Ten Criteria for Evaluating the Quality, Relevance and Usability of Generative AI Feedback to Students

By Elie ChingYen Yu & Carla Evans

Generative Artificial Intelligence (GenAI) tools are increasingly used to provide feedback in classrooms, yet their adoption has outpaced scrutiny of the feedback they produce. In our study, we examined how GenAI feedback is characterized and whether it aligns with research-based principles of effective feedback. These 10 criteria serve as a practical tool to help educators evaluate GenAI tools for the quality of their feedback and user interface before using them.

Who should use these criteria?

Educators or school system leaders who will use or purchase GenAI tools to provide feedback to students. The goal is to support educators in asking and answering the question: *How much does GenAI feedback help students move their learning forward—by showing them where they are in relation to the learning goal and how to improve—or is this feedback potentially doing harm by providing inaccurate, misaligned, and unactionable feedback to students?*[1]

Why evaluate the quality of the feedback provided to students?

As generative AI becomes more embedded in classrooms, educators need support not only in *using* these tools but in *evaluating* the quality, relevance and usability of the feedback they produce. According to a recent RAND report, one in four U.S. teachers—and nearly 40% of English language arts and science teachers—used AI tools for instructional purposes during the 2023–2024 school year.[2] Yet guidance on *how* to evaluate these tools remains limited and uneven.

This gap matters. Research shows that GenAI feedback often appears polished and persuasive, yet may rely on tricks of language—vague phrasing, unsupported claims, or praise that doesn't match the level of performance in the student's work.[3] In other words, the feedback can sound good without being accurate, aligned to grade-level expectations or standards, or actionable. Without clear evaluation criteria, there is a risk that teachers or school system leaders may unknowingly adopt tools that fail to meet the core purpose of feedback: to move learning forward.

These ten evaluation criteria were developed to address that need. The criteria offer educators a research-informed tool for examining whether GenAI feedback is doing what matters most—supporting student learning. This tool draws on and distills decades of research about the qualities and characteristics of effective feedback, and translates evidence-based practices into

[1] By seeking answers to this question, the tool aligns with international calls for human oversight, transparency, and educator agency in AI adoption (UNESCO; U.S. Department of Education, 2023).
[2] Kaufman et al., 2025
[3] Bergstrom & West, 2021; Liang et al. 2025

educator-friendly criteria for evaluating GenAI feedback to students,[4] while also aligning with broader student-centered AI principles.[5]

How to use the evaluation criteria?

The ten criteria are divided into two sections: **(A) Quality, Relevance and Usability of the GenAI Feedback** and **(B) User Interface Supports Classroom Use**. Each criterion includes a short explanation of the research basis and guiding questions. Educators can use these criteria in multiple ways: as a rubric for selecting GenAI tools, a checklist for reviewing GenAI feedback, and/or a discussion guide for professional learning and school- or district-level decision-making. This tool highlights core elements to make the evaluation process manageable—especially for educators new to GenAI tools.

[4] e.g., Brookhart, 2017; Hattie & Timperley, 2007; Shute, 2008
[5] Smarter Balanced Assessment Consortium & IBM Consulting (2025)

Criteria	Research Basis	Evaluation Question(s)	Reference(s)
A. Quality, Relevance and Usability of the GenAI Feedback			
1. Accuracy/ Hallucination	Hallucination occurs when the model produces information that is factually incorrect, irrelevant, or unverifiable—even when the language appears fluent and confident. Educators should be aware that hallucinated feedback may go unnoticed by students and could misguide learning if not detected.	Can I trust that the feedback is factually correct and matches what the student actually did, or do I need to double-check it for errors or made-up information?	Liang et al. (2025); Ji et al. (2023); Aleksandra et al. (2025)
2. Standards-aligned	Feedback should be grounded in clear, established grade-level expectations and success criteria (e.g., standards, curriculum-based learning objectives).	Is the feedback related to the standard(s) that the students are supposed to be mastering? Would a student understand what success looks like based on the feedback?	Andrade & Heritage (2017); Ruiz-Primo & Brookhart (2018)
3. Clarity	Feedback should use clear, age-appropriate language that students can understand. It should avoid jargon and be tailored to learners' developmental levels.	Is the feedback easy to interpret and understand for learners based on their reading levels and knowledge?	Ruiz-Primo & Brookhart (2018); Shute (2008)
4. Specificity & Descriptive	Effective feedback is descriptive in nature, providing information that helps students locate themselves relative to the learning goal. It should answer: (1) Where am I going? (goal/success criteria), (2) Where am I now? (current performance), and (3) How do I get there? (actionable next steps). The feedback should be specific enough to guide improvement without doing the work for the student.	Is the feedback specific, descriptive, and helps students understand what is the issue and how they might approach addressing it? Does the feedback help students see (1) what they're working toward, (2) where they currently stand, and (3) what they can do to improve? Does it give just enough detail to guide—not do—the work?	Andrade & Heritage (2017); Hattie & Timperley (2007); Ruiz-Primo & Brookhart (2018)
5. Prioritization	Effective feedback should highlight the most important points, helping students focus on a few key improvements without overwhelming them.	Is the feedback prioritized so that students are not overwhelmed, but can focus on what is most important?	Andrade & Heritage (2017); Ruiz-Primo & Brookhart (2018)

Criteria	Research Basis	Evaluation Question(s)
B. User Interface Supports Classroom Use		
6. Feedback Customization	This criterion examines whether educators can review or customize feedback before it reaches students—such as aligning it to rubrics, learning goals, or tone. These controls help ensure feedback is relevant, accurate, and instructionally appropriate.	Can I review/edit feedback before students see it? Can I link the feedback to our rubric or learning targets? Does the tool allow uploading or referencing rubrics, learning targets/goals, or standards?
7. Validity and Reliability of Feedback	This criterion asks whether developers or ed tech companies provide evidence that the feedback their tools generate is accurate, instructionally appropriate and beneficial for student learning. This includes documentation of how feedback was tested, validated, or evaluated—through educator review, expert judgment, student outcomes, or other forms of research. This criterion is especially relevant for school leaders or teachers considering adoption, as it helps surface whether feedback can be trusted at scale before classroom use.	Has this tool been tested or reviewed to ensure the feedback helps with learning? Can I trust it to give consistent, helpful feedback to students working on the same task, project, or assignment? Does the company share any evidence—not just claims—about the quality, relevance, and usability of its feedback?
8. Types of Student Input	Students demonstrate their learning in various ways—through written responses, oral explanations, diagrams, multimedia work, etc. This criterion focuses on whether the GenAI tool can accept and interpret a range of input types that reflect these common classroom tasks.	Can the tool analyze the formats my students typically use—typed text, audio recordings, or multimodal responses?
9. Student Data Use and Privacy	This criterion evaluates whether the GenAI tool transparently communicates what student data it uses and stores—both for generating feedback and for model improvement. It also considers whether educators and institutions have control over what data is shared, where it's stored, and how long it's retained. In an educational context, student data privacy is a legal and ethical imperative. Tools that let users opt out of certain data uses or limit the granularity of what is shared demonstrate stronger trustworthiness and accountability.	Do I know what student data is being used, stored, or modeled? Can I opt out of certain data uses? Can I opt out of data sharing or delete personally identifiable information? Are there clear policies on data retention, storage location, and third-party access?
10. Learner Personalization & Accessibility	This criterion examines whether the GenAI feedback tool allows educators to customize inputs based on student characteristics—such as language, reading level, learning needs, or accessibility requirements—before generating feedback. Robust systems should support differentiation by allowing educators to input relevant background information (e.g., grade level, cultural/linguistic context) and should offer inclusive design features, such as multilingual support, text simplification, and screen reader compatibility. These elements help ensure that feedback is appropriately tailored, developmentally suitable, and usable for diverse learners, including those with disabilities or learning differences.	Can I tell the system what to consider about my student—such as reading level, language, or learning needs—before feedback is generated? Are there features (e.g., simplified language, text-to-speech, screen reader compatibility) that support accessibility and developmental appropriateness?

References

Aleksandra, N., Bojana, J., Maryan, R., & Dimitar, T. (2025). Evaluating Trustworthiness in AI: Risks, Metrics, and Applications Across Industries. *Electronics, 14*(13), 2717. https://doi.org/10.3390/electronics14132717

Andrade, H., & Heritage, M. (2017). *Using assessment to enhance learning, achievement, and academic self-regulation.* Routledge.

Bergstrom, C. T., & West, J. D. (2021). *Calling bullshit: The art of skepticism in a data-driven world*. Random House Trade Paperbacks.

Brookhart, S. M. (2017). *How to give effective feedback to your students*. ASCD.

Hattie, J. A. C., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Ji et al. (2023). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys.

Kaufman, J. H., Woo, A., Eagan, J., Lee, S., & Kassan, E. B. (2025, February 11). Uneven adoption of artificial intelligence tools among U.S. teachers and principals in the 2023–2024 school year (Report No. RRA 134-25). RAND Corporation. https://www.rand.org/pubs/research_reports/RRA134-25.html

Liang, K., Hu, H., Zhao, X., Song, D., Griffiths, T. L., & Fisac, J. F. (2025). Machine Bullshit: Characterizing the emergent disregard for truth in large language models. *arXiv preprint arXiv:2507.07484*.

Ruiz-Primo, M. A., & Brookhart, S. M. (2018). *Using feedback to improve learning*. Routledge. https://doi.org/10.4324/9781315627502

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. https://doi.org/10.3102/0034654307313795

Smarter Balanced Assessment Consortium & IBM Consulting. (2025). *SmarterAI Think Tank: Student-centric design principles for responsible use of AI*. The Regents of the University of California. https://portal.smarterbalanced.org/hubfs/44715956/SmarterAI%20Think%20Tank.pdf

U.S. Department of Education (2023). Artificial intelligence and future of teaching and learning: Insights and recommendations. Office of Educational Technology, Washington, DC. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

United Nations Educational, Scientific and Cultural Organization. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535