do it.

States and their vendor partners can help build a research base for through-year by prototyping various designs, running small-scale pilots, gathering information from stakeholders about how well they worked and what needs to change, and widely sharing this information. In these ways, states might move forward with minimum impact as more clarity emerges.

REFERENCES

Badrinarayan & Dadey. (2023), in press

Bennett, R. E., Kane, M., & Bridgeman, B. (2011). Theory of action and validity argument in the context of through-course summative assessment. https://www.ets.org/Media/Research/pdf/TCSA_Symposium_Final_Paper_Bennett_Kane_Bridgeman.pdf

Cole, S. K., & Swanson, C. (2022). Content progressions & clustering across instructional materials: Viability for supporting the design of a through-year assessment model. Smarter Balanced. https://portal.smarterbalanced.org/library/en/sb_content-progressions_through-year-assessment.pdf

Clark, A. K., & Karvonen, M. (2021). Instructionally embedded assessment: Theory of action for an innovative system. Frontiers in Education, 6, 401. https://www.frontiersin.org/articles/10.3389/feduc.2021.724938/full

Dadey, N. (2019, March 7). Providing flexibility to schools and districts while meeting ESSA assessment requirements. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/considering-interim-assessments-and-summative-information

Dadey, N. (2018, August 24). When it comes to getting summative information from interim assessments, you can't have your cake and eat it too. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/assessment/when-it-comes-getting-summative-information-interim-assessments-you-cant-have-your

Dadey, N. & Badrinarayan, A. (2022, April 21). In search of the "just right" connection between curriculum and assessment. National Center for the Improvement of Educational 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/resource-library/using-interim-assessments-place-summative-assessments-consideration-essa-option

EducationFirst. (2022). What are through-year assessments? https://www.education-first.com/download/12842/?tmstv=1673362241

Evans, C. M. (2022, June 22). Can a state test inform instruction?: The typical, the improbable, and the possibilities in the messy middle. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/can-a-state-test-inform-instruction/

Faxon-Mills, S., Hamilton, L. S., Rudnick, M., & Stecher, B. (2013). New assessments, better instruction? Designing assessment systems to promote instructional improvement. Washington, DC: RAND Education. https://www.rand.org/pubs/research_reports/RR354.html

Gewertz, C. (2011, July 7). State consortium scales back common-test design. *Education Week.* https://www.edweek.org/teaching-learning/state-consortium-scales-back-common-test-design/2011/07

Gianopulos, 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. (2021a, March 3). Could two through year assessment designs provide both summative and instructional information? An exploration of feasibility for a through year assessment system. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/could-two-through-year-assessment-designs-provide-both-summative-and-instructional-information/

Gong, B. (2021b, February 23). Why has it been so difficult to develop a viable through year assessment? National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/state-testing/why-has-it-been-so-difficult-develop-viable-through-year-assessment

Lazarin, M. (2014). Testing overload in America's schools. Center for American Progress. https://eric.ed.gov/?id=ED561097

Lazarus, S. S., Hinkle, A. R., Liu, K. K., Thurlow, M. L., & Ressa, V. A. (2021). Using interim assessments to appropriately measure what students with disabilities know and can do: Advisory panel takeaways and NCEO recommendations. (NCEO Report 427). National Center on Educational Outcomes. https://nceo.umn.edu/docs/OnlinePubs/NCEOReport427.pdf

Lorié, W. (2021, May 5). Educating Pablo: A play about through course assessment systems. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/educating-pablo-a-play-about-through-course-assessment-systems/

Marion, S. (2023). Validity for through-year assessments. In R. J. Tierney, F. Rizvi, & K. Erkican (Eds.), *International Encyclopedia of Education* (Vol. 13, pp. 212–219). Elsevier https://doi.org/10.1016/B978-0-12-818630-5.09071-0

Marion, S. (2019, March 21). Five essential features of assessment for learning. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/five-essential-features-of-assessment-for-learning/

Marion, S., & Evans, C. (2021, August 20). Following their lead: Conversations with innovative assessment demonstration authority state leaders. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/following-their-lead/

Marion, S., Thompson, J., Evans, C., Martineau, J., & Dadey, N. (2018). A tricky balance: The challenges and opportunities of balanced systems of assessment. National Center for the Improvement of Educational Assessment. https://www.nciea.org/library/the-challenges-and-opportunities-of-balanced-systems-of-assessment-a-policy-brief/

Munyan-Penney, N., & Barone, C. (2020). Innovative assessment pilot: Louisiana's LEAP 2025 humanities assessment. Education Reform Now. http://edreformnow.org/wp-content/uploads/2020/11/Louisiana-IADA-LEAP-2025-Humanities.pdf

National Center for the Improvement of Educational Assessment. (2021). Through-year assessment virtual convening. https://www.nciea.org/library/?_resource_events=convening

- Session 1: Dadey, N., Gong, B., Lorié, W., & Marion, S. (2021). Through-year assessment virtual convening, Session 1: Definition, aims, and use cases. National Center for the Improvement of Educational Assessment. https://www.nciea.org/library/definitions-aims-and-use-cases/
- Session 2: Gong, B., Dadey, N., Lorié, W., & Marion, S. (2021). Through-year assessment virtual convening, Session 2: Claims, designs and evidence. National Center for the Improvement of Educational Assessment. https://www.nciea.org/library/claims-designs-and-evidence/
- Session 3: Lorié, W., Dadey, N., Gong, B., & Marion, S. (2021). Through-year assessment virtual convening, Session 3: Technical & logistical Issues. National Center for the Improvement of Educational Assessment. https://www.nciea.org/library/technical-logistical-issues/
- Session 4: Marion, S., Lorié, W., Dadey, N., & Gong, B., (2021). Through-year assessment virtual convening, Session 4: Threading the needle. National Center for the Improvement of Educational Assessment. https://www.nciea.org/library/threading-the-needle/

New Teacher Center. (2022). Through-Year Assessments: Practical considerations for LEA implementation. Smarter Balanced. https://newteachercenter.org/resources/through-year-assessments-2022/

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). Knowing what students know: The science and design of educational assessment. https://eric.ed.gov/?id=ED458233

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice,* 28(3), 5-13. https://www.researchgate.net/publication/227711414_Moving_Toward_a_Comprehensive_Assessment_System_A_Framework_for_Considering_Interim_Assessments

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). Universal design applied to large scale assessments. (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. https://nceo.info/Resources/publications/OnlinePubs/Synthesis44.html

U.S. Department of Education. (2010). *Catalog of Federal Domestic Assistance (CFDA) Numbers:* 84.395B (Comprehensive Assessment Systems grants) and 84.395C (High School Course Assessment Programs grants). 75 FR 18171. Washington, DC. https://www.federalregister.gov/d/2010-8176

Wise, L. L. (2011). Picking up the pieces: Aggregating results from through-course assessments. Center for K–12 Assessment & Performance Management at ETS. https://www.ets.org/Media/Research/pdf/TCSA_Symposium_Final_Paper_Wise.pdf

ADDITIONAL SOURCES

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Dadey, N., & Gong, B. (2021). Some framing for through year assessment (working draft). National Center for the Improvement of Educational Assessment.

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. Teachers College Record, 117(4), 1–26. https://eric.ed.gov/?id=EJ1056748

Farley-Ripple, E. N., Jennings, A. S., & Buttram, J. (2019). Toward a framework for classifying teachers' use of assessment data. AERA Open, 5(4), 1–18. https://doi.org/10.1177/2332858419883571

Goertz, M. E., Oláh, L. N., & Riggan, M. (2010). From testing to teaching: The use of interim assessments in classroom instruction (CPRE Research Rep. #RR-65). Consortium for Policy Research in Education. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1023&context=cpre_researchreports

Marion, S. (2021a, November 30). Collaboratively learning about through year assessments: It will take a village to fulfill the opportunities of through year assessment systems. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/collaboratively-learning-about-through-year-assessments/

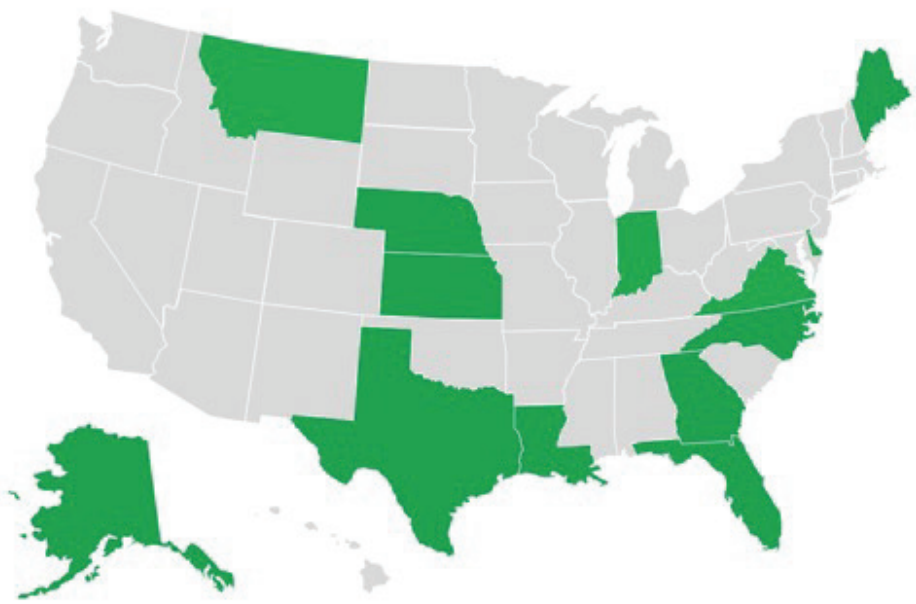
Marion, S. (2021b, September 29) Trying to serve multiple uses with through year assessments. National Center for the Improvement of Educational Assessment. https://www.nciea.org/blog/through-year-assessment/trying-serve-multiple-uses-through-year-assessments

Nabors-Olah, L., Lawrence, N., & Riggan, M. (2010). Learning to learn from benchmark assessment data: How teachers analyze results. Peabody Journal of Education, 85, 226–245. https://www.tandfonline.com/doi/abs/10.1080/01619561003688688

APPENDIX: MATRIX OF STATE THROUGH-YEAR ASSESSMENT PROGRAMS

At our November 2021 virtual convening[35], we noted 11 states that were developing or considering through-year assessment programs, and one alternate assessment consortium. Since then, that number has risen to 13 states, as shown in the map below, since Delaware and Indiana began to develop through-year programs.

1. Alaska
2. Delaware
3. Florida
4. Georgia
5. Indiana
6. Kansas
7. Louisiana
8. Maine
9. Montana
10. Nebraska
11. North Carolina
12. Texas
13. Virginia



In this appendix, we group these 13 state programs[36]—and one alternate assessment program, Dynamic Learning Maps, used in five states—into one of eight possible through-year assessment design models based on information currently available.[37] As shown in the table on the next page, the eight possible design models are derived from various mixtures of how the state plans to (a) organize content across test administrations and (b) use within-year assessment information in a student's annual summative determination of proficiency. We used these two decisions to create a matrix because they represent different theories of action and key design choices from which other decisions will necessarily follow.

If a state's theory of action is built around providing instructionally useful information, for example, it will likely choose to administer a subset of standards at each administration and those administrations will be required. Additional explanation follows the table.

[35] National Center for the Improvement of Educational Assessment. (2021).

[36] Some states (GA, LA, MT) have more than one through-year assessment program under development. In some cases these programs fall under different models based on how we've organized the design decisions (e.g., Georgia MAP and Georgia Navvy). In other cases, the programs fall under the same overall model, though they are distinct (e.g., Louisiana's curriculum-specific ELA and Louisiana's curriculum-relevant ELA and math). See EducationFirst (2022) for more detail on each state's through-year assessment program.

[37] These groupings are based on the information available at publication time. We note, however, that some states, such as Texas, have not finalized their design decisions. At least one, Florida, is bound by statute to explore the feasibility of an alternative design.

Score Aggregation Method *Is within-year information used to compute a student's annual determination of proficiency? [Yes or No] If not, what is the within-year information used for?*	**Content Organization at Each Administration** *How is the content in state standards organized across administrations?*	
	All State Content Standards	**Subset of State Content Standards**
Yes, within-year information is used.	**MODEL 1** [No current state designs]	**MODEL 2** Delaware DLM Georgia Navvy Louisiana Montana
Yes, within-year information is used if it would improve a student's cumulative score.	**MODEL 3** Texas	**MODEL 4** [No current state designs]
No, within-year information is used for other measurement purposes (create common scale for prediction or growth; create claim or subclaim scores; locate starting point in multi-stage adaptive).	**MODEL 5** Alaska Georgia MAP Kansas Nebraska Maine Virginia	**MODEL 6** Indiana North Carolina
No, within-year information is used only for informational purposes.	**MODEL 7** Florida	**MODEL 8** [No current state designs]

The two main ways to organize content across test administrations are to sample from all state content standards or to sample from a subset of state content standards.

In most (eight) current through-year assessment programs, the entire breadth of the state content standards is sampled at each administration within the year, likely because of the challenges associated with local curricular scope and sequencing and instructional pacing.

States that are more interested in connecting the through-year assessment program to curriculum and instruction tend to divide, cluster, or distribute state content standards in some manner at each administration (see, for example: Delaware, Georgia Navvy, Indiana, Louisiana, Montana, and North Carolina). As we noted earlier, these decisions represent distinct theories of action related to through-year assessment programs.

States also vary in how, if at all, they use within-year information to compute a student's annual determination of proficiency. For example, we have identified four main answers to the questions: *Is within-year information used to compute a student's annual determination of proficiency? [Yes or No] If not, what is within-year information used for?*

- Yes, within-year information is used. Student scores from earlier administrations are combined to produce a student's annual determination of student proficiency. Through-year assessment components are not optional.

- Yes, within-year information is used only if it would help a student's cumulative score. Otherwise, the final administration is used.
- No, within-year information is used for other measurement purposes. Through-year student scores from earlier administrations are used to (a) create a common scale for prediction or growth, (b) create claim or subclaim scores, and/or (c) locate each student's starting point more precisely in a delayed multi-stage adaptive test. Through-year assessment components are sometimes optional. Only the final administration is used to compute a student's annual determination of student proficiency.
- No, within-year information is used for informational purposes. Only the final administration is used to compute a student's annual determination of student proficiency.

Grouping current through-year assessment programs into eight possible models allows us to see that **the most common models right now are Model 5 (six states) and Model 2 (four states).**

- Model 5: Each through-year component is based on all state content standards and within-year information is not used to compute a student's annual determination of proficiency, but for other measurement purposes. In most cases, the other measurement purpose is to create a common scale for growth (Alaska, Georgia MAP, Nebraska, Maine, and Virginia). A couple states are currently planning to also use the within-year information for growth and to inform a delayed multi-stage adaptive test (Georgia MAP and Nebraska).
- Model 2: Each through-year component is based on a subset of state content standards at each administration ranging from one standard to a cluster of standards depending on the state design. Student scores from earlier administrations are combined to produce a student's annual determination of student proficiency. Through-year assessment components are not optional. Most states are focusing on English language arts and mathematics, though some states like Delaware are focusing on social studies. Only Louisiana's Innovative Assessment Pilot for English language arts is curriculum-specific; in all other cases, the connection to curriculum is not as tightly coupled.

The matrix also allows us to see that states are not exploring all of these possible models, likely due to considerations discussed in this paper (see, for example: Models 1, 4, and 8). Brief descriptions of each model that combines the content organization at each administration and score aggregation method can be found in the table below.

Model 1	All state content standards in a particular grade and content area are sampled at each test administration. Student scores from earlier administrations are combined to produce a student's annual determination of student proficiency. Through-year assessment components are not optional.
Model 2 (GA Navvy, DLM, DE, LA, MT)	Subset of state content standards at each administration ranging from one standard to a cluster of standards depending on the state design. Student scores from earlier administrations are combined to produce a student's annual determination of student proficiency. Through-year assessment components are not optional.
Model 3 (TX)	All state content standards in a particular grade and content area are sampled at each test administration. Student scores from earlier administrations may be weighted into a student's annual determination of student proficiency only if it would help their cumulative score. Otherwise, the final administration is used.

Model 4	Subset of state content standards at each administration. Student scores from earlier administrations may be weighted into a student's annual determination of student proficiency only if it would help their cumulative score. Otherwise, the final administration is used.
Model 5 (AK, GA MAP, KS, NE, ME, VA)	All state content standards in a particular grade and content area are sampled at each test administration (typically fall, winter, spring), but only the spring score is used to produce a student's annual determination of student proficiency that is required under federal law. The information from earlier through-year components (e.g., fall and winter) is used in different ways for measurement purposes depending on the state, including to: (a) create a common scale for prediction or growth, (b) create claim scores, and/or (c) locate each student's starting point more precisely in a delayed multi-stage adaptive test. Through-year assessment components are sometimes optional.
Model 6 (IN, NC)	Subset of state content standards are sampled at each test administration, but only the spring score is used to produce a student's annual determination of student proficiency that is required under federal law. The information from earlier through-year components is used in different ways for measurement purposes depending on the state, including to: (a) create a common scale for prediction or growth, (b) create claim scores, and/or (c) locate each student's starting point more precisely in a delayed multi-stage adaptive test. Through-year assessment components are sometimes optional.
Model 7 (FL)	All state content standards in a particular grade and content area are sampled at each test administration, but only the spring score is used to produce a student's annual determination of student proficiency that is required under federal law. Through-year results provide information only.
Model 8	Subset of state content standards are sampled at each test administration, but only the spring score is used to produce a student's annual determination of student proficiency that is required under federal law. Through-year results provide information only.

National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

www.nciea.org