BEST NC Business for Educational Success and Transformation

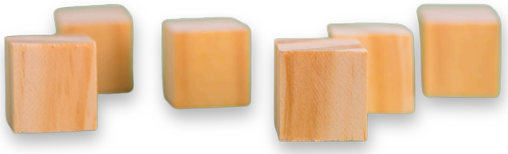
Evidence for Excellence

Building a Toolkit of Actionable Measures to Support and Advance Teachers



A Policy Brief by BEST NC

2026



Evidence for Excellence

Building a Toolkit of Actionable Measures to Support and Advance Teachers

Table of Contents

Executive Summary.....i

Introduction.....ii

Section 1. Treating Teaching as a Profession Requires Tools for Professional Growth..... 1-9

- 1.1 The Importance and Application of Professional Evaluation Systems
- 1.2 The Current System's Proxies Are Inadequate for the Needs of Teachers
- 1.3 How North Carolina Currently Evaluates Teachers
- 1.4 The Current System's Tools Are Inadequate for the Needs of North Carolina
- 1.5 Performance Data Must Inform Workforce Decisions
- 1.6 How a Multi-Measure System Can Strengthen North Carolina's Teaching Workforce
- 1.7 Why Past Evaluation Reforms Failed

Section 2. Understanding Teacher Performance and Why Measurement Matters.....10-13

- 2.1 Defining Teacher Performance
- 2.2 Why Measurement Is Hard – and Why Guardrails Matter
- 2.3 Why Improving Measurement Is Necessary

Section 3. The Measurement Toolkit.....14-21

- 3.1 How EVAAS Growth, Student Experience Data, and Video Observations Can Work Together
- 3.2 How the Three Measures Meet the Bar for Measuring Professional Practice
- 3.3 How the Toolkit Works Together
- 3.4 Testing the New Measures Before Broader Use

Section 4. Applying Validated Measures to Licensure, Advancement, and Preparation.....22-24

- 4.1 Licensure: Reimagining a Licensure-by-Performance Model
- 4.2 Advancement: Expanding the Reach of Effective Teachers
- 4.3 Preparation: Closing the Feedback Loop for Educator Preparation Programs

Section 5. Policy Recommendations.....25-26

Conclusion..... 27

References.....28-31

Acknowledgements

This brief is about teachers, and teachers both inspired and shaped it. We thank Ashley Thompson, who shared her story of teaching art in Cumberland County. Special thanks to Maureen Stover, along with many other educators, researchers, and policy experts who provided input. The recommendations are BEST NC's alone.

BEST NC Business for Educational Success and Transformation

BEST NC is a non-profit, non-partisan coalition of business leaders committed to improving North Carolina's education system by convening, informing, and engaging state leaders to identify and advance improved education policies, programs, and initiatives. BEST NC envisions a North Carolina in which every student graduates with the knowledge, skills and behaviors to succeed in a competitive global economy. Our mission is to unite an engaged and informed business perspective to build consensus toward dramatically transforming and improving education in North Carolina.

© Copyright 2026. BEST NC | Business for Educational Success and Transformation in North Carolina

Executive Summary

Every teacher deserves to know how they are performing, and every student deserves a teacher who has the feedback and support to keep improving.

In most professions, people receive regular, structured feedback on their performance, and that evidence shapes how they grow and advance. Most North Carolina teachers do not: about three in five have no individual measure of their impact on student learning, and at the high school level, more than four in five do not. The state's current evaluation tools were not built to fill that gap. Observation-heavy systems from earlier reform eras asked a single, ill-fitting tool to do too much, consuming principals' time without producing feedback teachers could use or evidence that reliably distinguished strong performance from weak.

Decades of research point to a different model consisting of multi-measure systems that combine evidence of instructional practice, student learning, and student experience. Multi-measure systems are more reliable, fairer, and more useful than observation-only models for both professional growth and formal decision-making. And yet, a multi-measure system is only as useful as the feedback it produces: teachers need results they can act on and a clear path from feedback to improvement. A teacher near the top of the effectiveness distribution produces roughly three times the learning gains of a teacher near the bottom in a single year – variation current tools mostly cannot see, let alone address.

This brief recommends North Carolina build an evidence-based multi-measure system governed by one principle: the state sets a rigorous menu of measures, each validated for the specific purpose it will serve. Principals and team leaders use every available measure to support a teacher's growth, and for career-impacting steps such as licensure and advancement, the teacher chooses which validated measures count. New measures begin as tools for professional feedback and inform career-impacting decisions only after an independent evaluation shows they meet the standard for that use.

Three recommendations carry this forward (see Section 5 on page 25):

Recommendation A: Establish and fund a "Multiple Measures of Teacher Practice" pilot.

Alongside existing EVAAS growth data, pilot student experience data (research-aligned surveys on instructional practices, not popularity) and a video observation option (using best practices and techniques learned from the Best Foot Forward study). Let districts opt in with state funding for the tools and training. An independent evaluation then determines which measures are ready for broader use, e.g., for Recommendations B & C.

Recommendation B: Tie career advancement and licensure to demonstrated performance.

Once validated, identify ways to use these measures for career pathways including paid mentoring opportunities and Advanced Teaching Roles (referred to as ATR throughout this brief).

Recommendation C: Close the educator preparation feedback loop.

Close the educator preparation feedback loop. Include these measures of teacher practice in the EPP Dashboard, as each is validated for that use, giving preparation programs the evidence to see which practices and pathways are actually producing effective teachers.

Together, these recommendations move North Carolina from fragmented, compliance-driven evaluation to a coherent, professional system – one that gives teachers the same kind of credible, actionable evidence of their impact that other professions have long taken for granted.

Introduction

Doctors get regular, multi-source feedback on their clinical performance. Attorneys build their careers on demonstrated results. Teachers deserve the same: credible evidence of their impact, real feedback, and a clear path to grow. Today, most do not have it. About three in five North Carolina teachers have no individual measure of their instructional impact. At the high school level, more than four in five teachers lack measures of impact.[1,2]

EX 0.1 – Feedback Gap: Doctors, Attorneys, and Teachers Have Very Different Methods of Building Professional Capacity



Research and best practice indicate that North Carolina should expand the evidence available to teachers about their own practice. The state should begin with a pilot of student experience data and video-supported observation that gives teachers more actionable feedback, then evaluate to determine whether those measures are reliable, valid, fair, and practical enough for broader use. Where measures that already meet that bar exist, the state should use them now to open career advancement and license renewal to teachers who demonstrate strong practice.

A modern system offers every teacher clearer feedback, more credible recognition, and a faster path to advancement based on demonstrated skill. Specifically: Teachers gain credible, actionable evidence of their own practice, with supports targeted to where they are in their career.

60%

of NC Teachers Have No Individual Measure of Their Impact on Student Learning

Source: *NC DPI (Data Request)*

- School and district leaders can match teachers' strengths to students' needs and spend less time on compliance-driven observation.
- Teachers who demonstrate strong practice gain a route to paid mentoring and ATR without leaving the classroom.
- Effective teachers can reach a continuing license sooner, based on demonstrated skill rather than seat time.
- Veteran teachers can be recognized and retained for demonstrated skill, not time in the profession alone.
- Policymakers can invest with confidence, and the North Carolina Department of Public Instruction and the state's educator preparation programs get feedback on which practices and pathways produce effective teachers.

No system is perfect, and no set of measures will satisfy every stakeholder on every point. This brief does not shy away from those tensions. Instead, it proposes a set of validated professional measures, opening up the possibility for teachers to choose which measure best represents their practice for key professional gateways such as licensure and pathways such as paid mentoring and ATR positions. The opportunity in front of North Carolina is real: to move from fragmented signals to a coherent, professional system that treats teaching as the complex, consequential work it is.



SECTION 1. Treating Teaching as a Profession Requires Tools for Professional Growth

1.1 The Importance and Application of Professional Evaluation Systems

Professional evaluation systems are a broadly accepted workforce tool that enables organizations to manage talent strategically, support continuous improvement, and retain high performers. This is true for both public-sector and private-sector organizations, particularly for high-skill professions. Typical business and public-sector organizations provide an annual formal performance review opportunity between a professional employee and their manager to set goals, review professional responsibilities, identify areas for strategic growth, and offer a chance for employees to ask for additional support toward their career goals.

A Note on Terminology: "Career-Impacting," Not "High-Stakes"

Decisions about licensure advancement, coaching, retention pay, and program feedback matter to a teacher's career, and this brief treats them that way. It describes them as career-impacting rather than high-stakes because the two models work differently.

High-stakes evaluation typically means one opaque number such as a test score or a single observation producing a possible punitive outcome. The system described here is designed the other way around: no single measure is determinative; teachers select which eligible measures they submit as evidence of their practice, and each measure must be validated and transparent before it informs a rating, licensure step, or pay decision. This word choice reflects the design.

In most professions, evaluation systems are not designed as compliance exercises or retrospective judgments. Instead, they function as a continuous improvement tool to develop the talent on their teams. When designed well, professional evaluation systems serve two primary purposes: identifying skill and knowledge gaps so employers can provide targeted support for professional growth, and recognizing strong performance to retain, promote, or expand the reach of effective professionals.[3]

Tools for Professional Growth

Valuable feedback answers three questions: where am I going (the goal), how am I going (progress toward it), and where to next (concrete next steps). It also addresses the task and the learning process rather than the person.[4] In this way, evaluation systems diagnostically identify where professionals can grow and how organizational resources can be effectively deployed to meet those needs.[5] This function is essential for supporting continuous improvement of practice, particularly in complex professions where performance depends on applied skill rather than discrete tasks.[6]

Tools for Retaining and Extending the Reach of Top Talent

Organizations rely on evaluation evidence to retain high performers, make promotion decisions, and identify individuals who are ready to take on expanded responsibilities or leadership roles.[7] Without credible performance information, organizations struggle to distinguish excellence from adequacy, weakening retention efforts and undermining leadership pipelines.[8]

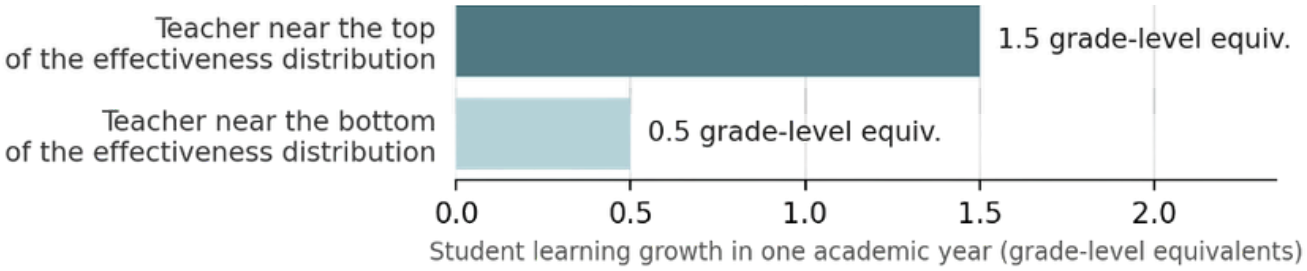
Because evaluation systems inform consequential – career-impacting – decisions, well-designed public-sector evaluation systems are expected to meet basic standards for promotion and professional development decisions.[9] Well-designed systems are expected to produce results that are reliable across observers and occasions, valid for the decisions they inform, credible to the educators being evaluated, and practical to operate at scale.[10] Systems that fail to meet these criteria risk losing credibility and cannot reliably inform workforce decisions.[11]

In addition to these core purposes, evaluation systems serve important secondary functions. In aggregate, performance data can inform recruitment strategies, identify which preparation pipelines are producing the strongest talent, and support long-term workforce planning.[12] Additional information about exceptional educators can help generate just-in-time solutions to North Carolina's current teacher pipeline challenges.

1.2 The Current System's Proxies Are Inadequate for the Needs of Teachers

Research over several decades demonstrates that teacher effectiveness varies substantially and that these differences have meaningful consequences for student learning.[13] Students taught by more effective teachers consistently make greater academic gains than those taught by less effective teachers, and these differences accumulate over time into clear lifetime benefits in both achievement and earnings.[14]

EX 1.1 – A Year's Worth of Difference in Student Learning Growth



Note: *A teacher near the top of the effectiveness distribution produces roughly 1.5 grade-level equivalents of student learning in a year, compared to about 0.5 for a teacher near the bottom.*[15]

Source: *Hanushek and Rivkin (2006)*

At the same time, the research is clear that commonly used proxies for teacher quality are weak predictors of instructional effectiveness. Credentials, advanced degrees, licensure exams, and experience beyond the early career years explain little of the observed variation in teacher contributions to student learning.[16,17,18] While such requirements may serve as gatekeeping or signaling functions, they do not reliably distinguish teachers who are more successful in advancing student learning from those who are less effective. License renewal follows the same logic. A teacher renews a continuing professional license every five years by completing eight units of renewal credit, and nothing in the renewal process considers whether that teacher's students are learning.[19]

As a result, regulatory systems that rely primarily on inputs rather than outcomes create the appearance of quality control without consistently identifying effective teaching. Eric Hanushek, an economist and senior fellow at Stanford University's Hoover Institution, argues that it is "extraordinarily difficult" to ensure teacher quality through regulatory entry requirements alone, because credentials, degrees, experience, and even teacher test scores are not consistently correlated with teaching skill.[20] Systems built on these assumptions therefore struggle to function as true professional workforce systems.

For most teachers, the current data provided to teachers about their own professional capabilities and contributions to student learning is not detailed, timely, or specific enough to support meaningful continuous professional improvement. When this support is also not accessible to administrators, a teacher has few opportunities to identify and improve on professional weaknesses, celebrate professional strengths, and learn from recognized peers.

1.3 How North Carolina Currently Evaluates Teachers

North Carolina currently uses three separate measurement tools for teachers, none of which were designed with the others in mind:

A) The North Carolina Educator Evaluation System (NCEES) generates a principal-issued rating of each teacher on the North Carolina Professional Teaching Standards, but as detailed below, nearly half of all teachers are observed briefly and rated on only two of the five standards. These results also rarely align with student growth data.

B) The Education Value-Added Assessment System (EVAAS), the state's growth model, only exists for teachers of tested subjects, and since 2016–17, when growth stopped being a stand-alone standard in the evaluation, it has been information a teacher sees rather than a separately rated part of their evaluation.

C) National Board Certification is voluntary and runs entirely outside the state system; the state helps cover the application fee and rewards it with a 12% salary supplement while the certification remains current, but it was not built to generate feedback and has no bearing on a teacher's annual rating.

None of these three tools feeds into the others. A principal's observation, an EVAAS growth score (for the minority of teachers who have one), and National Board Certification status come from different bodies on different timelines for different purposes, and nothing combines them into one measure of practice. That means a district trying to figure out who is ready for a leadership role, or whose growth data should shape a professional development plan, or which new teachers need the most support, cannot check one system for the answer – it has to check three, and for most teachers, at least one of those three is missing data.

Without a broader performance toolkit, most teachers receive:

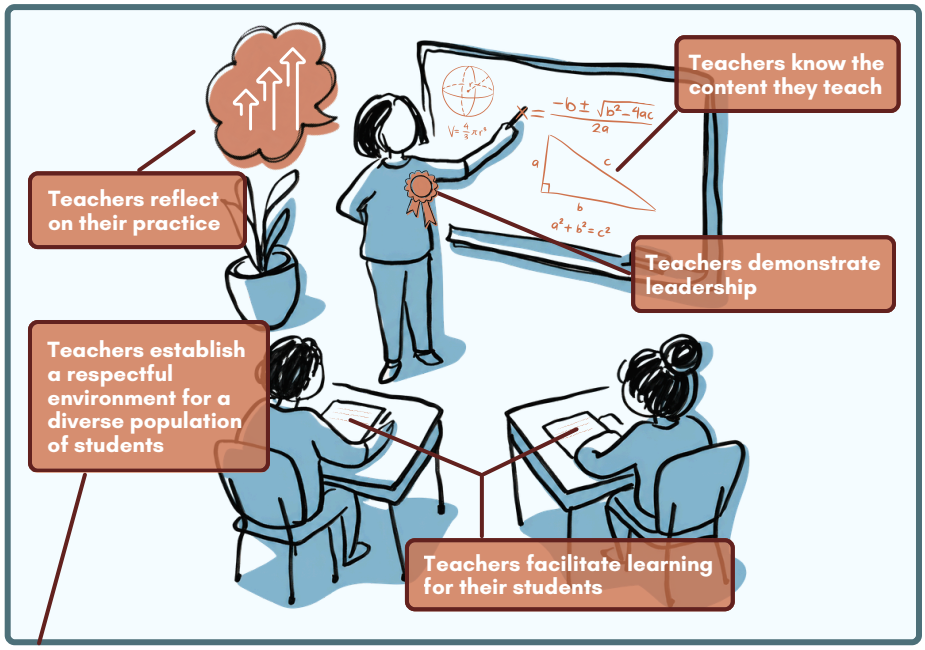
- Limited information about their instructional impact; and
- Feedback that is difficult to compare across classrooms, schools, or time.

The lack of reliable, comprehensive performance data for most teachers limits the career's professionalism both for individual teachers' career development and recognition, and for medium- and long-term workforce planning needs.[21]

A. NC Educator Evaluation System (NCEES): Principal Evaluations of Teachers

The North Carolina Educator Evaluation System (NCEES) relies on principal evaluations of teachers' practice through classroom observation. NCDPI (the North Carolina Department of Public Instruction) defines the system as a growth instrument. "It identifies the knowledge, skills, and dispositions expected of teachers, and measures the level at which they meet the standard as they move from ratings of 'not demonstrated' to 'distinguished.'"[22]

EX 1.2 – NCEES Standards for Teacher Evaluation



Educators are assigned effectiveness ratings based on their overall performance on the five criteria. The specific requirements depend on the teacher's tenure and the evaluator's discretion. Teachers employed by the local school district for fewer than three consecutive years are guaranteed a comprehensive cycle: three formal observations by the principal plus one by a peer, each lasting at least 45 minutes, followed by a summative conference and a summary rating.[23]

Teachers with three or more years may be placed on that same comprehensive cycle, a standard cycle, or an abbreviated cycle, at the evaluator's discretion. The standard cycle requires three observations in total – one formal observation with a pre- and post-conference, plus two more that may be formal or informal – and, unlike the abbreviated cycle, is not limited to two standards; it covers all five, the same as the comprehensive cycle.[24] The abbreviated cycle requires only two observations, which may be informal and as brief as 20 minutes each, and rates the teacher on only two of the five standards (Standards 1 and 4).[25] Of the 103,288 teacher evaluation plans North Carolina archived at the close of the 2025-26 school year, 45% were abbreviated, 32% standard, and 22% comprehensive.[26]

Regardless of which evaluation cycle is used, the resulting rating determines what happens next. A teacher rated proficient across all standards develops an Individual Growth Plan aimed at continued improvement. A teacher rated below proficient on one or more standards is placed on a Monitored or Directed Growth Plan, with a set timeline to reach proficiency; a teacher who is rated below proficient in a way that constitutes unsatisfactory performance may instead be placed on a Mandatory Improvement Plan under state statute.

Student growth was originally a sixth, stand-alone standard in this rating. The EVAAS section (Section B, below) explains why that changed and what followed.

A teacher on the abbreviated cycle who also teaches a subject without a statewide test is evaluated on two brief observations against two of five standards and has no outcome measure at all. Neither condition is unusual: 45% of teachers are on that cycle, and about three in five teach non-tested subjects.[27,28]

In the absence of stronger performance measures, principal evaluations are often expected to serve as the primary indicator of teacher quality. While principals play a critical role in instructional leadership, research on evaluation systems indicates that observation-only approaches are vulnerable to subjectivity and inconsistency.[29]

The competencies addressed in the NCEES evaluations are like those in other public- and private-sector evaluations, but the major differences are in the way the data is collected and used. In other professional environments, no single evaluator is responsible for as many evaluations as the average principal and assistant principal. With fewer evaluators responsible for more evaluations, differing views of effective teaching can limit accuracy between reviewers and the large number of required evaluations limits precision.[30]

Additionally, evidence shows that principals can distinguish more effective teachers from less effective ones in their observations, but they are limited in their abilities to recognize and address specific skill gaps through observations alone.[31] Observation feedback is also often too general to act on: in a study of written evaluation feedback given to roughly 1,200 early-career Tennessee teachers, nearly half received no actionable or goal-setting feedback over a full school year.[32]

Because the average principal has 50 direct reports, principal-led observation a few times per year is a poor proxy for the type of daily interaction that managers with six to eight direct reports can have with their team. Practitioners report that this approach places a heavy burden on school leaders, whose responsibilities already span instructional leadership, operations, staffing, and student discipline.[33]

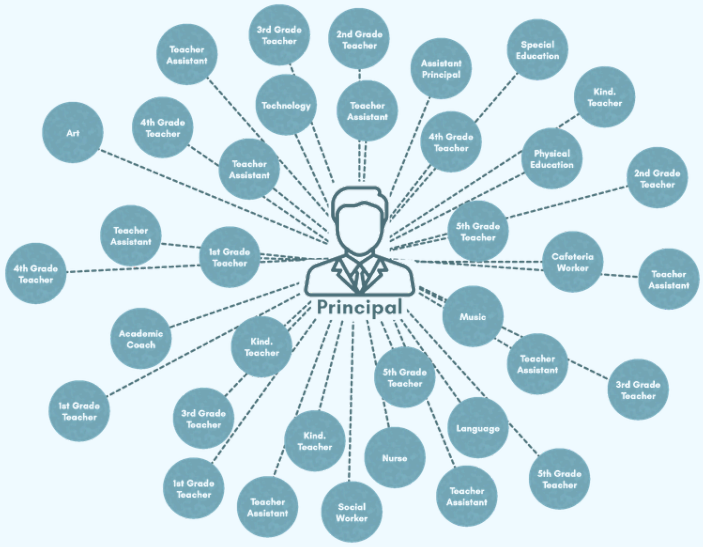


Strategic staffing initiatives offer clear alternatives for instructional leadership but have not yet been integrated into statewide data systems on measures of teacher practice.

Over time, these observation-only systems also struggled to meet basic standards of objectivity, reliability, and usefulness. Studies of evaluation design found that single-observer models are vulnerable to subjectivity, limited sampling of instruction, and inconsistent calibration across schools.[34,35] Principals observing a teacher only a few times per year cannot reliably capture the full range of instructional practice or student experience.

EX 1.3 – A Structural Mismatch

The average North Carolina principal supervises about 50 teachers. The entire evaluation load sits with one person, perhaps shared with an assistant principal.



Most critically, early systems failed to meaningfully differentiate performance. In a study of 12 districts across four states, more than 99% of teachers in districts with two-level rating systems were rated satisfactory, and 94% of teachers in districts with more rating levels received one of the top two ratings, despite well-documented variation in instructional performance.[36] This phenomenon – often described as ratings compression – produced evaluation results that were poorly aligned with student learning outcomes and offered little guidance for professional growth or support.[37]

Similarly, NCEES ratings and EVAAS growth results for the same teachers show a substantial mismatch. This is explored further in Section 2.3.

B. The Education Value-Added Assessment System (EVAAS): Student Growth

The Education Value-Added Assessment System (EVAAS), North Carolina's statewide growth model, uses a statistical method to estimate a teacher's contribution to student learning. The model is designed to isolate a teacher's effect from the many other factors that shape achievement – a student's starting point, background, and trajectory – by using the student's own past scores as the baseline for comparison.

The model, however, can only run where the underlying data exists: It requires standardized, comparable test scores in the same subject across consecutive years. As a result, EVAAS is generated only for teachers of state-tested subjects – End-of-Grade, End-of-Course, and a handful of other statewide assessments – not for every teacher or every course. The State Board's accountability task force is currently reviewing how growth is calculated for school accountability; that review and its relationship to this brief are described in the "A Movement to Adapt EVAAS" box in Section 3.

In North Carolina, EVAAS serves as the state's only statewide, outcome-based measure of instructional impact. Until the 2016–17 school year, growth was a stand-alone standard in a teacher's NCEES rating. The State Board removed that standard in 2016 (see box). A teacher's EVAAS score, where one exists, is no longer a separate component of the summary rating.[38] Growth data still carries real stakes outside the evaluation rating, however: the State's performance bonuses for reading and math teachers are awarded based on EVAAS growth scores, ranking teachers in the top 25% statewide or within their own school or district.[39]

While growth measures can provide useful information where available, they reach only the minority of teachers who teach tested grades and subjects. EVAAS is generated only for teachers of state-tested subjects, meaning approximately 40% of North Carolina's teaching force receives an EVAAS score.[40] This leaves three in five teachers – and at the high school level, more than four in five – with no individual, outcome-based measure of their instructional impact.[41]

How 3 in 5 Teachers Lost a Measure of Their Impact, 2016–2020

North Carolina did not decide that most teachers should go without a measure of their impact on student learning. It happened through three separate changes.

In April 2016, the State Board of Education removed student growth as a stand-alone standard in teacher evaluation, beginning in 2016–17. The reason was practical. Under G.S. 115C-333, a teacher in a low-performing school who receives a below-proficient rating must remediate it within 90 instructional days, and a growth-based rating arriving in November left no realistic window to do so. The Board's policy stated that student growth estimates would continue and would inform educator evaluation processes.[42]

Two later changes had a larger effect. Under North Carolina's Race to the Top grant, the state had built two ways to give growth measures to teachers without an EVAAS score: the Analysis of Student Work and the NC Final Exams. The General Assembly eliminated the Analysis of Student Work in 2017 and, in a 2019 law, the NC Final Exams, effective with the 2020–21 school year.[43]

S.L. 2019-212 directed the State Board to report to the Joint Legislative Education Oversight Committee on other ways to accomplish what the Final Exams had done. The Board's March 2020 report noted that roughly 12,000 teachers had relied on the NC Final Exams as their only source of student growth data, and it proposed several alternative measures, some of which appear in this brief. The COVID-19 pandemic arrived weeks later, and the alternatives were not implemented.[44]

The result is the situation described above: EVAAS remains, but it reaches only teachers of state-tested subjects.

C. National Board Certification (NBCT)

Unlike NCEES and EVAAS, National Board Certification is a voluntary professional credential administered by the National Board for Professional Teaching Standards (NBPTS), an independent nonprofit, not a state licensing body. To earn it, a teacher submits a portfolio that includes videos of their own classroom teaching, samples of student work, and written commentary analyzing their instructional choices, then separately completes a computer-based assessment of content knowledge. Most candidates spend a full year assembling their portfolio, and national pass rates run well below half on a given attempt. National Board Certification must be renewed every five years to remain current.

Although North Carolina has no role in administering or scoring this process, the state actively supports and rewards it: North Carolina helps cover the certification fee, and state law guarantees a 12% permanent salary supplement to any teacher who earns it. National Board Certification is an independent measure that can help accelerate a teacher's career pathway. It is a clear example of an independent, external evaluation already used in North Carolina's system that could be leveraged to identify teachers for leadership opportunities or career advancement.

However, National Board Certification was never designed to generate ongoing feedback. EVAAS and NCEES both produce a result that feeds back into a teacher's annual cycle: a growth score, a rating, a development plan. National Board Certification produces a five-year credential with no comparable feedback loop attached.

1.4 The Current System's Tools Are Inadequate for the Needs of North Carolina

The lack of reliable, scalable measures of teacher practice has concrete consequences for policy and practice. Without such measures, North Carolina cannot objectively or consistently perform several critical workforce functions.

First, the state cannot provide individualized professional growth support at scale. Current evaluation processes cannot reliably identify where support is most needed or which development strategies are most effective to improve instruction.[45]

Second, the state cannot match student learning needs with teacher strengths and expertise. Research shows that most of the variation in teacher effectiveness occurs within schools rather than between them, so assignments based on impressions or years of experience cannot ensure that students, particularly those with the greatest needs, have access to the most effective teachers.[46]

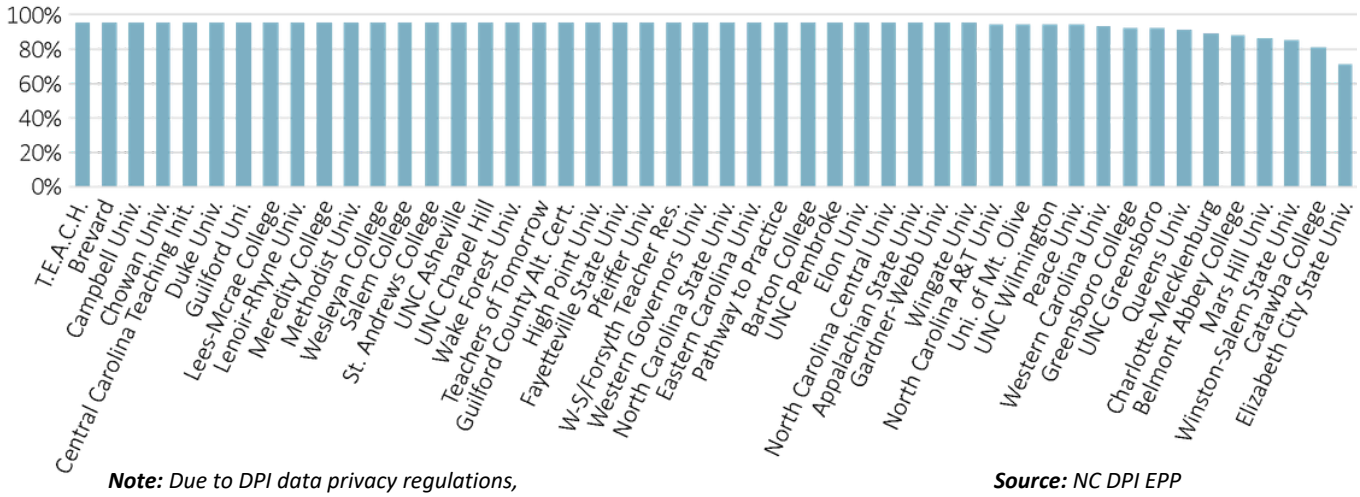
Third, without credible performance signals, the state struggles to strategically identify, reward, and retain high-performing teachers and school leaders. Strategic staffing initiatives such as Advanced Teaching Roles (ATR) are designed to develop these leaders, and an increased organizational capacity to identify and develop leadership candidates would improve teacher practice and student outcomes. Early results from Advanced Teaching Roles schools suggest these models can support stronger student outcomes. Systems that cannot reliably recognize strong performance are less able to deploy talent strategically or to retain top performers, particularly in a constrained fiscal environment.[47]

Finally, the state cannot identify educator preparation program (EPP) quality in a meaningful way. Research consistently shows that preparation quality cannot be inferred from program characteristics alone.[48] North Carolina's EPP Dashboard was designed to give programs feedback on their performance, but it lacks the teacher performance data needed to do so, which limits its value as a tool for continuous improvement.

The dashboard reports licensure examination pass rates, entering-candidate measures, employment and retention rates, and principal survey results.[49] But the two outcome measures drawn from the state's own evaluation system barely distinguish between programs. On the first, the share of a program's graduates' NCEES ratings at proficient or above, the dashboard's highest reporting category is "above 95%." In 2025, 33 of the 49 programs with reportable data landed in that top category, and only six fell below 90%.[50] The second measure, the share of completers whose principals judge them as effective as or more effective than other first-year teachers, ranges from 79% to 100%.[51]

In other words, almost every program looks strong on these limited metrics. The same ratings compression also limits what information is available to EPPs about their own graduates. The State Board returned the educator preparation accountability model to the Professional Educator Preparation and Standards Commission (PEPSC) for further review in August 2025, making this an opportune moment to strengthen the outcome side of the model.

EX 1.4 – Percent of Candidates Rated Proficient or Above on NCEES Evaluation, by EPP (2025)



Note: *Due to DPI data privacy regulations, any value over 95% is reported as >95%.*

Source: *NC DPI EPP Dashboard: Teaching Performance*

1.5 Performance Data Must Inform Workforce Decisions

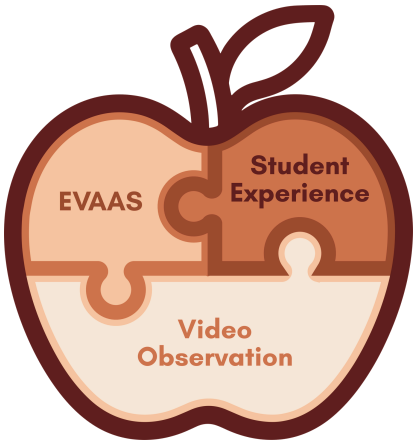
Nearly every major workforce decision in public education depends on understanding measures of teacher practice. School and district leaders, policymakers, EPPs, and other stakeholders need to know:

- What instructional knowledge and skills each teacher needs help developing;
- Which teachers on an Advanced Teaching Roles (ATR) team could be paired with specific student needs and which of their colleagues could observe and learn from them;
- Which teachers are ready to lead, mentor, or expand their instructional reach;
- When teachers are qualified for a full (Continuing Professional) license based on outcomes and could, perhaps, be accelerated to full teacher pay instead of following an experience- or input-based pay pathway;
- When to offer retention bonuses to experienced teachers;
- Which educator preparation programs are producing effective teachers in which subjects; and
- How to connect licensure progression to demonstrated performance over time, rather than to years of experience or credentials alone.

Today, none of these questions can be answered objectively or consistently at scale.

1.6 How a Multi-Measure System Can Strengthen North Carolina's Teaching Workforce

Decades of research make clear that teacher performance varies meaningfully, that this variation matters greatly for student outcomes, and that commonly used credentials and regulatory requirements are weak predictors of instructional impact.[52,53] Certification status can vary so widely across contexts that distinctions such as "certified" versus "not certified" do not reliably represent comparable preparation or performance.[54]



No single measure – EVAAS growth, observations, student experience data – captures teaching's full complexity or fits every teacher's context. The goal is not a perfect metric. It is a fair, coherent system built on multiple, complementary measures that together paint a clearer picture of student learning and teacher performance.

Coherence does not require a one-size-fits-all approach. North Carolina should still set the menu: which measures are valid, reliable, and available for a given teacher's assignment, and any minimum quality bar a measure must clear before it counts for anything. Within that menu, the purpose determines who chooses. For professional growth, the principal and the teacher's team leader should have access to every available measure, so that support is built on the fullest picture of the teacher's practice. Professional development is not punitive, and it works best when it draws on all the evidence available. For career-impacting steps such as licensure advancement, ATR selection, and pay, the choice of which validated measure counts should belong to the teacher. A teacher whose strength is classroom relationships might submit student experience data; one with strong instructional technique might submit video; one teaching a tested subject might rely on EVAAS growth data.

The state's job is to keep every option on the menu comparable in rigor and consistent with licensure law. A new measure starts in the professional growth application; it moves into evaluation or licensure only after an independent evaluation shows it meets the standard for that specific use. Section 3 develops what belongs on that menu and how the state keeps it rigorous.

Such choice within a rigorous menu responds directly to where past evaluation reforms went wrong.

1.7 Why Past Evaluation Reforms Failed

Calls for stronger measures of teacher practice are not new. In the 1990s and early 2000s, states across the country – including North Carolina – responded to concerns about instructional quality by expanding formal teacher evaluation. The core premise was sound and remains valid today: Teaching is complex professional work, and professional systems should include structured evaluation, feedback, and accountability.[55] Yet the tools and systems that were implemented often did not result in the type of valid and actionable feedback that was intended.

Most early evaluation processes relied almost exclusively on principal-led classroom observations, often requiring principals to evaluate every teacher multiple times per year.

The consequences were predictable:

- Evaluations provided limited actionable feedback for instructional improvement.
- High-performing teachers were rarely distinguished from their peers.
- Struggling teachers were difficult to identify early and support effectively.
- Evaluation became a compliance exercise rather than a professional growth tool.

As Hanushek notes, systems that rely on weak proxies or poorly designed evaluation tools can create the appearance of quality control without reliably identifying effective teaching.[56] Over time, these shortcomings led many educators and policymakers to conclude – incorrectly – that evaluation itself was the problem.

The lesson from this first generation of reforms was that early systems asked a single, ill-fitting tool to do too much. The resulting resource drain on principals actually reduced their ability to provide embedded supports for teachers.

Contemporary research points to a more effective approach: Multi-measure systems that combine evidence of instructional practice, student learning, and student experience are more reliable, fairer, and more useful than observation-only models for both professional improvement and employer decision-making.[57] By applying lessons from earlier reform efforts, the state can preserve the original goals of professionalism, growth, and accountability while modernizing the way it measures teacher practice.



SECTION 2. Understanding Teacher Performance and Why Measurement Matters

2.1 Defining Teacher Performance

Teacher performance – also referred to as teacher effectiveness – is often oversimplified or conflated with credentials, compliance, or observable effort. Decades of rigorous empirical research suggest a far more precise and consequential definition: Teacher performance is how well a teacher supports student learning, and one central source of evidence for that support is a teacher's consistent contribution to student learning growth, recognizing that learning itself is cumulative and influenced by multiple factors beyond the classroom.[58]

This definition reflects the reality that students arrive in classrooms with varied prior knowledge, skills, and supports shaped by family background, community context, and prior schooling. Effective teaching does not operate in a vacuum; rather, it alters the trajectory of student learning during the period a student is under a teacher's instruction. In this sense, performance is inherently relative, contextual, and dynamic, not absolute.

One way to account for these contextual differences is through "total teacher effects."[59] Total teacher effects represent the differences in student achievement growth that can be attributed to teachers after accounting for prior achievement and other observable influences. This concept underlies value-added and growth models, which seek to estimate the marginal instructional contribution of a teacher within a complex learning system. In North Carolina, the EVAAS metric is based on these concepts.

Importantly, defining performance in this way does not imply that teachers are solely responsible for student outcomes. Rather, it recognizes that while many factors shape learning, teachers are among the most powerful in-school influences on student achievement – as Hanushek puts it, "no other measured aspect of schools is nearly as important in determining student achievement."[60] The policy significance of this insight is profound: Since research shows that performance varies meaningfully among teachers, systems that treat all teachers as functionally interchangeable are misaligned with evidence.

Research consistently finds that:

- Differences in teacher performance are large relative to other school-based factors;
- Variation within schools often exceeds variation between schools; and
- Exposure to effective versus ineffective teachers has cumulative effects on student achievement over time.[61]

These findings challenge long-standing assumptions embedded in traditional licensure processes. They demonstrate that performance cannot be inferred reliably from inputs such as degrees, certification status, or years of experience beyond the early career stage. Instead, performance measures must be grounded in evidence of what students learn while in a teacher's classroom.

2.2 Why Measurement Is Hard – and Why Guardrails Matter

While research establishes how much teachers matter, it is equally clear that measuring their effects accurately is difficult. In a review of research on teacher quality, Hanushek and Steven Rivkin, an economist at the University of Illinois Chicago, identify three structural challenges: the nonrandom sorting of students across schools and classrooms, the limits of the assessments themselves, and year-to-year instability in estimates built on small numbers of students.[62] Section 3 examines each one in the context of the growth measures it most directly affects.

What matters for policy design is the conclusion it leads to; precision should not be mistaken for accuracy. A single metric can produce a precise-looking score that masks substantial uncertainty, and when that metric is used in isolation for career-impacting decisions, it risks unfairness and distortion.

Therefore, evaluation systems need guardrails that responsibly manage uncertainty. And decision-makers should clearly understand both the limitations of performance measures and how to use them alongside other evidence.[63]

Key guardrails include:

- Avoiding single-metric decision-making;
- Using multiple years of evidence where possible;
- Combining outcome-based measures with evidence of instructional practice; and
- Clearly distinguishing formative uses (growth and improvement) from summative uses (career-impacting decisions), and validating each measure for a specific summative use before it is applied to that use.

The goal of these guardrails is not to eliminate uncertainty – an impossible task – but to ensure that decisions are based on balanced, contextualized evidence rather than on any one imperfect indicator.

EX. 2.1 – Matching Measures to Purpose

Purpose	Key Requirements	Current Tools	Possible Next Steps
Professional feedback and growth	• Timely, specific, and actionable.	• NCEES observation • EVAAS reports where available	• Student experience data • Video-supported observation
Formal evaluation rating	• Reliable across raters and years • Transparent • Open to dispute	• NCEES observation	• Student experience data • Video-supported observation
Career advancement (ATR, mentoring, etc.)	• Multiple years of evidence • Valid for the role	• NCEES observation • EVAAS • National Board Certification (NBCT)	• Piloted measures, used alongside existing ones
License conversion and renewal	The highest bar: • Multiple years • Multiple measures • Consistent with licensure law	• EVAAS (2026 budget conversion route)	• Piloted measures for teachers without an EVAAS growth score
Educator preparation program (EPP) feedback	• Reliable in aggregate • Comparable across programs	• NCEES ratings • EVAAS • Retention rates • Principal survey	• Aggregated piloted measures

Defining Teacher Performance

Teacher performance is how well a teacher supports student learning. Evidence about it comes from three complementary sources: student learning growth, observations of teaching practice, and students' reports of their experience of instruction. These are not interchangeable measures; no single one captures the whole picture.



2.3 Why Improving Measurement Is Necessary

Given the complexity of teaching and the challenges of measurement, some systems respond by minimizing differentiation altogether. However, measurement processes that avoid real differences in performance fail to function as professional workforce systems.

Measurement that does not meaningfully differentiate performance cannot reliably:

- Identify instructional strengths and weaknesses;
- target professional development effectively;
- recognize and retain high-performing educators; or
- address chronic low performance in a fair and evidence-based manner.[64]

In such systems, measurement becomes largely symbolic. Most educators receive similar ratings regardless of effectiveness, limiting the system's usefulness for either improvement or accountability. This undermines trust in the evaluation system and weakens the professional signal of evaluation.

In contrast, multi-measure evaluation designs are widely viewed as more robust, fair, and useful. Jeff Schulz, Gunjan Sud, and Becky Crowe of Bellwether Education Partners highlight growing consensus around combining:

- Evidence of instructional practice (e.g., observations),
- evidence of student learning and growth, and
- stakeholder input, including student experience data.[65]

Each measure captures different dimensions of teaching. Observations provide insight into pedagogy and classroom interactions but are vulnerable to subjectivity and limited sampling. Growth measures anchor evaluation in student outcomes but are constrained by assessment quality and coverage. Student experience data surface aspects of instruction – such as classroom climate, expectations, and engagement – that may not be visible in brief observations but have demonstrated stability over time.[66]

When used together, these measures provide a more balanced perspective on a teacher's professional capabilities. Multi-measure systems can improve reliability, fairness, and educators' perceptions. They also produce richer information for professional growth, enabling more targeted coaching and support.

Improving evaluation is not about intensifying scrutiny or punishment. Rather, it is about aligning evaluations with their intended purposes: supporting improvement, recognizing excellence, and informing workforce decisions. Systems that lack credible performance measures cannot fulfill these purposes, regardless of how much effort is invested in evaluation processes.

Implications for State Policy

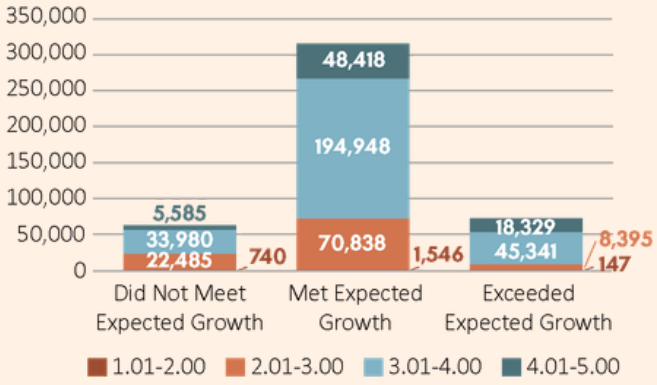
Taken together, the research makes several points clear. Teacher effectiveness is how well a teacher supports student learning. Evidence about effectiveness can come from several sources. One is a teacher's contribution to student learning growth, understood within a cumulative and complex learning system.[67] Others are observations of teaching practice and students' reports of their experience of instruction. These are complementary sources of evidence, not interchangeable measures; no single one captures the whole picture. Each is challenging to measure well and requires explicit guardrails. Together, multi-measure systems provide the most defensible foundation for strengthening the feedback teachers receive now and, as each measure is validated for a specific use, for evaluation and policy decisions.[68]

Improving how teacher performance is defined and measured is therefore not a technical exercise. It is a necessary step toward giving teachers better feedback on their practice now – and, if the evidence supports it, toward evaluation, licensure, advancement, and preparation-program reporting systems that treat teaching as a profession, support educator development, and ultimately improve outcomes for students.

NCEES vs. EVAAS – Do the Two Existing Systems Work Together?

When comparing EVAAS scores with NCEES average scores, large discrepancies emerge for the same teachers evaluated against both systems. Because of limitations (discussed above) with the populations of teachers receiving the NCEES evaluation, this data underrepresents more experienced teachers, those at better performing schools, and teachers previously receiving high NCEES evaluations. As EVAAS data is only available for tested subjects, only teachers in 3rd grade and above are represented in the data.

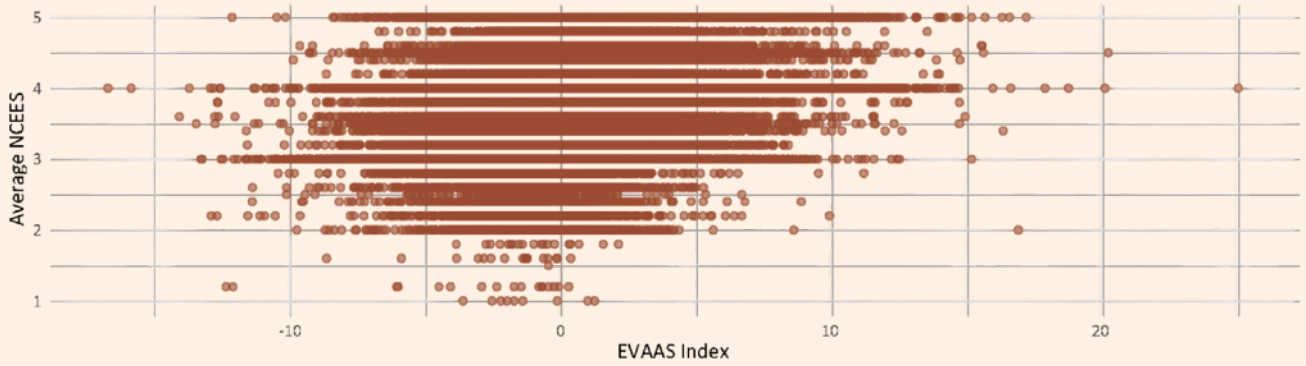
EX. 2.2 – Distribution of EVAAS Scores, by NCEES Average (2022-23, 2023-24, & 2024-25)



Source: *NC DPI (Data Request)*

If both systems were calibrated, teachers not meeting growth should receive overwhelmingly below average scores, while teachers exceeding growth should receive overwhelmingly above average scores. Instead, a majority of rated teachers in every EVAAS category received a NCEES score between 3.01 and 4. The NCEES evaluation generates little helpful information to teachers, communities, principals, and districts if there is a structural inability to meaningfully distinguish a teacher whose students are demonstrating high academic growth from a teacher whose students have not learned what was expected.

EX. 2.3 – Distribution of EVAAS Index, by NCEES Average (2022-23, 2023-24, & 2024-25)



Source: *NC DPI (Data Request)*

This data tells us:

- 9% of teachers in the dataset who did not meet expected growth received an NCEES rating of 4.01 or higher.
- 12% of teachers receiving an EVAAS rating of Exceeds Expected Growth did not receive an average NCEES score above 3.
- Just 25% of teachers exceeding growth expectations also scored between a 4.01 and 5 on the NCEES evaluation.
- Just 1% of teachers not meeting growth expectations received an NCEES score between 1 and 2.

Using the EVAAS Index instead of growth designations allows for a more detailed look at the data, which continues to support the same conclusion. In a dataset this large (over 450,000 observations), the two teacher rating systems have very little correlation. Any weak correlation between the two is most likely driven by overwhelmingly positive NCEES ratings.

The policy governing who received an NCEES evaluation means the teachers in this data are less likely to be the state's best teachers. If the NCEES evaluation gives the vast majority of teachers high ratings, how can the teachers actually doing the best job be identified and recognized?

SECTION 3. The Measurement Toolkit

The research points one direction: no single measure of teaching should be asked to stand alone, and in the research reviewed here, combinations of measures have outperformed the best of them individually. EVAAS growth, student experience data, and video observations each contribute unique information, provide a strong foundation for professional growth, and inform evaluation and licensure decisions, once validated.

North Carolina's challenge is not a lack of tools, but a need for intentional integration and clear use. The stakes of that integration are already concrete: state law now lets a positive EVAAS growth record substitute for the licensure examination, but only for the roughly two in five teachers who receive an EVAAS growth score at all (see Section 4). Building a coherent measurement toolkit is a critical step toward treating teaching as a profession, supporting educators, and improving outcomes for students.

A Movement to Adapt EVAAS

EVAAS North Carolina

North Carolina is actively working to improve how it measures student growth. The State Board of Education's accountability task force is reviewing how growth is calculated for school accountability, including alternatives such as student growth percentiles and growth-to-proficiency models.[69] Public reporting indicates that the review centers on the school-level calculation and that state leaders continue to view EVAAS as a useful measure for teachers.[70]

This brief is designed to complement that work. It uses EVAAS as its example of a teacher-level growth measure because it is the one North Carolina has today, and its recommendations do not depend on any particular growth model. Whatever model the state uses, growth data can show teachers of tested grades and subjects whether their students are growing, but not what to change in their practice, and it will not reach most teachers at all. The measures of practice this brief proposes serve both groups: they add actionable feedback for teachers who already have an EVAAS growth score and extend credible evidence of practice to those who do not.

3.1 How EVAAS Growth, Student Experience Data, and Video Observations Can Work Together

Measuring teacher performance is essential to a professional, improvement-oriented educator workforce system. The harder question is how. No single tool can capture the full complexity of teaching, nor should any one measure be asked to serve all purposes. Instead, research and practice increasingly point toward a measurement toolkit – a set of complementary tools that together provide a more accurate, fair, and actionable understanding of instructional effectiveness.

This section outlines how three core tools – educator value-added or growth measures (using EVAAS as an example), student experience data, and video-based classroom observations – can function together as an integrated system. Each tool provides distinct information, compensates for the limitations of the others, and supports different policy and practice goals. When used in combination, they offer a more robust foundation for professional growth, evaluation, and licensure decisions than any single measure alone.

3.2 How the Three Measures Meet the Bar for Measuring Professional Practice

The table below answers the standard set in Section 1: growth measures, student experience data, and independent observation each have real strengths against the reliable/valid/credible/scalable bar, and real gaps that the other two measures are best positioned to cover – which is the case for using all three together rather than any one alone.

EX. 3.1 – Comparing Three Different Methods of Measuring Professional Educator Practice in North Carolina

Criterion of Measure	EVAAS Growth	Student Experience Data	Independent Video Observation
Reliable across observers and occasions?	☑ Reliable when pooled across multiple years. ⊖ Less stable year-to-year for teachers with small numbers of tested students.	☑ Shows greater year-to-year stability than test-based measures in the MET research, and results pool the impressions of many students.	⊖ Depends on sustained rater calibration. ☑ A hybrid model pairing video with at least one in-person observation is more consistent than either alone.
Valid for the decisions it informs?	☑ The most direct evidence of contribution to student learning, and the state's only outcomes-based indicator. Covers about 40% of teachers, and estimates are sensitive to nonrandom student sorting.	☑ Predicts achievement gains in certain contexts when items focus on instructional practices (relationships, rigor, engagement, responsiveness) rather than popularity or personality.	Captures direct evidence in practice: ☑ Strong for teacher directives and lesson narrative. ⊖ Weaker for student activities, discussion, and group work.
Credible to the profession?	☑ Credible to teachers when reported across multiple years and paired with practice-based evidence. ⊖ Seen as reductive when a single year stands alone.	☑ More credible when tied explicitly to coaching and support rather than used only summatively.	☑ Most credible when teachers choose the lesson they submit, as increased professional agency improves responsiveness to reviewer feedback.
Practical to operate at scale?	☑ Already computed statewide at low marginal cost. ⊖ Structurally excludes the majority of teachers who do not teach a tested subject.	☑ Low-cost and scalable once a validated instrument exists but requires real investment in age-appropriate design and privacy protections.	☑ Builds on observation infrastructure districts already operate ⊖ The most resource-intensive of the three options: Camera and platform costs; Capacity to review footage; Ongoing rater calibration.

Measure 1: Educator Value-Added Growth/Impact on Student Learning

Purpose

Educator value-added or growth measures examine the impact of teachers, schools, and districts on students' learning – specifically, whether students made more or less progress than expected during a period of instruction. In North Carolina, EVAAS is the state's primary growth model and only statewide, outcome-based measure of teacher impact; it is reported to teachers rather than folded into their evaluation rating.

In a multi-measure system, growth data do not define teacher quality in isolation. They anchor the system to student outcomes and answer a question no other tool addresses directly: Are students learning more, less, or about as much as expected in a typical teacher's classroom? Because learning is cumulative and shaped by many influences,[71] growth measures do not claim sole causation; they estimate a teacher's marginal contribution to student achievement.

Design Principles and Implementation Considerations

Hanushek and Rivkin identify several reasons growth should be treated as one component of effectiveness, not the whole story:[72]

- Estimates are sensitive to nonrandom sorting of students. Even sophisticated controls cannot fully account for differences in student background, prior achievement, or classroom composition, so results should be read as estimates with uncertainty, not precise rankings.
- Measures depend on assessment quality and alignment. Tests may not rank teachers well when students start at very different levels or when assessments miss the full range of intended learning.[73]
- Estimates can be unstable year to year, especially for teachers with few tested students; multi-year or pooled data improve stability. In the MET study, student survey results were more stable year to year than test-based measures.[74]

Challenges

Growth measures apply to only about 40% of teachers. The majority – those in non-tested grades and subjects – receive no growth scores, leaving most of the workforce without outcome-based feedback. Growth data are a valuable part of the toolkit, but they cannot be the only tool in a coherent, fair system.

Policy Examples

States can embed these principles into policy by defining how growth evidence may be used. For licensure advancement or other career-impacting decisions, acceptable growth evidence could be defined as:

- Multiple years of growth data, when available, rather than single-year results; and/or
- Growth evidence combined with practice-based measures, such as classroom observations and student experience data.

This signals that growth matters while ensuring it is interpreted within a broader evidentiary framework and not asked to do more than it was designed to do.

Guardrails for Responsible Use

States should apply these principles consistently and exclude teachers in non-tested grades and subjects from growth-based determinations until comparable measures exist.



Measure 2: Student Experience Data and Instructional Quality from the "Direct Consumers of Instruction"

Scan QR Code to Learn More About Student Surveys

Purpose

Student perceptions offer a fundamentally different perspective on teaching: they capture instructional quality as experienced by the individuals most directly affected by daily classroom practice. Research increasingly recognizes student experience data as a low-cost, scalable, and informative measure of instructional effectiveness. Since the ultimate goal of effective teachers is for students' benefit, families can be encouraged that this tool directly reflects the primary purpose of a teaching professional – to ensure students leave their classroom with meaningful learning progress and the content knowledge they need to succeed.

This evaluation tool offers a direct feedback mechanism for students to inform teachers and school and district leaders of their experience in a teacher's classroom. When well designed, these tools elicit valuable information about the teacher's professional practice, while minimizing the identifiable student data collected. In addition to only collecting needed data, vendors chosen to conduct these surveys should have robust data protections in place to protect student and family privacy.

Student experience data is valuable because it predicts achievement gains in certain contexts and can be more reliable than observations or growth metrics.[75] Since students observe instruction continuously over time, they can provide insight into classroom dynamics that may be invisible in short administrator observations.

The Measures of Effective Teaching (MET) study played a central role in legitimizing student experience data as a valid aspect of teacher evaluation processes. According to Schulz et al., the MET study concluded that student perception data can be an effective and reliable tool for evaluating teaching.[76] Notably, student experience data demonstrated greater year-to-year stability than test-based measures, suggesting it captures enduring aspects of instructional practice rather than short-term fluctuations.

SAMPLE QUESTIONS FROM THE TRIPOD PROJECT SURVEY
(STUDENT ASKED TO RESPOND YES OR NO)

Elementary level: "In this class, we learn a lot almost every day."

Upper elementary level: "My teacher wants me to explain my answers—why I think what I think."

Secondary level: "My teacher takes the time to summarize what we learn each day."

Source: *Schulz, Sud, and Crowe (2014) Lessons from the Field*

Strengths

Student experience data is most effective when the questions focus on instructional practices correlated with learning, rather than popularity or personality. Schulz et al. emphasize survey domains such as:

- Teacher-student relationships;
- Academic rigor and expectations;
- Student engagement; and
- Teacher responsiveness to student needs.[77]

These dimensions align closely with research on effective instruction and complement both growth data and observation evidence. Student experience data can surface patterns – such as inconsistent expectations or uneven engagement – that might not be visible through other measures. Establishing this data collection on a wider scale also offers an additional feedback loop for students to communicate their experience in a teacher's classroom with clear guiding questions, anonymity, and oversight.

Dual-Use: Summative and Developmental

Research supports using student experience data in two complementary ways:[78]

- Summative evaluation, when student experience data is aligned with other measures and weighted appropriately. Used in this way, it contributes to a broader teacher performance picture without dominating it.
- Professional development, when student experience data is explicitly tied to coaching and support. In this role, it provides actionable feedback that teachers can use to refine practice.

Crucially, these uses should be clearly distinguished in policy and practice. When student experience data is linked to coaching and improvement, teachers are more likely to view it as credible and useful. When student experience data is used summatively, its role should be transparent and limited.

Challenges

One emerging challenge involves parent and community perceptions. On the one hand, parents increasingly expect transparency and assurance that teachers are being evaluated meaningfully. On the other hand, concerns about student and parent privacy – sometimes formalized through policies such as parents' bills of rights – can create tension around survey use. Transparent communication is essential to maintaining public trust while preserving the instructional value of student experience data.

Addressing this challenge requires good research practices in the collection of student data and effective messaging to families that these practices:

- Do not collect personal or sensitive information;
- Focus on instructional practices rather than individuals; and
- Are one part of a broader system designed to support quality teaching.

Guardrails for Responsible Use

Research cautions against over-weighting student experience data. Student experience data should remain one measure among many, aligned with observations, growth data, and professional learning supports. Over-reliance on student experience data risks distorting incentives and undermining trust in the evaluation system.[79]

Effective guardrails include:

- Age-appropriate tools and questions;
- Robust privacy protections in the questions that are asked and how the data is stored;
- Limiting the weight of student experience data in decisions that determine a teacher's rating or licensure status;
- Aggregating results over time;
- Pairing student experience data with coaching conversations; and
- Ensuring survey questions are research-based and instructionally focused.

Measure 3: Independent Classroom Observations

Purpose

Classroom observations remain a central component of measuring teacher performance, providing direct evidence of instructional practice. However, traditional in-person observations are often limited by time, capacity, and sampling. They also can lead to a disconnect between what the observer experiences and a teacher's perceptions of events during that lesson. Video-based observation models – such as those tested in the Best Foot Forward study – offer a way to strengthen observation quality, increase feasibility, and improve professional learning.

Scan QR Code to Learn More About the Best Foot Forward Study

Research on Best Foot Forward shows that when teachers record, review, and submit their own lessons, observation conversations improve markedly: Administrators were 28 percentage points more likely to report that teachers were never or rarely defensive, and teachers were 16 points more likely to identify a specific change in their practice resulting from the conversation and 18 points more likely to call it productive.[80] Because a video platform lets observers pause, rewind, and tag comments to specific moments, teachers and observers work from the same record, which makes feedback more specific and actionable. Where trained external reviewers score part of the observations against a consistent rubric, principals need not be the primary reviewer for every evaluation in their building.

A Cautionary Tale: Shelby High School Loses Future Teacher of the Year to Another School Because of the Praxis Exam

Chad Beam overcame dyslexia to become a beloved literature teacher at Shelby High School in Shelby, North Carolina. Even though he was eventually honored with a regional teacher of the year award, Mr. Beam almost left the teaching profession because he could not pass the Praxis exam due to his testing anxiety. Scan the QR code to watch a short video (originally produced for BEST NC's 2022 North Carolina Education Innovation Lab) to learn why he believes independent classroom observation would have been a better indicator of his teaching practice.

Advantages of video observation (continued on next page)

Independent, video-based observation offers several distinct benefits:

- Professional agency: Allowing teachers full control over video submission increases buy-in without compromising accountability.
- Expanded sampling: Teachers can submit multiple lessons across the year, reducing reliance on single snapshots.
- Improved feedback quality: Observers can cite specific moments in instruction, increasing precision.
- Consistency across schools and districts: Virtual submission offers additional opportunities for standardized feedback across schools and districts. Trained independent reviewers using a common rubric can make ratings more comparable across schools and districts.
- Reach and reallocated time: Video-based observation can substitute for a portion of the required in-person observations, and can extend a substantive, standards-based observation to the 45% of teachers on an abbreviated cycle who may currently be observed twice a year and rated on just two of five standards.
- Better use of principal time: In the Best Foot Forward study, administrators were able to prioritize other responsibilities during instructional hours, while fulfilling observation duties at less busy times of the day or week.[81]

Challenges

- Data security and consent: Video of minors requires media release, FERPA-compliant storage, and a secure portal accessible to the teacher and school administrators.
- Capacity: Building central capacity to review footage across a district is itself a significant investment, and rater calibration (training reviewers to score lessons consistently with one another) must be maintained over time.
- Context: Administrators rated video effective at capturing teacher directives and lesson narratives but poor at capturing student activities, class discussions, group work, and board work – the clearest argument for retaining at least one in-person observation.

Guardrails for Responsible Use

Video observation should complement, not replace, in-person observation. Research cautions that video may be less effective for understanding broader classroom context or school culture. Hybrid models – combining at least one in-person observation with multiple video submissions – balance context with evidence and are more likely to meet standards for fairness and credibility.

A credible system also needs a clear process for disputing a determination. Teachers and evaluators, including principals, should have a defined avenue to contest a rating they believe is inaccurate or unfairly derived before it becomes final, supported by documentation standards and an independent review step. Building such guardrails in from the start protects trust in the system for both teachers and the school leaders responsible for evaluating them.

3.3 How the Toolkit Works Together

Multi-measure systems are more robust, fair, and actionable than single-metric approaches.[82] By triangulating evidence, states can reduce bias, improve reliability, and generate information that supports both improvement and decision-making.

- Growth measures anchor evaluation in student outcomes.
- Student experience data capture lived instructional experience.
- Classroom observations, both video and in-person, provide detailed evidence of practice.

Together these tools provide multiple measures of teacher practice, forming a more coherent system where every teacher has access to at least two measures of practice that can help identify opportunities for professional growth.

A Menu of Options

Whether these tools are used for a teacher's accelerated licensure and pay, or for extended professional opportunities such as Advanced Teaching Roles or paid mentoring, a teacher should be able to choose the tool that they feel best represents their performance from the validated measures for that use. Empowering teachers in this way will also help them view the system as additive, not punitive.



3.4 Testing the New Measures Before Broader Use

Pilot Design

- **Participation:** a limited number of districts that opt in, varied by size, region, and school performance, with teacher participation voluntary within those districts.
- **Duration:** at least two full school years, so results can be compared across years.
- **Measures:** a research-aligned, age-appropriate student experience instrument with privacy protections; a video-supported observation option paired with at least one in-person observation; and existing EVAAS data where available.
- **Use During the Pilot:** professional feedback and coaching only. Piloted measures do not count toward a formal rating, licensure step, or pay decision during the pilot.
- **Supports:** funding for instruments, video equipment and platform, rater training and calibration, and communication with families.

Independent Evaluation and Decision Criteria

An independent evaluator should assess each piloted measure against the standards set out in Section 1:

- **Reliability:** consistency across raters, occasions, and years.
- **Validity:** relationship to student learning and to other evidence of practice.
- **Fairness:** whether results differ systematically by teacher assignment or student population in ways that do not reflect practice.
- **Credibility:** whether teachers and principals find the results accurate and useful.
- **Feasibility:** the time, cost, and capacity required to operate at scale.

The evaluation should reach one of three findings for each measure: retained for professional growth only; ready to be considered for a specific career-impacting use, with each use authorized separately; or refined and tested further. The findings should be public.

Legislative Interest: A 2026 Senate Proposal

Legislators have already explored a pilot much like the one described above. In 2026, Section 1.10 of Senate Bill 1006 proposed a Teacher Evaluation Pilot Program in which the State Board would select at least three evidence-based, replicable evaluation tools, drawing on evidence such as student outcomes, student surveys, or independent evaluations. Districts would have opted in over several school years, with results reported to the Joint Legislative Education Oversight Committee.[83]

The proposal named four purposes that match this brief's recommendations: improved professional development, opportunities for career advancement, insight for educator preparation program accountability, and licensure renewal. Whatever legislative vehicle carries this work forward, the design questions in this section apply.



SECTION 4. Applying Validated Measures to Licensure, Advancement, and Preparation

This section describes how measures that meet the standard for a given use – those North Carolina already has and those that clear the pilot evaluation – can potentially inform licensure, advancement, and retention. North Carolina can move beyond fragmented evaluation tools and input-based licensure rules toward a coherent system that links evidence of teaching quality to professional growth, advancement, and retention. Doing so requires more than selecting the "right" measures. It requires intentional system design – one that integrates multiple sources of evidence, establishes clear guardrails for fairness and validity, and is implemented with careful attention to change management and educator trust. How results are reported to teachers and administrators also shapes whether they lead to action.

Research consistently finds that no single measure can capture the full complexity of teaching, and that systems relying on a single indicator are vulnerable to error, bias, and loss of trust. In contrast, systems that use multiple, complementary measures are more likely to differentiate performance fairly and produce information that is useful for both improvement and decision-making.[84] Measures should also be designed so that results cannot easily be inflated.

Systems relying on a single measure or indicator of teacher influence are more vulnerable to error, bias, and loss of trust.



North Carolina should set the menu of valid, reliable measures available for a given assignment and the minimum threshold each must clear, consistent with state licensure law. For professional growth, principals and team leaders draw on every available measure to support the teacher. For career-impacting steps such as licensure advancement, ATR selection, and pay, the teacher chooses which validated measure or measures count. This keeps the state responsible for rigor and comparability while giving teachers a voice in the evidence that shapes their careers.

With these new tools, North Carolina can:

- Target and improve professional supports for teachers at every stage of their career;
- Leverage data to align the talents of teachers with the needs of students;
- Utilize performance data to identify, select, and support teachers in leadership, mentorship, or extended reach roles;
- Reimagine a licensure-by-performance model to accelerate licensure, and perhaps compensation, for effective teachers;
- Implement a teacher retention bonus strategy that addresses concerns about veteran teacher retention beyond only time in the profession; and
- Utilize performance measures to help EPPs improve practice.

4.1 Licensure: Reimagining a Licensure-by-Performance Model

The Pathways to Excellence (PTE) Proposal

The Pathways to Excellence for Teaching Professionals framework – developed by the Education Human Capital Roundtable, refined by PEPSC through 2022, and unanimously recommended by PEPSC to the State Board of Education in November 2022 – could help operationalize this approach. Rather than creating a new accountability system, PTE leverages existing and emerging performance measures and defines what evidence is required at each licensure stage, how long it must be demonstrated, and what supports accompany advancement.



North Carolina's current licensure structure reflects an earlier era of policy, when time, credits, and credentials were assumed to be reasonable proxies for effectiveness. Research now shows these proxies are weak predictors of instructional impact.[85] The result is a system that advances educators without consistent evidence that students are learning more.

That is not an argument against entry requirements. Initial licensure establishes that a new teacher has the foundational knowledge and skills to begin teaching; it is not, and should not be presented as, a predictor of later effectiveness. The question is what evidence should govern the steps after entry: conversion to a continuing license, renewal, and advancement.

Already in Law: Performance in Place of an Exam

North Carolina's 2026 state budget requires the State Board to convert a limited license to a continuing professional license for any teacher who has EVAAS growth data and a positive growth score in two of the three most recent years, waiving the examination requirement entirely for this group of educators.

This route is open only to teachers who hold a limited license and have an EVAAS growth score – have an EVAAS growth score – a small subset of the roughly two in five teachers who receive one. In this case only, the General Assembly has established that measured performance can stand in place of an examination.[86]

A simplified, outcomes-focused licensure model responds directly to this gap by proposing:

- Fewer, clearer licensure levels;
- Advancement based on demonstrated effectiveness, not seat time or accumulated coursework; and
- Alignment between licensure, evaluation, and professional growth so systems reinforce one another.

Under this model, licensure progression signals instructional performance and readiness for expanded responsibility, not just longevity. This benefits early-career teachers who demonstrate competencies ahead of an average timeline and veteran teachers who consistently demonstrate strong practice.

Allowing veteran teachers to use one of these measures for license renewal saves the time and cost of completing Continuing Education Units (CEUs), which may be unrelated to teaching practice, and ties licensure to demonstrated skill. Over time, the state could also consider whether validated measures should inform retention incentives at renewal, so veteran teachers are honored for both skill and experience, not time in the profession alone.

4.2 Advancement: Expanding the Reach of Effective Teachers

Advancement in teaching has traditionally meant leaving the classroom. The route upward has run through administration – coach, specialist, school leader – which removes the strongest teachers from the students who benefit most from them. Strategic staffing models such as Advanced Teaching Roles (ATR) create advancement opportunities that keep excellent teachers teaching. Some roles extend a teacher's reach to more students or to a team's content; others add coaching and mentoring of colleagues to a continuing classroom assignment.

Selecting teachers for these roles depends on credible evidence of practice. Today that evidence comes from NCEES ratings, EVAAS scores where they exist, and National Board Certification, three sources that do not connect and, for most teachers, are incomplete (see Section 1.3). Additional measures of teacher practice would widen the pool of potential Collaborative Impact Leaders to teachers whose strengths current tools do not capture, including the three in five teachers without an EVAAS growth score.

The same evidence serves the team once it is formed. It helps principals design ATR teams and match students to the teachers best positioned to serve them, gives Collaborative Impact Leaders specific information to guide their mentoring, and identifies teachers who would benefit from additional support. Paid mentoring roles, including TeachReadyNC supervisors, rest on the same judgment: which teachers are ready to help develop others.

4.3 Preparation: Closing the Feedback Loop for Educator Preparation Programs (EPPs)

Educator preparation programs (EPPs) improve only as fast as they learn how their graduates perform. North Carolina's EPP report card describes candidates on the way in and credentials on the way out, and the two outcome measures it draws from the state's evaluation system barely distinguish between programs (see Section 1.4). Its only measure of classroom impact, EVAAS, reaches the minority of graduates who teach tested grades and subjects.

Extending validated measures of teacher practice to program reporting would let programs see how their graduates perform by subject taught and by school setting, including graduates who teach non-tested subjects such as art or the early grades. Reported in aggregate, these measures would show which practices and pathways are producing effective teachers and where preparation needs to change. They would also give the EPP evaluations established in the 2026 budget actionable evidence on how graduates are prepared for the profession, at a time when PEPSC is reviewing the educator preparation accountability model.



SECTION 5. Policy Recommendations

Sequencing BEST NC's Policy Recommendations

1. Establish the pilot in law and recruit participating districts.
2. Run the pilot for at least two school years, with piloted measures used for professional feedback only.
3. Complete the independent evaluation and publish its findings.
4. Authorize specific uses measure by measure – for example, identifying Advanced Teaching Roles Collaborative Impact Leaders and TeachReadyNC supervisors, EPP reporting, and license renewal – only where the evaluation supports that use.
5. Recommendations B and C do not wait on this sequence where they rely on measures North Carolina already uses, such as EVAAS and National Board Certification.

Recommendation A: Establish and Fund a "Multiple Measures of Teacher Practice" Pilot

Establish a multiple measures pilot program, including practice evidence, student experience evidence, and student learning evidence, while allowing local flexibility in implementation within a common statewide framework that preserves validity and replicability.[87] Select an independent entity to evaluate the pilot using a rigorous research design (see Section 3, "Testing the New Measures Before Broader Use"). During the pilot, results from every piloted measure should be available to the teacher, the principal, and the teacher's team leader, so they can be used together for coaching and support. Piloted measures should not count toward a formal rating, licensure step, or pay decision until the independent evaluation finds they meet the standard for that use.

Allow districts to opt in to using student experience data and practice evidence evaluation tools. This program would provide resources for measurement tools as well as teacher engagement incentives. Successful plans would also include a communications strategy to inform families and communities of the implementation of these tools. In addition to the existing EVAAS measures, pilot these measures of teaching practice and student experience:

1. **Pilot student experience data collection with explicit coaching linkage, with statewide scaling considered if the pilot evaluation supports it.**

   Adopt research-aligned surveys focused on instructional practices correlated with learning – relationships, rigor, engagement, and responsiveness – and use them as stable, complementary evidence, not standalone determinative metrics.[88]

2. **Expand observation capacity through a Best Foot Forward–style video option.**

   Authorize teacher-collected video for a portion of observations, paired with at least one in-person observation, to improve feedback quality, reduce defensiveness, and enable time-shifting for administrators.

These observations can also be used to identify one or both of the following:

- ☑ **Classroom teachers who may benefit from additional instructional supports.**
- ☑ **Classroom teachers who may be qualified to provide additional instructional supports to other classroom teachers as mentors for apprentices, ATR Collaborative Impact Leaders, or other coaching structures.**

The pilot design and the independent evaluation's decision criteria are set out in Section 3, under "Testing the new measures before broader use."

Recommendation B:
Tie Career Advancement and Licensure to Demonstrated Performance

Tie career advancement and licensure progression after entry to demonstrated performance rather than time in role. Section 4.2 describes how validated measures of practice can widen the pool of teachers identified for Advanced Teaching Roles, paid mentoring, and TeachReadyNC supervisor roles, and support the teams those teachers lead. As each measure clears the independent evaluation for that use, districts should be able to use it to select and support teachers in these roles.

Some teachers already in the classroom demonstrate strong practice but struggle to pass a licensure examination. North Carolina has already accepted a performance-based alternative for one group: the 2026 budget lets a positive EVAAS growth record substitute for the examination when a limited license is converted to a continuing license (see Section 4). Extending that option to teachers without an EVAAS growth score, using measures that clear the independent evaluation, would apply the same principle to the three in five teachers the current route cannot reach.

A Cautionary Tale: Ms. Art Leaves Teaching – When One Exam Outweighs the Evidence

Ashley Thompson – "Ms. Art" to her students – taught art at Howard Hall Elementary in Cumberland County, the same school she attended as a child. Her principal called her "a top tier teacher" and said simply, "These kids love her." But she could not pass the Praxis art content exam required for a continuing license, missing the passing score narrowly on repeated attempts. The school wanted to keep her and, for a while, she was able to keep teaching at a substitute's pay.[89] A state budget provision later gave teachers in her position a short reprieve, but she has since left the classroom.[90]

Her story exemplifies the question of whether a single exam should outweigh years of evidence from the students, colleagues, and principal who saw her teach and valued her role in the school. As she put it, "Beyond the test, teachers can prove their worth with children."[91] And, because she taught art, she never had an EVAAS score, so she did not have the evidence of success that a student survey could have given her.

Scan the QR code to read more about Ms. Art's story. Article originally published on March 13, 2023, by WUNC News.

Additionally, consider creating a simplified licensure model that draws on the Pathways to Excellence proposal described in Section 4.1 to set evidence thresholds, timelines, and supports for licensure advancement.

Recommendation C:
Close the Educator Preparation Feedback Loop

Close the feedback loop between educator preparation programs and the classroom outcomes of the teachers they produce. Section 4.3 describes the gap in the current EPP report card and how validated measures of practice would fill it.

As measures are validated for program reporting, feed them back into the EPP Dashboard so preparation programs can see which practices and pathways are producing effective teachers, by subject.



Conclusion

Doctors get regular, multi-source feedback on their clinical performance. Attorneys build their careers on demonstrated results. North Carolina's teachers, the professionals responsible for the state's most consequential long-term investment, largely do not have that option: about three in five have no individual measure of their impact on student learning, and at the high school level, more than four in five do not.

This brief has argued for a specific way to close that gap. No single measure – EVAAS growth, observations, or student experience data – can carry the full weight of describing what a teacher does and how well they do it. But a coherent system that draws on more than one, with teachers choosing among validated measures for career-impacting steps, can. That is what the last two decades of research on teacher effectiveness, professional evaluation, and multi-measure system design consistently point toward, and North Carolina legislators have shown interest in the same approach: a 2026 Senate proposal included a teacher evaluation pilot built on many of the same elements.[92]

Three recommendations turn that principle into practice: establish and fund a multiple-measures pilot that pairs student experience data and expanded, video-enabled observation capacity with existing EVAAS growth data; tie career advancement and licensure progression to demonstrated performance rather than time in role; and close the feedback loop between educator preparation programs and the classroom outcomes of the teachers they produce. None of the three depends on the others to deliver value, but together they give North Carolina a system where evidence flows in both directions – to the teacher, and back to the state.

References

1. Tomberlin, T. R. (2026, June 22). Teacher preparation, licensure, & retention [Presentation to the Blue Ribbon Commission on Public Education], Slide 7. North Carolina Department of Public Instruction.
2. Bass, D. (2026, June 24). NC schools commission turns to teacher licensing, weighs scrapping exams. Carolina Journal. (Quoting Thomas R. Tomberlin, NCDPI: "fewer than 40% of teachers receive an individual EVAAS growth score – the state's value-added measure of teacher effectiveness – and under 18% at the high-school level.")
3. Aguinis, H. (2013). Performance management (3rd ed.). Pearson.
4. Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112, pp. 86, 96. https://doi.org/10.3102/003465430298487
5. Little, O., Goe, L., & Bell, C. (2009). A practical guide to evaluating teacher effectiveness. National Comprehensive Center for Teacher Quality, p. 15.
6. London, M., & Smither, J. W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. Personnel Psychology, 48(4), 803–839. https://doi.org/10.1111/j.1744-6570.1995.tb01782.x
7. Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. The New Teacher Project, pp. 13, 29.
8. Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness (2nd ed.). The New Teacher Project, p. 13.
9. Little, O., Goe, L., & Bell, C. (2009). A practical guide to evaluating teacher effectiveness. National Comprehensive Center for Teacher Quality, pp. 8, 15, 23
10. Little, O., Goe, L., & Bell, C. (2009). A practical guide to evaluating teacher effectiveness. National Comprehensive Center for Teacher Quality, Appendix A, pp. 23–24.
11. Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness (2nd ed.). The New Teacher Project, pp. 4–6.
12. Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness, p. 25; Weisberg et al., The widget effect, p. 29.
13. Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. A. Hanushek & F. Welch (Eds.), Handbook of the economics of education (Vol. 2, pp. 1051–1078). Elsevier. https://doi.org/10.1016/S1574-0692(06)02018-6
14. Hanushek, E. A. (2011). The economic value of higher teacher quality. Economics of Education Review, 30(3), 466–479. https://doi.org/10.1016/j.econedurev.2010.12.006
15. Hanushek & Rivkin, Teacher quality, p. 1068.
16. Hanushek, The economic value of higher teacher quality.
17. Bastian, K. C. (2019). A degree above? The value-added estimates and evaluation ratings of teachers with a graduate degree. Education Finance and Policy, 14(4), 652–678. https://doi.org/10.1162/edfp_a_00261
18. Libman, Z. (2009). Teacher licensing examinations–True progress or an illusion? Studies in Educational Evaluation, 35(1), 7–15. https://doi.org/10.1016/j.stueduc.2009.01.003

19 North Carolina Department of Public Instruction. (n.d.). Renew or update your professional educator's license. https://www.dpi.nc.gov/educators/educators-licensure/renew-or-update-your-professional-educators-license

20. Hanushek, The economic value of higher teacher quality.
21. Tomberlin, Teacher preparation, licensure, & retention, Slide 7.
22. North Carolina School Report Cards. (2026). Educator Effectiveness. https://ncreports.ondemand.sas.com/src/state?year=2025
23. N.C. State Board of Education, Policy EVAL-004: North Carolina Educator Evaluation Process (effective July 9, 2026), §2(a) (comprehensive cycle) and §3(e) (observation minimums and tenure-based cycle assignment).
24. N.C. State Board of Education, Policy EVAL-004: North Carolina Educator Evaluation Process (effective July 9, 2026), §§2(b), 3(e).

25. N.C. State Board of Education, Policy EVAL-004: North Carolina Educator Evaluation Process (effective July 9, 2026), §2(c) and Table 2 (Abbreviated Evaluation Cycle Standards: Teacher – Standards 1 and 4).
26. North Carolina Department of Public Instruction. (2026, June 30). NCEES 2025–2026 end of year process completion notice [Internal communication].
27. North Carolina Department of Public Instruction, NCEES 2025–2026 end of year process completion notice.
28. Tomberlin, Teacher preparation, licensure, & retention, Slide 7.
29. Little, O., Goe, L., & Bell, C. (2009). A practical guide to evaluating teacher effectiveness. National Comprehensive Center for Teacher Quality, pp. 7–9.
30. Little, O., Goe, L., & Bell, C. (2009). A practical guide to evaluating teacher effectiveness. National Comprehensive Center for Teacher Quality, pp. 7, 23.
31. Little, O., Goe, L., & Bell, C. (2009). A practical guide to evaluating teacher effectiveness. National Comprehensive Center for Teacher Quality, pp. 8–9.
32. Hunter, S. B., & Springer, M. G. (2022). Critical Feedback Characteristics, Teacher Human Capital, and Early-Career Teacher Performance: A Mixed-Methods Analysis. Educational Evaluation and Policy Analysis, 44(3), 380-403. https://doi.org/10.3102/01623737211062913
33. Best Foot Forward Project. (2015). Best foot forward: A toolkit for fast-forwarding classroom observations using video. Center for Education Policy Research, Harvard University, pp. 4, 6.
34. Kane, T. J., Gehlbach, H., Greenberg, M., Quinn, D., & Thal, D. (2015). The Best Foot Forward Project: Substituting teacher-collected video for in-person classroom observations – First year implementation report. Center for Education Policy Research, Harvard University, p. 15.
35. Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness, pp. 7, 23.
36. Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness (2nd ed.). The New Teacher Project, pp. 6, 10–11



37. Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness (2nd ed.). The New Teacher Project, pp. 12, 14
38. N.C. State Board of Education, Policy EVAL-006: Evaluation Standards and Criteria: Teachers (adopted Apr. 7, 2016).
39. North Carolina Department of Public Instruction, Teacher Legislated Performance Bonuses – LEAs and Charter Schools FAQ (bonuses for 3rd Grade Reading, 4th/5th Grade Reading, and 4th–8th Grade Math awarded to teachers ranked in the top 25 percent by EVAAS growth score, statewide or within their LEA/school).
40. Tomberlin, Teacher preparation, licensure, & retention, Slide 7.
41. Bass, NC schools commission turns to teacher licensing.
42. N.C. Gen. Stat. § 115C-333; N.C. State Board of Education, Policy EVAL-006: Evaluation Standards and Criteria: Teachers (adopted Apr. 7, 2016).
43. North Carolina General Assembly. (2017). Session Law 2017-57 (Senate Bill 257), § 7.23E; North Carolina General Assembly. (2019). Session Law 2019-212 (Senate Bill 621), § 1.
44. N.C. State Board of Education & N.C. Department of Public Instruction. (2020, March). Alternative for measuring teacher effectiveness: Report to the North Carolina General Assembly, S.L. 2019-212 (SB 621), § 1. https://webservices.ncleg.gov/ViewDocSiteFile/15525; EdNC. (2020, March 5). NC Final Exam elimination poses problem for State Board. https://www.ednc.org/nc-final-exam-elimination-poses-problem-for-state-board/
45. Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness, p. 15.
46. Hanushek & Rivkin, Teacher quality, p. 1068.
47. Weisberg et al., The widget effect, p. 13; Hanushek, The economic value of higher teacher quality.
48. Hanushek, The economic value of higher teacher quality.
49. Required under G.S. 115C-269.35(b) and G.S. 115C-269.50.
50. North Carolina Department of Public Instruction. (2026). EPP teaching performance: Ratings on NC Educator Effectiveness System (NCEES) standards, 2025 [Data dashboard]. Retrieved October 7, 2026, from https://bi.nc.gov/t/DPI-EducatorRecruitmentandSupport/views/EPPTeachingPerformance/NCEESCharts
51. Tomberlin, T. R. (2026, June 22). Teacher preparation, licensure, & retention [Presentation to the Blue Ribbon Commission on Public Education], Slide 16 (Data from the EPP Dashboard, 2019–2025). North Carolina Department of Public Instruction.
52. Hanushek & Rivkin, Teacher quality.
53. Hanushek, The economic value of higher teacher quality.
54. Hanushek & Rivkin, Teacher quality, p. 1065.
55. Weisberg et al., The widget effect, p. 7.
56. Hanushek, The economic value of higher teacher quality.
57. Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness, pp. 15, 17; Schulz, J., Sud, G., & Crowe, B. (2014, May). Lessons from the field: The role of student surveys in teacher evaluation and development. Bellwether Education Partners, p. 1.
58. Hanushek & Rivkin, Teacher quality, pp. 1058, 1065–1066.
59. Hanushek & Rivkin, Teacher quality, p. 1065.
60. Hanushek, The economic value of higher teacher quality.
61. Hanushek & Rivkin, Teacher quality.
62. Hanushek & Rivkin, Teacher quality, pp. 1066–1067.
63. Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness, pp. 5–6, 17.
64. Weisberg et al., The widget effect, pp. 6, 13–17.
65. Schulz, Sud, & Crowe, Lessons from the field, p. 1.
66. Schulz, Sud, & Crowe, Lessons from the field, p. 4.
67. Hanushek & Rivkin, Teacher quality.
68. Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness, pp. 15, 17.
69. Bass, D. N. (2026, June 23). NC public schools task force considers moving on from EVAAS growth metric. Carolina Journal. https://www.carolinajournal.com/nc-public-schools-task-force-considers-moving-on-from-evaas/; Walkenhorst, E. (2026, September 13). NC education leaders look to recalculate key school-performance metric. WRAL. https://www.wral.com/news/education/nc-education-leaders-look-to-recalculate-key-school-performance-metric-september-2026/



70. Walkenhorst, NC education leaders look to recalculate key school-performance metric."
71. Hanushek & Rivkin, Teacher quality.
72. Hanushek & Rivkin, Teacher quality; Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness, pp. 5–6.
73. Hanushek & Rivkin, Teacher quality, p. 1066.
74. Schulz, Sud, & Crowe, Lessons from the field.
75. Schulz, Sud, & Crowe, Lessons from the field, p. 2.
76. Schulz, Sud, & Crowe, Lessons from the field, p. 4.
77. Schulz, Sud, & Crowe, Lessons from the field, p. 2.
78. Schulz, Sud, & Crowe, Lessons from the field, p. 1.
79. Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness, p. 14; Schulz, Sud, & Crowe, Lessons from the field, p. 10.
80. Kane et al., The Best Foot Forward Project, pp. 17–18, Tables 4–5.
81. Kane et al., The Best Foot Forward Project, p. 19, Table 10.
82. Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness.
83. North Carolina General Assembly. (2026). Senate Bill 1006, Second Edition, § 1.10(a)–(e).
84. Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness.
85. Hanushek, The economic value of higher teacher quality.
86. North Carolina General Assembly. (2026). Session Law 2026-41 (Senate Bill 257), § 7.10.
87. Schulz, Sud, & Crowe, Lessons from the field.
88. Schulz, Sud, & Crowe, Lessons from the field; Little, Goe, & Bell, A practical guide to evaluating teacher effectiveness, p. 14.
89. Schlemmer, L. (2023, March 13). 'Ms. Art' wants to keep teaching, but one exam stands in her way. WUNC/WHQR. https://www.whqr.org/2023-03-13/ms-art-teaching-exam-stands-cumberland-thompson-fayetteville
90. Schlemmer, L. (2023, December 15). North Carolina budget lets teachers who struggle with licensing exams keep their jobs. WUNC/BPR. https://www.bpr.org/2023-12-15/north-carolina-budget-teachers-struggle-licensing-exams-jobs-cumberland
91. Schlemmer, North Carolina budget lets teachers who struggle with licensing exams keep their jobs.
92. North Carolina General Assembly. (2026). Senate Bill 1006, Second Edition, § 1.10.

BEST NC

Business for Educational Success and Transformation

Evidence for Excellence

Building a Toolkit of Actionable Measures to Support and Advance Teachers



https://BESTNC.org

NCEdFacts.org

info@BESTNC.org

/BESTNorthCarolina

bestnc_org

company/BESTNC

© Copyright 2026. BEST NC | Business for Educational Success and Transformation in North Carolina