educationfirst

Rethinking the Test Pile:

A National Study of K–8 Academic Assessments

February 2026

Table of Contents

About The Report

Education First conducted this national study to shine a light on local K–8 academic assessment systems in U.S. school districts. The last fieldwide analysis of local assessments was completed more than a decade ago, leaving policymakers and educators without an updated view of how assessment systems have evolved. This report begins to fill that gap. By uncovering patterns, pressures and points of overload, we aim to spark conversation and action about how to streamline and declutter academic assessment systems to focus on assessments that provide the most valuable data to better serve students, educators and families.

About the Authors

Education First staff Khaled Ismail, Kelly Jasiura, Yeonjae Lee and Jennifer Vranek 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

Study Overview

Drawing on a diverse sample of districts, **we examine what academic assessments students are taking in grades K–8, why they are administered, how much time and money they require, and how the data are being used across classrooms, schools and systems.** We also explore how state policy contexts may be shaping district choices and how practices differ across district types.

The findings offer a system-level view into current assessment practices and highlight critical opportunities for strategic coherence and improvement to rebuild trust in standardized testing and ensure academic assessments are purposeful, efficient and aligned to instructional priorities.

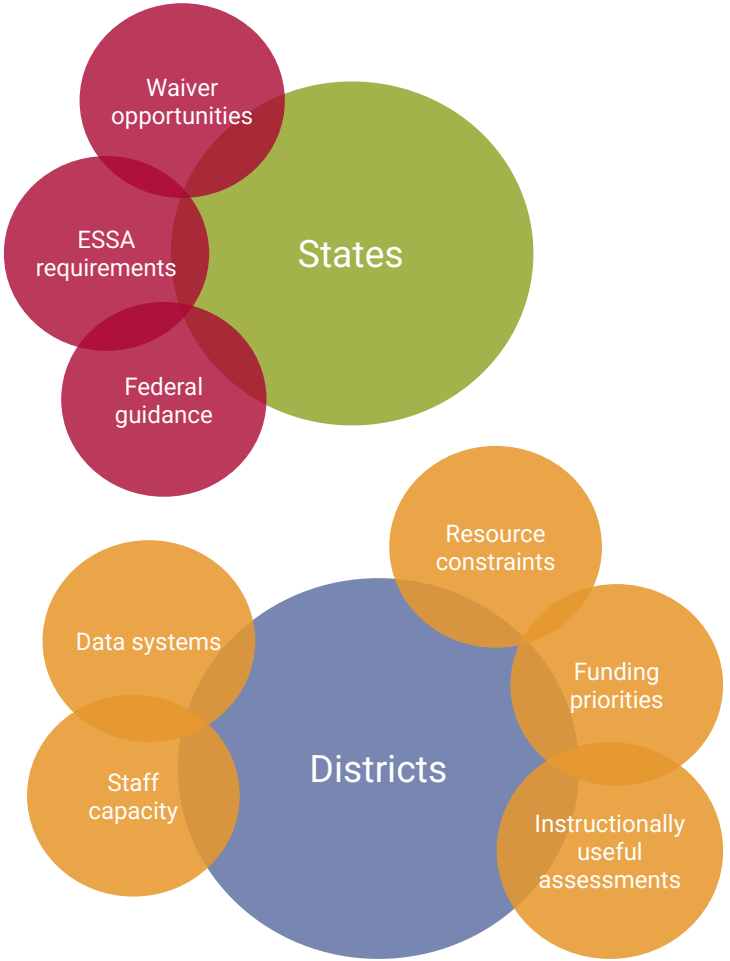
We prioritized K–8 academic assessments administered or required by states and districts

We included	We did not include
■ **State procured and administered academic assessments (Grades K–8):** Summative tests in ELA, math, science, and in some states, social studies; English learner screeners and proficiency tests; and other state-level requirements (e.g., kindergarten readiness) ■ **District procured and administered academic assessments (Grades K–8):** Gifted and dyslexia screeners, diagnostics, entrance exams, interims, benchmarks, end-of-course tests and district-wide classroom-embedded assessments across core subjects	■ Assessments that individual schools or teachers may choose to administer ■ Curriculum-embedded assessments that are not mandated district-wide ■ Non-academic assessments such as social-emotional learning surveys, school climate or condition surveys, student perceptions and other feedback instruments ■ Alternate assessments for students with significant cognitive disabilities ■ Observation-based assessments (used for kindergarten readiness)

This assessment study comes at a critical moment for U.S. education systems

States are bound by the Every Student Succeeds Act (ESSA) testing requirements—even as shifting federal guidance and new waiver opportunities create both openings for innovation and a climate of uncertainty. Districts also are navigating resource constraints, shifting priorities and pressure to deliver assessments that are instructionally useful, within the limits of procurement cycles, data systems and staff capacity. At the same time, state legislative proposals (e.g. Texas HB8) and waiver requests (e.g. Oklahoma) are reshaping the landscape.

In this context, our study shows evidence of persistent misalignment across levels—state, district, school and classroom. **With trust running thin among states, districts, vendors, educators and families, the challenge of building coherent, responsive systems is even more urgent.**



After a decade of progress on implementing high-quality instructional materials, incoherent assessments now threaten to stall the momentum

Instructional materials have come a long way.

Over the past decade, states and districts have invested heavily in High-Quality Instructional Materials (HQIM) adoption and professional learning to strengthen teaching and learning.

But assessment systems haven't kept pace.

Many school systems still rely on layered, redundant and misaligned tests that offer limited instructional value and send mixed signals about what matters for learning.

This disconnect creates noise, not insight.

Teachers and leaders are left with conflicting data from multiple sources—obscuring what students actually know and can do within the context of HQIM.

As a result, time, trust and coherence are lost.

This study begins to scratch the surface of these challenges. It offers an early look at how local assessment ecosystems are evolving and where the next wave of coherence work must focus.

Amid these mounting pressures, states and districts have an opportunity to examine their assessment systems to advance coherence and reclaim time and resources for instruction

School systems across the country are facing:

Declining academic proficiency:

Many districts are grappling with stalled post-pandemic recovery and pressure to raise student performance.

Policy and funding volatility:

Leadership changes, shifting priorities and declining enrollment and revenue add to uncertainty and instability.

Technology Advances:

Rapid advances in AI, edtech and next-generation assessments offer promise, but require careful planning and use to protect quality.

Instructional Shifts:

Many states and districts are rolling out evidence-based literacy and math strategies and continuing to adopt HQIM, creating opportunities for better alignment with assessments.

Key Findings

Insights from District Academic Assessments Analyses

1	**88 assessments before high school**	**In some districts, students take as many as 88 assessments before entering high school.** The testing load is greater in districts serving more students of color and low-income students, potentially reinforcing existing inequities.
2	**47 extra hours for ELLs**	**For English Language Learners, testing volume is especially high**, averaging about 47 additional hours devoted to ELL-specific assessments.
3	**No relationship between volume and proficiency**	**In our sample of districts, there is no relationship between assessment volume 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. This may suggest districts may be expending time and resources on redundant or low-value tests that don't meaningfully support student progress.
4	**Assessments not aligned with design**	**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.

Key Findings

Insights from District Leader Interviews

1	**Desire to streamline assessments**	**District leaders have a strong desire to streamline assessments** into a cohesive system, but they often face challenging decision points as they navigate tensions among local needs and state policies while trying to balance overtesting.
2	**Assessments should support teaching and learning**	**District leaders emphasize that one of the primary purposes of assessment should be to support teaching and learning**, prioritizing tools and practices that provide timely, actionable insights to guide instruction and improve student outcomes.
3	**Commitment to results used meaningfully**	**District leaders are committed to ensuring that assessment results are used meaningfully** and are investing in professional development and tools to help school leaders and educators to effectively interpret and act on assessment data.

Summary of Recommendations

Addressing a cluttered assessment landscape isn't as simple as removing unpopular or costly tests. It's influenced by tensions among state mandates and local flexibility, varied assessment purposes and uses, and the distinct needs of administrators and educators. To overcome these challenges, a district's assessment suite must prioritize coherence and alignment across instruction, assessment and accountability.

States and districts should clarify the purpose of each assessment and prioritize tools that demonstrate clear instructional value. Procurement processes should demand evidence of instructional utility from developers and ensure this evidence reflects reliable, valid and fair uses of an assessment.

Developers must be transparent about intended uses, design for interoperability between multiple assessment and data systems, and provide teachers with actionable, instructionally useful and standards-aligned information. Developers should work with districts to replace—not add to—tests with a focus on instructional utility and coherence at the center.

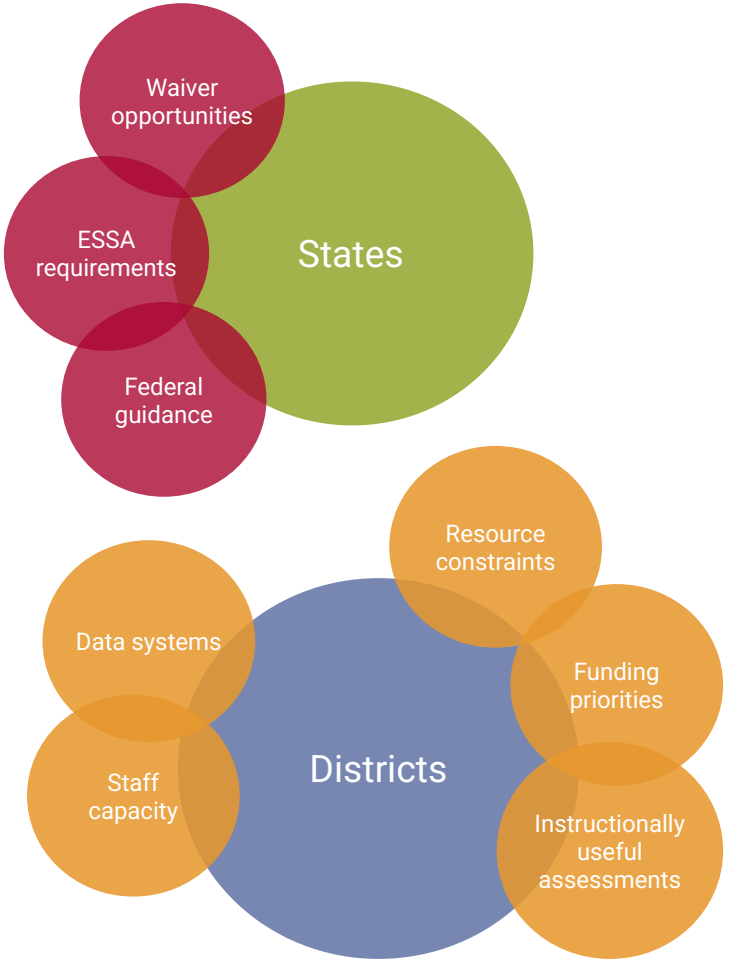
Advocates and policymakers should press for policies that align assessment policy and local selection to High-Quality Instructional Materials and grade-level instruction, call out duplicative mandates and demand transparency through public calendars and cost analyses. Hold vendors accountable and build public momentum for coherence.

Introduction & Methodology

This assessment study comes at a critical moment for U.S. education systems

States are bound by the Every Student Succeeds Act (ESSA) testing requirements—even as shifting federal guidance and new waiver opportunities create both openings for innovation and a climate of uncertainty. Districts also are navigating resource constraints, shifting priorities and pressure to deliver assessments that are instructionally useful, within the limits of procurement cycles, data systems and staff capacity. At the same time, state legislative proposals (e.g. Texas HB8) and waiver requests (e.g. Oklahoma) are reshaping the landscape.

In this context, our study shows evidence of persistent misalignment across levels—state, district, school and classroom. **With trust running thin among states, districts, vendors, educators and families, the challenge of building coherent, responsive systems is even more urgent.**



After a decade of progress on implementing high-quality instructional materials, incoherent assessments now threaten to stall the momentum

Instructional materials have come a long way.

Over the past decade, states and districts have invested heavily in High-Quality Instructional Materials (HQIM) adoption and professional learning to strengthen teaching and learning.

But assessment systems haven't kept pace.

Many school systems still rely on layered, redundant and misaligned tests that offer limited instructional value and send mixed signals about what matters for learning.

This disconnect creates noise, not insight.

Teachers and leaders are left with conflicting data from multiple sources—obscuring what students actually know and can do within the context of HQIM.

As a result, time, trust and coherence are lost.

This study begins to scratch the surface of these challenges. It offers an early look at how local assessment ecosystems are evolving and where the next wave of coherence work must focus.

Amid these mounting pressures, states and districts have an opportunity to examine their assessment systems to advance coherence and reclaim time and resources for instruction

School systems across the country are facing:

Declining academic proficiency:

Many districts are grappling with stalled post-pandemic recovery and pressure to raise student performance.

Policy and funding volatility:

Leadership changes, shifting priorities and declining enrollment and revenue add to uncertainty and instability.

Technology Advances:

Rapid advances in AI, edtech and next-generation assessments offer promise, but require careful planning and use to protect quality.

Instructional Shifts:

Many states and districts are rolling out evidence-based literacy and math strategies and continuing to adopt HQIM, creating opportunities for better alignment with assessments.

But advancing coherence in assessment systems means grappling with critical design tensions…

How do we balance technical rigor and high-quality measurement with classroom utility and meaningful information for teachers?

Commercial assessments generally demonstrate strong technical quality and are standardized, making them easier to aggregate and report, but they are often misaligned with curriculum and instructional goals.

District-developed and curriculum-embedded assessments tend to be tightly connected to instruction, grounded in local context and more attuned to how students learn, though their technical quality is often uncertain or uneven.

Single-purpose assessments offer higher validity and reliability and allow for clear interpretation and use, but often require multiple assessments to meet different needs, leading to higher overall costs and testing time.

Multiple-purpose assessments can leverage evolving technology to reduce testing time and resource demands and enable more efficient data collection, but can also risk misinterpretation or being stretched beyond their intended purpose.

…and finding a way to meet different needs and priorities across multiple levels of the education system

How do we co-create assessment systems that balance tensions between standardization and flexibility, and between data for accountability and data for instruction?

Aligning state assessment requirements with policy priorities can ensure compliance, consistency and comparability across districts. This can be seen as less adaptable to local priorities.

Enabling local flexibility and tailoring assessment requirements to local contexts helps prioritize relevance to district goals, but can lead to wide variation and less alignment with state policy.

Prioritizing administrator and public needs in assessment requirements can meet demands of central offices, state leaders and caregivers for aggregated data, but often lacks granular insights for classroom use.

Prioritizing educator needs in assessment requirements can provide actionable information to teachers and school leaders for instruction, but are often difficult to aggregate.

This study aims to examine the current assessment demands on students in order to highlight critical opportunities for strategic coherence and improvement

In a diverse sample of school districts spanning different sizes, geographies and student demographics, we explored:

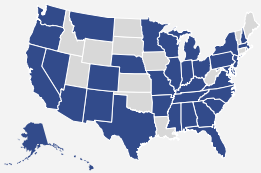
- What academic assessments K–8 students are taking and why
- How much time and money they require
- How system leaders decide what to administer
- How different stakeholders use the data
- How state policies may be shaping local choices
- Where patterns and variations appear across district types
- Whether districts are working to streamline and reduce testing

We chose a diverse set of districts from across the country to reflect system complexity and surface a range of leader perspectives

We focused on districts that primarily serve students of color and students living in poverty, spanning a range of geographies, locales, enrollment sizes and revenues.

Our sampling methodology is outlined in the Appendix.

Sample at a Glance



67 Districts from 38 states

21% Midwest
13% Northeast
45% South
21% West

61% City
27% Suburb
12% Town or Rural

Serving 3,967,140 Students
Average Enrollment: 59K
Min. Enrollment: 2K
Max Enrollment: 336K

Serving, on average, 78%+ Students of Color
Average % Black: 35.3
Average % Hispanic: 31.5
Average % Living in Poverty: 21.4

~20K average Revenue Per Student
Average Total Revenue: 1.2B
Min. Total Revenue: 32M
Max Total Revenue: 8.4B

In each district, we documented the assessments administered and explored how they are used; where available, we examined district spending on assessments

We gathered information on student demographics, district spending and revenue and academic performance from public sources.

We reviewed state requirements, district testing calendars, assessment websites and strategy documents to catalog the assessments used in each district.

For a subset of our districts, we examined district spending on assessments. We also looked at the revenue generated by the most common assessment vendors.

We also interviewed district leaders to better understand the factors guiding assessment adoption

We interviewed **31 assessment and academic leaders across 17 districts** to better understand their approach to assessment.

When possible, we spoke with staff from the assessment and academic offices to better understand the relationship between assessment adoption and instructional strategy.

We used a consistent approach to selecting and recording data to support clear comparisons across districts

We standardized how we categorized assessments and their primary uses to enable cross-district analysis. To do so, we…

- Imposed a single type label per assessment and a minimum use for each based on its intended design
- Focused on state- and district-required assessments in grades K–8 to understand system-level testing patterns
- Concentrated on assessments used with the general student population, including English learners, to capture the core testing landscape
- Applied consistent time estimates to allow comparisons across districts
- When available, used the developer's recommended time for assessment administration rather than a district's reported time to standardize comparisons

The research catalogued state- and district-administered academic assessments given to all students

State Assessments

State procured and administered academic assessments in grades K–8 including:

- Summative assessments in ELA, math and science, and for some states, in additional subjects like social studies
- English language learner screeners and end-of-year proficiency tests
- Additional requirements procured and administered at the state level (e.g. kindergarten readiness assessments, screeners)

District Assessments

District procured and administered academic assessments in grades K–8 including:

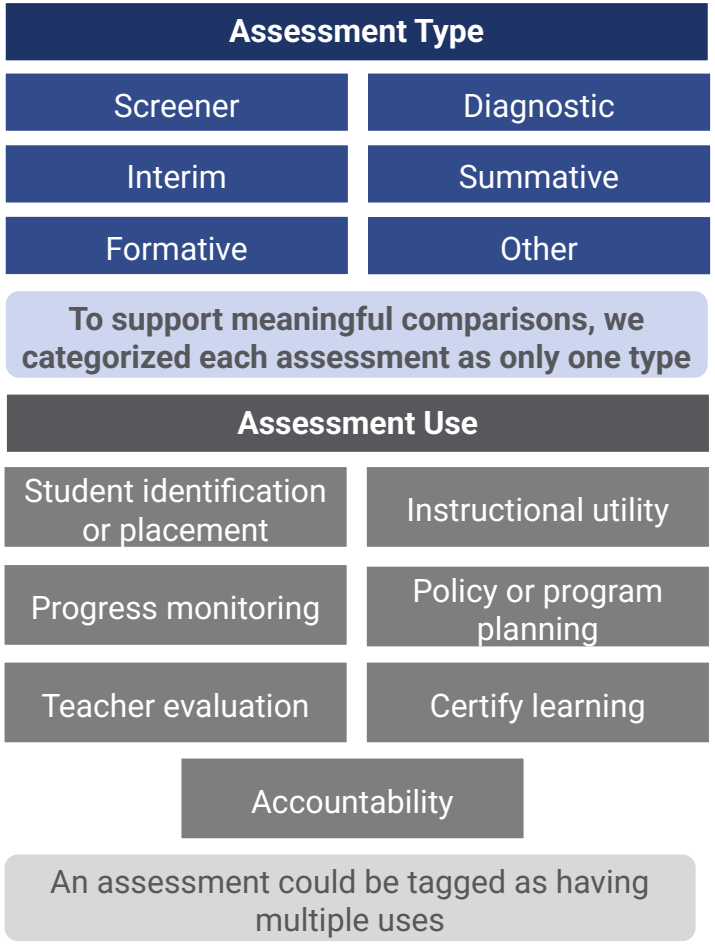
- Gifted screeners, dyslexia screeners, diagnostic assessments and high school entrance exams
- Interims, benchmarks and end-of-course assessments in core subjects including ELA, math, science and social studies
- District-wide classroom-embedded assessments
- District assessments for different student populations including English language learners, students enrolled in dual language programs and/or advanced course work, and students who failed or scored low on previous assessments
- State-required assessments that are selected and procured at the district level (e.g. early literacy screeners)

For each of the state- and district-administered assessments, we assigned a single primary type and a minimum use based on its design

Across the field, there is limited agreement upon set of definitions for assessment types, and products labeled as the same type (e.g., screener, diagnostic, interim) can vary widely in design and actual uses.

Districts also often describe and use the same assessments in different—and sometimes overlapping—ways, which created ambiguity about how to classify them.

To enable consistent analysis, we assigned each assessment a single primary type and primary intended use, based on a review of vendor documentation and district-reported purposes.



Find the definitions for each type and use in the Appendix.

Our analysis likely underestimates both the total number of assessments students take and the time they spend taking them

Assessments Not included In This Research

- Non-academic assessments, such as social-emotional learning surveys, school climate or condition surveys, student perceptions and other feedback instruments
- Assessments that individual schools or teachers may choose to administer
- Curriculum-embedded assessments that are not mandated district-wide
- Alternate assessments for students with significant cognitive disabilities
- Observation-based assessments (used for kindergarten readiness)

Also, the data we collected on assessment duration does not include time expended on (1) other aspects of test administration, such as reading instructions or implementing breaks; (2) preparing students for assessments; or (3) grading of assessments by teachers. The true time spent on assessments by students and teachers likely greatly exceeds our findings.

Landscape Overview & Findings

What academic assessments do students take, and how much do they cost?

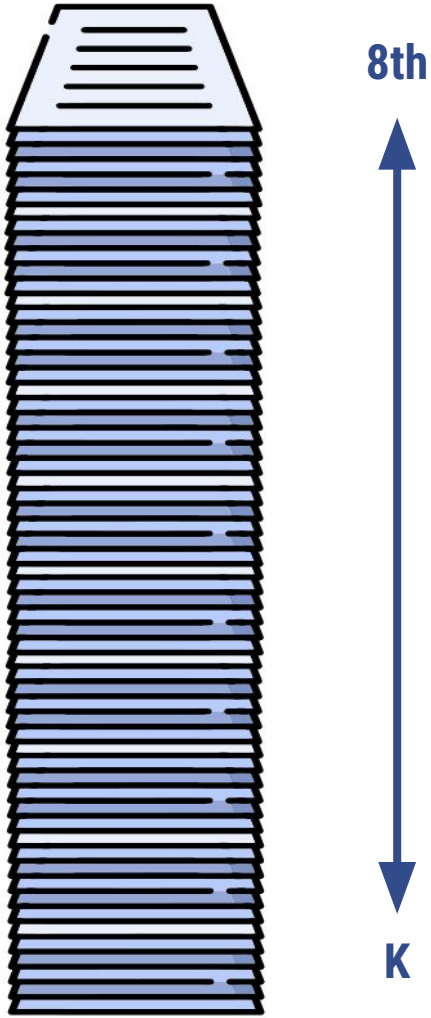
In some districts, students can take as many as 88 assessments over the course of their K–8 experience

In our sample, students in the highest-testing districts took an average of 7 different assessments* per grade, each given 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.

Number of state and district assessments administered to all students in grades K–8 for highest-testing districts

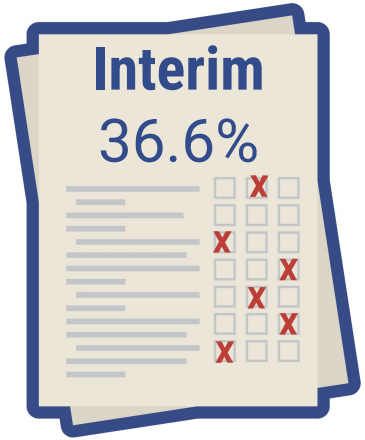
Minimum	Maximum	Median	Average
35	88	57	60

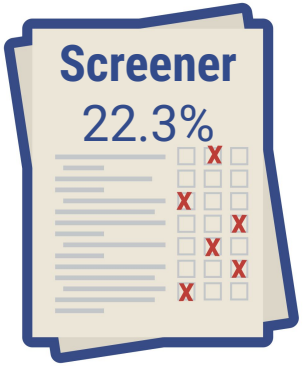
*Number of assessments counts distinct assessments per grade across K–8, not test administrations (e.g., i-Ready ELA in grades K–8 = nine assessments), so actual testing frequency and time burden are likely higher.



Districts in our sample most commonly administer screeners, diagnostics and interims…

often using them in overlapping ways to identify and place students, guide instruction and monitor progress—even when, in some cases, those uses do not align with the assessments' intended uses.





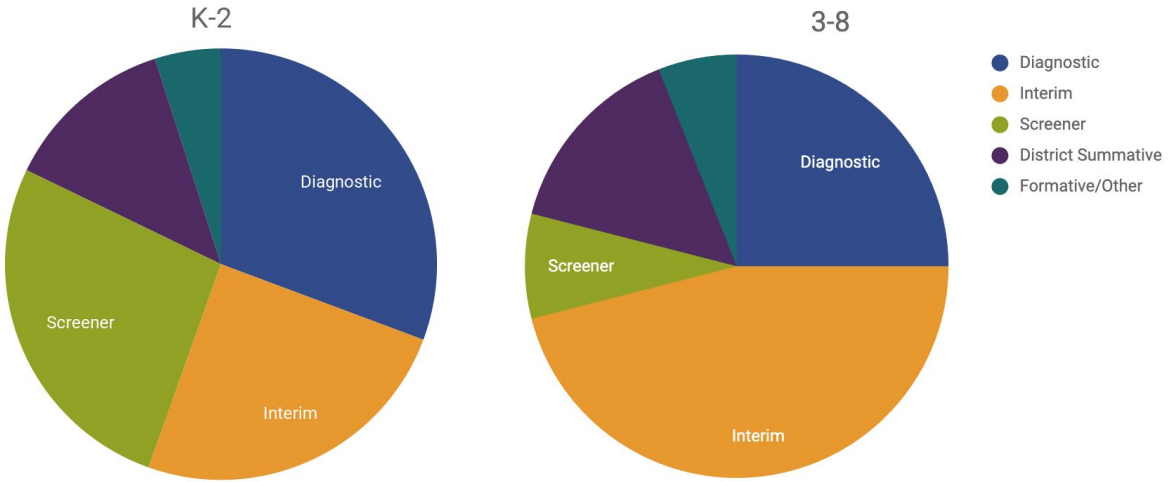
Summative 16.6%
Other 4.3%

Find the definitions for each assessment type in the Appendix.

In K–2, the most-used assessments are screeners and diagnostics, while grades 3–8 rely heavily on interim assessments to progress monitor and predict summative test performance—potentially reducing the instructional value of 3–8 tests

This difference in assessment types by grade band can likely be attributed in part to state requirements mandating universal screeners in early grades, and district decisions to administer interims for older students in preparation for end-of-year state summative exams.

District-Administered Assessments Type Breakdown by Grade Band



Districts tend to administer more literacy/ELA than math assessments, likely a result of stronger research evidence and growing state emphasis on early literacy improvement

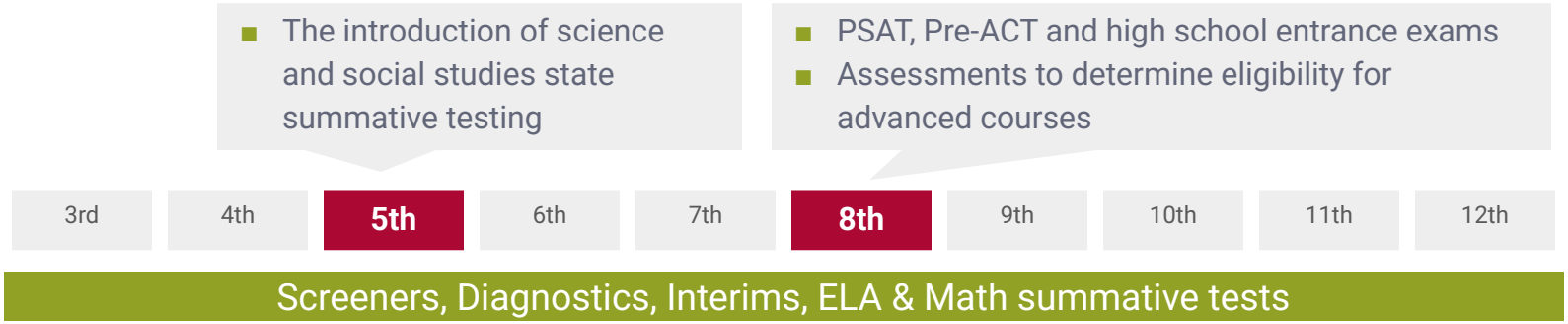
Number of Assessments Administered to All Students in Grades K–8 by subject

	Min.	Max.	Med.	Avg.
Math	6	39	18.5	18.8
Literacy and/or ELA	5	42	21	22.1

State policies that mandate assessments or may influence local assessment decisions:

- Universal reading screeners for all K–3 students
- Beginning-of-year grade 3 reading test to capture baseline data
- Requiring proficiency on a reading assessment for promotion to grade 4
- Statewide adoption of HQIM materials, which may include embedded assessments, to align with science of reading practices

Fifth and eighth grade are key academic milestones and also the most tested—amplifying the pressure on students at critical junctures



In addition to screeners, diagnostics, interims and traditional ELA and math summative tests, students in these grades face a layered assessment load that can include:

- 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

District and state spending on assessments represents a multi-billion-dollar annual investment

$2.5M
Per district
(over the past two fiscal years)

Across 20 districts of varying size and geography, average spending on assessments and implementation reached up to **$2.5M per district** over the past two fiscal years **($23–46 per pupil annually).**

$40M
Per state
(over the past two fiscal years)

In a separate analysis of 15 states, average state spending reached up to **$40M per state** over the past two fiscal years.

$3-5B
National annual assessment spending

Extrapolated nationally, this suggests **$3–5B** in annual assessment spending, likely an underestimate given sample limits and indirect costs.*

*Education First conducted an analysis of district and state assessment-related contracts from the two most recent fiscal years available in GovSpend. Inconsistent procurement reporting and incomplete fiscal year attribution mean that annual spending totals are approximate.

Hi, I'm Maya. Let me tell you about all the tests I've taken this year. First, I took a WIDA screener since my first language is Spanish and I was placed into an English learner class. I also take an i-Ready diagnostic on the computer in Spanish, which had questions that didn't really match what we were learning in class. That test happens three times. After that, I started doing i-Ready lessons on the computer every week to practice skills it said I was missing. Then came another test, NWEA MAP, to check how I was doing halfway through the year. I think we do that test two or three times, too. Most of these tests don't seem connected to the stories we read in class or the writing we do with our teacher. My mom says she's not sure what all the scores mean or how to help me. She gets a bunch of reports, but they all look different and don't say the same things. My MAP report said I was growing but my i-Ready report said I was really behind. It's kind of confusing for all of us.

In my English learner class, we take the WIDA Access at the end of the year to determine if we're ready to be placed out of this class. We also do quizzes and exit tickets our teacher makes to check if we understood the lesson. So, most weeks, we're doing some kind of test, either from the district, the computer or our classroom. And, we still have a big end-of-year test coming up. It's my first time taking that one and my teacher says it's really, really, really important. She says it will take about two hours to finish so we have to practice it a few times before to make sure I can get it right. If I do good on that one, I can go to fourth grade! All of this is just for reading. We haven't even started the math tests yet.

Note: Profile is a composite based on our stakeholder interviews and district research.

Meet

Maya

3rd grade
English Learner student
Midsize ex-urban district.

Maya lives in a state with required early literacy screening and a third-grade retention law.

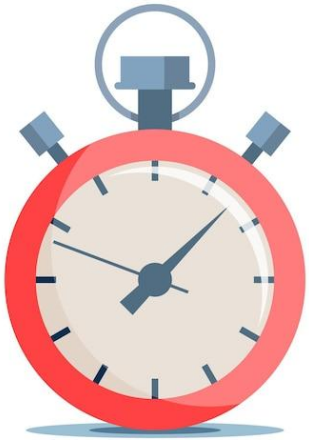
How much instructional time do students spend taking academic assessments?

On average, students in the highest-testing districts in our sample spend about 128 hours on assessments over their K–8 years

The total hours spent taking assessments in K–8 can vary significantly from district to district, and is primarily influenced by district choices rather than state requirements

Hours spent on assessments in grades K–8 for highest-testing districts	Minimum	Maximum	Median	Average
	100	222	113	128

These estimates likely undercount the total amount of time students spend taking assessments, as they do not include non-academic, school-mandated, curriculum-embedded or teacher-created assessments. They also do not account for the time teachers spend preparing students for testing or grading the assessments.



128
hours

The amount of time spent on assessments also increases by grade, with the max amount of hours spent on assessments in grades 6–8 more than double the amount spent in K–2

Max number of assessments and hours spent on assessments by grade band

	K–2	3–5	6–8
Max number of assessments	33	41	42
Max hours spent on assessments	49	88	108

Students in grades 3–8 spend more time taking assessments in large part due to state summative tests in ELA, math, science and social studies.

Districts may also choose to administer benchmark and interim assessments in these grades to prepare students for the end-of-year assessments.

Assessments in grades K–2 tend to focus more on shorter screener and diagnostic assessments.

Districts serving more students of color and more students living in poverty administer more assessments

In our sample, **districts with higher percentages of students of color administer, on average, 23 more hours of testing per year across K–8** than districts with lower percentages.*

Average Hours Spent Taking Assessments K–8	
Districts with higher % students of color enrolled	96
Districts with lower % students of color enrolled	73

*Districts classified as having a "higher % of students of color" are those above the sample median—more than 82% students of color, per NCES data.

In our sample, **districts with higher percentages of students living in poverty administer, on average, 9 more hours of testing per year across K–8** than districts with lower percentages.**

Average Hours Spent Taking Assessments K–8	
Districts with higher % students living in poverty enrolled	89
Districts with lower % students living in poverty enrolled	80

**Districts classified as having a "higher % of students living in poverty" are those equal to or above the sample median—equal to or more than 20% of students in poverty, per Urban Institute data.

English Language Learners (ELLs) can take on average about 47 hours of ELL-specific assessments from kindergarten through 8th grade

Some of this testing time is in addition to the general assessments non-ELL students take, while others replace or overlap with them, depending on students' instructional settings. Because not all ELLs are in sheltered settings, many likely experience both sets of assessments. ELLs appear to experience the same mix of screeners, diagnostics and interims—sometimes in their native language.

Districts in states that are transitioning to through-year assessments are currently administering more assessments, possibly because districts haven't yet eliminated redundant tests

Districts in our sample in states piloting or implementing through-year summative assessment systems administer, on average, **about 27 more hours of assessments per year** to K–8 students than districts in other states. Several factors likely contribute to this pattern:

- **State policy shifts have outpaced local adjustments:** Districts may still be running legacy assessments alongside the new system until they gain confidence that state changes will meet their local data needs.
- **Pilot years often drive higher testing volumes:** Early implementation typically requires extra administrations to generate comparability data, calibrate new tests and meet research/evaluation requirements.
- **Districts need support to streamline:** Without clear guidance and tools from the state, districts may struggle to retire redundant assessments or integrate new assessments into existing instructional cycles.

We analyzed the amount of time devoted to assessments in districts from our sample of states that are piloting or implementing through-year models, including Alaska, Florida, Indiana, Montana, North Carolina, Nebraska and Texas.

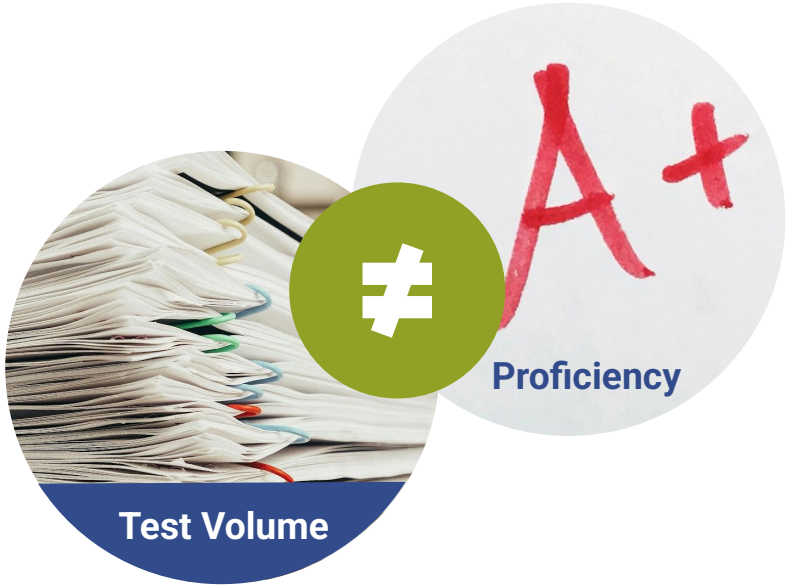
Through-year states:
~106 hours per year

Non-through-year states:
~79 hours

We found no relationship between assessment volume and academic proficiency or growth, suggesting an opportunity to better align assessments with instructional priorities

We found no correlation in our sample between assessment volume (a combined variable of number of assessments and time spent on assessments) 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. The absence of a relationship between these variables still held when we disaggregated the data by grade band (K–2 and 3–8).

This finding highlights an opportunity for districts to better align their assessment strategies with their instructional and academic goals, to ensure they use assessments strategically to improve student outcomes. It also highlights the need to provide districts with the tools and resources to enhance the value and use of their assessment data.



I'm Ms. Nguyen, a 7th grade ELA teacher. Our state is piloting a new through-year assessment model, which is supposed to replace the end-of-year test in a few years. This year we also started piloting an AI-based formative assessment platform. It's exciting to be part of something new that could eventually give us better, more timely data.

But right now, we're still giving the regular end-of-year test—plus the through-year tests, district benchmarks and the curriculum-embedded assessments we've always used. Each system has its own testing windows, practice sessions and scoring requirements, which have already eaten up almost a full month of instructional time.

And every test produces a different set of reports, with different scales and metrics, that don't line up with one another. I spend hours trying to reconcile contradictory results just to figure out what they mean for my students. Sometimes it feels like I have to be a data wizard before I can even be their teacher.

Note: Profile is a composite based on our stakeholder interviews and district research.

Meet

Ms. Nguyen

7th grade ELA teacher

Excited by new state innovations but overwhelmed by the implementation in her district

How do districts use the academic assessments they administer?

Districts use assessments for a range of purposes, prompting questions about how clearly those purposes align with their design and validity evidence

Three dimensions of evidence form the foundation for valid and appropriate uses for assessments:

1. **Intended Design/Use(s)**: What the vendor or developer states the assessment is designed to measure or support.
2. **Validated Use(s)**: The uses for which the assessment has been empirically validated through technical documentation and evidence.
3. **Actual Use(s) in Practice:** How districts and educators are applying the assessment in real settings, which may align with or diverge from the intended design or validated use(s).

Other factors, such as implementation quality, policy context and data culture also influence whether assessments are used effectively in practice.

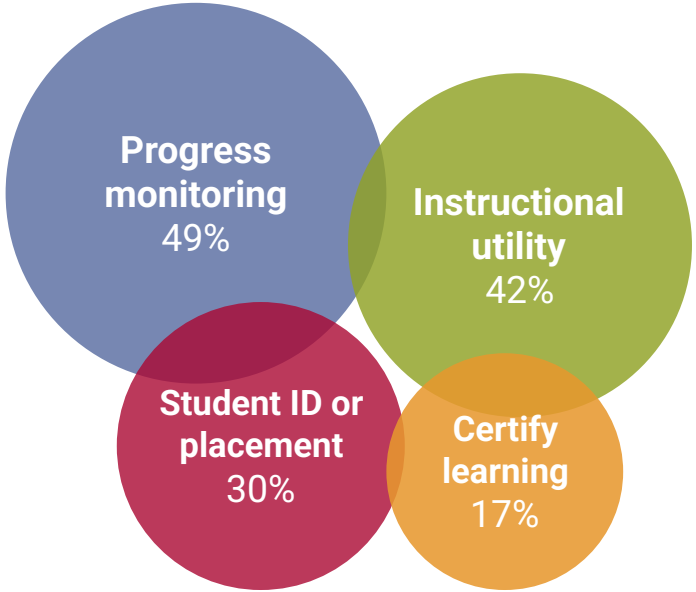
Our study explored the intended uses articulated by vendors and the actual uses articulated by districts. We did not review technical documentation or validity evidence for each assessment.

Source: American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2014)

Districts most commonly use their assessments to monitor student progress, guide instruction, identify and place students in specific programming or interventions

Among the districts in our sample, we found that assessments are most commonly used for progress monitoring (49%) and instructional decisionmaking (42%), followed by student placement (30%) and certifying learning (17%).

District assessments were often assigned multiple, overlapping—and sometimes conflicting—uses.



Percentages in this graph add up to over 100% because assessments are used for multiple purposes.

Many districts appear to stretch assessments beyond their intended purpose and design

In several instances, vendor descriptions were unclear about intended uses or did not include publicly available evidence demonstrating validity for those uses. Additionally, the uses that vendors describe often do not match how districts actually use the assessments. This suggests **potential overextension of assessments beyond their intended design or validity evidence**, though the study did not examine validity evidence directly. Ultimately, this disconnect can lead to misinterpretation of results and misaligned instructional and systemwide decisions.

20%
of assessments in our sample are **diagnostics** that may be appropriate for guiding daily instruction

but

42%
of assessments are identified as being used for **instructional utility**

37%
of assessments in our sample are **interims** that may be appropriate for monitoring student progress

but

49%
of assessments are identified as being used for **progress monitoring**

In some cases, assessments designed to screen or monitor progress are being used to guide daily instruction, reteach content and/or evaluate teachers

Despite being designed to identify risk, some screeners are often expected to provide fine-grained data to drive day-to-day instruction. Similarly, interim assessments, designed to progress monitor students' performance to inform district-level strategy or predict summative performance, are often used for reteaching decisions at the classroom level, with limited evidence of instructional or curriculum alignment.

For Example:

- **MAP Growth** is used in many districts to progress monitor student performance, identify students for interventions, guide instruction, and in some cases for teacher evaluation and accountability purposes.
- **DIBELS** is often used to identify and place students in programs or interventions, but some districts are also using it to guide instruction and monitor student progress.

Districts use the same assessments in different ways, including uses the assessments may not have been designed or validated to support

	District A	District B	District C	District D
map GROWTH	"MAP Growth measures what your student knows and how they are growing academically, allowing you to track your student's progress throughout the school year and across multiple years…[and is also used to] **appraise leaders/teachers**."	"MAP Growth creates a personalized assessment experience that accurately measures performance. Timely, easy-to-use reports **help teachers teach,** students learn, and administrators lead."	Primary uses for MAP Growth include "student instruction, **teacher evaluation** and school performance ratings."	"Teachers can use MAP Growth scores to monitor the growth of individual students and **develop classroom-level strategies for equitable instruction** that help maximize every student's learning potential."

Note: This study did not examine validity evidence for any of the district stated uses.

When used as intended, assessments can provide meaningful information for instruction, but weak links to curriculum can leave teachers with fragmented guidance

	District A	District B	District C	District D
DIBELS 8TH EDITION	"The District administers DIBELS to assist in making **data-driven instructional decisions for literacy.**"	DIBELS is used "to collect early childhood data, **monitor progress**, group students for interventions & support."	"DIBELS allows teachers to carefully monitor students' foundational reading development and reading comprehension. It also **helps teachers design reading instruction in individualized, targeted ways**."	"**Progress monitoring** throughout the year [through DIBELS] provides information on students' growth and the success of the intervention."

Note: This study did not examine validity evidence for any of the district stated uses.

High-quality, multi-purpose assessments can streamline testing, but weak curriculum connections or high-stakes uses like promotion can limit their impact

	District A	District B	District C	District D
i-Ready	"The i-Ready K-2 Assessments are a comprehensive suite of universal screening, diagnostic, and progress monitoring assessments…[i-Ready helps] **determine early literacy foundational skill levels and early numeracy skill levels** and identify students needing additional supports."	"Results of the end-of-year diagnostic test can also be used to determine **retention and promotion eligibility** for students in grades K–2 and 4–8."	"i-Ready is an online program that will help us determine your child's strengths and areas for growth, personalize their learning, and **monitor their progress** throughout the school year."	"Students will be provided the opportunity to meet **good cause promotion** criteria through an alternative assessment (i-Ready)"

Note: This study did not examine validity evidence for any of the district stated uses.

Meet

Ravi

Assessment specialist in a state department of education trying to make sure the state's policies, guidance and implementation are aligned

I'm Ravi, a state assessment specialist. A few years ago, our legislature passed a law requiring universal early literacy screening for all K–3 students, but we left it up to districts to pick their own tools and provided them with 30+ screeners on our state-approved list. On paper, it sounded like flexibility, but in practice most districts added new screeners on top of what they were already using. They were hesitant to let go of the old tools because they'd already built data systems, professional learning and intervention structures around them, and replacing anything would take years of procurement cycles and system reconfiguration.

Meanwhile, our agency has launched new grants to support HQIM adoption in early literacy, and we're just starting to talk about what kinds of assessments best align with those materials. But because we never gave guidance on how to retire older tools, districts are layering new curriculum-aligned assessments on top of the interim assessments they already have.

What we hoped would accelerate instructional improvement has instead left many districts with more tests, more data to reconcile and more confusion about what to trust.

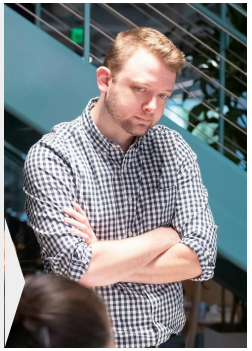
Note: Profile is a composite based on our stakeholder interviews and district research.

Learnings from District Leaders

District leaders continue to balance the need for meaningful, actionable data with the risk of overtesting students

"In the most simplistic way, I think our intent was to ensure we had a way to measure what we were asking our educators to spend time doing. And I think what feels hard is… we definitely don't want to overtest kids."

"We want to make sure that we're not overburdening schools and teachers and students with additional assessments that aren't already captured in the results of our state-required assessments."



"So that's maybe my one lone concern and something that…we need to be very particular about as far as which kids do we really want to retest and which ones are better if we just leave well enough alone and not overtest them, because if we overtest them…I think test fatigue does become an issue."

"I think we're always going to get challenges because testing is never a fun topic for anyone to talk about. They've had bad personal experiences with testing. So the challenge we always get is we're overtesting kids."

As they navigate this balance, districts must also respond to state policies that strongly shape local assessment decisions

"We want to make sure that we're not overburdening schools and teachers and students with additional assessments that aren't already captured in the results of our state-required assessments. And so that's kind of how we filled in the gaps with our benchmark assessment to make sure that once a quarter we have a measure of students how they're doing not only with the standards in the classroom, but in preparation for our state assessment."

"With the advent of this new [state] assessment where we're doing this three times a year, we've scaled back a bit on i-Ready. So we still utilize i-Ready as an instructional tool, but we've dropped the third diagnostic because we feel that there's a concern that we're overtesting students. As the state has made their testing program more robust, we've scaled back on assessments that we've administered."

Alongside state policy pressures, leaders must balance political, public and superintendent expectations tied to assessment performance

"We are trying to manage the fact that we are a large urban school district, and state assessments are a big thing both politically and in terms of district perception and family appetite. And we have a superintendent who has goals interconnected to state assessment…I was ready to kill the district interims. I was like, just throw 'em away. But [my colleague] pushed me on it, it'd be irresponsible for the superintendent to not have insight into how kids were progressing before the state assessment data came back.

"We have shifted assessments more frequently than I would want us to have done. And whenever you shift an assessment at the state level, it's almost like, I don't want to say three wasted years, but it's almost like a three-year process where they have to level set, determine if the assessment really works, see what real norms look like, see what you actually anticipate…And so we've actually invested in MAP because we haven't had that stable assessment. And I think if the state assessments were serving that function, we probably wouldn't need that as much."

In response to these pressures, districts prioritize assessments that are tightly aligned to learning standards

"We want to make sure that the rigor of the assessment is there, that it's coherent with the learning standards. We want to make sure depending on the assessment, it's as comprehensive as possible and gives the right aperture to make the interpretations."

"And so often the question is around crosswalk to our standards, which most vendors have now, but that's important that we can tie it to our specific standards so that teachers aren't doing a lot of extra work to understand how this connects to what they're required to be doing in the classroom."

Beyond alignment, district leaders seek assessments that provide timely, actionable data to inform instruction

"We try to find assessments…that are aligned to our curriculum [and also] assessments that help us best understand not only how students are doing on our specific standards but what next steps could be that instructional piece, how much information and how friendly is the information it gives us around the instructional next steps after the assessment."

"The number one thing is alignment. So all of our assessments are aligned and they provide us the opportunity to measure our impact on students. So if it's an assessment that we're administering, it should be an assessment that tells us whether our instruction has been impactful and if it hasn't, it will inform us as to what we need to do next."

To make assessment data more useful, districts invest in analysis and tools that help educators easily track and interpret student progress

"For most of our assessments, the key is that they are all aligned to our curriculum and they're used to inform and to provide next steps as to what we should do as far as instruction…We run prediction models from those interim assessments, which are typically administered during the middle of the year, and they're based on regression analysis that we do on previous assessments."

"One of the [report] in particular that comes to mind is called standard mastery. So it tracks at the student level the mastery of a standard from all the assessments where that standard has been tested. I could be a fourth grader and by middle of the year, I have seen that standard three, four times. Different places. I've seen it on a curriculum based assessment, I've seen it on an exit ticket, I've seen it on a benchmark. And these reports allow us to put them side by side cumulatively. How did the kid progress on that standard?"

Alongside the technical support, leaders emphasize building school-based staff capacity to interpret and lead meaningful data conversations

"We provide extensive training on data-driven instruction. We have almost perfected our PLCs so that we have protocols on how to process the data… and reteach."

“Our school test coordinator and our data coaches are our primary audience…we focus on more collaborative data conversations and how they are getting more people to feel more confident with the data.”

Districts also use assessment data to guide how resources are allocated to support schools and students

"The data that we have, particularly those that are based on [state] accountability measures, is used for lots of purposes. It's used for staffing purposes. Schools who are most fragile get first priority when it comes to staffing. They also get priority with staffers that are trying to recruit and find teachers. We also use that data to inform budget practices. Our most fragile schools tend to get more support."

"This year, we're focused on our multilingual learner populations and our students with disabilities, really teasing out the data from the different assessments on how those populations are performing and what supports need to be in place to help them to excel specifically. So if that's additional resources, if that's additional staffing, if there's additional training that's needed by our teachers to make sure that they understand how to meet their needs, and that students, again, are showing growth regardless of their designation."

Ultimately, district leaders aim to build a cohesive, instruction-focused assessment system that balances state demands, local needs and testing load

District leaders have a strong desire to streamline assessments into a cohesive system, but they often face challenging decision points as they navigate tensions among local needs and state policies while trying to balance overtesting.

District leaders emphasize that one of the primary purposes of assessment should be to support teaching and learning, prioritizing tools and practices that provide timely, actionable insights to guide instruction and improve student outcomes.

District leaders are committed to ensuring that assessment results are used meaningfully and are investing in professional development and tools to help school leaders and educators to effectively interpret and act on assessment data.

I'm Shante, the Director of Assessment in a large urban district. Over the years, different central office teams have layered on assessments to meet their own priorities—our MTSS office added a universal screener, the curriculum team adopted a diagnostic to guide instruction and intervention, and the office of accountability pushed for an interim to track progress toward state goals. Each decision made sense on its own, but they weren't coordinated. We've tried more than once to streamline, but every change sets off a chain reaction—procurement cycles are long, our student data systems take years to reconfigure and teachers need time and training to shift practice. It's often easier to just keep what we have, even when we know it's duplicative.

State requirements add another layer: we can't risk losing the data the state expects, so we end up adding new assessments on top rather than replacing old ones. Now our students are taking multiple overlapping tests that weren't designed to work together, and none of the systems speak the same data language.

Teachers are left trying to reconcile conflicting reports from all of them just to plan instruction, and the sheer volume of data is overwhelming. What was meant to support better decision-making has turned into noise—and it's hard to see a path out of it.

Note: Profile is a composite based on our stakeholder interviews and district research.

Meet

Shante

Director of assessment in a large urban district working to make sense of all their critical data

Implications and Recommendations

The assessment ecosystem is weighed down by clutter, mistrust and misuse

Pain Points			
Letting go is harder than adding on. Districts and states are far more comfortable layering new assessments on than they are retiring old tools. The result is a landscape that feels cluttered and often incoherent.	**Trust is in short supply.** There's a trust deficit between policymakers, states, districts and vendors. Each group questions whether the system is serving their needs, making it harder to build consensus around new solutions.	**Innovation is both promising and risky.** AI, adaptive testing and through-year models all hold enormous potential, but without guardrails, they risk amplifying misuse and complicating the landscape.	**Assessment purposes and uses are misaligned.** Tests are frequently stretched beyond their intended purpose leading to confusion, misuse and a lack of confidence in results.
Solutions			
Streamline and implement replacement strategies that reduce burder and inequities.	**Create intentional structures for shared decisionmaking** and cross-stakeholder collaboration to rebuild confidence and trust.	**Redirect investments toward assessments with clear instructional value** and adopt thoughtful guardrails for emerging tools.	**Require clearer evidence from vendors** and **guidance from states** to ensure assessments are used as intended.

To shape more coherent assessment systems, we must navigate the different realities that shape decisions for states, districts, developers and funders

States must balance federal accountability mandates with political pressure to improve and reduce testing. Limited budgets, resistance from educators and families and the push for equity without overburdening schools all influence what states adopt and require.

States

Districts are caught between ensuring equity and performance while reducing test fatigue and increasing instructional value. Procurement cycles, data system limitations and the need to train staff on new tools often slow down or complicate innovation.

Districts

Developers must reconcile profitability, scalability and usability. They face the challenge of building assessments that are rigorous yet flexible enough for classroom use, with data that serves multiple users (teachers, parents, students) and adapts to varied district needs.

Developers

Funders strive to drive innovation and close equity gaps, but face tensions between scaling quickly and supporting high-impact, resource-intensive solutions. Grant timelines and limited infrastructure in high-need areas often constrain sustainable success.

Funders

Recommendations for
State Leaders

Advance purpose-driven, instructionally aligned assessment systems that reflect state policy priorities. Establish a clear vision tying state assessments to instructional priorities.

Revisit state policies and mandates (e.g. screening) to ensure they are necessary, non-duplicative and aligned to current instructional and accountability priorities.

Conduct regular audits and reviews of the state assessment landscape to identify overlapping or redundant tests, especially where state mandates may have prompted layering.

Issue clear guidance on the roles and appropriate uses of assessments, and offer examples that show how state, district and curriculum-embedded assessments can work together.

Raise the Bar for Assessment Quality and Instructional Utility. Use procurement processes to require test vendors to provide independent evidence of reliability, validity and fairness for each intended use of an assessment. Require vendors to make such evidence publicly available and subject it to independent expert review before purchase or renewal decisions.

Establish cross-departmental teams across academic and assessment leaders to reduce fragmentation and clarify roles for adoption and oversight.

Incent coherence and innovation through funding, procurement flexibility and pilots to encourage districts and vendors to adopt coherent, instructionally useful systems.

Recommendations for
District Leaders

Invest in educator capacity so educators can interpret and use assessment data without becoming "data wizards." Leverage AI to integrate data sources and support evidence-based decisionmaking.

Audit and regularly inventory all tests administered, eliminate redundant or low-value ones and map remaining assessments to instructional priorities.

Create cross-departmental teams to clarify roles and responsibilities, coordinate adoption decisions and reduce siloes.

Strengthen coherence with state policy by aligning district assessments to state guidance and streamline sequencing of state, district and curriculum-embedded assessments.

Integrate assessment and instructional strategies to ensure local assessment choices strengthen, rather than compete with, high-quality curriculum and instruction. Districts should prioritize assessments that mirror the rigor, content, and pacing of HQIM to avoid fragmented signals, redundant testing, and instructional detours.

Monitor testing loads by school type and student group to ensure that assessments don't disproportionately burden students in higher-poverty schools or English learners.

Engage stakeholders including teachers, families and students in evaluating the purpose and value of assessments to build trust and reduce fatigue.

Recommendations for
Product Developers

Be transparent about purpose and limits. Clearly state the intended uses and boundaries of your products—and discourage districts and states from stretching assessments beyond their design (e.g., screeners as daily lesson guides, interims as teacher evaluation tools).

Design for interoperability, not lock-in. Build products that integrate seamlessly with district data systems so leaders can track student progress across schools, grades and programs without piling on new tests.

Prove instructional value. Demonstrate—through evidence and transparency—how your assessments inform instruction in real classrooms. Provide teachers with timely, actionable information that directly supports grade-level, standards-aligned instruction, rather than overwhelming them with data they cannot use.

Replace, don't add. Partner with districts and states to design implementation plans that show explicitly how your product reduces, not increases, the total testing burden.

Co-create with practitioners. Work alongside teachers, principals and system leaders to ensure products meet real instructional needs, not just market demand.

Recommendations for
Advocates and Policymakers

Champion instructional coherence.
Push for state policies that align assessments with HQIM and grade-level standards, ensuring tests reinforce—not fragment—teaching and learning.

Raise the bar for quality assurance.
Call for independent evaluations of evidence to support the technical quality of commercial tests and transparency from the vendor community to publish this information.

Promote transparency and clarity.
Advocate for states to publish clear assessment calendars, cost-benefit analyses and guidance on the intended uses of both state and local assessments.

Hold vendors accountable.
Insist that states require developers to demonstrate alignment to instructional priorities, interoperability and actionable information for teachers before approving products for adoption.

Influence funding priorities.
Push philanthropy and state policymakers to tie funding and grantmaking to assessments that deliver instructional value and streamline, rather than layer, the system.

Build public will.
Translate technical policy debates into accessible messages for parents, educators and communities, building pressure on legislators and state boards to act.

Recommendations to the
Gates Foundation

Review and align your portfolio to by mapping current investments across policy, curriculum, data and assessment teams. Surface where grantees' work may be duplicative, disconnected or even competing (e.g., HQIM adoption vs. standalone assessments).

Require product developer grantees to demonstrate clear alignment to instructional priorities. Tie grantmaking to evidence that new assessments support HQIM and local/state curriculum goals.

Prioritize grants for products that provide instructionally relevant, timely and actionable information to teachers—not just measures growth, compliance or predictive analytics.

Support tools for coherence. Invest in guidance protocols, cost–benefit analyses and exemplar assessment calendars that districts can use to sequence and select assessments strategically.

Enable local "assessment audits." Provide resources for districts to map their current assessment ecosystem—including costs, time burden and redundancy—and identify which assessments can be retired.

Incent coherence across grantees by building expectations into grant agreements for organizations to coordinate efforts and demonstrate how their work connects to, rather than fragments, district and state systems.

We are still learning and more research is needed to guide the work ahead

Our study surfaces important insights, but much remains to be explored to anchor the work ahead for leaders interested in improving academic assessment systems. Areas for future research include:

Conduct a comprehensive accounting of how much money districts and states spend on assessments, including hidden costs such as software licenses, test prep, professional development and data systems.

Explore more deeply the assessments chosen by schools and teachers, and how they interact with district and state requirements—revealing where "the pile" grows.

Conduct further analysis of state mandates (e.g., early literacy policies) and how they shape local assessment decisions and contribute to layering.

Develop evidence on whether and how assessments are supporting HQIM and where they fall short or hinder effective implementation.

Provide more comprehensive data on how often students of color, English learners and students in higher-poverty schools are tested and for what purposes.

Conduct comparative studies of which assessments deliver the most instructional value, and under what conditions.

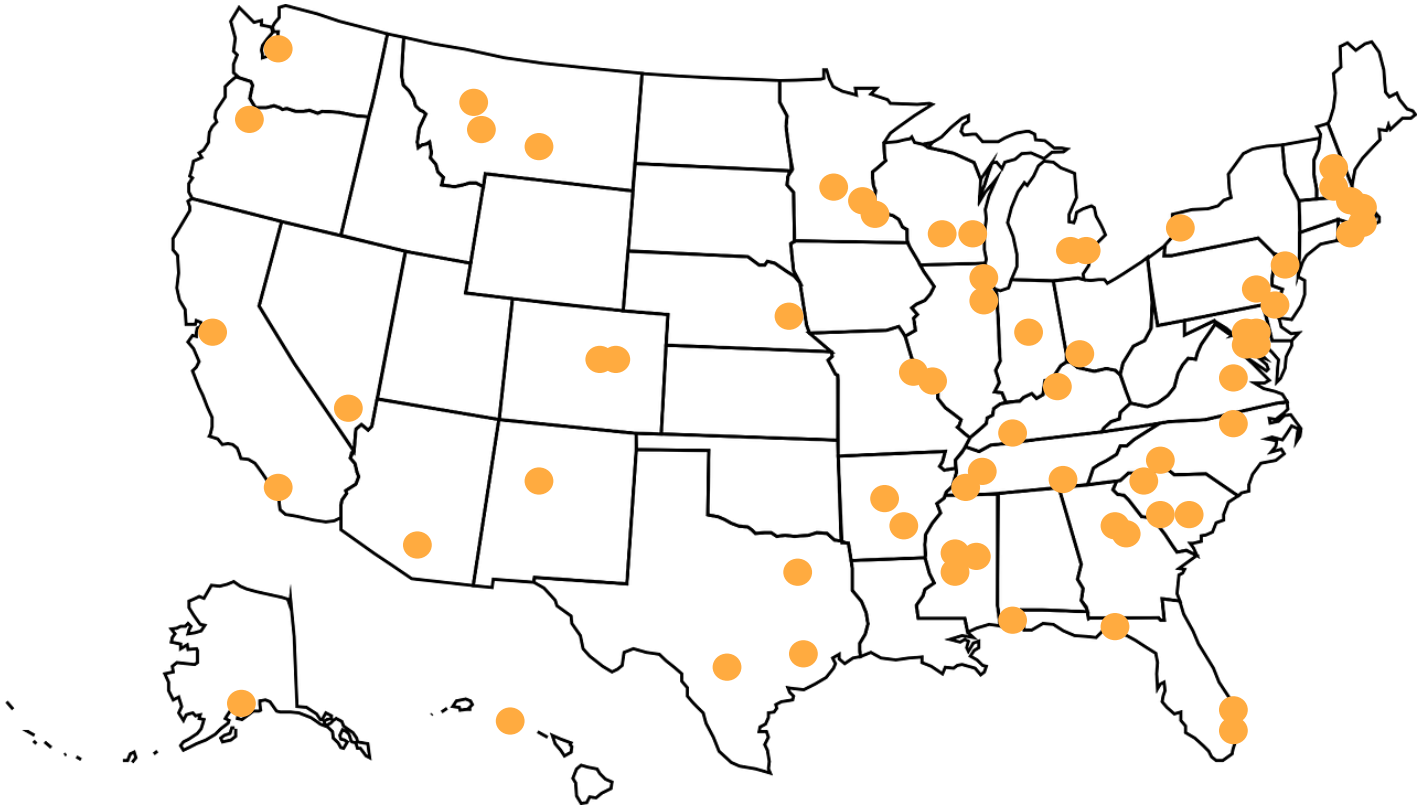
Appendix

Overview of Sampling Method

1. We started with a list of 600+ districts that met the following criteria:
   a. Serve 6+ schools
   b. Enroll 2000+ students
   c. Have greater than 2x % of Latino or Hispanic students compared to 2x national % (58%) OR Greater than 2x % of Black students compared to 2x national % (30%) OR have greater than 2x % of children under 18 in families living in poverty than 2x the national average (32%), according to NCES data for 2022 (poverty, race)

2. We then used the following parameters to ensure a diverse set of our final sample of districts:
   a. Ensure that districts from the same state are not included more than three times, and that we have at least 35 different states represented in our sample
   b. Ensure representation of districts from varied regions (e.g., midwest, northeast, south, west) and locales (e.g., city, suburb, town, rural) as determined by Census and NCES classifications
   c. Ensure a relatively balanced political representation of districts from as determined by the 2024 presidential election results

Research Sample of Districts



Assessment Type Definitions

Screener Assessment is designed to identify students to determine which are at risk or need additional support to meet learning goals. Often brief, standardized and does not provide enough details to guide instruction alone.

Diagnostic Assessment is designed to identify specific knowledge or skill gaps in a targeted content area. Typically help pinpoint specific learning needs (e.g. calculation vs. math).

Formative Assessment is designed to provide real-time feedback to support learning and inform instruction. Typically informal and embedded in daily instruction use.

Interim Assessment is designed to measure progress towards standards or predict performance on summative assessments. Most are curriculum agnostic, standardized and computer based. Some are curriculum-embedded and are aligned to the order in which students encounter content in the curriculum.

Summative Assessment is designed to evaluate cumulative learning at the end of a period of instruction. Often high-stakes and standardized, and may be used to grade or promote students and certify achievement as well as measure growth. In addition, results from these assessments may be used to inform effectiveness of programs or practices and/or inform resource allocation.

Sources: National Center on Educational Outcomes (2025); Center for Assessment (2019); Iowa Reading Research Center (2024); AIR (2025)

Assessment Use Definitions

Student identification or placement	Assessment results are used to determine placement of students in specialized status, coursework or programming, including, but not limited to, EL status, IEP, gifted and exceptional abilities and secondary school entry.
Instructional utility	Assessment results are used to inform district, school or classroom approaches to curriculum and instruction.
Progress monitoring	Assessment results are used to measure students growth or progress toward academic outcomes and standards.
Policy or program planning	Assessment results are used to inform changes to academic programs, evaluate program effectiveness and inform system-wide shifts in professional learning, resources and supports.
Teacher evaluation	Assessments results are used as part of a formal system of evaluating teachers.
Certify learning	Assessment results are used to measure academic proficiency of a student at the end of a learning period (i.e. unit, course, grade).
Accountability	Assessment results are used as a part of a state and/or federal accountability system.

Sources: National Center on Educational Outcomes (2025); Center for Assessment (2019); Iowa Reading Research Center (2024); AIR (2025)

Rethinking the Test Pile:
A National Study of K–8 Academic Assessments

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, Kelly Jasiura, Yeonjae Lee and Jennifer Vranek led this research and authored the report. Austin Wechter designed the report.

We are grateful to the many individuals who provided guidance, time and shared their knowledge with us to inform this report through interviews, participation in sensemaking sessions, critical friend reviews and 1:1 discussions.

Thank you to Education First staff Sheila Briggs, Ila Deshmukh Towery, Aneesha Badrinarayan and Tara Czupryk for their support, collaboration and contributions that made this report possible. In addition, we express our sincere gratitude to Chris Domaleski, Kristen Huff, Michelle Odemwingie and Carter Romansky for sharing valuable critical friend feedback.

Suggested Citation (APA): Education First. (2026). Rethinking the test pile: a national study of K–8 academic assessments.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.
- D'Brot, J. (2019, June 11). *Improving Interim Assessments.* Center for Assessment. https://www.nciea.org/blog/stuck-in-the-middle-with-interim-assessments/
- *Interim, Formative, and Diagnostic Assessments.* (n.d.). NCEO. https://nceo.info/Assessments/formative
- National Center for Education Statistics. (2024). *Characteristics of Children's Families. Condition of Education.* U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cce/family-characteristics
- National Center for Education Statistics. (2024). *Racial/Ethnic Enrollment in Public Schools. Condition of Education.* U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cge/racial-ethnic-enrollment
- *Reading Assessments and Their Purposes.* (2024). Iowa Reading Research Center - the University of Iowa. https://irrc.education.uiowa.edu/blog/2024/09/reading-assessments-and-their-purposes
- *Tools Chart Overview*. (n.d.). National Center on Intensive Intervention. https://intensiveintervention.org/tools-charts/overview
- Topol, B., Olson, J., Roeber, E., & Hennon, P. (2013). *Getting to higher-quality assessments: Evaluating costs, benefits, and investment strategies.* Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education. https://edpolicy.stanford.edu/sites/default/files/publications/getting-higher-quality-assessments-evaluating-costs-benefits-and-investment-strategies.pdf