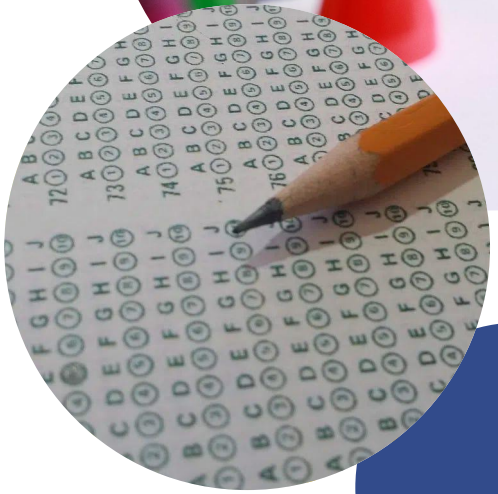
educationfirst

Rethinking the Test Pile

A Scan of State Assessment Policies



February 2026

Table of Contents

About The Report

This state policy scan analyzes statutes, regulations, guidance, and instructional frameworks to understand how state policy influences local assessment practice. We identify opportunities for state leaders to reduce assessment clutter, strengthen quality and improve coherence with instruction.

The report builds on Education First's national study of district assessment systems, which found that many sources of local overload and incoherence trace back to state policy pressures. Together, the state and district analyses are intended to spark action by clarifying how leaders at both levels can streamline assessment systems to focus on the measures that provide the most meaningful information for students, educators and families.

About the Authors

Education First staff Khaled Ismail, Senna Lamba, and Olivia Kelly led this research and authored the report. Austin Wechter designed the report.

educationfirst

About Education First

Education First is a national, mission-driven strategy and policy organization with deep expertise in education improvement. We help system leaders, advocates, funders and policymakers think bigger and work smarter to create the conditions that drive equity, coherence and excellence for all students. Our mission is to deliver exceptional ideas, experience-based solutions and results so all students—and particularly Black, Indigenous and other students of color and students living in low-income communities—are prepared for success in college, career and life.

For questions about this report or to learn more about Education First's work on issues related to assessment and accountability, contact Khaled Ismail.

Executive Summary

State policies—even those not explicitly labeled “assessment policies”—are increasingly defining the architecture of local assessment systems

In Rethinking the Test Pile, A National Study of K–8 Academic Assessments (Education First, 2026), we found:

In some districts, students take as many as 88 assessments before entering high school, with particularly high testing loads for English Learners, 5th and 8th grade students, and districts serving more students of color and low-income students.

District leaders report a strong desire to streamline assessment into a cohesive system, but shared how they have to navigate balancing local testing needs while responding to state policies that strongly shape assessment systems.

This report examines how state policy shapes assessment systems and surfaces critical questions for future study

What This Scan Examined

To identify opportunities for state leaders to support local coherence and assessment decluttering efforts, we examined:

- State statutory requirements that shape assessment and instruction
- State assessment and instructional guidance with implications for assessment
- State-provided assessment tools and resources
- State funding strategies and incentives that influence local assessment systems

Important Questions Beyond This Scan

Several questions surfaced during our research that fall outside the scope of this report and warrant further study.

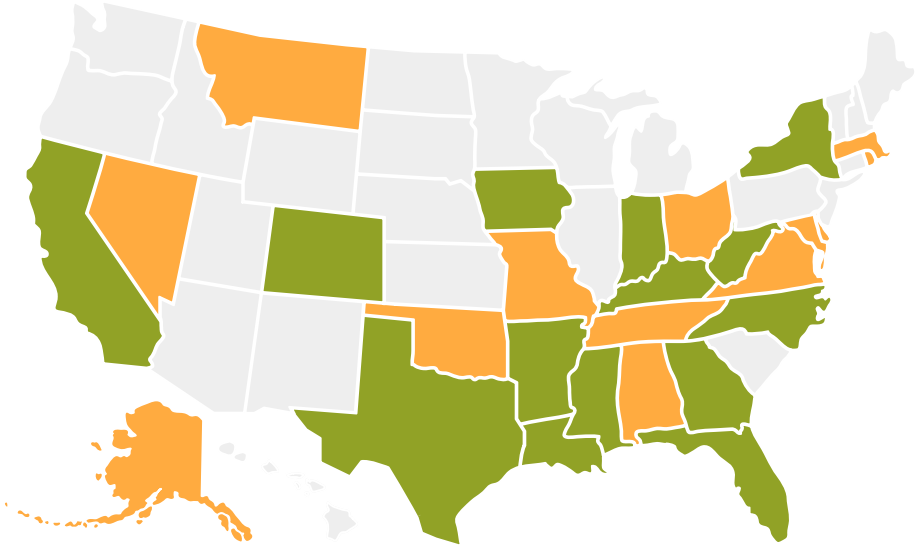
- What state infrastructure, review processes and governance models can support assessment quality?
- What evidence of quality for curriculum-embedded assessments in HQIM is needed to support adoption?
- Which assessment types should align to curriculum and which should remain curriculum-agnostic?

We explored multiple, overlapping state policy levers that shape local assessment systems across 27 states

Formal requirements
Statutes and regulations governing assessment practices

Guidance and expectations
State assessment or instructional guidance, frameworks and strategy documents

Tools and supports
State-developed tools, templates, incentives and optional planning resources



Depth of review varied by state based on availability and relevance of public materials. Findings should be interpreted as a composite policy landscape rather than a comprehensive evaluation of each individual state or a 50-state comparison scan.

- 14 states reviewed across all three levers
- 13 states reviewed across two levers

Three core problems are emerging:

Across both the district and state policy scans

1. Assessment clutter

Districts often administer too many assessments, many of which serve overlapping purposes, produce conflicting data or do not meaningfully inform instruction. States contribute to this accumulation by having overlapping testing requirements and guidance that can confuse district leaders and educators, divert time from teaching and learning and add to the testing burden of students.

2. Limited evidence on assessment quality and instructional utility

Many assessments claim to provide instructionally actionable information, but vendors rarely supply the evidence needed to verify those claims. In practice, very few tools align tightly enough to curriculum or developmental trajectories to meaningfully guide instructional decisions. The lack of transparent, high-quality evidence makes it difficult for states and districts to assess whether tools actually support teaching and learning or simply add noise.

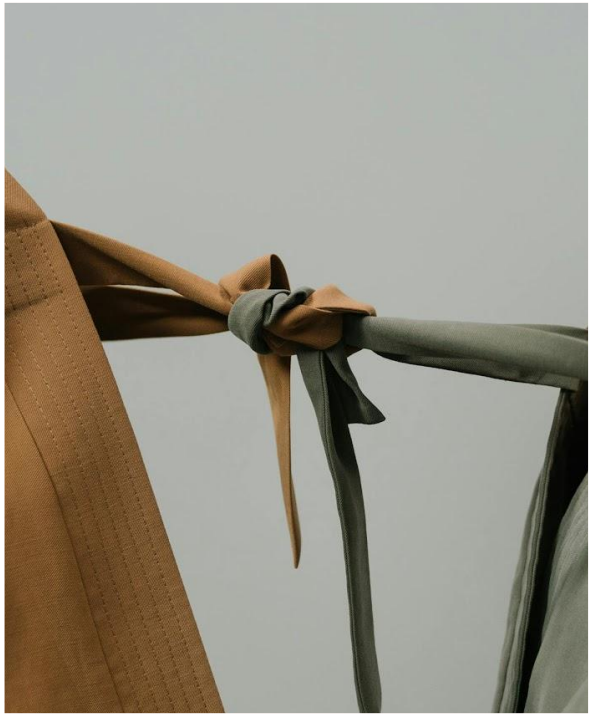
3. Incoherence with teaching and learning

Many local assessments are disconnected from state standards, instructional expectations and High-Quality Instructional Materials. As a result, teachers receive data that does not map cleanly to what they teach, which reduces instructional value and contributes to fragmented improvement efforts.

Our analysis surfaces clear patterns and potential risks within state policies that contribute to the architecture of districts' assessment systems

State policies and agency guidance often pull in different directions, creating significant overlap across math and literacy screening, interim and diagnostic assessments and the MTSS and instructional guidance intended to support them.

State assessment policies also increasingly determine how districts allocate time, staff and instructional attention.



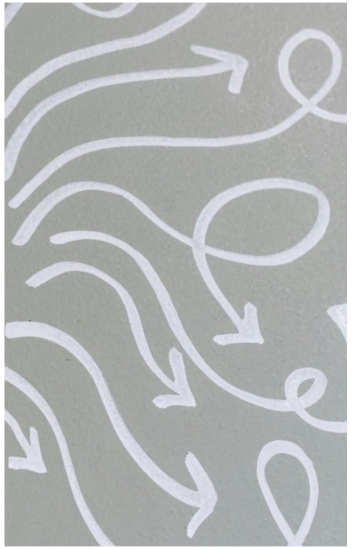
The policy trade-offs states make about prescriptiveness, tool selection, data expectations and incentives directly determine the coherence and complexity of local assessment systems

Prescriptiveness vs flexible choice

States set how tightly they define assessment purpose, timing and administration, shaping consistency and local flexibility.

State-provided tools vs. approved lists vs. local autonomy

States choose whether districts use state tools, select from approved vendors or make local choices, affecting alignment and implementation burden.



Frequency and types of data required

States determine how often districts collect and report various data, influencing calendar crowding and instructional focus.

Incentives vs. restrictions

States use incentives and constraints such as time caps, bans and links to intervention, funding and professional learning to steer district behavior.

KEY FINDING #1

State screening mandates, diagnostic requirements, MTSS rules and interim assessment policies collectively dictate large portions of local assessment calendars, often leaving districts little room to simplify or rebalance assessment use

While different assessments can serve distinct purposes, many districts stretch tools beyond the purposes for which developers validated them, such as using screeners for instructional planning or interim assessments for high-stakes decisions.

Overlapping and sometimes unclear state requirements reinforce this pattern by sending mixed signals about assessment purpose, frequency and use, making it harder for districts to design coherent assessment systems.

KEY FINDING #2

Incoherent policy layers create instructional noise for teachers

Across the year, teachers must track and respond to multiple assessment requirements layered on top of one another, including state mandated screeners, district selected diagnostics, state or district interim assessments and instructional frameworks tied to MTSS or curriculum implementation.

Each assessment produces different data, uses different benchmarks and arrives on different timelines. Teachers must interpret these signals while planning lessons, grouping students and responding to intervention requirements, often without clear guidance on how the data should connect or which signals matter most.

Instead of reinforcing instruction, these overlapping demands compete for attention and time, making it harder for teachers to focus on teaching and supporting students.

KEY FINDING #3

Incoherent state and district structures produce fragmented assessment systems

In many states and districts, separate offices oversee assessment, curriculum, MTSS and school improvement, each issuing guidance and requirements aligned to their own goals and timelines. These siloed structures lead to layered assessment expectations that districts must reconcile on their own.

Breakdowns between research, policy design and implementation further compound the problem. States often base policies on sound research but lose coherence as guidance moves from statute to regulation to practice, leaving districts and schools to interpret how assessment tools should work together in real classrooms.

As a result, local assessment systems reflect accumulated compliance rather than intentional design.

KEY FINDING #4

Without guardrails, assessment policy can undermine coherent teaching and learning

Teachers often face pressure to prepare students for assessments that do not reflect the content, sequence or cognitive demand of the materials they use in the classroom. Many commercial assessment products lack alignment to state standards or to specific curricula, yet policy signals can elevate these tools through mandates, approved lists or accountability links. As a result, teachers adjust pacing, lesson focus or instructional strategies to meet assessment demands rather than following the instructional arc of HQIM.

These misalignments weaken fidelity of implementation by fragmenting instructional time, introducing competing definitions of mastery and eroding the coherence that HQIM aims to provide across classrooms and grade levels.

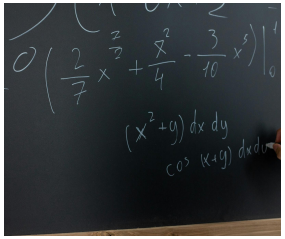
Policy Recommendations to State Leaders

To address overlapping mandates, unclear state guidance and weak evidence for many locally selected assessment tools, states should:

1. Develop purpose-driven, state-vetted lists of high-quality assessments with flexibility pathways and clear incentives for adoption
2. Require a periodic district assessment inventory and streamlining cycle supported by state-provided tools, templates and capacity-building
3. Invest in high-quality, standards-aligned progress monitoring options while allowing opt-outs for districts using strong HQIM-embedded measures
4. Right-size literacy screening frequency and use appropriate tools for targeted follow-up assessments based on student need
5. Strengthen math screening policy by clarifying key constructs and evidence requirements, while proceeding cautiously in a developing research field
6. Set clear expectations for how districts use data to enter, monitor, and exit student interventions in alignment with IDEA, MTSS and instructional pathways

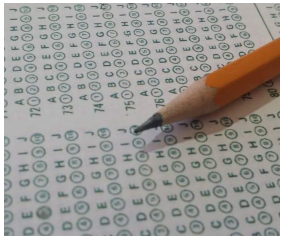
Key Findings

We summarize findings across five key policy areas



Math Policies

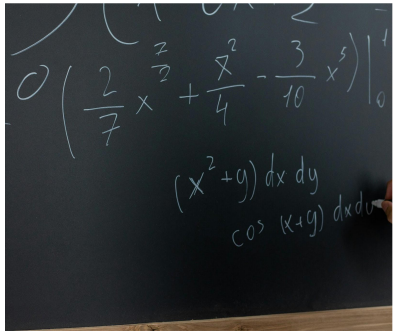
Early Literacy Policies

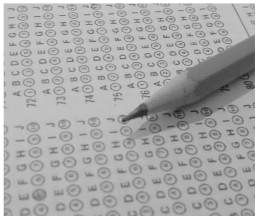


Interim and Through-year Assessment Requirements

Instructional Guidance and Frameworks

Third Grade Retention and Math Acceleration Policies





Math Policies

Math screening mandates are expanding rapidly, with at least 18 states now requiring screening, adding a new layer of required assessments to already crowded local systems

- Many of these screening mandates are included as part of larger efforts to improve math outcomes or state strategies for instruction (i.e., legislation includes support for math professional development, math coaches or high dosage tutoring).
- Many states are still in the early stages of implementation; some have recently passed legislation and not yet started planning for implementation.
    - + While 18 states have mandated, about half of these states are currently implementing their screening mandates
- Given the early nature of this work, there is an opportunity to intentionally think through how screening fits within an overall system of assessment

As evidenced by a surge in legislation between 2022 and 2025, states are increasingly codifying a commitment to early identification of students at risk for mathematics difficulties

States with screening mandates: Alabama, Arkansas, Florida, Indiana, Iowa, Kentucky, Louisiana, Mississippi, New Mexico, North Dakota, Ohio, Oklahoma, Rhode Island, Tennessee, Texas, Utah, West Virginia, Maryland

State policies rarely clarify the intended purpose of math screeners, generally framing them as tools for early identification

- Unlike literacy, where reading screeners tend to measure a small set of well-defined foundational skills, our research suggests that states are currently relying on a wider variety of math screeners to capture different constructs.
- Mathematics spans multiple domains, and there is no single general-outcome measure that can capture proficiency across number sense, operations, geometry, measurement, probability, or algebra, making math screening more complex and nuanced than literacy screening.
- This complexity means that different measures are often required to assess different aspects of mathematics proficiency at various educational levels (early, elementary, middle, secondary). For example, early math screeners focus on number sense, counting objects, and quantity discrimination, while middle school screeners focus on algebra readiness.

State policies provide little detail on what math screeners should measure, or how educators should use the results

The evidence base for math screening is growing and supports their use, particularly for early grades, but lags behind that for literacy

What the existing research does say…

- Early mathematics skills may have the strongest predictive power for later academic achievement compared to other earlier academic skills.
  - A meta-analytic study demonstrated that early math skills are as predictive of later reading achievement as are early reading skills.
- Children's early mathematics skills are linked to later mathematics outcomes.
  - Kindergarten number competence and number sense strongly predict later mathematics outcomes, including proficiency in third grade.
  - Early numerical competencies, or the lack thereof, uniquely predict future math difficulties.
- Research indicates that gaps in math skills are particularly difficult to close, and without early interventions, they tend to worsen over time.

Much of the research is siloed in academic journals, and it is further siloed in psychology and special education research, rather than education policy or broader math research

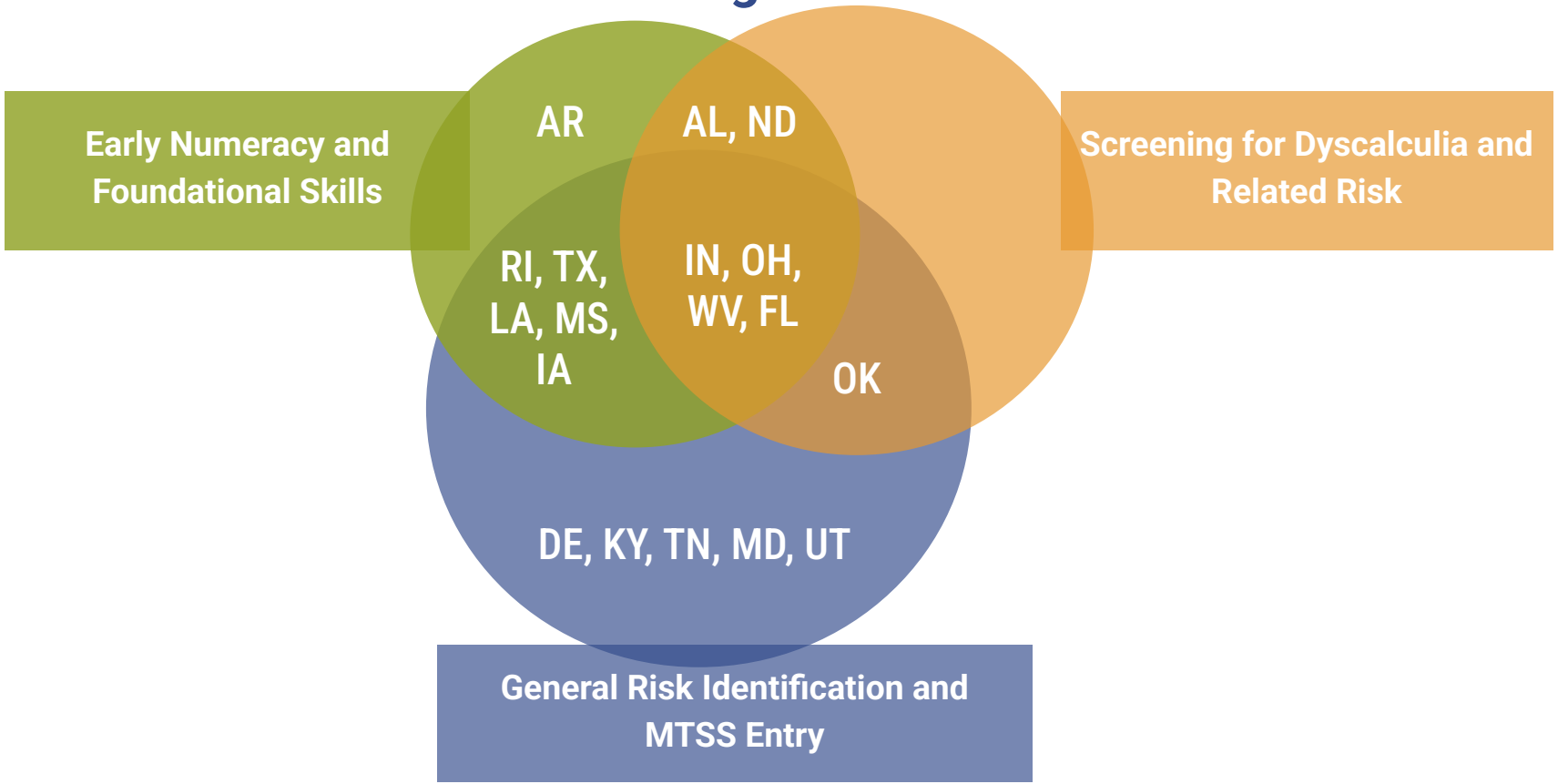
States differ sharply in what math screeners measure—risk for falling behind, specific skills, dyscalculia* or broad numeracy

States are currently screening students in mathematics across various grade levels for a broad range of purposes

General Risk Identification and MTSS Entry	Early Numeracy and Foundational Skills	Screening for Dyscalculia and Related Risk
Universal screening serves as a critical first step within a Multi-Tiered System of Supports (MTSS) or Response to Intervention (RTI) framework. The overarching purpose is to rapidly detect students experiencing academic difficulties to provide timely intervention.	In the primary grades (PK–3), screening is heavily focused on measuring early numeracy and foundational quantitative concepts. Specific skills mentioned in policy and guidance vary from counting to number sense.	Several states mandate screening processes designed to specifically identify students who exhibit characteristics of dyscalculia and dysgraphia or who have a related substantial math deficiency.

**Dyscalculia* is a term used to describe specific learning disabilities that affect a child's ability to understand, learn and perform math and number-based operations.*

How different states are using math screeners



With math screening requirements still emerging, states offer limited information on screener approval processes—creating an opportunity to clarify criteria and strengthen guidance

Eight states* have publicly available criteria for choosing math screeners, and these vary in specificity. Below are the most common criteria states have used for approving, selecting or utilizing math screeners and diagnostic assessments

Technical Adequacy	Efficiency and Administration Feasibility	Instructional Utility and Data Management	Alignment with State Standards	Dyscalculia Identification

*Kentucky, Indiana, Ohio, Iowa, Alabama, West Virginia, Oklahoma and Rhode Island

Overview of Screener Approval Requirements

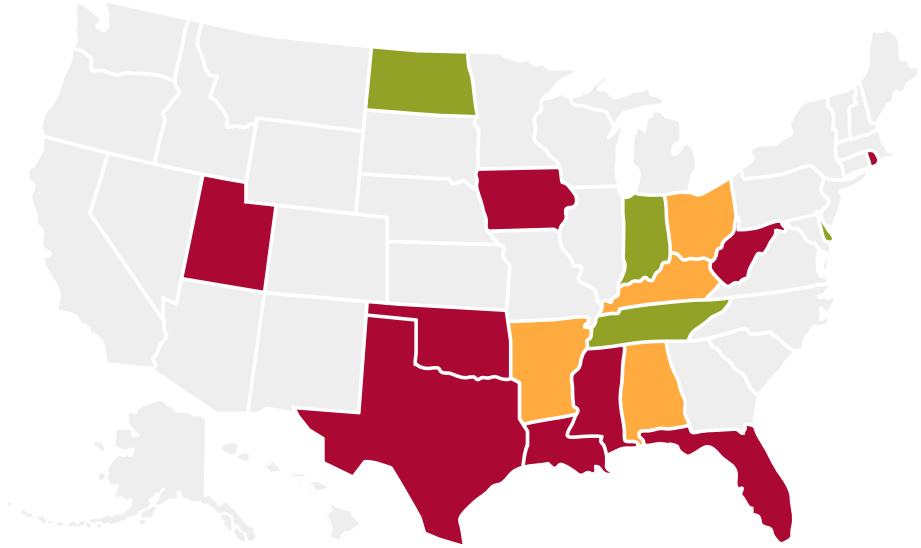
Technical Adequacy	KY, IN, OH, IA, WV, OK, RI	Require quantitative evidence (validity, reliability and accuracy) to confirm that a screener is psychometrically sound and accurately identifies students who are genuinely at risk for academic failure.
Efficiency and Administration Feasibility	KY, IN, IA, AL, OK, RI	Require measures that are practical for routine use in school settings, minimizing disruption to instructional time.
Instructional Utility and Data Management	IN, OH, AL, RI	Require instructionally relevant information and/or contribute to an MTSS process by guiding instruction and providing relevant data for systemic analysis.
Alignment with State Standards	IN, OH	Require screeners be fundamentally tied to the core academic content students are expected to learn.
Dyscalculia Identification	OH	Require a screening process capable of the early identification of mathematics deficiencies and/or the characteristics of dyscalculia.

Most commonly used screeners across states we reviewed: i-Ready; aimswebPLUS; STAR Math; FastBridge; Acadience Math; Istation Indicators of Progress (ISIP) Math

State approaches to selection vary, from a single required tool to approved lists or open choice, each with distinct risks

	Benefits	Risks	Example States
Mandated Screening *Districts must use a math screener; state does not provide a list of approved screeners.*	Fosters local ownership and innovation	Places a greater burden on districts to conduct their own due diligence and can lead to wide variability in the quality of data collected across the state	Iowa
Mandated Screening + State-Approved List *Districts must choose from a list of state approved screeners.*	This approach ensures a minimum standard of practice and can simplify procurement for districts	A one-size-fits-all solution that may not be optimal for every local context	Alabama, Florida, Oklahoma
Single Mandated Screener *The state prescribes one screener statewide.*	This approach ensures a universal measure	May not be optimal for all contexts and may severely limit what data is captured in screening	Arkansas

More frequent screening significantly increases district and teacher workload and signals a growing conflation of screeners with diagnostic tools in both policy and practice



The frequency of required screening varies. Many states mandate or recommend that screeners be used more than once a year

No Explicit Guidance on Frequency

Delaware
Indiana
North Dakota
Tennessee

1x Year

Arkansas
Alabama *(for K only, 2x a year for 1-5)*
Kentucky
Ohio *("more than once per year")*

At Least 3x Year

West Virginia
Oklahoma
Mississippi
Iowa
Rhode Island
Florida
Louisiana
Utah
Texas

The follow-up actions tied to math screening results, including MTSS steps, additional diagnostics, and improvement plans, substantially increase district workload and operational complexity

- **Mandated Inclusion in MTSS/RTI Structure:** While guidance varies, several states (e.g., West Virginia, Alabama, Tennessee, Rhode Island, Arkansas) uniformly tie the required math screening to mandated intervention or support systems.
- **Identification and Planning:** In some states, screening results trigger specific actions, such as identifying "priority students" (e.g., scoring at or below the 40th percentile in Tennessee) and require the creation of individualized improvement plans for struggling students (e.g., West Virginia). This is either part of an MTSS structure or a similar system.
- **Diagnostic Follow-up:** Arkansas mandates the use of diagnostic testlets following the initial screener to provide teachers with detailed information on skill gaps, allowing for tailoring of Tier 2 and Tier 3 interventions. Rhode Island requires diagnostic assessments to follow up on the screener results before intervention begins.
- **Standalone Tool:** In a few states, math screeners operate outside formal MTSS, though policies still require planned interventions based on screening results (e.g., Oklahoma, North Dakota).

Math screening mandates also have direct implications for staffing and professional learning, requiring districts to build new expertise and capacity to interpret and act on results

Training

- Educators need training on recognizing the characteristics of dyscalculia, dysgraphia and other mathematics “deficiencies” in students (e.g., West Virginia specifically mentions dyscalculia and Florida dysgraphia).
- Kentucky funds professional learning opportunities for teachers and administrators through its Numeracy Counts Academies.

Staffing

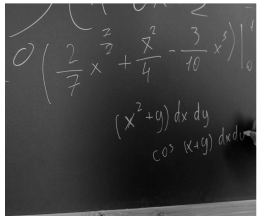
- Some policy requires hiring additional staff (e.g., Alabama explicitly mandate the allocation and employment of K-5 mathematics coaches to support teacher training, model lessons, monitor progress, and assist in interpreting screening and diagnostic data).
- Screening may require additional intervention support staff (e.g., tutors, intervention specialists, or highly qualified teachers).
- West Virginia's Third Grade Success Act explicitly phases in reduced pupil-to-teacher/assistant ratios for Grades 1–3.

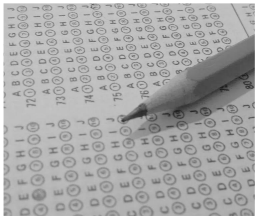
Instructional Time

- The process of administering math screeners and analyzing the results, while deemed crucial for early intervention, consumes valuable instructional time.

Impact

Creation of substantial professional development needs for educators across all relevant roles, including teachers, interventionists and school leaders





Early Literacy Policies

While requirements, implementation and supports vary, 42 states require a universal reading screener and 36 states require screening for characteristics of dyslexia

- States vary in prescriptiveness with screening policies:
  - Some mandate a single tool or narrow list; others allow choice among approved options.
  - States vary in the grades included and frequency of screening (e.g., K-2 vs. K-3, x2 a year vs x3 a year).
- Some states (e.g. TN and GA) provide a free, state-approved screener, financially incentivizing districts to use the state tool, likely reducing duplication in local assessment practices.
- Other states (e.g. TX) allow LEAs to use one screener for both universal screening and dyslexia screening requirements for certain grades, potentially reducing assessment clutter.
- States differ in how much guidance they provide around the purpose and uses of screeners and diagnostics, and the level of implementation support they provide (staff training, data submission and monitoring etc.).

States vary in their literacy screening mandates, requiring districts to adapt to different tools, frequencies and follow-up expectations

We reviewed universal literacy and dyslexia screening across a sample of states.

Dyslexia Screener Requirements Key

- Currently mandates separate dyslexia screener
- Dyslexia screener part of universal reading screener
- → Future dyslexia screener mandate
- Ⓧ No dyslexia screening mandate; state provides guidance

Mandated Screening + State-Approved List *Districts must choose from a list of state approved screeners.*	Colorado → Georgia ✓ Maryland (beg. 26-27 SY Ⓧ Massachusetts Ⓧ Mississippi ✓ Tennessee ✓ Texas ✓
Single Mandated Screener *The state prescribes one screener statewide.*	Arkansas ✓ Louisiana ✓
No state specific screener requirement *State does not mandate a universal reading screener, but provides screening guidance to districts.*	New York Ⓧ Illinois Ⓧ

States often pair screeners with mandated diagnostics, progress monitoring or intervention plans

Mandated Diagnostics	General Intervention/progress monitoring guidance	More prescriptive guidance
Require diagnostic assessments for students who flag “at risk” to pinpoint specific skills deficits and inform intervention planning.	Link screening data to interventions and progress monitoring. Districts must ensure they align assessment results with intervention systems and infrastructure.	Require schools to develop specific reading plans, but vary on the amount of support and guidance given
TN and AR require additional screening for students flagged on the initial universal screener.	LA, GA and MA require schools to provide targeted interventions and/or progress monitoring for students depending on screener results.	CO, AR and MS require and provide templates for districts/schools to develop individualized reading plans outlining goals and progress monitoring strategies for students demonstrating reading difficulties.

States' processes for selecting and mandating screeners vary but typically begin with baseline criteria outlined in statute

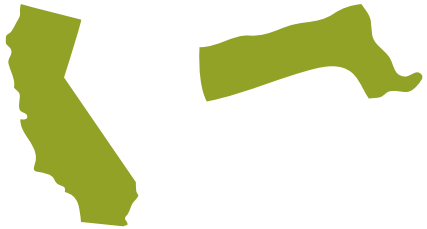
States typically develop scoring rubrics to evaluate vendors based on baseline criteria outlined in statute. States often consider the following domains in their rubrics:

- Technical quality: validity, reliability, classification accuracy, norming sample representativeness
- Literacy construct coverage: coverage of core early literacy skills (e.g. phonological and phonemic awareness)
- Accessibility and equity: bias review, accommodations
- Usability and reporting: efficiency, ease of use, timeliness of reporting
- Feasibility and support: PD/support, cost-effectiveness, progress monitoring linkage

Most common screeners on approved lists across reviewed states:

- DIBELS 8th Edition
- Amira ISIP
- i-Ready
- STAR
- MAP Growth

States' processes for selecting and mandating screeners vary but typically begin with baseline criteria outlined in statute



California and Massachusetts call out attention to linguistic diversity and "culturally and linguistically appropriate screening instruments," respectively.

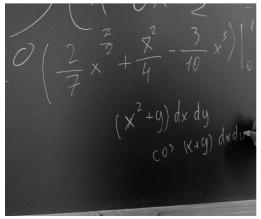
Mississippi includes specific content criteria coverage broken out by grade level.

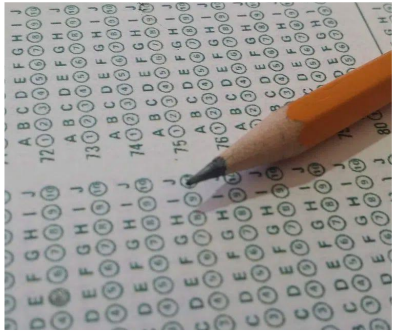
Maryland will require consideration of the length of the assessment, the PD needs and the timeliness of results.

While most states require literacy screening three times a year, implementation gaps may limit instructional impact

Recent studies examining the implementation of universal literacy screeners across 39 states found that:

- **Educators often lack clarity on the distinct purposes of screeners vs. diagnostics** and how each should inform instruction.
- **Training and implementation vary widely-** many educators aren't adequately prepared to administer screeners, interpret results, or use diagnostic data to plan targeted interventions.
    - In one recent study, **nearly half of educators felt only slightly or not at all confident administering a screener** after minimal training, and **half had no chance to practice** before testing their first student.
    - Educators frequently report **uncertainty interpreting screening results**, especially for multilingual learners.
- Preparing for, administering and scoring screening takes **substantial time** for educators and schools.
- Even when screening occurs, systematic procedures for developing intervention plans for struggling readers are often missing, creating a **"leaky pipeline" between identification and support.**





Interim and Through-year Assessment Requirements

Many states now embed interim and through-year assessments models into statewide systems

In addition to screeners and diagnostics, many states now offer or require statewide interim or through-year assessments as part of their assessment systems. States often frame these assessments as ways to reduce testing by replacing district interims and to provide more instructionally useful data by aligning to state standards.

States requiring statewide interim or through-year assessments in ELA and Math

Florida FAST
Indiana Checkpoints
Montana MAST
Alaska AK STAR
Maine Through-Year Assessment

States offering statewide interim or through-year assessments on an optional basis in ELA and Math

North Carolina NC Checks 2.0
Arkansas ATLAS
Georgia Beacon
Ohio Readiness Assessments
Texas STAAR Interim Assessments

In practice, how states design interim policies determines whether they streamline or add burden

Required models
States such as Florida, Indiana and Montana require districts to use statewide interim or through-year assessments. These policies force districts to adjust pacing, scheduling and instructional routines regardless of existing local systems.

Optional models
States such as Georgia, Ohio, North Carolina and Arkansas allow districts to opt into state interims. These policies preserve local choice but often add complexity as districts decide whether and how to integrate state tools alongside existing assessments.

Alignment to summative assessments
Some state interims closely mirror the state summative assessment like in North Carolina and Arkansas, while others, like Georgia's and Ohio's, operate as stand-alone tools with different purposes and signals, further shaping district decisions.

Our district scan shows that some districts adopt state interims to reduce local testing, while others decline because the tools do not align with local curricula, instructional pacing or assessment strategies.

Interim and through-year designs make different bets about purpose: accountability, instructional utility or both

Many states use these models to address a single core challenge: supporting instruction at scale. These design choices determine whether districts receive data teachers can use to support instruction or whether districts must rely on parallel assessments to obtain actionable instructional information.

Curriculum-agnostic vs curriculum-relevant

Curriculum-agnostic models prioritize comparability, scalability and alignment to state standards across districts, as seen in Florida and Indiana. Curriculum- relevant approaches attempt closer alignment to instructional materials and pacing, as seen in Montana's approach.

Role in accountability

Some states integrate interim or through-year assessments into accountability, increasing pressure on districts and schools, as seen in Florida and Indiana. Other states position interims as informational tools that do not factor into accountability calculations, as seen in Georgia, Ohio, North Carolina and Arkansas.

Item transparency and reporting granularity

Some states provide limited item-level transparency and aggregate reporting, as seen in Florida and Indiana. Other states offer greater item visibility and more granular reporting intended to support instructional use, as seen in Montana, Georgia and Ohio.

Instructionally useful and curriculum-relevant assessments have the potential to strengthen instruction and reinforce use of high quality materials

Instructionally useful assessments provide timely, specific and interpretable information that can shape instructional decisions. The information provided surfaces what students know, what they partially understand, where they are getting stuck and why. Feedback from an instructionally useful assessment is actionable at the level of lessons, units, and materials, not just standards or domains.

What information needs to be provided

- Clear signals about student thinking and misconceptions, not just right/wrong
- Results that are mapped to curriculum sequences and tasks, enabling next-step planning
- Reporting that supports teachers, coaches, and leaders differently, aligned to their roles in the instructional cycle

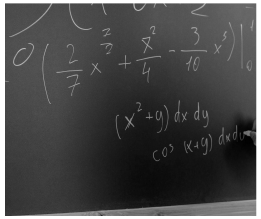
States can reinforce and accelerate adoption of high-quality instructional materials (HQIM) by ensuring interim and through-year designs provide instructionally useful information

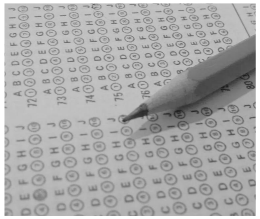
Strengthening the instructional utility of statewide interim offerings can:

- Provides a coherent alternative to commercial interims, reducing duplication and burden
- Strengthens the connection between state standards, local curricula and classroom practice without prescribing a single curriculum (*e.g. Indiana integrates data from state checkpoints into approved tier 1 and tier 2 platforms*)
- Disincentivize districts from continuing to layer on commercial interim assessments to fill instructional gaps (*e.g. Texas HB8 prohibits benchmark testing that is not embedded in curriculum*)
- Provide clear signals that state priorities for assessment support teaching and learning

Districts often weigh familiarity, perceived signal quality and usability against the promise of coherence. Whether statewide interims reduce district clutter depends less on **policy intent** and more on whether the model delivers **instructional utility strong enough to replace existing tools.**

42





Instructional Guidance and Frameworks

State instructional frameworks increasingly guide educator, school and district assessment practices

We looked at instructional guidance and frameworks from a number of states to develop a better understanding of how **these documents coming from state various offices influence district assessment choices**. Instructional frameworks and guidance as documents commonly articulate:

- Core instructional practices, pedagogical approaches and conditions that support effective implementation of the state's academic standards.
- How the state's instructional vision should be implemented leveraging standards, curriculum, instruction, professional learning and assessment.
- How instruction should be designed, delivered and supported to ensure all students meet grade-level expectations.

States often embed expectations for formative assessments and screeners directly into instructional frameworks

Most instructional frameworks reviewed emphasize developing a **comprehensive assessment system,** with formative, interim and summative assessments all playing unique roles. A number of states **specifically emphasize and encourage districts, schools and educators to use formative assessments and screening tools to drive instruction.**

Emphasis on formative assessments part of the daily teaching and learning process

Emphasis on screening and diagnostic tools

State guidance tends to highlight the role of formative assessments, screening and diagnostic tools

Emphasis on formative assessments part of the daily teaching and learning process:

Some state frameworks and guidance encourage educators to use formative classroom level assessment data as part of the teaching and learning process.

- Ohio's Plan for K-12 Math, Rhode Island's curriculum frameworks, California's ELA/ELD Framework and New York's K-3 Literacy Curriculum Review Guide all particularly emphasize using formative assessments to differentiate core instruction, guide scaffolding decisions and identify student misconceptions and learning gaps.

Emphasis on screening and diagnostic tools:

Other states also emphasize and encourage using assessment tools like screeners and diagnostics:

- Nevada's PreK-12 State Literacy Plan emphasizes using both screening and diagnostic assessments across elementary and secondary grades, and emphasizes the need for staff training.
- New York's Guide also emphasizes the use of universal screeners for all students in K-3 and diagnostic assessments for those performing below expected benchmarks.

46

States vary in how specific their guidance around assessments is within instructional frameworks

Less specific guidance: Some states encourage assessment-informed instruction, but guidance is more general.

- Georgia's ELA Instructional Framework encourages teachers to use varied assessment methods to gauge learning and to plan instructional next steps based on student data, but does not specify the exact types of assessments.
- The Wisconsin Standards for Math encourage teachers to use a variety of purposeful assessments before, during and after instruction to measure math proficiency, and to use this data to adapt instruction, but does not specify the exact types of assessment or data sources.

More specific guidance: Other states outline structured assessment cycles and specific required assessment types:

- Arkansas's Academic Playbook provides a structured cycle for how assessments should inform instruction: a pre-assessment to determine prerequisite skills students are missing, a common formative assessment given during a unit to guide Tier 1 adjustments, and an end of unit assessment to measure student mastery.
- Beginning in the 26-27 school year, Maryland will require LEAs to identify and implement common K-8 math unit or modular based assessments aligned to grade-level standards and develop systems to monitor and report progress on these assessments, and include support for teachers to interpret and respond to student performance.

State instructional frameworks encourage district and school leaders to use multiple types of data, potentially contributing to a more crowded assessment landscape

Across the frameworks, leaders are encouraged to use the following data sources to evaluate program effectiveness, make resource, staffing and budget decisions, and ensure alignment of curriculum, instructional materials and assessments:

- **Student achievement data**
  (state tests, interim benchmarks, unit assessments) (OH, RI, NY, MD, AZ, WI, AR, NV, GA, CA)
- **Fidelity and implementation data**
  (walkthrough tools, HQIM usage) (OH, RI, NY, MD, VA, NV)
- **Tiered/diagnostic data**
  (screeners, diagnostics, K–3 literacy results, MTSS placement data) (OH, RI, NY, AZ, AR, NV, CA)
- **Equity and climate data**
  (attendance, discipline, course access, subgroup performance) (OH, NY, MD, AR, NV)
- **Program and system health indicators**
  (chronic absenteeism, intervention data, curriculum-based measures) (OH, WI, NV)

States emphasize these different data sources to different extents, but overall **the trend is towards encouraging districts to use multiple data sources across the student, teacher and school level.**

Despite widespread emphasis, states provide limited guidance on how to evaluate the quality of curriculum-embedded assessments - an area that remains under-examined

A number of state frameworks emphasize curriculum-embedded assessments as essential to high-quality instruction. These states encourage the use of curriculum-embedded formative and summative assessments to provide ongoing, actionable feedback and support progress monitoring. However, states vary in how they position these assessments and the degree of guidance they offer to districts - both in selecting high-quality criteria and in using the embedded assessments within them. State guidance differed in terms of the system level decision districts have to make:

Curriculum selection	Use of curriculum embedded assessments	Alignment within a broader assessment system

Instructional frameworks encourage districts to think intentionally about selecting, using and aligning curriculum-embedded assessments.

Curriculum selection

Some states encourage districts to consider the quality of embedded assessments when selecting curricula and instructional materials.

New York and New Jersey recommend that **districts consider the embedded assessments when selecting a curriculum** and encourage selecting a curriculum with valid embedded formative and summative assessments.

Use of curriculum embedded assessment data

Some states provide guidance on how curriculum-embedded assessment data should be used.

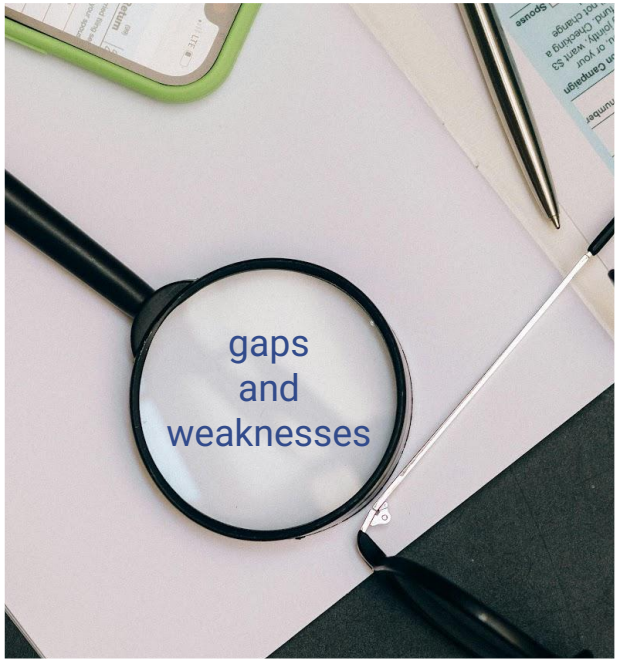
Maryland's PreK-12 Math Policy recommends teachers use and interpret curriculum-embedded assessments to monitor progress toward numeracy milestones.

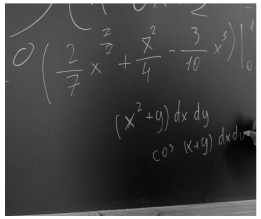
Ohio's Plan for K-12 Math **recommends districts to prioritize and track the use of high-quality instructional materials-embedded assessments**, align assessment calendars to curriculum scope and sequence and use curriculum-embedded assessments in data meetings and student work analysis.

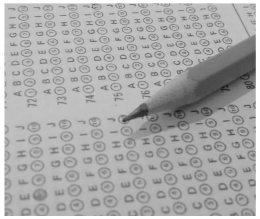
Alignment within a broader assessment system

Some states provide guidance on considering how curriculum-embedded assessments fit alongside other district or school assessments.

Rhode Island's Curriculum Frameworks also encourage districts to use HQIM-embedded assessments, but **provide guidance to districts to review these assessments to identify gaps and weaknesses** and determine whether additional assessment are needed.







Third Grade Retention and Math Acceleration Policies

There are several additional policy areas that influence local assessment practices worth highlighting

Additional policies with implications for local assessments include:

- Third grade reading laws and retention policies, including requiring proficiency on a reading assessment for promotion to grade 4
- Accelerated math policies, especially in the middle grades

Our district scan highlighted students in grades five and eight face a layered assessment load that can be a result of:

- The introduction of science and social studies state summative testing
- PSAT, Pre-ACT and high school entrance exams
- Assessments to determine eligibility for advanced courses

54

Third Grade retention policies substantially raise the stakes of K-2 local assessment and data collection requirements

- States with retention laws (e.g., IN, TN, AZ, OH) encourage LEAs to use local assessments in grades K-2 to predict readiness for the third-grade state test.
- Retention laws often mandate local assessment systems to provide continuous, valid data for promotion decisions and exemptions (e.g., North Carolina, Nevada).
- Local assessment results are also used for students seeking exemptions or mid-year promotion.
  - Districts are required **utilize local assessments to demonstrate sufficient progress for retained students to be promoted mid-year** (e.g., AZ).
  - Nevada and Florida encourages LEAs to **use alternative standardized reading assessments or portfolios to demonstrate proficiency** when a good-cause exemption is sought.

55

States also increasingly use assessment-driven policies to identify students for accelerated math pathways

State-driven Autoenrollment

State and/or local assessment results are used as the mandatory trigger for placement in advanced math coursework. Districts must automatically enroll eligible students into advanced math courses based on their scores.

- In Texas, students who performed in the top 40% on the Grade 5 Math STAAR assessment or a local measure (that includes student's ranking or demonstrates proficiency in grade-level coursework) are automatically enrolled.
- In Illinois, high school students who exceed state standards in math on a state assessment will automatically be enrolled into the next most rigorous level of advanced coursework.

District-determined/Multiple Measures Autoenrollment

Other states mandating (or encouraging) autoenrollment include multiple criteria for qualifying students.

- In Washington, districts can include their own additional eligibility criteria so long as the district criteria doesn't create inequities among student groups in the advanced course or program.
- In Virginia, students who score in the upper quartile on the statewide end-of-year Standards of Learning math assessment.
  - + The policy includes opportunities for school boards to establish additional criteria (i.e., teacher and parent recommendations, student preferences).

Automatic placement expands access but risks misidentification; multiple-measure systems reduce errors but increase complexity

Wins

Increased Course Access: Research in Washington indicates that students belonging to groups historically underrepresented in advanced courses showed a greater increase in enrollment rates in advanced math courses than students in districts without the policy

Subgroup Reporting: North Carolina mandates an annual report on the policy's implementation, including tracking enrollment data by subgroups (racial/ethnic groups, gender), requiring local systems to maintain and report this specific data for state monitoring.

Challenges

Increased middle school EOCs: These policies collectively push more students to complete Algebra I by grade 8, which results in an increase in middle-school End-of-Course (EOC) testing.

Pacing and content coherence: Accelerated pathways (like Georgia's) require local systems to compact standards to complete Algebra I in grade 8, creating curriculum and content alignment challenges. Acceleration decisions also often rely on assessments that are not aligned to curriculum progressions or HQIM pathways

In both policy areas, districts have to collect, interpret and respond to additional data requirements

- Districts must **maintain assessment data that can document exempted students' status** (e.g., ELL students, students with disabilities) to justify non-retention as well as acceleration and placement (e.g., disaggregated data requirements in North Carolina).
- In many states, **districts are required to notify parents, involve them in decisions and document the retention, promotion and acceleration processes.**
  - This requires LEAs to build assessment and data-tracking systems capable of supporting this level of communication.
- Many states (e.g., AZ, IN, NV, OH) with retention laws require LEAs to **provide intervention services to retained students, and track their progress.**
  - Ohio requires schools to create a reading improvement and monitoring plan and provide high-dosage tutoring for any retained students as well as any students promoted to 4th grade on an exemption.

Key Policy Trade-Offs & Recommendations

TRADE-OFFS

We see key trade-offs and decision points for states interested in streamlining and driving coherence across their state assessment policies

Prescriptiveness vs flexible choice balancing consistency and quality with local autonomy

State-provided tools vs. approved lists vs. local autonomy determining what role the state should play in curating or supplying assessments

Frequency and type of data required by the state clarifying what evidence is truly necessary for accountability, improvement and instruction

Incentives vs. restrictions choosing whether to shape behavior through supports (PD, funding, tools) or limits (time caps, bans, intervention)

RECOMMENDATION #1

States should develop purpose-driven, state-vetted lists of high-quality assessments with flexibility pathways and clear incentives for adoption

- To addresses inconsistent quality, misuse of assessments and the accumulation of overlapping tools driven by weak evidence and uneven district capacity, states should create a short, purpose-driven vetted set of high-quality assessments that meet clear evidence and technical-quality thresholds.
- States should begin with screeners and diagnostics, where variation in quality is greatest and state oversight is most needed. Interim assessments and HQIM-embedded measures should be included in alignment with instructional strategy and/or a through-year model.
- States should pair vetted lists with incentives (funding, training, technical assistance) to support tool adoption, implementation and educator training.

RECOMMENDATION #2

States should require a periodic district assessment inventory and streamlining cycle supported by state-provided tools, templates, and capacity-building

- Assessment clutter may be accumulating at the district level due to lack of structured routines, and incoherence between local assessment systems and state priorities.
- States should require districts to complete an assessment inventory and streamlining process every 3–4 years, aligned to procurement cycles and district capacity. A rolling statewide schedule can distribute SEA workload and allow for targeted support.
- States should provide a standard, easy-to-use template, purpose categories and an instructional utility rubric aligned to HQIM and state standards. Inventories should focus on clarifying purpose, identifying redundancies and ensuring instructional usefulness, not simply cataloging tools.
- States should pair this expectation with incentives and supports, such as grants, sample inventories, training modules and coaching. Public summaries can strengthen transparency and family understanding.

RECOMMENDATION #3

States should invest in high-quality, standards-aligned progress monitoring options while allowing opt-outs for districts using strong HQIM-embedded measures

- Incoherence created by interims and progress-monitoring tools produce data disconnected from curriculum, forcing teachers to triangulate conflicting signals.
- States should invest in or negotiate access to high-quality, standards-aligned progress-monitoring and instructional assessment tools that are instructionally meaningful and reduce reliance on misaligned commercial interims.
- Districts using HQIM-embedded or HQIM-aligned assessments that meet state evidence thresholds should be allowed to opt out of state-provided tools to avoid duplication and maintain coherence with curriculum and instruction.

RECOMMENDATION #4

Right-size literacy screening frequency and use appropriate tools for targeted follow-up assessments based on student need

- Overuse and misuse of literacy screeners contributes to clutter, reduces instructional time and blurs the distinction between screeners, diagnostics and progress monitoring tools.
- States should interrogate blanket requirements for multiple screening windows to a right-sized model that prioritizes additional checks only for students flagged for risk or whose classroom performance warrants follow-up.
- Clear guardrails should ensure screeners are not used as diagnostics or progress monitoring tools, preventing misuse and reducing unnecessary testing.
- States should pair this shift with strong guidance and capacity-building to help teachers interpret and act on screening results. Messaging should anticipate concerns from stakeholders accustomed to frequent data points.

Simply mandating universal literacy screening three times a year will not necessarily translate to improved literacy outcomes without effective follow-up (diagnostics, interventions, progress-monitoring), sufficient district and school capacity, and adequate teacher and leader training and support.

In an ideal scenario….

Teachers are trained on how to administer, interpret and analyze screener results

All students are screened at the beginning of the year using a valid, reliable screening instrument to identify students needing extra support

Students flagged by the screener are given a diagnostic assessment to pinpoint their specific reading difficulties

Teachers use results from the diagnostic to plan appropriate interventions and progress monitoring supports

RECOMMENDATION #5

Strengthen math screening policy by clarifying key constructs and evidence requirements, while proceeding cautiously in a developing research field

Premature mandates could lock districts into weak or inconsistent tools in an emerging area of research. Given the emerging and still unsettled evidence base for math screeners, states should proceed cautiously.

Instead of mandating specific tools, states should:

- Clarify priority early-math constructs (e.g., number sense, relational thinking);
- Align those constructs to standards and HQIM learning progressions;
- Establish evidence expectations for validity, classification accuracy and intended use;
- Provide guidance on selecting tools and interpreting results;
- Build educator capacity in foundational math development and dyscalculia and dysgraphia risk identification.

RECOMMENDATION #6

Set clear expectations for how districts use data to enter, monitor, and exit student interventions—aligned with IDEA, MTSS and instructional pathways

Duplication and misuse of assessments within MTSS, reinforces instructional incoherence and promotes inequitable variation in district decisionmaking.

States should define a limited, coherent set of data sources (e.g., screeners, progress monitoring, classroom evidence) that districts must use to identify students for interventions and determine when they should exit.

Guidance should:

- Emphasize using the right assessment for the right purpose;
- Ensure teachers are not responding to multiple “progress monitoring” signals from different assessment tools to understand and respond to student needs;
- Prevent redundant or duplicative testing to “prove” eligibility;
- Ensure interventions are time-limited, flexible, and instructionally coherent;
- Align state expectations with IDEA, dyslexia laws and MTSS frameworks.

Rethinking the Test Pile: State Policy Scan

Education-first.com

Acknowledgements

This report is based on research funded by the Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect the positions or policies of the foundation.

Education First staff Khaled Ismail, Senna Lamba and Olivia Kelly led this research and authored the report. Austin Wechter designed the report.

Suggested Citation (APA): Education First. (2026). Rethinking the test pile: State A Scan of State Assessment Policies.

References

Research and Technical Literature

- Codding, R. S., Nelson, G., Kiss, A. J., Shin, J., Goodridge, A., & Hwang, J. (2025). A meta-analysis of relations between curriculum-based measures in mathematics and criterion measures. School Psychology Review, 54(3), 275–290.
- Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics. The Journal of Special Education, 41(2), 121–139.
- Gareis, C. R., McMillan, J. H., Smucker, A., & Huang, K. (2021). MAP Growth validation study. William & Mary & Virginia Commonwealth University.
- Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for primary grades. Exceptional Children, 78(4), 423–445.
- Hall, G. J., Mejias Vazquez, E., & Wackerle-Hollman, A. K. (2025). Predictive strength of math screening across English language proficiency. Learning Disabilities Research & Practice, 40(2), 71–85.
- Ketterlin-Geller, L. R., Shivraj, P., Basaraba, D., & Schielack, J. (2019). Universal screening for algebra readiness. Investigations in Mathematics Learning, 11(2), 120–133.
- Kiss, A. J., & Christ, T. J. (2020). Screening for math in early grades. Assessment for Effective Intervention, 45(1), 38–50.
- Klingbeil, D. A., Maurice, S. A., Van Norman, E. R., et al. (2019). Improving mathematics screening in middle school. School Psychology Review, 48(4), 383–398.
- McNeil, N. M., Jordan, N. C., Viegut, A. A., & Ansari, D. (2025). Arithmetic fluency and the science of learning. Psychological Science in the Public Interest, 26(1), 10–57.
- Nelson, G., Kiss, A. J., Codding, R. S., et al. (2023). Curriculum-based measurement in mathematics. Journal of School Psychology, 97, 1–42.
- Ozernov-Palchik, O., Elizee, Z., Catania, F., et al. (2025). (Not so) universal literacy screening. Annals of Dyslexia, 75(3), 489–523.
- ten Braak, D., Lenes, R., Purpura, D. J., Schmitt, S. A., & Størksen, I. (2022). Early mathematics skills and later achievement. Journal of Experimental Child Psychology, 214, 105306.
- VanDerHeyden, A. M., & Codding, R. S. (2020). Belief-based versus evidence-based math assessment. Communiqué, 48(5), 20–25.
- Xu, Z. (2024). Predicting math proficiency using K–3 diagnostic assessments. American Institutes for Research.

References

State Legislation and Regulatory Guidance

- Alabama State Legislature. (2022). Senate Bill 171: Alabama Numeracy Act.
- Arkansas Department of Education. (2024). Commissioner's Memo COM-25-008: Launch of K–2 assessment system.
- Arkansas Department of Education. (2024). Commissioner's Memo LS-25-017: ATLAS K–2 math screener guidance.
- Delaware Department of Education. (2021). 14 DE Admin. Code 508 Multi-Tiered System of Support.
- Florida State Legislature. (2025). Florida Statutes §1008.25 Public school student progression.
- Indiana General Assembly. (2025). House Bill 1634.
- Kentucky General Assembly. (2005). Kentucky Revised Statutes §158.791 Read to Succeed Act.
- Mississippi Legislature. (2020). Miss. Code Ann. §37-23-16 Early literacy and numeracy screening.
- Nebraska Legislature. (n.d.). Nebraska Reads Act statutes.
- North Dakota Legislative Assembly. (2025). Senate Bill 2213.
- Oklahoma State Legislature. (2025). Senate Bill 140 Math Achievement and Proficiency Act.
- Rhode Island Board of Education. (2023). Regulation 12863 amendment.
- West Virginia Legislature. (2023). House Bill 3035 Third Grade Success Act.

Media, Advocacy, and Contextual Sources

- DeMauro Mullins, A. (2023). West Virginia invests in third grade success. ExcelinEd in Action.
- DeMauro Mullins, A. (2025). State actions updates: June 2025 education reforms. ExcelinEd in Action.
- ExcelinEd. (n.d.). Literacy matters: State policy guidance and resources.
- ExcelinEd. (n.d.). Math matters: State policy guidance and resources.
- Education Week. (2025). Schools prioritize reading intervention. But what about math?
- Education Week. (2025). What is a math screener? Three things to know.

References

State Guidance, Policy, and Instructional Frameworks

- Arizona Department of Education. (2025). Middle school mathematics guidance.
- Arkansas Department of Education. (2024). Commissioner's Memo COM-25-008: Launch of K–2 assessment system.
- Arkansas Department of Education. (2024). Commissioner's Memo LS-25-017: ATLAS K–2 math screener guidance.
- Arkansas Department of Education. (n.d.). Academic playbook: Assessments.
- California Department of Education. (2023). Mathematics framework for California public schools.
- Colorado Department of Education. (2022). READ Act advisory list assessment review process.
- Colorado Department of Education. (2022). Assessment review application: Part II.
- Delaware Department of Education. (2021). 14 DE Admin. Code 508 Multi-Tiered System of Support.
- Georgia Department of Education. (n.d.). K–12 ELA instructional framework companion.
- Kentucky Department of Education. (n.d.). Numeracy screening assessments selection criteria.
- Kentucky Department of Education. (2025). Numeracy screening assessments.
- Maryland State Department of Education. (2025). High-quality instructional materials identification framework.
- Maryland State Department of Education. (n.d.). PreK–Grade 3 reading screening assessments.
- Maryland State Department of Education. (n.d.). PreK literacy policy.
- Maryland State Department of Education. (n.d.). Students at risk for reading difficulties implementation guide.
- Massachusetts Department of Elementary and Secondary Education. (2017). ELA and literacy curriculum framework.
- Massachusetts Department of Elementary and Secondary Education. (n.d.). Screening assessments guidance.
- Massachusetts Department of Elementary and Secondary Education. (n.d.). Screening assessment criteria.
- Minnesota Department of Education. (n.d.). Literacy screening guidance.
- Mississippi Department of Education. (n.d.). Instructional planning guides for English language arts.
- Mississippi Department of Education. (n.d.). Universal screener and diagnostic assessment guidance.
- Mississippi Legislature. (2020). Miss. Code Ann. §37-23-16 Early literacy and numeracy screening.
- Nebraska Department of Education. (2021). Literacy assessment rubric.
- Nebraska Department of Education. (n.d.). Nebraska Reads approved assessments.

References

State Guidance, Policy, and Instructional Frameworks Cont.

- Nevada Department of Education. (2025). PreK–12 Nevada State Literacy Plan.
- New Jersey Department of Education. (2025). New Jersey literacy framework.
- New York State Education Department. (2025). District assessment procurement guidance.
- Ohio Department of Education and Workforce. (2025). Ohio's plan for K–12 mathematics.
- Oklahoma State Legislature. (2025). Senate Bill 140 Math Achievement and Proficiency Act.
- Oregon Department of Education. (n.d.). Approved universal screeners.
- Rhode Island Department of Education. (2021). ELA/Literacy curriculum framework.
- Rhode Island Department of Education. (2024). Literacy screening guidance.
- Early Literacy Matters. (2024). Rhode Island implementation report.
- Tennessee Department of Education. (2024). Reading 360.
- Tennessee State Board of Education. (2021). Universal screeners policy 3.302.
- Tennessee State Board of Education. (2022). Universal student screeners rule 0520-01-03-.15.
- Tennessee Department of Education. (2022). Comprehensive universal screener matrix.
- Virginia Department of Education. (2024). Mathematics playbook implementation guidance.
- West Virginia Department of Education. (2024). Approved universal screeners: Measures of Academic Progress.
- West Virginia Department of Education. (2025). Third Grade Success Act approved screener list.
- West Virginia Legislature. (2023). House Bill 3035 Third Grade Success Act.
- Wisconsin Department of Public Instruction. (2021). Wisconsin standards for mathematics.