

4 Reasons Teacher Evaluation Still Falls Short—and How to Make It Worth the Effort

Why Teacher Evaluation Still Falls Short

Good intentions aren't enough. These common barriers keep evaluation from driving the growth teachers and students deserve.



BETTER EVALUATION → STRONGER TEACHING → GREATER STUDENT OUTCOMES

By: [Martha Hightower, Ed.D.](#)

October 2026

Topics Covered

1. What Is Teacher Evaluation?
2. Why Does Teacher Evaluation Often Fall Short?
3. Reason 1: Compliance Over Growth
4. Reason 2: Weak Link to Student Evidence
5. Reason 3: Tools Not Grounded in Research or Standards
6. Reason 4: Inconsistent Scoring
7. How Can the Marzano Evaluation Center Help?
8. Frequently Asked Questions

Key Takeaways

- **Evaluation should build expertise.** Teacher evaluation works best as a growth process, not an annual rating.
- **Make observations about growth.** Treat each rating as the start of a feedback conversation, not the end of a compliance task.
- **Use student evidence.** What students say, do, think, and produce shows whether instructional changes are working, often before test scores arrive.
- **Choose a research-based model.** Specific, standards-aligned competencies give teachers clear next steps. In a study of 12,000+ Florida teachers, Marzano observation scores were the largest predictor of value-added measures.
- **Calibrate evaluators continuously.** Consistent scoring depends on a shared language, ongoing calibration, and tools that ease administrator workload.

What Is Teacher Evaluation?

Teacher evaluation is the process schools and districts use to assess teachers' instructional practice, provide feedback, and inform professional growth and personnel decisions. It is intended to be a growth-oriented process for developing teacher expertise and improving student outcomes. The goal of teacher evaluation should not be to create better documentation of instruction. It should be to create better instruction.

Why Does Teacher Evaluation Often Fall Short?

Teacher evaluation falls short not because educators do not need feedback or accountability, but because many systems disconnect evaluation from its central purpose: building professional expertise. It was never supposed to be simply an annual judgment of a teacher's performance. Yet too often, the evaluation process becomes something educators endure rather than something that meaningfully improves their practice.

Reason 1: Observation Becomes a Compliance Exercise Instead of a Growth Process

Although teacher evaluation is intended to support professional growth, research suggests that implementation can sometimes prioritize compliance and accurate completion of the evaluation process over using evaluation evidence to improve instruction (Sartain & Morris, 2024).

In practice, evaluation can become centered on completing observations, entering ratings, meeting deadlines, and fulfilling required processes. When that happens, the evaluation process may feel like an activity to complete rather than an opportunity to improve instruction. Leaders may ask, *"How do we get everyone on board with evaluation?"* A more productive question is, *"How do we help teachers experience success as they improve their practice?"*

Sustainable growth occurs when teachers have opportunities to implement instructional practices, receive targeted feedback, reflect on evidence, and try again. This creates a cycle of **practice, feedback, reflection, and refinement** rather than a single observation followed by a score. Evaluation must be connected to the ongoing supports that help teachers interpret feedback, practice new approaches, and improve instruction (Shyjka, 2022).

Research and guidance from the Marzano Evaluation Center describe [focused feedback](#) and [deliberate practice](#) as critical conditions for increasing teacher expertise (Marzano Evaluation Center, n.d.). When teachers can see the connection between focused practice and their own improvement, evaluation becomes more meaningful. They begin to see evaluation not as something being *done to them*, but as something that helps them become more effective in the classroom. Over time, successful experiences with improvement can build confidence and commitment to continuous growth.

Solution: How Can Classroom Observations Focus on Growth, Not Compliance?

The first step is to shift the focus of classroom observation from compliance to quality. Instead of centering the conversation on, “What rating did I get?” the conversation should move toward, “What aspect of my practice am I intentionally developing, and what student evidence will tell me I am getting better?”

The evidence that matters is not simply whether a teacher performed a particular strategy during an observation. It is whether that practice produced the intended effect for students. Are students more engaged? Are they thinking more deeply? Are they collaborating more effectively? Are they demonstrating greater ownership of their learning?

Focusing teacher evaluations on growth requires a corresponding shift in how school and district leaders use classroom observations. At the district level, that shift starts with what principals are asked to report and whether feedback quality, not just observation completion, is part of how evaluation is monitored. The focus should be moving from compliance to coaching.

Principals should not view observations simply as a requirement to complete the minimum number of visits during the year, accompanied by a checklist of basic expectations. Instead, classroom observations should be opportunities to examine the quality and impact of instruction and to provide teachers with specific feedback they can use. In other words, **the observation should not end with a rating. The rating should be the beginning of a conversation about growth.**

When teachers can see their own growth, connect that growth to student learning, and understand what to do next, evaluation moves beyond compliance. It becomes a meaningful process for developing expertise, and that is what makes it worth the effort.



Reason 2: Evaluation Lacks a Meaningful Connection to Student Evidence

Evaluation feedback is most useful when it connects instructional practice to what students are [experiencing and learning](#). Teacher actions are important, but student evidence tells us whether those actions are making a difference.

Student learning behaviors can offer timely evidence of how students are responding to instructional changes. Teachers can note:

- Are students asking more thoughtful questions?
- Are they able to explain their thinking?
- Are more students engaging in productive discussion?
- Are they applying what they have learned independently?

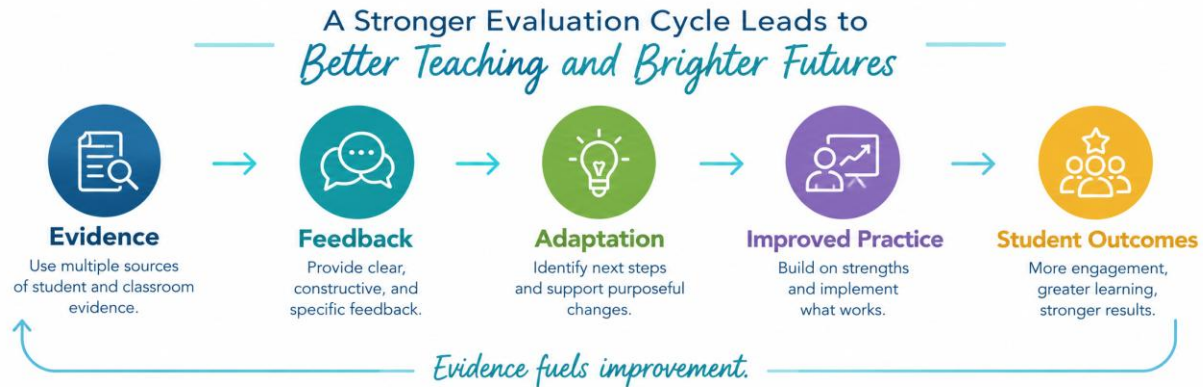
These observable changes can give teachers timely information about the impact of an instructional change, often well before the results of an end-of-unit assessment or standardized test are available.

The focus on student evidence is an important distinction in a meaningful teacher evaluation. **A rating by itself cannot close an instructional gap. The focus, therefore, should not be simply on what the teacher did, but on what happened to students as a result.**

Solution: How Can Student Evidence Strengthen Teacher Feedback Conversations?

The goal is not to determine whether a teacher is “good enough.” The goal is to create the conditions for teachers to become increasingly expert at what they do—and to help them recognize whether their instructional decisions are producing the intended results. This shifts the purpose of evaluation, and three practices can help:

1. **Make feedback specific and anchored in evidence of student learning.**
Classroom observations are only as useful as the feedback that follows. If an observation ends with a rating but does not help the teacher understand the relationship between instructional decisions and student responses, an important opportunity for growth has been missed.
2. **Provide one-on-one coaching conversations.** For these [coaching discussions](#), teachers bring student work, formative assessment results, or other evidence of student learning. Rather than simply reviewing whether the teacher implemented a particular strategy, the conversation can explore questions such as:
 - What did students demonstrate?
 - What did they not yet demonstrate?
 - Where did learning break down?
 - What instructional decision may have contributed to that result?
 - What might the teacher try differently?
 - What evidence will tell us whether the adjustment worked?
3. **Make student evidence matter!** Once student evidence becomes part of the evaluation conversation, teachers are better positioned to see evaluation as a tool for learning rather than a judgment of their worth as educators. They can begin to answer the question that matters most: *“Is what I am doing actually working for my students, and what can I do next to make it work even better?”*



Reason 3: Evaluation Methods and Tools Are Not Always Research-Based or Connected to Standards

Teacher evaluation models must do more than measure instructional performance. They must provide a methodology that supports teacher growth as educators make the instructional shifts necessary to support students in rigorous, standards-based classrooms (Marzano Evaluation Center, n.d.).

The evaluation process is only as strong as the framework and tools used to carry it out. Effective evaluation systems should provide administrators with a fair and objective way to distinguish among levels of teacher performance using specific, research-based competencies rather than broad or subjective judgments. Systems that recognize the full range of each teacher's strengths and weaknesses provide a more precise picture of teacher effectiveness (Kraft & Gilmour, 2017).

Just as importantly, those research-based competencies should reflect the instructional practices and student outcomes that matter most in standards-based classrooms. When the language and processes used for evaluation are disconnected from standards-based instruction, teachers may receive feedback that is technically accurate but not necessarily useful for improving their practice. The result is a system that measures what is easiest to document rather than what is most important to develop.

Solution: What Does an Evaluation Model That Develops Teacher Expertise Look Like?

An effective teacher evaluation model should provide more than a rubric and a rating scale. It should provide a **coherent system for identifying, developing, and measuring teacher expertise**. For district leaders choosing the most effective evaluation tools, the key question is whether they provide enough specificity to help teachers identify a meaningful area for growth and determine the next instructional move. A score tells us where performance falls on a continuum; it does not necessarily tell us what to do next. This is what transforms evaluation from a measurement system into a professional learning system.

The [Marzano Focused Teacher Evaluation Model](#) was designed to measure instructional effectiveness while also providing a structure for improving instruction in standards-based classrooms. The model organizes instructional expertise into four domains and uses [23 specific elements](#), desired effects, and student evidence to provide a common language for examining and improving practice. In a peer-reviewed research study of more than 12,000 Florida teachers, observation scores from the original Marzano model were the largest predictor of teachers' value-added measures (Basileo & Toth, 2019).

What Makes the Marzano Focused Teacher Evaluation Model Effective?

- **It pairs a clear framework with protocols for applying it.** The framework identifies the instructional knowledge and skills teachers need to develop. The protocols create a consistent process for observing practice, examining student evidence, providing feedback, and determining next steps. The protocols are what help turn the framework into action.
- **It increases the specificity and accuracy of observation.** Rather than relying on general impressions of instructional quality, an effective model gives observers specific competencies and observable indicators. The Focused Teacher Evaluation Model further emphasizes evidence of the desired effect on students, helping observers look beyond whether a strategy was used to consider whether students demonstrated the intended response.
- **It connects evaluation to the instructional shifts required by rigorous academic standards.** Standards-based instruction requires more than teachers covering content. Students must have opportunities to engage in increasingly complex thinking and demonstrate their understanding. An evaluation model should therefore help teachers and leaders recognize and develop the instructional practices necessary to move students toward those expectations.

Providing Specific Examples for Teachers



“When teachers come to you and they’re asking for protocols, sample teacher evidences, or sample student evidences, the Focused Teacher Evaluation Model has samples that a principal or an assistant principal can work from to support that teacher.”

– BJ Price

Director of Human Resources, Santa Rosa County School District, FL

Reason 4: Administrator Inconsistency Can Undermine Scoring Reliability

A central challenge in teacher evaluation is ensuring consistency among evaluators. When administrators interpret and score classroom practice differently, the evaluation process can unintentionally create different expectations for teachers across schools or even within the same school. When evaluators are not calibrated around what effective practice looks like and how evidence should be interpreted, differences in scoring can undermine both the credibility of the evaluation system and its ability to support teacher growth (Hill et al., 2012).

A district may invest significant time and resources in an evaluation system, only to discover that the implementation is producing inconsistent results. The issue may be that evaluators have not developed a sufficiently common understanding of what to look for, how to interpret evidence, and how to translate that evidence into actionable feedback.

Solution: How Can Districts Improve Inter-Rater Reliability Through a Model, Calibration, and Technology?

Inter-rater reliability is the degree to which different evaluators reach the same rating when observing the same instruction. Improving inter-rater reliability requires more than training administrators once at the beginning of implementation. Districts need an ongoing approach that builds a shared language, develops evaluator expertise, and provides tools that make accurate implementation manageable. Consistency does not mean every evaluator sees every classroom the same way. It means evaluators share a framework for interpreting evidence, making sound judgments, and helping teachers move forward.

1. Establish a common language for instruction.

To create greater consistency in instructional expectations, everyone involved in supporting teaching and learning needs a [common understanding](#) of what effective instruction looks like and how it contributes to student learning. When educators are using the same language and examining the same types of evidence, feedback

becomes more specific and less dependent on individual preferences or interpretations.

The [Marzano Focused Teacher Evaluation Model](#) provides a common language for professional conversations about instruction. Its specific elements, desired effects, and evidence-based approach give teachers, administrators, and instructional coaches a shared framework for discussing practice.

2. Give principals meaningful opportunities to calibrate.

Principals cannot effectively develop teacher expertise without first developing expertise in using the evaluation system themselves. Calibration should be an ongoing process that gives administrators opportunities to examine evidence together, discuss differences in interpretation, practice scoring, and strengthen evaluator judgment. In the same peer-reviewed study referenced earlier, the link between observation scores and student learning gains appeared stronger for principals who had received training and side-by-side coaching (Basileo & Toth, 2019).

More importantly, calibration is not just about achieving scoring agreement. It is about improving the quality of feedback teachers receive. When administrators become more skilled at interpreting evidence, they are better equipped to help teachers identify meaningful next steps and improve instruction.

The Focused Teacher Evaluation Model supports this through its five-step observation process, providing a common sequence for examining instructional strategies, desired effects, student evidence, teacher adaptation, and scoring. This helps reduce reliance on general impressions and promotes greater consistency (Marzano Evaluation Center, n.d.).

3. Use technology to support accuracy without creating additional burden.

Inter-rater reliability can also be affected by the practical demands placed on administrators. Cumbersome systems for scripting, documenting, and reconciling evidence can make consistent implementation difficult for time-constrained leaders.

Technology should not replace professional judgment, but it can make evaluation more efficient, consistent, and evidence-based. The goal is not simply to make evaluations faster; it is to make the time invested in evaluation more valuable.

[iObservation](#), the Marzano Evaluation Center's teacher evaluation software, provides a digital environment for documenting observations, tracking teacher growth, organizing evidence, and supporting reflection over time, allowing administrators to spend less time managing evaluation information and more time using it to support teachers.

How Can the Marzano Evaluation Center Help You?

Ultimately, the strength of an evaluation model lies not in how sophisticated its rubric appears on paper, but in whether it helps educators see instruction more clearly, understand its impact on students, and know what to do next. The goal is not simply to produce a more accurate rating, but to help teachers develop the knowledge, skills, and habits of reflection necessary to become increasingly expert in their practice.

Whatever change your school or district is asking teachers to implement this year, keep returning to one question: **Did the process produce evidence that the teacher's practice improved and that students experienced a better learning outcome?**

And perhaps an equally important question for leaders is: How do we create experiences that allow teachers to discover the value of this work for themselves?

If your district is asking these same questions about teacher evaluation, you don't have to navigate the work alone. The Marzano Evaluation Center helps schools and districts examine their current educator evaluation practices, strengthen evaluator calibration, develop teacher expertise, and build a more meaningful connection between evaluation and student learning. [Contact us](#) to talk through your district's evaluation system.

<https://marzanoevaluationcenter.com/contact-us/>

Frequently Asked Questions About Teacher Evaluation

What is the purpose of teacher evaluation?

The purpose of teacher evaluation is to build teacher expertise and improve student outcomes. Ratings support accountability, but the real value comes from specific, evidence-based feedback teachers can act on.

What is the difference between formative and summative teacher evaluation?

Formative evaluation happens throughout the year through observations, feedback, and coaching focused on growth. Summative evaluation is the end-of-year rating used for accountability and personnel decisions. Strong systems connect the two so the final rating reflects documented growth.

What student evidence should observers look for?

Observers should look for evidence that students are asking thoughtful questions, explaining their thinking, engaging in productive discussion, and applying what they have learned independently.

What makes teacher evaluation feedback effective?

Effective feedback is specific, anchored in evidence of student learning, and focused on a clear next step. Rather than ending with a rating, it opens a coaching conversation about what students demonstrated, where learning gaps may have formed, and what the teacher might try differently.

How can districts improve inter-rater reliability?

Districts can improve inter-rater reliability by using a shared instructional framework, giving administrators ongoing opportunities to calibrate by scoring and discussing evidence together, and providing tools that make documentation manageable.

Is the Marzano Focused Teacher Evaluation Model research-based?

Yes. The Focused Model is a streamlined version of the original Marzano Teacher Evaluation Model. In a peer-reviewed study of more than 12,000 Florida teachers, observation scores from the original model were the largest predictor of teachers' value-added measures (Basileo & Toth, 2019).

About the Author

Martha Hightower, Ed.D.

Dr. Martha Hightower serves as the Product Development Specialist for Marzano Evaluation Center. In addition, she delivers professional development to both teachers and school leaders in various schools to improve planning, instruction, and evaluation.

She has extensive experience in education ranging from special education to general education in primary and secondary schools. She also was a school leader for over ten years in Orlando, Florida where she resides. She received a doctorate degree in Organizational Leadership from Nova Southeastern University.

Dr. Hightower is affiliated with several local non-profit organizations and is committed to leading change and helping others develop systems and plans to sustain better living after trauma.

References

- Basileo, L. D., & Toth, M. (2019). A state level analysis of the Marzano Teacher Evaluation Model: Predicting teacher value-added measures with observation scores. *Practical Assessment, Research, and Evaluation*, 24(1), Article 6. <https://doi.org/10.7275/cc5b-6j43>
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56–64. <https://doi.org/10.3102/0013189X12437203>
- Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 46(5), 234–249. <https://doi.org/10.3102/0013189X17718797>
- Marzano Evaluation Center. (n.d.). *The Marzano Focused Teacher Evaluation Model: A focused, scientific-behavioral evaluation model for standards-based classrooms*. <https://marzanoevaluationcenter.com/library/the-marzano-focused-teacher-evaluation-model/>
- Sartain, L., & Morris, W. T. (2024). Are teachers satisfied with their evaluations? The importance of teaching context and trends over time. *AERA Open*, 10. <https://doi.org/10.1177/23328584231221539>
- Shyjka, A. (2022). Reaching for improvement: Teacher evaluation and its role in instructional improvement. University of Chicago Consortium on School Research. <https://consortium.uchicago.edu/sites/default/files/2022-05/Reaching%20for%20Improvement-May2022-Consortium.pdf>