education sciences

MDPI

Article

The Emergence of GenAI Feedback in Teacher Evaluation: Challenges and Questions

Helen M. Hazi

College of Human Services, West Virginia University, Morgantown, WV 26506, USA; helen.hazi@mail.wvu.edu

Abstract

The use of AI in teacher evaluation is one of the five priorities of the U.S. Department of Education in 2025. Yet, delivering feedback in teacher evaluation is a problematic school practice, and AI adoption is happening without guidance, research and foresight of its impact. The purposes of this essay are to report on GenAI's use in teacher evaluation, to address its challenges, and to offer questions as guidance when considering the purchase of GenAI feedback systems, since these systems appear to address both teacher and principal problems with the evaluation of teaching. Principals lack time, and teachers find principal feedback not to be useful, transparent, accurate, fair, or safe. Yet developers claim GenAI saves time, collects large amounts of data for documentation, and requires limited oversight. Are these claims illusory? Are we technomyopic? How will human judgment matter in an age of GenAI automated conclusions? The author of this study offers one perspective on a watershed moment: AI is a ubiquitous, dynamic, worldwide phenomenon being adopted at an unprecedented rate, while the use of GenAI in teacher evaluation is nascent and just emerging.

Keywords: teacher evaluation; AI; feedback; GenAI feedback; automated feedback

1. Introduction

With news about artificial intelligence (AI) bombarding educators daily, it would be easy to overlook its looming use in teacher evaluation. AI's use in teacher evaluation became one of five priorities of the U.S. Department of Education in 2025 to advance artificial intelligence in schools "to improve classroom efficiency, reduce administrative burdens and *improve teacher* training and *evaluation* [emphasis added]" (Executive Order No., 2025, 2025, Sec. 7(ii)). However, its potential use in teacher evaluation appeared on the author's radar when Langreo (2023) reported on the use of AI-enabled feedback in the Spokane School District's coaching program. While both provide teacher feedback, coaching is typically performed by trusted peers, and teacher evaluation is conducted by principals.

Artificial intelligence (AI) has arrived at an important moment in education with unprecedented attention. Fifty-four countries have published national AI plans (Dawson et al., 2025). One educational platform reports AI presence in more than 1 million classrooms in over 80 countries (Leiva et al., 2025). An email alert on AI from Policy Commons (personal email, 9 February 2026) reported that 360 AI news items and reports were available worldwide in one week.

In the U.S., 5 states required their schools to develop policies to regulate AI, while another 21 attempted to pass such bills in 2025 (Bushweller, 2025; Langreo, 2026). After

Check for updates

Academic Editors: Leslie W. Grant and James H. Stronge

Received: 26 March 2026
Revised: 18 May 2026
Accepted: 30 May 2026
Published: 10 June 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

the President issued his Executive Order, the U.S. government embedded AI in hundreds of projects in 2025 to automate routine tasks, make high-impact decisions, and surveil (Duncan et al., 2026). The U.S. Senate conducted hearings in 2026 with a coalition of 17 educational organizations advocating for continued funding of technology (Coffey, 2026; Senate-Commerce-Hearing-Letter, 2026). Three big tech companies, Microsoft, Google and Anthropic, are providing teachers with free software and training both in and outside of the U.S. and forging partnerships with teachers to collaboratively design AI (Morrone, 2026). Educational news outlets (e.g., *Education Week*), centers (e.g., Brookings Center for Technology Innovation, the Center for Democracy & Technology, Center on Reinventing Public Education) and individuals (e.g., Bushweller, 2025) develop databases, track policy and use, conduct research, issue reports, review products, and post news. Teachers are regularly surveyed to gauge increases in use, training, and concerns (e.g., Ward, 2025). AI is embedded in many software tools that teachers already use. Since then, "dozens of tools for generating language, code, images, video, and music have captured public and economic attention" (DeLuca et al., 2025, p. 1). AI has become ubiquitous.

What makes this moment consequential for educators is that the use of AI is happening without guidance and foresight of its impact. Most adoption is fragmented and precedes school policy, training and research on its effectiveness. The November 2022 launch of ChatGPT 3.5,[1] characterized as a "watershed moment" that changed the world, spurred Generative AI (GenAI) interest and adoption in many sectors of society, including education (Leiva et al., 2025). GenAI "technologies [such as ChatGPT] have gained millions of users within months of their release into the world" (Mishra et al., 2024, p. 205).

There is a difference between AI and GenAI. "*AI* is a term broadly referring to any system designed to perform tasks typically requiring human intelligence," while GenAI is "a type of AI that is able to learn patterns in massive amounts of raw unstructured data and then generate new outputs based on those learned statistical patterns (I. Hughes & Samo, 2026, pp. 186–187). Teacher evaluation requires GenAI. Each term is used specific to its meaning in the text that follows and are not treated as synonyms.

The purposes of this essay are to report on GenAI's use in teacher evaluation in the U.S., to address its challenges, and to offer questions as guidance in selection. After a brief explanation of the materials and methods used, the sections include Background; Findings, which describe the broader research and writings on GenAI and those specific to teacher evaluation; Challenges; and, finally, Summary, Questions, and Conclusions.

2. Materials and Methods

GenAI use in teacher evaluation is nascent. A systematic all-dates EBSCOhost search of 17 databases was conducted using the following keywords: "teacher evaluation or appraisal or personnel evaluation" and "employee performance or job performance or work performance" and "artificial intelligence or ai or a.i. or generative ai or generative artificial intelligence." Of 1390 records from 1986 to June 2026 that included dissertation abstracts, conference proceedings, books, and peer-reviewed periodical and news articles, at least 1 addressed teacher evaluation in the schools (e.g., Hazi, 2026). When the keyword "teacher" was added to keywords of "teacher evaluation. . ." and "artificial intelligence. . .," 180 records primarily focused on teacher attitudes, teacher feedback to and assessment practices of students, and teacher use of AI tools, but not teacher evaluation using GenAI. Thus, GenAI's use in teacher evaluation is best learned about in reports, surveys, and news reports. These were obtained from daily scans of subscribed new sources (e.g., Education Week, ISTE + ASCD Smartbrief, Brookings Brief, Tech&Learning) from 10 May 2023 to the present. Relevant results are reported in Findings.

3. Background

Teacher evaluation is the annual process of rating teachers to inform personnel decisions.[2] It has two purposes: the helping or formative function to improve and the evaluative or summative function for the end-of-year decision for teacher retention, dismissal or improvement. In its early history, the two functions were vested in different people: male principals, who were generalists, evaluated, and female content specialists helped teachers improve. But as schooling expanded and the Depression of the 1930s forced schools to eliminate positions, the two functions were combined into that of the principal. Since then, the summative has tended to overshadow its formative purpose. For some, teacher evaluation has become an annual ritual—time-consuming for the principal and not helpful to the teacher—despite much policy emphasis, training, and effort (Hazi, 2022a, 2022b).

In the U.S., states vary, and each has its own teacher evaluation policy that is governed by state statute and department of education regulations, which school districts can exceed, but cannot violate. Most states use principal observation to evaluate teachers, usually with a standards-based instrument that includes a rubric. Popular ones that have been used or modified include those by Charlotte Danielson and Robert Marzano. In the 2000s, due to Race to The Top, a federal initiative, state and federal policymakers invested billions of dollars in high-stakes teacher evaluation that included more robust observations and the use of student test scores. It was believed that test scores could be used to measure teacher quality with value-added measures, showing the achievement gains students made and attributing them to a specific teacher. This practice proved especially controversial, since 70% of the teachers did not have appropriate tests to make such judgments, and teachers were dismissed in some states. Despite its many flaws, including that of validity (Amrein-Beardsley et al., 2023) and evidence that its use has not resulted in increased student achievement (Bleiberg et al., 2021), 40 states continue to use value-added measures in annual teacher evaluation (Hewitt, 2015).

Artificial Intelligence was founded on the belief that a machine could simulate every aspect of human learning. The concept of AI can be traced back to the 1840s, when Ada Lovelace imagined that machines could do things like humans, such as compose music; then, a century later, Alan Turing in 1936 built the codebreaking, top-secret Turing machine, considered to be the first modern computer (Boden, 2016/2018). AI in education, however, is only decades old (DeLuca et al., 2025; see Trumbore, 2025 for a history).

GenAI is prevalent in business but limited in education. GenAI use is widespread in 89% of the firms worldwide and used in every aspect of the employee lifecycle, including selection, onboarding, instruction, monitoring, evaluation, and retention of workers (OECD, 2026).

> The algorithms of data analytics, natural language processing, and predictions are the foundations of the AI-based talent management systems with identifying the high-potential candidates, predicting the trends within the workforce, and tailoring the employee experience. The innovations have assisted organizations in making fast and evidenced based decisions and minimize biases that are part of the human judgment. (Umarani & Ramu, 2025, p. 3350)

In business, it is called AI-based talent management, designed to enhance efficiency and decision making.

In education, AI is a field of study at the crossroads of cognitive psychology, computer science, educational technology, and philosophy, a mix of discourses informing its theory and practice. Many in the field believe that

> AI tools can reduce the amount of time teachers need to spend on administrative minutiae. . ., personalize assessment and learning opportunities to students' specific needs. . ., gather real-time feedback about student learning. . ., support

> cognitive and emotional engagement. . ., and foster student metacognition. . . Yet these promises are intertwined with safety and security concerns, bias issues, and general reluctance. . . .These issues are well founded because AI tools can. . . invite incorrect anthropomorphic comparisons. . .; sidestep teachers' professional expertise. . .; and facilitate undetectable cheating that undermines student learning and assessment validity. (DeLuca et al., 2025, p. 2)

4. Findings

Most of the focus of educational research has been on the K-12 teacher and student use of AI in the classroom. However, the emerging use of GenAI in teacher evaluation appears in sources such as educational news, surveys, and research related to teacher coaching.

In the U.S., two in five schools have a reading coach, one in four have a math coach, and about one-third have a general coach (Hill & Papay, 2022).[3] Peer coaches provide the subject expertise that principals often cannot, but it is time-intensive and costly, with some districts spending as much as USD 2–3 million dollars on salaries. Because of cost, many look to GenAI to make coaching scalable and affordable (Hazi, 2026).

Langreo (2023) was one of the earliest reported uses of GenAI coaching in the Spokane School District in the state of Washington. The Title I and special programs' coordinator established a professional development course with 30 teachers for use with peer observation and video reflection. The Edthena online platform was used with an AI "coach" who provided feedback and walked teachers through reflection with questions using videos of their teaching. Teacher response has been positive so far, since feedback is customized based on their goals, it is voluntary, and teachers can do it on their own time. In addition, its coaching program is not connected to teacher evaluation.

Videos and webinars provide in-depth descriptions of some systems. One example, TeachFX, can be downloaded to a phone or laptop. It can then be used by the teacher for "recording data points and offering potential actions to help improvements" (Edwards, 2025, para 14). For example, it can tell the teacher the amount of teacher talk vs. student talk "to find the right balance that best suits the lesson and the students" (para 12). Once the teacher sets a goal, TeachFX can "track progress as they work toward that, and ultimately, see how they achieve that goal and what affect it has" (para 15).

TeachFX offers workshops and resource kits for new teachers. A small school "will typically pay about $10,000 for a subscription, with a medium school paying $20,000, and the top-end larger schools getting charged $30,000" (Edwards, 2025, para 20). Developed by a former English and math teacher, TeachFX earns USD 2.5 million dollars per year.

News articles have reported on the use of AI-assisted coaching and teachers' comfort with it. TeachingLab, a platform with tools for both teacher and coach,[4] is being piloted in New York City schools, and Edthena, a video-coaching platform, is used in a school district in Laremont, Colorado.[5] According to Will (2025), more than one-half of the teachers reported in an Education Week Research Center survey that they would feel "somewhat" comfortable or "very" comfortable in AI-assisted coaching.

News sources also provide opinions of experts. Advocates report these apps and platforms are efficient, scalable and can be driven by the teacher. They claim that they can supplement what human coaches do: analyze data, deliver feedback, guide teachers through self-reflection and goal setting, and offer readings and resources "without risk or fear of being judged" (Will, 2025, para 8). Developers claim that it only has to be used for five hours and gives feedback "immediately."

Attendance at a webinar of another platform, GoReact, provides a specific example of feedback on teaching. Once the teacher or supervisor selects the focus (e.g., praise, questioning, pacing) and the teaching is recorded, relevant segments are timestamped so

they can be viewed on the platform, with summary data graphed and GenAI-enabled feedback. The GenAI assistant might say:

> Your pacing was superb when you smoothly transitioned into the writing activity. You stated "So, let's see. Grasshoppers are insects. Do you want to say they are?" This quick transition kept the lesson moving while maintaining student engagement.

This expresses a judgment about pacing. Another example of feedback includes a suggestion for improvement:

> To further engage students, you could consider incorporating more collaborative learning opportunities. For instance, when you said "Ready? So show me your hand like this. We're gonna say they are black and brown." You might try pairing students and having them discuss the sentence structure together before sharing with the class. This could promote peer interaction and deeper engagement with the material.

These examples seem to be based on the assumption that users know what pacing means and what to do with the feedback.[6]

One randomized, controlled trial study of, and in partnership with TeachFX, was reported by Demszky et al. (2025), who asked 200 K-12 Utah science and math teachers to assess its impact. Teachers used the app to record their lessons, which they then transcribed and used to provide feedback on teacher talk, wait time, instructional techniques, and focus questions. Engagement with what they call "automated feedback" was low, with teachers opening emails about 53–65% of the time. Some questioned feedback accuracy and reported privacy issues and time constraints. Those who received feedback increased their focus questions by 20 percent ($p < 0.01$), more than the control group. However, no other behaviors changed.[7]

In addition to providing feedback in coaching, one survey reported that principals are using AI in teacher evaluation. Kaufman et al. (2025) wanted to know how many principals were using it and for what purposes. They did not differentiate between AI and GenAI in this study. This was part of a larger RAND study about AI use by both teachers and principals. While they expected teacher use to increase, "it was a bigger surprise" about its widespread use among principals. In an analysis of data obtained in one open-ended question of the 2024 American Mathematics Education survey, principals reported that they used AI for different purposes. While the most common use was to draft communications to parents, staff and others, another use was to support teacher evaluation and provide feedback. About 150 (out of 1942) principals reported this use. More specifically, they created evaluation templates, summarized observation notes, and drafted observation reports and improvement plans. All principals reported concerns about data privacy, potential bias, lack of professional development, accuracy, and quality, which influenced its use.

5. Challenges

Most of the research about the adoption of educational technology has not necessarily led to improved teaching and learning (Trumbore, 2025). Current opinions, as reflected in national and international reports, echo this and warn of a risk for both teachers and students (Burns et al., 2026; Center for Democracy & Technology, 2025; Forsström et al., 2025, 2026). In the most comprehensive review of research, GenAI promises to

> boost teacher productivity and the quality of teaching. It can already quickly write summaries, design exercises and even offer real-time tutoring support. But there is a risk that overreliance on GenAI could lead to the loss of skills and teaching expertise. (Forsström et al., 2026, p. 11)

Furthermore, researchers conclude that the use of GenAI makes students "lazy," failing to develop the cognitive effort to learn content. Will they say the same about principals who use GenAI for teacher feedback? Since research about the use of GenAI feedback in teacher evaluation is limited, speculations about its challenges are warranted.

By next focusing on developer promises about the use of GenAI feedback, challenges are offered that reveal more about educator vulnerabilities and why such systems appeal to both teachers and principals. However, first, a flawed assumption about "delivering feedback" to teachers in teacher evaluation is addressed.

Feedback is defined as the transmission of evaluative or corrective information about a person's performance on a task. According to Hattie (2012), feedback should be frequent, timely, consistent, specific, and private. However, this advice is based on research about giving students feedback. Elsewhere, this argument is presented in detail (Hazi, 2024) but presented briefly herein. Student feedback research was applied to teachers during a time of limited research about how teachers, as adults, learn and within the context of the workplace. This assumption was accepted because teachers were learners like students. In addition, most of this research focused on the feedback giver, not the receiver. By focusing on the giver's delivery, the learner has the potential to be ignored and treated as a passive recipient.

> Researchers have only begun to understand recently that [learners] require skills of feedback literacy. . . Such literacy is a process within learners, requiring them to notice evidence of their learning, interpret and make sense of it, transform it into something usable, and then act on it. (Hazi, 2024, p. 94; Winstone et al., 2017)

Once feedback is delivered, it requires teachers to be open to it, interpret it within multiple contexts, and make it actionable if understood and agreed upon (Hazi, 2026). Thus, thinking about "delivering feedback" to teachers and its research is becoming contested by some (see Hazi, 2024, 2026).

Claims about GenAI feedback are both interrelated and complicated. What is touted as a promise can also be a challenge (e.g., Hazi, 2014). Furthermore, Turkle (2017), a clinical psychologist, professor of Science and Technology at M.I.T., and insider critic, claims that technology emerges to address our needs and vulnerabilities. A review of developer claims that help "sell" a platform also reveals educator vulnerabilities about teacher evaluation.

GenAI saves time. Timesaving—in the name of efficiency—is the major appeal of GenAI feedback. However, time saved is time spent after hours or on other things. Teachers may review feedback outside of school time. Coaches, who may also be teachers, spend time checking feedback accuracy, which may require them to watch a video, review feedback, and correct it. Educators should consider time-saving as a seductive claim made by companies to encourage purchase and *an illusory promise* difficult to deliver (see Buck, 2026).

GenAI can collect large amounts of data. This capacity for detail allows the system to collect low inference teacher and student behaviors that are *countable*. This capacity also makes GenAI feedback *appear to be useful*. This is especially helpful for documentation required to meet legal requirements for personnel decisions. Developers and educators may assume that the data can "speak for itself" and, thus, imbue the system with *transparency*. In addition, when the system totals behaviors and reports patterns, some may assume its judgments are *accurate and fair*. However, because "their inner workings are often hidden from users", teachers may also consider them to be inaccurate and not credible (Mishra et al., 2024, p. 206). In addition, GenAI is known to have "hallucinations," but its *error rate* is not yet reported or known.

Gen AI requires limited oversight. Since teachers can operate the system intuitively and review feedback on their own, feedback *appears to be safe*. However, when a teacher uses a personal phone to record their teaching, as one app suggests, it becomes publicly

accessible, and privacy becomes a concern. Furthermore, developers claim that supervisors can easily correct GenAI feedback before it goes to the teacher. However, when there are complex teaching problems, human judgment can conflict with AI judgment. This potential conflict can result in grievances and costly litigation.

Thus, Gen AI feedback reveals educator vulnerabilities about principal evaluation of teaching. It is time-consuming for both the teacher and principal. But for teachers, it is not useful, transparent, accurate, fair, or safe—all features that teachers look to GenAI feedback to remedy.

The use of a GenAI feedback system is further complicated by *technomyopia*, i.e., ascribing more power to technology than it deserves (Hazi, 2013, 2026). Indeed, because the system can "talk back" to users in a voice, this tool appears to be real. Then, practitioners use language that invests it with human qualities. For example, AI agents are being called "thought partners," "co-pilots," "digital teammates" and "labor with talent" (Stave et al., 2025). AI "learns," "must be able to understand. . .," "exaggerates biases," is "theoretically trained," and "can use interpersonal behaviors" (e.g., Lavigne & Ocran, 2026). Educators say GenAI "makes judgments," "makes sense. . .," and "rates teachers," when the system merely totals behaviors, reports patterns and provides findings. When educators anthropomorphize AI, they give it power it does not deserve—psychologically as well as legally.

In addition, a GenAI feedback system cannot allow for the uncertain, the unexpected, and the unique (Wang, 2025). Yet teaching is uncertain and unpredictable (Jackson, 1986). Since there is no consensus on a definition of good teaching, multiple conceptualizations complicate how educators think about teaching and its improvement (Russ et al., 2016). However, teaching has to be simplified for it to be improved in a GenAI feedback system. Nor can a GenAI feedback system allow for infinite explanation, variation, and differentiation. Classes have 30+ students, each unique. Because it is generative, it can "generate unique outputs even when given the same prompt or input" and gives the impression it can address the needs of individual students (Mishra et al., 2024, p. 206). While some claim personalization is possible, others question its limits (e.g., National Educational Policy Center, 2026). Finally, prescriptions to change teaching practice are guesses meant to be tried out. What might work in one context or for one student may not work for another. Context matters in teaching. Lesson objective, skill or concept learning, grade level, and student ability are some of the many factors that shape recommendations of expert supervisors. GenAI feedback systems cannot accommodate this great variability.

6. Summary, Questions, and Conclusions

AI is a ubiquitous, dynamic, worldwide phenomenon and everyone is concerned about it. Even Pope Leo has addressed it in his annual report to his diplomatic corps, promoting and at the same time tempering its use (Castel-Branco, 2026). GenAI in teacher evaluation is a nascent topic just emerging in education and is reported more frequently in educational news. While an educator's stance is easy to recommend, a precise level of guidance is more difficult, since our level of AI understanding continues to grow, depending upon what we access and digest and when. Thus, a summary, questions to guide decisions about use, and conclusions offered here are those deemed reasonable rather than overly ambitious.

In summary, educators have great hope that a technology tool designed for efficiency and profit will also enhance teaching and learning in the schools. School adoption precedes policy and research, but not unanticipated problems that educators believe they can confidently solve. While they may be able to see AI and GenAI's promises and challenges, they may not yet understand they are simultaneously both, and that their promises, appealing to our vulnerabilities, may be illusory. The high-profile claims to save time,

capture large amounts of data, and require limited oversight overshadow the uncertain and unpredictable nature of teaching. Our long-held technomyopia makes us look to technology as our measure of educational progress (Trumbore, 2025).[8]

Bushweller (2025) recommends educators adopt a healthy middle ground position—be curious and experiment—especially when it is designed to enhance teaching and learning. However, most AI and GenAI are not educational but recreational, designed for the consumer and profit.

In addition, Kaufman et al. (2025) recommend that an AI tool should be guided by leaders, backed by rigorous research and trials, and introduced with guardrails, training, and best practices. They conclude that

> [s]tates and school systems have an extremely challenging job: to guide the use of AI in schools in a sea of resources with very little evidence on what uses of AI are most effective for improving the quality of instruction and student learning. . . [where] one quarter of U.S. teachers are using AI. (p. 20)

They also recommend that educators partner with developers in AI development as well as research. Elsewhere, I address the complications of AI design and pedagogical ideologies (Hazi, 2026). GenAI design should be a long and circumspect process, requiring careful definition by teaching experts to ensure validity, pilot testing to establish reliability, and further research to assess its impact and effectiveness. This takes time that tech giants do not have as they rush to market. AI is designed to work autonomously and for profit. It may take small companies, such as Humans&, to be concerned about "human-centric AI," i.e., empowering rather than replacing, and requiring humans to work with AI in real-time, providing oversight as it goes, not operating autonomously and making errors with post hoc oversight (Metz, 2026).

Recommendations to mitigate risk (e.g., Forsström et al., 2026) and conduct research (Lavigne & Ocran, 2026) are becoming available. However, questions to think about GenAI feedback design and use are absent. To this end, GenAI's feedback system features are presented in Table 1, grouped by context, design and use. It is based on a review of opinions that dominate the writings and my experiences as a practicing public school supervisor.

Table 1. Features and questions about GenAI feedback to teachers regarding its design and use.

Groups of System Features	Feature	Potential Questions
I. Context	Can result in teacher ratings that can be used in personnel decisions	Is it legally permissible in state statutes, collective bargaining, local policy? Does it meet federal requirements of an employment test
	Teacher participation	Required vs. voluntary?
II. Design Delivery	Designed by AI expert	How have teaching experts been used? Is it valid, reliable, fair and not discriminate?
	Accessed by personal phone or laptop	Is it confidential, alterable? Who owns the data? Who has access to it?
	Platform can centralize services, e.g., recording, analysis, communications, professional development	Does it facilitate or hinder decisions about teacher?
	Accessible 24/7	Does it save time or just allow time to be spent on other things?

Table 1. *Cont.*

Groups of System Features	Feature	Potential Questions
III. Content	Designed for change in teaching, efficiency and school cost savings	Improvement in practice requires time, resources and learning. Is system designed for teacher learning? Could a summary of teacher thinking be helpful? How is progress defined?
	Research-based, which makes it appear credible	Is it based on process-product research that favors skill-based learning in the elementary? Is its content subject-specific? What is missed?
	What text was used and scanned?	What educational experts were used? Does it favor one conceptualization of teaching over others?
	Details make it appear transparent	Is it accurate, missing items, outdated?
	Can be used with all teachers yet customized	Is teaching generic? Is there a balance between standardized and individualized?
	Can be easily counted	Is it biased to trivial behaviors? What's missed?
	Can address multiple variables, e.g., veteran vs. novice teacher, subject, grade, student ability(s), lesson objective	How are variables combined to result in a finding? Is it clear AI is "guessing"?

Context. Since teacher evaluation is a law-enabled and constrained practice in schools, GenAI feedback use requires enabling state legislation and/or regulation, and compliance with all state and federal regulations regarding this personnel practice. It would be fatal to recommend otherwise.[9]

Design. Writings are limited in revealing the design principles of GenAI feedback (Mishra et al., 2024). Yet teacher learning, thinking and reflection should be at the heart of improving teaching practice—not changing behaviors. Kennedy (2016) argues that we need to better understand teacher mental knowledge structures, i.e., what teachers do and why they do it. Improvement takes time and resources, and we are still learning what it takes for teachers to improve their practice (Kennedy, 2019).

Content. Due to the enduring influence of the 1980s process-product research, teacher evaluation instruments and their developers have continued to consider teaching to be generic. Yet teachers report that principals lack subject-specific knowledge that they find useful to improving their practice. This, in large part, explains the use of peer coaches with subject expertise as a form of professional development and outside the bounds of teacher evaluation. Will GenAI feedback be generic, or will content integration be too costly? While some advocate for the power of GenAI to individualize feedback, how content feedback will and can be used receives limited coverage. Thinking about how subject experts will be used in content design has yet to be reported.

Use. And while we can imagine clashes in judgment between AI agents and supervisors, we have yet to consider how much time and what teaching problems are allotted to AI agents and which ones are assigned to humans. Nor have we yet to consider who gets the AI agents: the beginning or veteran teachers, and if participation is voluntary or required. We can only hope that human judgment is essential to these decisions, not GenAI-automated conclusions.

The debates about AI and GenAI use and the assumptions that follow them are complicated. A portion of that debate goes as follows:

- It will help educators do "time consuming tasks such as creating exams, presentation slides, and rubrics for grading in seconds. AI can help provide feedback on student assignments; write individualized programs for students in special education; and provide translation services" (Prothero, 2026, para 5).

However:

- When educators offload busywork, they avoid deliberate reflection and the time required to think through important issues or to learn about how students think (Buck, 2026).
- What makes it useful also makes it easy to overuse (Prothero, 2026).
- Some of the large language models result in remarkably similar responses called the *artificial hivemind*, where "[t]he process favors safe, consensus-based responses and penalizes risky, unconventional ones. Originality gets stripped away" (Barshay, 2026, para 11).

It seems educators may be trapped in a Catch 22. Educators are feverishly adopting AI and GenAI technology. Have we considered why, in the absence of professional development, policy and research? How much is driven by those who develop and finance the technology and a government that has the authority to regulate it (Trumbore, 2025)? Teachers may have to advocate that teacher learning and improvement are the primary purposes of GenAI feedback.

To that end, Kennedy's (2019, p. 147) work can help us understand where we are in our understanding of teaching and teacher learning:

> As we have become more aware of the many contingencies involved in teaching, we have shifted our focus away from telling teachers what to do and toward deepening their understanding of their students so that they can make better in-the-moment judgments about how to respond to students.

We can only hope that the purpose of GenAI feedback is not just to "deliver feedback," but first and foremost, promote teacher learning. If not, then we may have to challenge its misuse in court when teachers are dismissed.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1 ChatGPT is a free generative AI app that is accessible to anyone with an internet connection and no prior training, allowing a user to converse with it by asking questions or making requests and getting an instant response (Schleicher, 2026). GenAI is "a subset of AI focused on producing new content such as text, pictures, videos, songs, mathematical equations, computer programmes, typically in response to a question or command" (Forsström et al., 2026, p. 13).

2 While it is beyond the scope of this article to address the status of teacher evaluation in the U.S. in detail, it is important to understand its context for GenAI's emergence. While principals are still expected to be instructional leaders to improve teaching in teacher evaluation, teachers still fail to find their feedback useful. Coaching (as professional development, not evaluation) has emerged to provide the content feedback that principals cannot. Value-added measures continue to be used primarily in noncollective bargaining states in the southeast and mid-west. Generic teaching and rubrics continue to dominate observation systems, while subject matter content, student ability, and grade level continue to be ignored.

3 While Showers and Joyce (1996) are often cited as originators of coaching, Hill and Papay (2022), Ridge and Lavigne (2020) and Woulfin et al. (2023) offer the best summaries of its research.

4 Teaching Lab has two relevant tools. First, the Relay GSE Teaching Simulator is an on-demand teaching practice experience with AI-powered virtual students. Teachers participate in several voice-based scenarios that are recorded and become the basis of feedback. It was piloted in Fall 2025 and is being expanded to include video input of the teacher, lesson plan uploads, and AI generated student work (D. Hughes & Waldron, 2026). The second is Nisa an AI coach that provides coaching cycles to teachers,

models instructional best practices, tracks goals and action steps, sends customized podcasts aligned with teacher goals, and follows up with reminders (Nisa: The Coach in Your Corner, 2026).

[5] Two systems were identified during the writing of this article but not investigated due to time and space constraints. TalkMoves was designed to provide feedback on effective strategies to encourage student reasoning and talk about mathematics (Suresh et al., 2021). M-Powering Teachers was designed to provide feedback on student discourse contributions and teacher and student talk time (Demszky & Hill, 2022) These two teams of researchers have been conducting research on these feedback systems (to include TeachFX).

[6] This researcher received no response to emails requesting an interview after attending a GoReact webinar. For further research, users might be a better source of data, if developers appear guarded.

[7] Van Camp et al. (2025) note that systems such as TeachFX can be developed that identify specific coded behavior attached to time stamps within audio and video recordings, using Automatic Speech Recognition and Natural Language Processing software such as that used to run Amazon's Alexa and Google's Home, and transcripts of teaching coded by expert coders. However, noisy classrooms, rapid turn-taking of speakers, transparency in how the system works and makes decisions, and teacher trust in the feedback pose challenges in their use. While it is beyond the scope of this essay to address preservice teaching, GenAI feedback is being considered by some to be a "promising solution" to address the problem of limited opportunities to teach with personalized feedback. However, research is limited. In a quasi-experimental study of microteaching, Zhang et al. (2026) provided "AI feedback" to 20 preservice mathematics teachers and human feedback to 17 considered the control. While both groups rated teaching self-efficacy high, teacher feedback was rated more credible and useful in the control group that received feedback from human supervisors. Thus, the credibility of GenAI feedback has been questioned in this early research.

[8] Experience with the Florida Performance Measurement System (FPMS) in the 1980s provides lessons about the use of technology in teacher evaluation. Educators believed back then that judgment about teaching for a career ladder award that was calculated by a computer was error-free and more sound that principal judgment, and that educators were willing to defer judgment about teacher quality to a computer (Hazi, 1989).

[9] While coaching and teacher evaluation are similar enough to apply this analysis to both with some stipulations. Because teacher evaluation is a legally based practice, it must be authorized by law. Coaching, when done by peers and not associated with teacher evaluation, only requires voluntary participation, teacher consent and a trust relationship. Because teacher evaluation is a legally based, high stakes, personnel situation that affects teacher employment decisions, use of GenAI feedback would require statute and must comply with all federal, state, and local law. When feedback results in errors, conflicting judgments, and breach of confidentiality, legal consequences occur.

References

Amrein-Beardsley, A., Lavery, M., Holloway, J., Pivovarova, M., & Hahs-Vaugh, D. (2023). Evaluating the validity evidence surrounding the use of value-added models to evaluate teachers: A systematic review. *Education Policy Analysis Archives*, *31*, n117. [CrossRef]

Barshay, J. (2026, March 23). The AI 'hivemind': Why so many student essays sound alike. *The Hechinger Report*. Available online: https://hechingerreport.org/proof-points-ai-similarity/ (accessed on 25 March 2026).

Bleiberg, J., Brunner, E., Harbatkin, E., Kraft, M., & Springer, M. (2021, December). *Taking teacher evaluation to scale: The effect of teacher evaluation on achievement and attainment* (EdWorkingPaper: 21-496). Annenberg Institute Brown University. Available online: https://www.edworkingpapers.com/ai21-496 (accessed on 25 March 2026).

Boden, M. (2018). *Artificial intelligence: A very short introduction*. Oxford Press. (Original work published 2016).

Buck, D. (2026, March 18). Why teachers shouldn't offload their busywork to AI. *Education Week*. Available online: https://www.edweek.org/technology/opinion-why-teachers-shouldnt-offload-their-busywork-to-ai/2026/03?utm_source=nl&utm_medium=eml&utm_campaign=eu&M=17297954&UUID=5db99379423bd8f2b72457b68b37431b&T=22520178 (accessed on 25 March 2026).

Burns, M., Winthrop, R., Luther, N., Venetis, E., & Karim, R. (2026, January). *A new direction for students in an AI world: Prosper, prepare, protect*. Center for Universal Education at Brookings. Available online: https://www.brookings.edu/wp-content/uploads/2026/01/A-New-Direction-for-Students-in-an-AI-World-FULL-REPORT.pdf (accessed on 25 March 2026).

Bushweller, K. (2025, September 23). Which states require schools to have AI policies? *Education Week*. Available online: https://www.edweek.org/technology/which-states-require-schools-to-have-ai-policies/2025/09 (accessed on 25 March 2026).

Castel-Branco, N. (2026, January 25). Why the pope is right to weigh in on AI. *The Washington Post*. Available online: https://www.washingtonpost.com/opinions/2026/01/25/pope-leo-ai-history/?utm_campaign=wp_todays_headlines&utm_medum=email&utm_source=newsletter&carta-url=https%3A%2F%2Fs2.washingtonpost.com%2Fcar-ln-tr%2F469275d%2F6977493a130502561e5dea76%2F596a90fa9bbc0f0e09ec970c%2F35%2F69%2F6977493a130502561e5dea76 (accessed on 25 March 2026).

Center for Democracy & Technology. (2025, October). *Hand in hand: Schools' embrace of AI connected to increased risks to students. Survey research on parent, student, and teacher experiences*. Available online: https://cdt.org/wp-content/uploads/2025/10/FINAL_CDT-Polling-2025-Supporting-Slides.pdf (accessed on 25 March 2026).

Coffey, L. (2026, January 28). Will school cellphone bans morph into wider screen time regulations for kids? *EdSurge*. Available online: https://www.edsurge.com/news/2026-01-28-will-school-cellphone-bans-morph-into-wider-screen-time-regulations-for-kids (accessed on 25 March 2026).

Dawson, G., Desouza, K., Denford, J., & Picavet, M. (2025, August 18). How different states are approaching AI. Brookings. *Commentary*. Available online: https://www.brookings.edu/articles/how-different-states-are-approaching-ai/ (accessed on 25 March 2026).

DeLuca, C., Volante, L., & Holden, M. (2025). The AI 3 model: Future directions for artificial intelligence, assessment innovation, and academic integrity. *Educational Researcher*, *55*(2), 125–133. [CrossRef]

Demszky, D., & Hill, H. (2022). The NCTE transcripts: A dataset of elementary math classroom transcripts. In *Proceedings of the 18th workshop on innovative use of NLP for building educational applications*. Association for Computational Linguistics. [CrossRef]

Demszky, D., Liu, J., Hill, H., Sanghi, S., & Chung, A. (2025). Automated feedback improves teachers' questioning quality in brick-and-mortar classrooms: Opportunities for further enhancement. *Computers & Education*, *227*, 105183. [CrossRef]

Duncan, I., Merrill, J. K., & Dou, E. (2026, February 9). Trump set off a surge of AI in the federal government. See what happened. *The Washington Post*. Available online: https://www.washingtonpost.com/technology/2026/02/09/trump-administration-ai-push/?utm_campaign=wp_todays_headlines&utm_medium=email&utm_source=newsletter&carta-url=https%3A%2F%2Fs2.washingtonpost.com%2Fcar-ln-tr%2F46b5091%2F6989be371055606afe77eede%2F596a90fa9bbc0f0e09ec970c%2F17%2F66%2F6989be371055606afe77eede (accessed on 25 March 2026).

Edwards, L. (2025, June 20). What is TeachFX and how can I use it to teach? *Tech & Learning*. Available online: https://www.techlearning.com/how-to/what-is-teachfx-and-how-can-i-use-it-to-teach (accessed on 25 March 2026).

Executive Order 14277. (2025, April 28). *Advancing artificial intelligence education for American youth*. 90 Fed. Reg. 17519. Available online: https://www.govinfo.gov/app/details/FR-2025-04-28/2025-07368 (accessed on 25 March 2026).

Forsström, S., Morten, N., Munthe, E., Álvarez-Galván, J., & Houldsworth, L. (2025). *The impact of digital technologies on students' learning: Results from a literature review* (OECD Education Working Paper No. 335). OECD Publishing. Available online: https://www.oecd.org/en/publications/the-impact-of-digital-technologies-on-students-learning_9997e7b3-en.html (accessed on 25 March 2026).

Forsström, S., Morten, N., Munthe, E., José-Luis, Á.-G., & Houldsworth, L. (2026, September 29). *OECD Digital education outlook 2026: Exploring effective uses of generative ai in education*. OECD Publishing. [CrossRef]

Hattie, J. (2012). Feedback in schools. In R. Sutton, M. Hornsey, & K. Douglas (Eds.), *Feedback: The communication of praise, criticism, and advice* (pp. 265–277). Peter Lang.

Hazi, H. M. (1989). Measurement vs. Judgment: The Case of Sweeney v. Turlington. *Journal of Curriculum & Supervision*, *4*(3), 211–229.

Hazi, H. M. (2013). Digitized judgment: The seductions of teacher evaluation technologies. *Ubiquitous Learning: An International Journal*, *5*(1), 57–67. [CrossRef]

Hazi, H. M. (2014). The marketing of teacher evaluation: The seductive claims of instruments. *The WERA Journal*, *6*(1), 2–9. Available online: http://www.wera-web.org/wp-content/uploads/2014/04/TheWERAEd32014.pdf (accessed on 25 March 2026).

Hazi, H. M. (2022a). The restrictive concepts of teacher evaluation and their discourse communities. *Journal of Educational Supervision*, *5*(1), 44–67. [CrossRef]

Hazi, H. M. (2022b). Reconsidering the dual purposes of teacher evaluation. *Teachers and Teaching*, *28*(7), 811–825. [CrossRef]

Hazi, H. M. (2024). The trouble with 'delivering feedback": Reflections of a supervision scholar. *Journal of Educational Supervision*, *7*, 85–103. [CrossRef]

Hazi, H. M. (2026). The lure of automated feedback. *Theory into Practice*, *65*(1), 140–152. [CrossRef]

Hewitt, K. (2015). Educator evaluation policy that incorporates EVAAS value-added measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives*, *23*, 76. [CrossRef]

Hill, H., & Papay, J. (2022, October 25). Building better PL: How to strengthen teacher learning. *Research Partnerships for Professional Learning*. Available online: https://rpplpartnership.org/wp-content/uploads/2024/10/rppl-building-better-pl.pdf (accessed on 25 March 2026).

Hughes, D., & Waldron, A. (2026, February 2). Practice makes progress: Building and testing an AI-powered teaching simulator. *Teaching Lab Studio*. Available online: https://www.teachinglabstudio.com/blog/practice-makes-progress (accessed on 25 March 2026).

Hughes, I., & Samo, A. (2026). AI and the employee lifecycle: What we know and what may come. In P. Mangos, & J. Ferraro (Eds.), *AI and gamification technologies for complex work simulation and training* (pp. 185–205). CRC Press. [CrossRef]

Jackson, P. (1986). *The practice of teaching*. Teachers College Press.

Kaufman, J., Woo, A., Eagan, J., Lee, S., & Kassan, E. (2025, February). *Uneven adoption of artificial intelligence tools among U.S. teachers and principals in the 2023–2024 school year*. RAND. Available online: https://www.rand.org/pubs/research_reports/RRA134-25.html (accessed on 25 March 2026).

Kennedy, M. (2016). Parsing the practice of teaching. *Journal of Teacher Education*, *67*(1), 6–17. [CrossRef]

Kennedy, M. (2019). How we learn about teacher learning. *Review of Research in Education*, *43*(1), 138–162. [CrossRef]

Langreo, L. (2023, May 10). Can AI do teacher observations and deliver PD? In some schools, it already does. *Education Week*. Available online: https://www.edweek.org/leadership/can-ai-do-teacher-observations-and-deliver-pd-in-some-schools-it-already-does/2023/05 (accessed on 25 March 2026).

Langreo, L. (2026, January 23). States put 'unprecedented' attention on AI's role in schools. *Education Week*. Available online: https://www.edweek.org/technology/states-put-unprecedented-attention-on-ais-role-in-schools/2026/01 (accessed on 25 March 2026).

Lavigne, A. L., & Ocran, P. (2026). The promises and pitfalls of artificial intelligence in instructional leadership: A focus on teacher supervision. *Journal of School Leadership*, 1–28. [CrossRef]

Leiva, B., Ribeiro, A., Agnew, C., & Loeb, S. (2025). *How K-12 educators are actually engaging with AI: Early findings from an edtech platform*. Stanford Scale Initiative AI Hub for Learning. Available online: https://scale.stanford.edu/research-in-action/how-k12-educators-engaging-with-ai (accessed on 25 March 2026).

Metz, C. (2026, January 20). An A.I. start-up says it wants to empower workers, not replace them. *The New York Times*. Available online: https://www.nytimes.com/2026/01/20/technology/humans-ai-anthropic-xai.html (accessed on 25 March 2026).

Mishra, P., Oster, N., & Henriksen, D. (2024). Generative AI, teacher knowledge and educational research: Bridging short- and long-term perspectives. *TechTrends*, *68*, 205–210. [CrossRef]

Morrone, M. (2026, January 21). Anthropic, Google and Microsoft fight to win teachers. *Axios*. Available online: https://www.axios.com/2026/01/21/google-anthropic-microsoft-education (accessed on 25 March 2026).

National Educational Policy Center. (2026, March 6). Can AI really personalize learning? *Newsletter*. Available online: https://nepc.colorado.edu/publication/newsletter-barbour-030526 (accessed on 25 March 2026).

Nisa: The Coach in Your Corner. (2026, March 13). *Teaching Lab studio*. Available online: https://www.teachinglabstudio.com/projects/nisa (accessed on 25 March 2026).

OECD. (2026). *Governing algorithmic management tools: What are employers doing? A session at the international conference on AI in work, innovation, productivity and skills*. Available online: https://www.oecd-events.org/e/ai-wips-2026/session/d4aca82b-ec08-f111-8330-6045bd9548dc/governing-algorithmic-management-tools-what-are-employers-doing- (accessed on 31 March 2026).

Prothero, A. (2026, February 26). What's holding educators back from adopting AI? *Education Week*. Available online: https://www.edweek.org/technology/whats-holding-educators-back-from-adopting-ai/2026/02 (accessed on 25 March 2026).

Ridge, B., & Lavigne, A. (2020). Improving instructional practice through peer observation and feedback. *Education Policy Analysis Archives*, *28*, 1–28. [CrossRef]

Russ, R. S., Sherin, B. L., & Gamoran Sherin, M. (2016). What constitutes teacher learning? In D. Gitomer, & C. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 391–438). American Educational Research Association.

Schleicher, A. (2026, January 19). How to effectively use Generative AI in education. *OECD Education and Skills Today. Blog*. Available online: https://oecdedutoday.com/how-to-effectively-use-generative-ai-in-education/ (accessed on 25 March 2026).

Senate-Commerce-Hearing-Letter. (2026, January 13). Available online: https://www.cosn.org/wp-content/uploads/2026/01/Senate-Commerce-Hearing-Letter-2.pdf#:~:text=January%2013%2C%202026.%20Dear%20Chairman%20Cruz%20and,and%20non%2Dprofit%20leadership%20organizations%20is%20writing%20to (accessed on 25 March 2026).

Showers, B., & Joyce, B. (1996). The evolution of peer coaching. *Educational Leadership*, *53*(6), 12–16.

Stave, J., Kurt, R., & Winsor, J. (2025, May 22). Agentic AI is already changing the workforce. *Harvard Business Review*. Available online: https://hbr.org/2025/05/agentic-ai-is-already-changing-the-workforce (accessed on 25 March 2026).

Suresh, A., Jacobs, J., Lai, V., Tan, C., Ward, W., Martin, J., & Sumner, T. (2021). Using transformers to provide teachers with personalized feedback on their classroom discourse: The TalkMoves application. *arXiv*, arXiv:2105.07949. [CrossRef]

Trumbore, A. (2025). *The teacher in the machine: A human history of education technology*. Princeton University Press.

Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other.*. Basic Books.

Umarani, M., & Ramu, K. (2025). AI-powered talent management: Redefining HR practices in the 21st century. *European Economic Letters*, *15*(3), 3350–3362.

Van Camp, A., Vitale, J., & Lloyd, B. (2025). Next generation classroom observations: Leveraging AI to maximize the scalability and effectiveness of performance feedback for teachers. In S. Kelly (Ed.), *Handbook of research on classroom observation* (pp. 366–381). Edward Elgar Publishing.

Wang, Y. (2025). Artificial intelligence in educational leadership: A symbiotic role of human-artificial intelligence decision-making. *Journal of Educational Administration*, *59*(3), 256–270. [CrossRef]

Ward, M. (2025, July 10). New partnership with tech giants could shape AI use in schools. *District Administration*. Available online: https://districtadministration.com/article/new-openai-partnership-aims-to-shape-ai-use-in-schools/?utm_source=omeda&utm_medium=email&utm_campaign=Newsletter-20250712&om_id=1106685108&om_eid=6244I9839312I5A&oly_enc_id=6244I9839312I5A (accessed on 25 March 2026).

Will, M. (2025, February 14). Can AI effectively coach teachers? *Education Week*. Available online: https://www.edweek.org/leadership/can-ai-effectively-coach-teachers/2025/02 (accessed on 25 March 2026).

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, *52*(1), 17–37. [CrossRef]

Woulfin, S., Stevenson, I., & Lord, K. (2023). *Making coaching matter: Leading continuous improvement in schools*. Teachers College Press.

Zhang, C., Hu, M., Wu, W., Chen, Y., Wang, K., Gai, T., & Fu, Q. (2026). AI versus teacher feedback in developing pre-service teachers' teaching self-efficacy: A quasi-experimental study in simulated teaching. *Teaching and Teacher Education*, *172*, 105355. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.