EDUCATIONAL ACCOUNTABILITY REMADE

Authentic Assessment and Accountability as an Engine of Educator Renewal

center for innovation in education
leading with learning

FairTest
The National Center for Fair & Open Testing

AUTHENTIC ASSESSMENT AND ACCOUNTABILITY AS AN ENGINE OF EDUCATOR RENEWAL

THE EAR TENTPOLE PAPER SERIES

01 From Compulsion to Mutual Assent: Pitching a New Tent for Educational Accountability

02 Elevating Community Trust As the First Principle of Accountability

03 Proving Education Inadequacy Without the Test Score Trap: How Inputs and Meaningful Outcomes Make the Case

04 Comparability Without the Harms of Standardization: How Construct Comparability Promotes Authentic and Rigorous Evidence of Learning

05 **Authentic Assessment and Accountability as an Engine of Educator Renewal**

INTRODUCTION

For longer than standardized testing has dominated American education, pioneering states, districts, networks and schools have built assessment systems that include more authentic demonstrations of student learning. Using performance-based (or authentic) assessments, these systems ask students to demonstrate a broad range of knowledge, skills, and competencies, foster student agency, and can reflect the diversity of individual, community, and cultural priorities. When well-designed with community input, performance-based (PA) systems support a full range of students, including those who are preparing to attend higher education and those interested in work; and students in rural communities, those living in poverty, those who are English language learners, and those with IEPs. And when embedded in an aligned policy landscape, performance-based assessment systems build and sustain professional cultures of dignity and collaboration among educators while offering transparency to communities and families.

Yet while they are superior to standardized testing in many regards, performance-based systems haven't become widely embedded in state assessment and accountability models due, in part, to one recurring concern: *workforce readiness*. Indeed, we do not claim that authentic assessment is easy, automatic or free. Designing rich tasks, closely scoring student work, and calibrating with colleagues takes real time and real resources. In an era marked by high turnover and staffing shortages, sprawling demands on educators, and widespread educator fatigue, education leaders may question whether their already-stretched workforce has what is needed to

successfully implement a performance-based assessment system for local use, let alone for state accountability.[1]

Despite the demands, we think the educator workforce already has what it takes–if given the right support from policymakers. Admittedly, in a high-stakes, "results-now" policy environment, the investment required to build capacity for PA systems carries both financial and political considerations. Our argument is that policymakers must weigh these considerations against the risk of *not* shifting to accountability systems based on more authentic assessments of student learning. When viewed this way, it becomes clear that the status quo is draining educator capacity more than building it, all while sending public dollars to outside testing vendors instead of using those funds to increase educators' knowledge and skills where it matters the most. If policymakers attempt to layer authentic assessments on top of an unchanged system of standardized assessments, they will indeed find a shortage of time, educator capacity, and financial resources. Instead, **we urge policymakers to consider their investments in assessment, accountability, and educator capacity as parts of the same strategy; and to align those investments with what teaching and learning should be**.

Lastly, we suggest that the question, "*Can teachers do it?*" is premised on a flawed assumption that systems must wait for full capacity before transitioning. A better question is, "*How can systems build educator capacity* through *authentic assessment?*"[2] Thus, we position capacity as a developmental outcome generated through the process of creating a more authentic system of teaching, learning, assessment, and accountability – one that is more responsive to students and families, and that engenders greater community trust. In doing so, we show how such a system resolves many of the issues that plague student alienation as well as the educator workforce under the current assessment and accountability model.

1. Throughout this paper, the writers employ the terms "educator" and "teacher" to refer to education system employees who help students achieve academically and learn the skills they need to succeed in schools, community, and life. The term "educator" is broad and encompasses classroom teachers as well as Education Support Professionals and Specialized Instructional Support Personnel (such as occupational therapists, library media specialists, and others providing services for students) who collaborate with others to support individual students' learning needs. The term "teacher" is used to refer to an individual whose primary focus is to provide direct instruction to students in a particular grade level and/or academic content area(s). Our decision to use the term "educator" or "teacher" in various sections of the paper closely connects to the use of terms in resources cited and whether or not we felt a broader or more narrow term was appropriate.

2. World Economic Forum (2025). There is another key question beyond the scope of this paper that naturally arises from an examination of educator capacity and preparedness to implement high-quality authentic assessment systems, "*How much longer can we afford not to make this change*?" The World Economic Forum's 2025 Future of Jobs Survey asked employers from a diverse array of large multinational and national companies to identify core skills for the workforce.7 of the top 10 most in demand core skills were related to self-efficacy (resilience, flexibility, motivation, self awareness, curiosity) and working with others (leadership, social awareness, empathy, active listening), while reading, writing, and mathematics were identified as core skills by only 21% of employers. EAR is by no means positing that reading, writing, and mathematics skills are not critical objectives for organized education systems. We would point out, however, that regardless of students' pathways after graduation (college, career, technical education, apprenticeship, military, etc.), the most in-demand skills that students will need in order to access and succeed in the labor market are–at present–not measured or cultivated by systems that rely on standardized tests and that performance assessments are far more capable of ascertaining the full breadth and depth of students' knowledge, skills, and dispositions.

HOW THE CURRENT SYSTEM OF STANDARDIZED TESTING HARMS THE EDUCATOR WORKFORCE

The introduction of standardized test-based accountability was motivated by a civil rights commitment: to ensure that all children would be held to the same standards, and to provide policy makers, communities and tax payers a transparent way to assess how well their schools were delivering equitable education.[3] However, thirty years on, we see that high stakes testing and accountability has resulted in numerous negative consequences. Among them: alienation for students, disparate impact on restricted curriculum for students of color and students in poverty, and measurable harms to the educator workforce.

Currently, teacher turnover rates average 15% a year, at a cost of $25,000 per educator lost (Tan et al., 2024, 2026). The most severe impacts are among teachers of color, teachers in high poverty districts, and those in urban schools. By all accounts, structural conditions create high teacher turnover, resulting in an estimated 411,549 teaching positions either vacant or filled with underqualified educators– that's 1 in 8 (Learning Policy Institute, 2025).

While one could blame many factors for these trends, several researchers have linked them to one clear cause: the current paradigm of state standardized testing and accountability. Ryan and colleagues (2017) found that among 1,866 teachers across three states, test-based state accountability significantly predicted higher stress, which fueled burnout and ultimately predicted teachers' intent to leave. Using data from the National Center for Education Statistics' Schools and Staffing Surveys and Teacher Follow-up Surveys, Wronowski and Urick (2019) found that while teacher anxiety was widespread, worry and stress only predicted moving schools or leaving the profession when teachers self-cited accountability policies as a factor in their employment decisions.

Teacher turnover is also tied to the consequences of test-based accountability publicized in school ratings. Studying an abrupt change to Florida's test-based school-grading formula, Feng, Figlio, and Sass (2018) found causal evidence that teachers were substantially more likely to leave schools that were suddenly downgraded on test-based state metrics. Departures were concentrated among the highest-rated teachers – based, ironically, on student test scores – representing a departure of the very teachers whom students in struggling, high-poverty schools can least afford to lose. Educators have indicated that the motivation to leave schools was not primarily based on running away from poor school ratings or low test-scores, but a response to the narrowly prescriptive, top-down improvement interventions implemented in res-

3. Skinner (2026). The Elementary and Secondary Education Act of 1965 was amended by the No Child Left Behind Act of 2001 and implemented the requirement that states and public school systems must use statewide assessment results as part of educational accountability systems that identify academically underperforming schools as a condition of receiving the largest source of federal grant funds for education.

ponse; these interventions sharply erode educator autonomy, constrict the curriculum, deprofessionalize educators, undermine teachers' creativity and divert resources from the deeply human work of teaching and learning to focus on raising test scores (Kraft & Lyon, 2024, National Education Association, 2015)[4].

Even among teachers who decide to stay, standardized testing and accountability negatively impact teaching and learning, lowering teacher satisfaction and morale. In study after study involving thousands of public school teachers, teachers report the negative effects of high-pressure standardized tests, reporting that they:

- narrow the curriculum to focus primarily on tested subjects at the expense of other subjects and cross-cutting skills,
- siphon important instructional time to test preparation and administration,
- cause severe test anxiety in students,
- are not valid measures of student learning,
- do not accurately reflect good or bad teaching, and
- lower staff morale.[5]

Critically, despite the original civil rights aims of the current assessment and accountability system, its negative impacts on the educator workforce have disproportionately affected teachers serving historically disadvantaged students. Under surveillance and pressure to "improve" on test scores, those schools have become pressure cookers for test-prep, often sacrificing art, music, recess and community service, to focus on high stakes testing. Those educators endure and embody the pressure. The National Education Association (NEA) surveyed thousands of educators and found that concern about the current assessment system was highest among those who work with multilingual learners, educators in high-poverty schools, and educators serving higher percentages of students of color (NEA, 2022). School poverty level and the racial make-up of the student body also influence how educators feel about assessment. Educators teaching in schools with higher proportions of students who qualified for Free and Reduced Meals (FARM) or with higher proportions of students of color were more likely than their counterparts to want to get rid of standardized assessments altogether. These practitioners know the harms of high stakes testing, and as you will see below, many are ready to engage with new, more authentic and performance based systems of assessments.

4. National Education Association, 2015. The National Education Association's (NEA) Accountability Task Force reported that the consequences of test-based accountability and "relentless testing, labeling, and punishing" failed students due to "narrowing the curriculum to meet narrow standards of success."
5. NEA, 2022, Gunn et all, 2016; Rebora, 2012; Davis, 2011; Gordon and Rice 2004; Taylor et al, 2003.

This problem is not new; it has been with us for decades. Teachers' low opinions of standardized testing haven't changed much in the last two decades. Early on, when the modern era of standardized testing was just taking hold, Jones and Egley (2004) surveyed elementary school teachers in Florida and found that 80% felt that "high stakes testing was taking schools in the wrong direction." That sentiment persists: in 2022, the NEA surveyed thousands of parents and teachers, finding that 67% of parents and 85% of teachers identify standardized testing as a "serious issue" in public schools.[6]

THE WORKFORCE BENEFITS OF BUILDING CAPACITY FOR AUTHENTIC ASSESSMENT

Rectifying the numerous harms of the current test-based assessment and accountability system begins with two fundamental shifts:

1. Relying on more authentic evidence of learning, such as performance-based assessments, and
2. Re-investing in educators as capable co-designers of the system.



These shifts, in turn, catalyze a virtuous cycle in which educator capacity increases hand-in-hand with systemic capacity, intentionally cultivated through a multi-pronged approach consisting of strong instructional leadership, high-functioning professional learning communities, deep educator collaboration, sustained access to time and resources, and system calibration (Ronfeldt, et al, 2015; Villavicencio, et al, 2025).

6. In 2022, NEA conducted internal research surveying over 2,000 parents and 1,900 active NEA members about issues connected to assessment and accountability systems. The term "standardized testing" was defined throughout the internal research as statewide summative assessment (often conducted at the end of the year or over the course of multiple interims that roll up into an end of year score) in English Language Arts and Mathematics administered to all students in grades 3-8 and at least once in high school for the purposes of measuring student, school, and district performance. The survey was conducted as part of the NEA's Task Force on the Future of Assessment work, which resulted in the Principles for the Future of Assessment (NEA, 2022). The Principles were created with input from thousands of NEA members and outline NEA's priorities for transforming assessment systems to be more equitable, just, and student-centered.

It is already well-documented how investments in educator agency and collaboration increase student achievement and teacher retention, including and especially in schools serving historically underserved populations such as English language learners (Villavicencio et al, 2025).[7] Coupled with systemic shifts to more authentic assessments including project-based learning and performance-based assessments, these investments have been shown to simultaneously improve educator capacity, promote effective implementation, and increase student learning. For example, a federally funded review of seven states' performance-assessment implementation independently identified investment in the capacity of teachers and leaders, together with teacher collaboration and involvement, as decisive conditions for effective implementation (Howard Terrell & Barzee, 2023). Villavicencio, Jaffe-Walter and Cherng (2025) further show that the combination of high autonomy, interdisciplinary educator collaboration, performance assessments, and organizational focus on students' needs culminated in a culture of what these researchers consider "collective responsibility," resulting in higher levels of student achievement and greater educator retention among New York City schools.

Virginia offers a state policy illustration. In 2014, the Virginia General Assembly eliminated five statewide standardized tests and required districts to develop local alternative assessments, including performance-based assessments, in their place (Howard Terrell & Barzee, 2023). The state supported its mandate with statewide scoring tools and annual technical-assistance reviews. State officials report that the transition ignited educator capacity: collaborative scoring sessions deepened teachers' command of the rubrics and built support for the work, and teachers who engaged in assessment design became "ambassadors," leading colleagues within and beyond their districts.

STRATEGIC INVESTMENTS TO BUILD EDUCATOR CAPACITY FOR AUTHENTIC ASSESSMENT

How might a state or a district initiate a move toward more performance based assessment systems? We offer below a few routes.

States and districts can begin to build educator capacity for authentic performance-based assessments through several strategic investments and systemic levers, including:

7. The hallmark of performance assessment is application of students' "knowledge, skills, and abilities to authentic problems." (EducationWeek, 2019). With performance based assessment, a student might design and carry out an original science experiment and defend the findings to a panel; write an analytic essay on a work of literature and discuss it with outside evaluators; research a social studies question and present the paper; or apply higher-level mathematics to a real-world problem. Others might build and test a robot and explain their design choices and what failed along the way, assemble a portfolio of work in a career and technical field, or complete a capstone that addresses a community need and present it to educators, families, and local partners. In each case, students produce real work, apply what they know, and explain and defend their thinking, and teachers score the work against shared rubrics.

1. **Regulatory Waiver Pathways and Statutory Flexibilities:** States can grant statutory waivers or innovation zone status to districts or networks of schools allowing them to use authentic student work and capstone portfolios for graduation and accountability purposes. This provides essential regulatory cover, establishing a clear policy bridge toward local assessment design.
2. **Shared Resources: Construct Definitions, Performance Tasks, and Scoring Rubrics.** Rather than leaving capacity-building to individual sites, states and regional networks can build and share resources such as definitions of the standards or competencies to be measured through performance tasks, curating a bank of sample performance tasks, and creating rubrics describing student performance at various levels of sophistication. Involving educators in creating these resources and enlisting them to train others will further build workforce capacity.
3. **Regional Moderation Networks and Inter-School Calibration:** To build educator capacity for authentic assessments while also ensuring a level of reliability and comparability needed for accountability purposes, states or state intermediary organizations can routinely gather educators within or across sites to read, evaluate, and score student performance tasks against shared quality benchmarks (Leather et al, 2026).
4. **Micro-Credentialing Pathways:** States and districts can establish competency-based pathways where teachers earn advanced credentials by conducting classroom action research and defending portfolios of practice. Salary bumps offset program fees for participating educators, generating a financially self-sustaining mechanism for scaling assessment literacy across entire regions.
5. **Systemic Assessment Audits**: To create "space on the plate" for developing more authentic assessments, district leaders can review and streamline their local assessment calendars to clear away redundant interim tests and other materials that are not serving deeper learning goals. Reclaiming these instructional hours gives educators the dedicated time needed to design meaningful tasks, review student work, and collaborate without adding to workload strain.
6. **Practitioner-Led Peer Review and Advisory Infrastructure:** States and intermediary organizations can establish peer review boards of veteran educators who conduct school visits, observe classroom defenses, and offer constructive feedback to refine local practice. Tapping practicing educators as technical leaders builds institutional trust and accelerates peer-to-peer capacity growth far more effectively than top-down compliance monitoring.
7. **Public Exhibitions and Community Defenses of Learning:** One of the surest ways to deepen learning and to build community trust in public schooling is to make the work - of students and educators - public. Schools can organize capstone presentations where students

al civic or industry partners. This allows educators to experience quality learning, deepening their own understanding of what is possible and facilitating broader community buy-in.

Shared rubrics and cross-school moderation studies – formalized structures through which educators collaboratively evaluate student work against common benchmarks and calibrate their professional judgment – are especially critical for ensuring high student expectations throughout the PA system.Peer collaboration may appear costly and time-consuming, but it achieves more than just improving inter-rater reliability: it increases teachers' knowledge of standards and their competence and confidence to assess accurately, to provide feedback, and to encourage revision, all of which are essential to high-quality teaching and learning.



Importantly, moderation studies should be embedded in an accountability culture that prioritizes peer collaboration, trust, and mutual support. When this occurs, as the following case studies will show, educators are willing to engage in rich discussion and even disagreements that ultimately strengthen their knowledge and skills. Not only that, when openness and honest discourse are encouraged, the assessment system itself learns and improves: rubrics are strengthened, new tasks may be introduced, and overall system transparency increases.

CASE ONE: BUILDING CAPACITY WITH EDUCATORS AT THE NEW YORK PERFORMANCE STANDARDS CONSORTIUM

The New York Performance Standards Consortium (Consortium), a network of 38 public high schools (36 within New York City), has operated for more than twenty-five years under a State waiver permitting students to demonstrate readiness and graduate through performance-based assessment tasks (PBATs) rather than most Regents examinations. Critically for our argument, Consortium schools do not serve an advantaged population. Rather, compared to New York City High schools, Consortium schools enroll higher proportions of English Language Learners (ELLs), students in temporary housing, and students with disabilities; register a higher Economic Need Index; and admit students with lower incoming eighth-grade proficiency in both ELA and math (Table 1).

Table 1. Consortium Student Demographics (2024–25)[8]

Metric	Consortium	City	Difference
Economic Need Index	80.3%	77.9%	+2.4pp
English Language Learners	20.9%	14.3%	+6.7pp
Students with IEPs	23.0%	18.6%	+4.4pp
Students in temporary housing	17.8%	13.8%	+4.1pp
Avg. incoming grade-8 ELA (1–4 scale)	2.83	3.06	−0.23
Avg. incoming grade-8 Math (1–4 scale)	2.61	2.86	−0.25

As noted earlier, performance-based assessment asks more (or asks different) of teachers, so the question is whether that added labor depletes them or sustains them. If authentic assessment were simply "more work" that overburdened teachers, we would expect its most concentrated implementation -- in schools serving precisely the higher-need students the testing regime claims to protect -- to produce educator strain, disengagement, and flight. The evidence, however, points in the opposite direction. On the School Survey[9] domain of *Teacher Influence* -- the measure most directly capturing educators' sense of autonomy over curriculum and instruction -- Consortium teachers respond positively at 83.3%, compared to 67.5% of high school teachers citywide, a gap of nearly sixteen percentage points. This gap sits at the heart of what distinguishes performance-assessment schools: teachers are trusted to design curriculum and assessment in a community of educators.



8. NYC DOE 2024–25 School Quality Reports; enrollment-weighted across the 36 NYC Consortium schools vs. all city high schools. 36 of the 38 schools located in New York City are used for comparison purposes.
9. NYC School Survey, 2024–25 administration (teacher file). The NYC School Survey is conducted by the NYC Department of Education, with no statewide equivalent. Therefore, only 36 Consortium schools within NYC are included in the analysis.

This agency is embedded in a broader professional culture. Compared to NYC teachers in general, Consortium teachers report higher levels of *Innovation and Collective Responsibility* (83.3% vs. 75.9%, +7.4pp), *School Commitment* (91.1% vs. 84.0%, +7.1pp), Teacher-Principal Trust (90.8% vs. 84.2%, +6.6pp), *Peer Collaboration* (90.2% vs. 85.1%, +5.1pp), and *Professional Development* (86.3% vs. 81.6%, +4.6pp) (Table 2).

Table 2. Educator Experience — NYC School Survey 2024–25 (% positive responses)[10]

Domain	Consortium *(n=923)*	City *(n=18,799)*	Gap
Teacher Influence	83.3%	67.5%	+15.8pp
Innovation & Collective Responsibility	83.3%	75.9%	+7.4pp
School Commitment	91.1%	84.0%	+7.1pp
Teacher–Principal Trust	90.8%	84.2%	+6.6pp
Peer Collaboration	90.2%	85.1%	+5.1pp
Professional Development	86.3%	81.6%	+4.6pp
Teacher–Teacher Trust	90.1%	86.2%	+3.9pp

Notably, the Consortium's strongest advantages fall on conditions the literature identifies as protective – teacher autonomy, collective responsibility, collegial trust, and collaboration. Consortium teachers take on the added labor of authentic assessment, and in return report more influence over their work, deeper trust and collaboration with colleagues, and longer tenure in their schools.

10. *Note*. From New York City Public Schools Info Hub (2026). NYC School Survey Archives. https://infohub.nyced.org/reports/students-and-schools/school-quality/nyc-school-survey/survey-archives 2. This graphic reproduces data from the 2024–25 NYC School Survey teacher file (respondent-weighted percent-positive; Consortium N = 923, city N = 18,799).

This professional culture is not left to chance; the Consortium has built network-wide infrastructure that develops and sustains educator and school leader capacity and transparency. During twice-annual moderation studies, educators from all Consortium schools convene to review selected schools' PBAT tasks against shared rubrics and Webb's Depth of Knowledge framework, reading and scoring student work together (Consortium, 2025).[11] These conversations strengthen instructional and assessment practices and create openings to discuss more challenging examples of student work, or student history. Schools whose work was discussed receive a summary of the task feedback and the moderated rating. Subject-area teacher exchanges - in mathematics, science, literature and social studies - bring educators across schools together regularly to share pedagogy and examine student work, and an annual conference features more than thirty practitioner-designed and facilitated workshops. These network structures, along with internal school structures, help newer teachers grow into this work, by connecting them to a community of practice larger than any single school. A Performance Assessment Review Board of experienced educators from these schools and from local universities then visits schools, observes classes, and provides feedback that informs school-wide planning. The same structures that hold schools accountable to shared standards are the structures through which teachers build their capacity to do the work.

These conditions are reflected in the stability of the teaching-force in Consortium schools.[12] Eighty-six percent of Consortium teachers in non-transfer schools (4-year high schools) have remained at their school for three or more years, compared to 78.2% citywide,[13] and Consortium principals at these schools have served a median of eleven years, compared to a median of six years citywide. Consortium educators at both 4-year Consortium high schools and transfer schools have an average of 11.3 years in the profession and 7.2 years at their current school.[14] Set against the turnover costs documented in Section I -- the Learning Policy Institute’s estimate of roughly $25,000 per educator lost, with the steepest losses concentrated in high-poverty urban schools -- this retention advantage is not only a measure of morale but a fiscal one.

Of course it is possible that there is a self-selection bias. Indeed, that is true in every kind of school. Teachers choose to work in the kind of school they believe will satisfy their pedagogical desires. And yet almost half don’t stay. The retention data for Consortium teachers, a now subs-

11. New York Performance Standards Consortium (2025). *Where Thinking Thrives: 2025 Report,* 20–23. Describing the Consortium's moderation studies, teacher exchanges and professional learning groups, annual conference, and Performance Assessment Review Board.
12. Educator-stability and workforce figures reflect only the 36 New York City Consortium schools; the network's two upstate schools are excluded because school-level staffing data are available only through the NYC Department of Education.
13. NYC DOE 2024–25 School Quality Reports (4-year high schools).
14. NYC DOE Report on School-Based Staff Demographics (Local Law 226), FY 2024–25, via DOE InfoHub.

tantial and diverse population of educators, suggests that educators may be drawn to but they also remain committed to Consortium schools in large part because of the culture they engage/create/sustain.

The retention story extends to the educators the testing regime most often drives out.[15] According to the New York City Department of Education's Report on School-Based Staff Demographics (2025), Consortium schools employ a higher proportion of Hispanic teachers than the city (24.6% vs. 20.1%), and Black Consortium teachers' average years in the profession (11.7) are nearly identical to those of their White colleagues (12.3).[16] In schools serving predominantly Black and Latine student populations -- Consortium enrollment is 54.2% Hispanic and 23.2% Black -- this is consequential: an authentic-assessment culture appears to sustain, rather than deplete, a racially diverse and experienced teaching force (NYC DOE, 2025).[17]

Table 3. Educator Stability and Workforce Demographics[18]

Metric	Consortium	City	Difference
Teachers with 3+ years at school	86.1%	78.2%	+7.9pp
Median principal years at school	11.0	6.0	+5.0
Mean principal years at school	10.0	7.4	+2.6
Teachers — Hispanic	24.6%	20.1%	+4.5pp
Teachers — Black	19.7%	20.9%	−1.2pp
Teachers — White	49.6%	48.3%	+1.3pp
Teachers — Asian	12.1%	13.2%	−1.1pp

These figures are descriptive and cannot separate the effect of the Consortium's culture from the question of who chooses to teach there. A network known for teacher authority may attract educators who want that work and are more likely to stay. But Consortium schools are non-selective (in terms of student admission) and serve more students with characteristics that typically predict turnover (Table 1), so one can not argue that they are inherently easier places to teach. Drawing in as well as developing educators who want professional authority is part of how such a system builds capacity and sustains a consistent teaching force.

15. See Section I. The Learning Policy Institute (Tan, Wei, Carver-Thomas & García, 2026) documents that teacher turnover is most severe among teachers of color and in high-poverty urban schools.
16. New York City Department of Education (NYC DOE). (2025). Report on School-Based Staff Demographics (Local Law 226), FY 2024–25. NYC Public Schools Info Hub. Accessed November 4, 2026 at https://infohub.nyced.org/reports/government-reports/report-on-school-based-staff-demographics
17. NYC DOE 2024–25 School Quality Reports.
18. NYC DOE 2024–25 School Quality Reports (4-year high schools).

What these numbers describe, Consortium educators name in their own words, consistently framing their work in the language of professional dignity, intellectual ownership, creative engagement, and agency, rather than compliance or burden. Jeremy Copeland, a social studies teacher at the School of the Future, reflects: "I have a deep appreciation for the opportunity to develop rigorous, meaningful curricula -- a deep appreciation for what we're entitled to as professionals, learners, citizens."[19]

Anson Dean, a graduate of a Consortium school now teaching in the network, explains that what draws him and other teachers is "the freedom that teachers have... the autonomy teachers have to teach the things that they feel most passionate about," which, in his telling, "drives education forth" and "makes teachers really excited about coming in to teach."[20]

Charan Morris, an English teacher at Vanguard High School, traces her own route into the profession, highlighting the yearning teachers have for agency and creativity:

> *I walked into NYU as an undergraduate without the skills I needed. Many of my high school classes used rote memorization, copying down notes, and testing—without space for debate or critical thinking. I've since learned that my predicament is all too common among students of color entering college. This knowledge motivated me to become an educator. Fortunately, I met teachers from the Consortium and realized that the type of education I had yearned for was happening in these schools. We work together to design curriculum and educational spaces that engage students where they are. And this is what every public high school student deserves.*[21]

The Consortium's twenty-five years of practice shows that educator capacity for authentic assessment is neither rare nor fixed; it is developed, and sustained, by the structures schools and networks build around it. Districts do not need to wait for schools to have the capacity to do this work to implement change toward authentic assessment; as the Consortium example shows, capacity is an outcome of sustained investment.

19. New York Performance Standards Consortium, Where Thinking Thrives: 2025 Report (2025), p. 10. https://publuu.com/flip-book/967685/2125008/page/1?cover
20. New York Performance Standards Consortium, It's Where I Belong: Teachers Reflect on Why They Teach at Consortium Schools (video). https://www.youtube.com/watch?v=mfn58wQrt4o
21. New York Performance Standards Consortium, Where Thinking Thrives: 2025 Report (2025), p. 14. https://publuu.com/flip-book/967685/2125008/page/1?cover

CASE TWO: BUILDING EDUCATOR CAPACITY FOR RE-IMAGINED ACCOUNTABILITY AT SCALE THOUGH KENTUCKY CHANGEMAKERS PATHWAY

The Consortium demonstrates what educators' professional lives look like inside a mature authentic-accountability system. Most teachers, however, do not work in such a system – which raises the question of how this capacity gets built where it does not yet exist. Kentucky offers one answer. As Kentucky moves toward competency-based, performance-oriented learning statewide, anchored in the Kentucky Portrait of a Learner and the Kentucky United We Learn initiative, the University of Kentucky's Center for Next Generation Leadership has developed ChangeMakers, a year-long, state-approved rank-change pathway through which practicing teachers upgrade their certification (and their compensation) by transforming their own practice (UK Center for Next Generation Leadership, 2024).

The program is itself an authentic assessment system for teachers. As educators move through competency-based modules, including Student-Centered Learning, Learner Agency, Inquiry Learning, and Assessment for Learning, they plot their own learning paths, and travel in "crews" supported by experienced Crew Chiefs and Co-Pilots. Each ChangeMaker conducts a year-long action research project in their own classroom, assembles a portfolio of evidence that their practice has changed, and presents a final defense of their learning before a panel. Teachers learn student-centered assessment as they experience authentic assessment of their own progress through the program. Rather than being told about performance assessment, they experience what it demands and what it makes possible.

Across the program's first three cohorts (2021–2024), 193 teachers enrolled and 175 completed the rank change, a 91% completion rate. No participant who continuously engaged with the program failed to achieve it. Withdrawals were concentrated in personal and family circumstances, not inability to meet program criteria (UK Center for Next Generation Leadership, 2024).[22] Demand has grown from two founding districts (Jefferson and Shelby counties) to more than ten, spanning Kentucky's largest and



22. UK Center for Next Generation Leadership's ChangeMakers impact report details program outcomes and shares their approach to continuous improvement. 91% (175) educators who entered the program have achieved rank change with ChangeMakers' job-embedded professional learning pathway within a three-year period. Educators demonstrate their learning through a variety of performance assessments including capstone projects and a final defense of learning.

most diverse districts, including Jefferson County (Louisville) and Fayette County (Lexington), as well as small rural systems. The program reports that teachers of color complete at rates equal to their white colleagues. Nor is this a "cream of the crop" phenomenon: participants span content areas, career stages, and school contexts, suggesting that the capacity for this work is broadly distributed in the teaching force when structures support it.

What ChangeMaker participants report closely parallels what Consortium educators describe. Participants consistently named the experience as reinvigorating rather than burdensome: one Cohort 3 ChangeMaker reflected that the program "reinvigorated my teaching passion and encouraged me to push the boundaries of my instruction"(UK Center for Next Generation Leadership, 2024). In annual surveys, participants highlight collaboration with colleagues, the immediate classroom applicability of the modules, and the personalized support of crew leaders. Program leadership anecdotally observes a telling pattern in where the work proves hardest: not where teachers lack skill or will, but where they lack agency -- where prescribed curricula leave educators with little autonomy over their own instruction or classrooms. Capacity, in other words, is a function of the autonomy and support systems around teachers.

Two design features deserve particular note for districts weighing costs. First, the program is financially self-sustaining for participants: because rank change carries a salary increase, teachers effectively recoup the program's cost within roughly a year. Enrollment demand persists even though teachers pay, which program leadership reads as an indicator of educators' hunger for purposeful professional learning. Second, the model is built to scale: the program breaks even at roughly 60 participants per cohort, currently averages 61, and can expand to 200 or more as demand grows, with district partnerships embedding the work in teams rather than isolated individuals. Kentucky demonstrates that the labor of authentic assessment, far from exceeding teachers' capacity, can become the engine of their professional advancement. Hundreds of teachers are now skilled in designing the kind of teaching and authentic assessment the state seeks to scale. Whether this capacity-building translates into the retention advantages documented in Consortium schools is the subject of ongoing research by the program and its district partners.

CONCLUSION

Assessment and accountability systems built around standardized tests have negatively impacted teaching and learning, harming the retention of the educator workforce by undermining educators' professionalism and autonomy, depressing morale and motivating turnover. Recent polling by the NEA shows that educators, parents, and opinion leaders all prefer a new system built around more authentic, performance-based assessments that provide students with opp-

ortunities to develop and hone "real world" skills like collaboration, problem-solving, communication, leadership, and creativity (NEA, 2022). Further evidence shows that teachers are hungry for such a challenge and capable of thoughtfully undertaking it under the right conditions.

These conditions can be met by redirecting resources from the currently massive test-based accountability infrastructure toward systems that value and cultivate educator expertise and support expansive opportunities for collaboration and capacity-building. Systems like the New York Performance Standards Consortium, and other districts like Anaheim Union High School District in California,[23] demonstrate how intentional diversion of investment away from standardized assessments and test prep and toward the development of PA systems can improve both student and educator outcomes. These systems are showing gains in student achievement – not only in test scores but also in critical success metrics like graduation rate and college readiness – while boosting educator satisfaction and retention. Moreover, they offer proof that educators are capable of collaborating and co-developing systems to advance educational equity and ensure student access to high-quality performance based assessment.

Any school, network, district or state leader can begin to build capacity for performance-based assessment and accountability by pursuing the following questions with administrators, educators and the families and communities they serve:

1. ***Readiness:*** Where is there appetite to shift from predominantly test driven to PA? Who would be likely champions, and where might resistance emerge? How can concerns be addressed such that no group is excluded from the start?
2. ***Policy audit:*** Which local policies or practices could undermine a systematic move to PA, and what enabling policies and practices need to be built? For example, do structures exist that promote teacher collaboration? Are standardized test scores the primary metric for teacher evaluations, student tracking, school ranking, or other communications about school quality by realtors or journalists? AAre professional development funds spent on outside contracts and speakers or invested in teacher collaboration?
3. ***Participatory budgeting and finance reconfiguration:*** Given the current picture of system readiness and policy alignment, what will it cost to build capacity for PA, and where might savings accrue from reducing the footprint of standardized testing? Such questions need

23. Anaheim Union High School District (AUHSD) serves approximately 25,000 students grades 7-12, 74.7% of whom are economically disadvantaged. Under the leadership of Superintendent Michael Matsuda, AUHSD intentionally stepped away from rigid, optional interim assessments and traditional standardized test prep. Instead, the district developed a careful framework built around project-based learning (PBL) and localized performance-based assessments. In addition to state assessment gains, students have demonstrated incredible growth in achievement across a diverse array of indicators all of which speak to school quality and students' learning opportunities. For example, in 2022, AUHSD students earned 88% of Orange County's total number of civic badges and 28% of all civic badges in the state of California (Feder, 2024). These gains extend beyond academics students' connection to their school communities . According to the 2025-2026 California Healthy Kids Survey, AUHSD continues to outshine even its own performance in previous years, finding 92.5% of AUHSD students feel connected to their school (+5.5% from the 2023-2024 survey administration); 85.1% of students care about other students at their school (+21.1%) (California School Dashboard, 2026).

not be answered by education leaders alone but can be approached through participatory budgeting, which invites core constituencies to anticipate costs and savings and to recalibrate school and district budgets to align with shared priorities.

In summary, the greatest risk to building this new system is not educators' lack of interest, commitment, or innate capabilities; it is that investments in capacity-building are idiosyncratic, organized as an 'add-on' and/or are contradicted by long standing system policies and teacher evaluation systems. Success requires policies and investments that prioritize professional learning and collaboration, create "space on the plate" for meaningful learning and performance assessment, and supports rigorous expectations system-wide. **Thus, the question is not, "*Do teachers have what it takes?*" but "*How can we build a system of support through which educators co-construct and develop capacity alongside a more authentic system of assessment, accountability, and equity?*"**

By involving educators in the creation of new assessment and accountability systems as described in this paper, system leaders simultaneously invest in the design of a high-quality system and fortify the ability of the educator workforce to carry it out. Some educators may choose to leave the new system rather than adapt to it; others may join for the first time because of it. But overwhelmingly, we have found that when teachers are entrusted and supported to assess student learning in more robust and authentic ways, they deepen their sense of purpose, fueling commitment and retention while expanding the knowledge and skills of the educator workforce. Research has shown that increasing educator capacity further fuels student confidence, promoting academic gains and reducing achievement gaps by race, class, language, and disability. And, upon seeing the tangible results of students' performance-based demonstrations of learning, families and community members are more likely to join in supporting them rather than opting-out (Moran and Perry, 2026). These improvements further heighten student achievement and graduation in a virtuous cycle.

REFERENCES

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267. https://doi.org/10.3102/0013189X07306523

Andrew Carnegie Foundation. (2024). Anaheim collaborative: New ways to measure students' progress. https://carnegie.org/article/anaheim-collaborative-new-ways-to-measure-students-progress/

California School Dashboard. (2026). District Performance Overview: Anaheim Union High. https://www.caschooldashboard.org/reports/30664310000000/2026

David, J. L. (2011). High-stakes testing narrows the curriculum. *Educational Leadership*, 68(6), 78–80. https://www.ascd.org/el/articles/high-stakes-testing-narrows-the-curriculum

EducationWeek (2019). What is performance assessment? EducationWeek special report. https://www.edweek.org/teaching-learning/what-is-performance-assessment/2019/02

Feder, H. (2024). Performance assessment transforming a district: The Anaheim approach. FairTest. https://fairtest.org/performance-assessment-transforming-a-district-the-anaheim-approach/

Feng, L., Figlio, D., & Sass, T. (2018). School accountability and teacher mobility. *Journal of Urban Economics, 103*, 1–17.

Galla, B. M., Shulman, E. P., Plummer, B. D., Gardner, M., Hutt, S. J., Goyer, J. P., D'Mello, S. K., Finn, A. S., & Duckworth, A. L. (2019). Why high school grades are better predictors of on-time college graduation than are admissions test scores: The roles of self-regulation and cognitive ability. *American Educational Research Journal*, *56*(6), 2077-2115. https://doi.org/10.3102/0002831219843292

Gunn, J., Al-Bataineh, A., & Abu Al-Rub, M. (2016). Teachers' perceptions of high-stakes testing. *International Journal of Teaching and Education, 4*(2), 49–62. https://www.researchgate.net/publication/325966896_TEACHERS'_PERCEPTIONS_OF_HIGH-STAKES_TESTING

Inflexion. (2025). Empowering Students and Transforming School Culture Through Self-Reflection at Anaheim Union High School District. https://inflexion.org/case-study/empowering-students-and-transforming-school-culture-through-self-reflection-at-anaheim-union-high-school-district/

Intercultural Development Research Association. (2002). The dangerous consequences of high-stakes testing. *IDRA Newsletter*, 29(7) https://www.idra.org/resource-center/the-dangerous-consequences-of-high-stakes-testing/

Howard Terrell, J., & Barzee, S. (2023) *Graduation requirements and measures: A review of performance assessment implementation in select states for the New York State Education Department.* Region 2 Comprehensive Center, WestEd. https://www.wested.org/resource/ny-grad-measures-performance-assessment/

Jones, B. D. ., & Egley, R. J. . (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives,* 12, 39. https://doi.org/10.14507/epaa.v12n39.2004

Kraft, M. A., & Lyon, M. A. (2024). The Rise and Fall of the Teaching Profession: Prestige, Interest, Preparation, and Satisfaction Over the Last Half Century. *American Educational Research Journal, 61*(6), 1192-1236. https://doi.org/10.3102/00028312241276856

Learning Policy Institute. (2024). 2024 update: What's the cost of teacher turnover? [Interactive tool]. https://learningpolicyinstitute.org/product/2024-whats-cost-teacher-turnover

Leather, P., Rathmann-Noonan, A., and Rothman, M. (2026, November). *Educational Accountability Remade*. https://www.eduaccountability.org/comparability.html

Moran, P., and Perry, K. (2026, October). Elevating Community Trust As the First Principle of Accountability. *Educational Accountability Remade*. https://www.eduaccountability.org/community-trust.html

Musa, S. and Cataldo, K. (2024). Finding a balance: Navigating high stakes testing and standardized curricula for joyful teaching and learning. *New Jersey English Journal 13*, Article 11. https://digitalcommons.montclair.edu/nj-english-journal/vol13/iss2024/11

National Education Association (2015). A New Vision for Student Success: A Report from NEA's Accountability Task Force. https://www.nea.org/sites/default/files/2021-10/A%20New%20Vision%20%28Accountability%20Task%20Force%29%202015%5B12%5D.pdf

National Education Association (2022). Principles for the Future of Assessment. Resource Library. https://www.nea.org/resource-library/principles-future-assessment

New York City Department of Education (NYC DOE). (2025). Report on school-based staff demographics (Local law 226), FY 2024–25. NYC Public Schools Info Hub. Accessed November 4, 2026 at https://infohub.nyced.org/reports/government-reports/report-on-school-based-staff-demographics

New York City Department of Education (2025). NYC school survey archives. NYC Public Schools Info Hub. Accessed November 5, 2026 at https://infohub.nyced.org/reports/students-and-schools/school-quality/nyc-school-survey/survey-archives

New York Performance Standards Consortium. (2025). Where Thinking Thrives. https://www.performanceassessment.org/our-impact-22

Nichols, S.L. . and Berliner, D.C.. (2008). Why has high stakes testing so easily slipped into contemporary American life? *Phi Delta Kappan, 89*(9), 672-676. https://doi.org/10.1177/003172170808900913

Rebora, A. (2012) Teachers place little values on standardized testing. *Education Week*, 31, 26, 14. https://www.edweek.org/teaching-learning/survey-teachers-place-little-value-on-standardized-tests/2012/03

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, *52*(3), 475-514. https://doi.org/10.3102/0002831215585562

Ryan, S. V., von der Embse, N. P., Pendergast, L. L., Saeki, E., Segool, N., & Schwing, S. (2017). Leaving the teaching profession: The role of teacher stress and educational accountability policies on turnover intent. *Teaching and Teacher Education, 66*, 1–11.

Schaeffer, B., Neill, M., & Guisbond, L. (2012). Resistance to high stakes testing spreads. *District Administration*, 48(8), 34. This article was republished by The National Center for Fair & Open Testing and is available at https://fairtest.org/resistance-high-stakes-testing-spreads/

Skinner, R. R. (2026). *The elementary and secondary education act (ESEA), as amended by the Every Student Succeeds Act (ESSA): A primer.*(CRS Report No. R45977) https://www.congress.gov/crs-product/R45977

Taie, S., & Lewis, L. (2023). *Teacher attrition and mobility. Results from the 2021–22 Teacher Follow-Up Survey to the National Teacher and Principal Survey* (NCES 2024-039). U.S. Department of Education. Washington, DC.: National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2024039

Tan, T., Wei, W., Carver-Thomas, D., & García, E. (2026). Teacher turnover in the United States: Who moves, who leaves, and why. Learning Policy Institute. https://doi.org/10.54300/248.479

Tan, T. S., & Patrick, S. K. (2024). When teachers leave: Understanding the financial impacts on schools [Blog post]. *Learning Policy Institute*. https://learningpolicyinstitute.org/blog/when-teachers-leave-understanding-financial-impacts-schools

Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost (CSE Report 588). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). https://cresst.org/publication/a-survey-of-teachers-perspectives-on-high-stakes-testing-in-colorado-what-gets-taught-what-gets-lost/

Terada, Y. (22, October 14). *The psychological toll of high -stakes testing*. Edutopia. https://www.edutopia.org/article/psychological-toll-high-stakes-testing/ Discussing

University of Kentucky Center for Next Generation Leadership. (2024). CEO II rank change program review: Comprehensive quantitative & qualitative summative 3 year review report (2021-2024). https://docs.google.com/document/d/1nd4EDQTEm4F2NyDQqsTAGLUabJHuTiB_7AJzFyL4R8A/edit?tab=t.0#heading=h.yn3syzf53s74Docs

Villavicencio, A., Jaffe-Walter, R., Cherng, H.-Y.-S.. (2026). Relationship between professional culture and outcomes of immigrant English learners: A mixed methods study. *American Educational Research Journal, 63*(1)*, 38*-71. https://doi.org/10.3102/00028312251389220

World Economic Forum. (2025). The future of jobs report 2025. World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2025

Wronowski, M. L., & Urick, A. (2019). Examining the relationship of teacher perception of accountability and assessment policies on teacher turnover during NCLB. *Education Policy Analysis Archives, 27*(86). https://doi.org/10.14507/epaa.27.3858

Wronowski, M., & Urick, A. (2021). Teacher and school predictors of teacher deprofessionalization and demoralization. *Educational Policy, 35*(5), 679–720.

This paper was authored by Mica Baum-Tuccillo, Director of Research at the New York Performance Standards Consortium, Christine DonFrancesco, Senior Policy Analyst, Assessment and Accountability at the National Education Association, and Michelle Fine, Distinguished Professor at City University of New York, with editing by Jenny Poon and contributions from the Project EAR team.

SPONSORING ORGANIZATIONS

The Center for Innovation in Education supports education leaders to create effective, durable, community-owned solutions to complex problems.

The National Center for Fair & Open Testing (FairTest) works to end the misuses and flaws of standardized testing and to ensure that evaluation of students, teachers and schools is fair, open, valid and educationally beneficial.

SUGGESTED CITATION

Baum-Tuccillo, M., DonFrancesco, C., & Fine, M. (November 2026). *Authentic assessment and accountability as an engine of educator renewal.* Education Accountability Remade.

PROJECT EAR TEAM

MICA BAUM-TUCCILLO
Director of Research, New York Performance Standards Consortium; Senior Research Fellow, Public Science Project, City University of New York

CHRISTINE DONFRANCESCO
Senior Policy Analyst, Assessment and Accountability, National Education Association

HARRY FEDER
Executive Director, National Center for Fair & Open Testing (FairTest)

MICHELLE FINE
Distinguished Professor, Psychology, Urban Education, Liberal Studies, Women's and Gender Studies, American Studies, Public Scholarship, LGBTQ Studies, City University of New York

PAUL LEATHER
Partner, Education Policy and State and National Relationships, Center for Innovation in Education

TONY MONFILETTO
Fmr Director, Instituto del Puente at Future Focused Education

PAM MORAN
Executive Director, Virginia School Consortium for Learning

KAREN PERRY
Executive Director, University of Kentucky Center for Next Generation Leadership

JENNIFER POON
Partner, Center for Innovation in Education

ALEXANDRA (AJ) RATHMANN-NOONAN
Executive Director, NYPSC

MICHAEL ROTHMAN
Senior Director, P20 Pathways Development, City University of New York

DOANNIE TRAN
Managing Partner, Center for Innovation in Education