



Rethinking the Test Pile:

Executive Summary



February 2026

Assessment systems are meant to strengthen teaching, learning and system improvement. When designed well, high-quality assessments help educators focus instruction, help leaders make informed decisions and build shared understanding about what students know and can do. Instead, in many states and districts, they have grown into a bloated, incoherent pile of tools that consumes time, creates confusion and too often erodes trust without improving learning.

Rethinking the Test Pile is a new national study by Education First examining how academic assessment systems are evolving across states and districts. The research includes a national study of K–8 district assessment practices, a scan of state policies shaping local assessment architecture and a review of past assessment audit and literacy efforts.

We examined what academic assessments students take and why, how much time and money they require, how leaders decide what to administer and how state policy choices shape local systems. We catalogued assessments in 67 districts in 38 states and interviewed district and state leaders, analyzed assessment statutes and guidance and reviewed past audit and assessment literacy initiatives in 27 states. All together, we looked across 42 states and discovered remarkably consistent challenges.



Three core problems are emerging across findings

1. Assessment clutter

Layered state policy requirements and district adoption across screening, interims, MTSS and instructional guidance create overlapping testing systems that districts struggle to streamline.

2. Limited evidence of quality and instructional value

States and districts lack clear and transparent evidence to determine whether tools are high-quality and demonstrate valid evidence of serving instructional purposes.

3. Incoherence with teaching and learning

Assessment expectations often develop separately from curriculum and instructional strategy, leaving educators to juggle conflicting data signals that don't match what they're teaching.

These challenges do not originate from a single policy or level of the system. They reflect how multiple policies and levels of the system interact and fragment over time.

The roots of incoherent assessment systems are deep and intertwined. Even when states emphasize local control, policy choices around screening, interim expectations, accountability and instructional frameworks increasingly define what districts administer and how often.

Addressing the test pile requires more than removing unpopular or costly tests. It requires thoughtful decisions about how policies, guidance and instructional expectations work together and a stronger focus on assessments that provide clear instructional value. The opportunity ahead is not simply to reduce testing, but to build assessment systems that are coherent by design, where each measure has a clear purpose and supports teaching and learning.

Key Findings

Assessment volume and complexity of local systems continues to grow

- In some districts, students take as many as 88 academic assessments before entering high school, many requiring multiple administrations per year.
- Students in the highest-testing districts took an average of 7 different assessments per grade. Each assessment was typically administered 1–3 times per year. Most of these assessments are district selected and administered (60%) with the remainder (40%) required and administered by the state.
- English learners experience higher testing volume than their peers, with an average of 47 additional hours of testing across K–8.
- Many districts may be extending the use of assessments in ways that are inconsistent with their intended design, increasing the risk of misinterpreting results and limiting the usefulness of the data to inform instruction or drive improvement efforts effectively.
- District leaders consistently reported a desire to streamline assessment systems, but described state policies as a major driver of accumulation.

We found no relationship between assessment volume and academic proficiency or growth

- We found no statistically significant relationship between testing volume (number of assessments and testing time) and proficiency on ELA and Math end-of-year summative assessments, nor growth in ELA or Math proficiency as shown in the Education Recovery Scorecard.
- Districts may be expending time and resources on redundant or low-value tests that don't meaningfully translate to action to improve student learning.

State policy is increasingly defining local assessment systems

- States are making policy choices around screening requirements and frequency, interim expectations, accountability uses, MTSS and intervention requirements, retention and acceleration policies and graduation rules.
- These policies influence which assessments districts and schools feel compelled to administer, how frequently they are used and the purposes they are expected to serve.
- Even policies not labeled as assessment policy, such as literacy frameworks or intervention rules, shape district testing calendars and tool adoption.

Screening policies are expanding rapidly without consistency

- 42 states now require universal reading screening, often paired with/related to dyslexia screening and support requirements.
- 18 states have introduced math screening policies. Roughly half of the states have implemented the requirements, leaving room to clarify purpose and ensure that new math screening policies don't add to the test pile.
- States diverge on:
 - What gets measured and how assessment quality is defined
 - Frequency and timing of screening
 - Whether tools are consolidated or layered onto existing systems
 - How prescriptive states are about selection, approval and adoption

This variation increases the risk that districts adopt multiple tools that serve overlapping purposes.

Statewide interims are increasingly influencing local assessment practice

- Statewide interim models, whether required or optional, are expanding and reshaping local assessment systems.
- At least 20 states now provide statewide interim options, including through-year models that increasingly signal new expectations for districts.
- States diverge in how interim models are designed and used, including whether they are curriculum agnostic or curriculum relevant, whether they are tied to accountability and how much instructional reporting they provide.
- In some states, statewide interim assessments and through-year models help replace lower-quality commercial tools and reduce overall testing. In others, especially when statewide interims aren't instructionally useful or aligned to HQIM, districts keep using their own interim products, creating parallel systems that increase burden and weaken coherence.

As states expand requirements and vendors position each tool as essential, teachers are left navigating layered data streams that send mixed signals for implementation

- Teachers often receive data from multiple assessments that don't align clearly with the curriculum or grade-level instruction, yet they're still expected to adjust their teaching based on it.
- Many vendors market their assessments as useful for instruction, yet offer little transparent evidence, even to purchasers, about how they actually achieve that purpose.
- States often approve or mandate tools without clearly defining evidence expectations, leading districts to extend tools beyond their intended use.
- District leaders reported difficulty determining which assessments truly support teaching and learning versus compliance needs.

Past audit efforts raised awareness but did not resolve structural drivers

- State and district assessment audits consistently surfaced the same issues seen today: clutter, weak evidence of quality and misalignment with curriculum.
- Many audit efforts focus on inventories rather than policy alignment, limiting long term impact.
- Current assessment literacy tools help districts navigate complexity but do not address underlying policy pressures.



In less cluttered assessment systems:

The most promising examples of state and local actions share common features:

- Clear prioritization of assessment purpose and use
- Alignment between assessment policy, design and instructional priorities
- Greater transparency about required quality evidence and intended uses
- Efforts to consolidate rather than layer new measures over time

Complex problems demand multiple solutions

Across the district study, state policy scan and review of audit efforts, these patterns surfaced again and again. Rather than a single fix, the research points to a set of reinforcing actions that states and districts can take to build more coherent systems. Detailed, audience-specific actions for state leaders, district teams and advocates are included in the full Rethinking the Test Pile reports.

Clarify the role and purpose of assessments across systems

States and districts should more clearly define the intended role of different assessment types and how they fit together. Screening, interim and formative expectations are often layered without prioritization, leading to duplication and mixed signals for educators. Clearer decision rules about purpose and use can help reduce clutter while preserving necessary measures.

Strengthen evidence expectations for instructional use

Across analyses, leaders reported limited transparency about whether tools meaningfully support teaching and learning. States and districts should establish clearer expectations for evidence of validity, reliability and instructional utility, particularly for tools positioned as supporting daily instruction. Procurement and approval processes should reinforce these expectations so districts are not left to interpret quality alone.

Align assessment policy with instructional strategy and HQIM

Assessment systems should reinforce, rather than compete with, investments in high quality instructional materials and grade level learning. Policy signals that elevate misaligned tools can pull instruction away from intended curriculum pacing and coherence. States and districts should prioritize curriculum-relevant designs and clarify how assessment data connects to instructional decisions.

Consolidate rather than layer new requirements

As literacy and math screening policies expand and statewide interim models evolve, states have an opportunity to design policies that replace existing measures rather than add new ones. Clear consolidation strategies and incentive structures can reduce burden and improve coherence across state and local systems.

Move beyond one time audits toward ongoing system design

Past assessment audit efforts increased awareness of testing volume but rarely produced sustained change because they were not connected to broader policy shifts. Future approaches should link audit tools, assessment literacy supports and policy design so that districts are supported in building coherent systems over time, not simply cataloguing assessments.



Access all of the
Rethinking the Test Pile resources

[click here to access](#)