



How Do We Know Students Are Learning?

Rethinking the evidence behind the degree in the age of AI

Charlena Miller, Co-Founder and CEO of HumanWise

Every year, the public school I attended held a used-textbook sale, with tables of books arranged by subject and grade level. I was the eight-year-old with saved-up chore money, choosing books I would take home and work through on my own. Something I read would send me wondering, and a problem would grab hold of me for minutes, hours, or sometimes days. Then came the rush of recognition: *Oh. Now I see.* I wanted that feeling again and again.

In college as a single mother, I needed the degree to earn more than minimum wage and provide for my family. I still wanted to learn, and I also needed to finish. I remember this when I hear conversations about students taking shortcuts. Understanding their pressures does not excuse cheating, but it shapes the question I want to ask here: How do we know students have learned what we claim when we give them a grade and a degree?

What the work can and cannot tell us

This question is not new, but generative AI has made it harder to answer because AI can produce artifacts that look like learning. Picture a student asked to recommend whether a community center should extend its evening hours. The proposal includes staffing costs, demand estimates, and a sensible recommendation. Suppose AI supplied the assumptions, and the student edited the prose and checked the arithmetic. If the assignment allowed that help, nothing dishonest happened. We still do not know whether the student can judge if the assumptions make sense, or what happens to the recommendation if attendance runs low.

Learning research offers a related caution. Soderstrom and Bjork (2015) distinguish performance during instruction from learning that lasts and can be used later. Better performance does not mean learning has occurred. That is why I separate two conversations we often combine. Academic integrity concerns honesty and the conditions under which work is completed. Learning concerns what students become able to do. A student can follow an AI policy completely and still hand over too much of the thinking, and

permitted assistance can make demanding work more reachable and support real learning. Policy needs to address both.

Give students reasons to learn

Many students come to college for practical reasons, and I do not read that as a lack of seriousness. A student who wants to become a competent professional has a reason to care about the knowledge and judgment the work will require. Interest can also be developed. Hidi and Renninger (2006) describe how a spark of interest in one situation can grow into a lasting interest in a subject.

Asking students to teach can help. In a meta-analysis of 28 studies, Kobayashi (2019) found learning benefits from preparing to teach, and from expecting to teach, compared with studying without that expectation. When students teach, they decide where a newcomer would begin, what might be misunderstood, and which example would help. Their preparation gets an audience, and their revision gets a purpose.

Make difficulty productive

A student who has struggled repeatedly without reaching understanding has little reason to expect the next struggle to pay off, and calling an assignment “productive struggle” does not make it feel achievable. Research on productive failure suggests that attempting a problem before instruction can work well for novice learners, as long as the instruction that follows is deliberately designed (Sinha & Kapur, 2021).

Effective methods can also feel worse to the people using them. Deslauriers et al. (2019) found that students in active-learning physics classes learned more while feeling they had learned less than students in lectures. So I talk with my students about what felt difficult, revisit an earlier explanation, and look at what they can now do with it. Small, visible progress makes the next attempt feel worth making.

When AI supports learning and when it does too much

The same tool that can scaffold a student’s thinking can also finish the work for them. In a field experiment in Turkish high school mathematics, Bastani et al. (2025) compared practice with a general GPT-4 chatbot, practice with a GPT-4 tutor built with teacher guidance, and practice without AI. Both AI tools improved performance during practice. When students later worked without them, those who had used the general chatbot did worse than the control group, while the tutor group largely avoided that drop. It is one school and one subject, so it is not a verdict on students’ cognitive abilities. It does show that outcomes depend on how the assistance is designed.

In my own classes I use custom tutor bots configured with my teaching philosophy and examples of my feedback. They are constrained from giving answers or examples that would do too much of the thinking. Students work through part of an assignment with the bot, then talk with me about what they figured out, where they are still stuck, and what help they need.

From assignments to the degree

Individual faculty can ask whether an assignment gives meaningful evidence of the capability it is meant to develop. Academic leaders can ask a larger question that no single course can answer: whether the evidence gathered across a program supports what the institution claims when it awards the degree.

A practical start is to choose one capability graduates are expected to demonstrate and trace where students first encounter it, where they practice it, and what evidence supports the final judgment of competence, including what students can do with AI and without it. Students can contribute to that review. They can say where AI helped them understand something, where they accepted an answer they could not explain, and what made it easier to ask a person for help.

AI accelerated a conversation higher education was already having: how to judge learning from work completed under conditions that offer only a partial view of what students know. I understand the student who wants to finish and improve a family's life. I also remember the girl who bought more used textbooks than she could carry. I hope students keep room for both, and that we find ways to assess learning without reducing teaching to policing it.

To bring this conversation to your campus, visit humanwiselearning.com/contact.

Sources

- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122.
- Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, 116(39), 19251-19257.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111-127.
- Kobayashi, K. (2019). Learning by preparing-to-teach and teaching: A meta-analysis. *Japanese Psychological Research*, 61(3), 192-203.
- Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*, 91(5), 761-798.
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176-199.