

The Tool is Not the Pedagogy:

AI, Literacy Instruction, and the Science of Learning



Russ Kliman
AIM Institute
Head of Innovation, Product, & Digital Strategy

Artificial intelligence is moving into literacy instruction faster than most schools, districts, researchers, and professional learning systems can fully process. In a recent RAND study, the share of students who reported using AI for homework rose from 48% in May 2025 to 62% just seven months later. At the same time, students were also asked to reflect on their use of AI, with 52% of middle schoolers and 60% of high schoolers reporting concern that their own critical thinking was being harmed by their increased use of AI for schoolwork.

Education leaders navigating this rapid pace of AI adoption alongside concerns about student learning, need a framework to support them in making good instructional decisions on how AI is used in their classrooms. Unfortunately, the AI conversation has too often been pulled into a predictable debate: "AI is good. AI is bad. AI is transformational. AI is dangerous." That framing is understandable, but it does not get to the heart of what education leaders should be asking. The better question is not whether AI belongs in literacy instruction. The better question is what

© 2026 AIM Institute for Learning & Research

AI is being asked to do. Is it doing the work for students? Is it supporting students while they do the work? Or is it helping teachers and students see the work more clearly? That is the leadership question. And it is also an instructional question.

AI is not a pedagogy. It is not a curriculum. It is not an intervention. It is not a theory of learning. AI is simply a set of technological capabilities. Those capabilities can support learning, accelerate access, personalize practice, and give teachers more insight into student thinking. And unfortunately they can also bypass the very work students need to do to become stronger readers, writers, speakers, and thinkers.

The issue is not the acronym. The issue is the application. The way we use AI to support learning, not the tool itself, is pedagogy. That should be the starting point for any meaningful conversation about AI and literacy instruction. For leaders, this means AI adoption should begin with five practical moves: name the learning problem before naming the tool, decide whether AI is acting as a substitute, scaffold, or mirror, protect the student's cognitive work, preserve teacher visibility, and pilot with evidence before scaling.

> *"AI adoption should not begin with the tool. It should begin with the learning we are trying to protect, strengthen, and make more visible."*

Those moves do not require leaders to reject AI or rush toward it. They require something harder and more useful: disciplined optimism.

We are asking the wrong first question

A lot of conversations around AI are still focused on the tool itself. Leaders are rightly concerned about accuracy, bias, data privacy, hallucinations, student misuse, and the quality of AI-generated content. Those issues matter, and they cannot be ignored. But they are not the only questions that matter.

Leaders and educators are also asking whether AI can generate a lesson plan, summarize a text, create questions, draft feedback, adapt content, translate language, or produce student-facing materials. And in many cases, the answer is yes. But just because AI can do something does not mean the use of it is instructionally sound. This is where human judgement, experience, and oversight is critical.

In education, and especially in literacy instruction, the intervention is not the tool by itself. The intervention is the intentional use of a tool or instructional routine with a specific student, toward a specific learning goal, within a coherent instructional design, with a clear feedback and adjustment loop. That is where the conversation needs to move. For example, an AI-enabled writing tool can generate a stronger

paragraph for a student. But does it produce a stronger writer? A reading tool powered by AI can summarize a passage. But does it help the student build comprehension, vocabulary, background knowledge, inference-making, or stamina with complex text? As Shanahan (2026) states it, when a student summarizes a text, they may or may not learn something, but when AI does the summarizing, "a lack of learning is guaranteed."

> *"The better question is not whether AI can do the task. The better question is whether the task still belongs to the learner."*

These are not anti-AI questions. They are instructional questions. They force us to decide what we will allow, what we will model, what we will measure, and what we will improve.

Literacy is not the production of fluent output

One reason these instructional questions matter so much in literacy is that AI can create misleading work products. A fluent summary of a book is not always evidence of comprehension. A grammatically correct paragraph is not always evidence of writing development. A polished essay is not always proof that a student can plan, organize, reason, use evidence, revise, and make decisions as a writer. AI can make student work look literate before the student has become more literate. Said simply, literacy is deeper than the appearance of fluent output.

The science of reading gives us a strong foundation for thinking about this distinction. Skilled reading depends on the integration of word recognition and language comprehension, as described in the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990; Castles et al., 2018; Foorman et al., 2016). This means students need explicit instruction in foundational skills such as phonemic awareness, phonics, decoding, word recognition, and fluency, while also developing vocabulary, syntax, background knowledge, inference-making, comprehension monitoring, and meaningful engagement with text (Castles et al., 2018; Foorman et al., 2016). Phonics is central in alphabetic systems, but skilled reading extends well beyond phonics alone. The implication for AI is clear: a tool that produces fluent language may create the appearance of literacy without showing whether students are developing the component skills and knowledge that skilled reading and writing require.

© 2026 AIM Institute for Learning & Research

> *"AI can make student work look literate before the student has become more literate."*

As Hoover and Tunmer (2022) emphasize, skill in one component cannot compensate for any limitations in the other. In other words, skilled reading requires both accurate, efficient word recognition and strong language comprehension. Strength in one area cannot make up for weakness in the other.

A better leadership frame: substitute, scaffold, or mirror

A more useful way to evaluate AI in literacy instruction (or any instruction for that matter) is to ask whether the tool is acting as a substitute, a scaffold, or a mirror.

As a **substitute**, AI ***does the work*** for the student. It writes the response. It summarizes the text. It answers the comprehension question. It revises the sentence so completely that the student no longer has to make decisions. The artifact may look better, but the teacher will know less about what the student can actually do.

As a **scaffold**, AI can support the student while keeping the student ***in the work***. It gives multiple hints and cues before an answer. It asks the student to reread. It prompts the student to find evidence. It helps the student clarify a claim. It supports vocabulary or background knowledge. It gives feedback that requires revision. The student still has to think, choose, explain, and revise.

As a **mirror**, AI can help the student and teacher ***see the work*** more clearly. It may surface patterns in writing, identify where evidence is weak, show where a response drifts from the question, or help a teacher notice common misconceptions across student work. In that case, AI is not replacing teacher judgment. It is making student thinking more visible so the teacher can respond more precisely.

Consider the same AI writing tool use in three different classrooms. In one classroom, the student asks the tool to write the response. That is substitution. In another, the student writes first, then uses AI to ask where the evidence is weak or where the claim needs to be clarified. That is scaffolding. In a third, the teacher uses AI to review a set of student drafts and identify common patterns in evidence use, sentence control, or organization. That is mirroring. The tool did not change. The pedagogy did.

© 2026 AIM Institute for Learning & Research

> *"The same tool can replace the work, support the work, or reveal the work. The difference is not the technology. The difference is the pedagogy."*

This is why the "AI-good vs. AI-bad" conversation tends to miss the mark. The same tool can function in all three ways depending on the task design. An AI writing assistant can substitute if it writes the essay. It can scaffold if it helps the student strengthen a claim after the student has made an attempt. And it can mirror if it helps the student notice patterns in sentence structure, word choice, or evidence use.

The issue is not AI in isolation. The issue is what role AI is playing in the learning journey.

Productive struggle matters

This is also where the science of learning becomes critically important. Learning is not just exposure to information. It involves attention, memory, retrieval, practice, feedback, motivation, self-regulation, prior knowledge, and transfer (Dunlosky et al., 2013; National Academies of Sciences, Engineering, and Medicine, 2018). How People Learn II reminds us that learning happens through the interaction of the learner, the task, the environment, and the supports around the learner (National Academies of Sciences, Engineering, and Medicine, 2018). So if AI makes something faster, that does not automatically mean it makes learning stronger. And if AI makes something easier, that does not automatically mean it makes learning more accessible in the right way. There is a difference between reducing barriers and removing the work.

RAND's researchers frame this as the difference between cognitive offloading, using AI to do the mental work for students, and cognitive augmentation, using AI to spur deeper or more robust work (Schwartz & Diliberti, 2026). Productive struggle lives inside that distinction.

Productive struggle does not mean we leave students confused or unsupported. It means we preserve the meaningful cognitive work of learning while giving students enough support to keep moving (National Academies of Sciences, Engineering, and Medicine, 2018). In literacy, that work may include decoding an unfamiliar word, rereading a sentence, revising a claim, finding stronger evidence, explaining an inference, reorganizing a paragraph, or monitoring comprehension when a text becomes difficult. AI can support that work. But it can also erase it.

© 2026 AIM Institute for Learning & Research

> *"Faster is not always better, and easier is not always more accessible. Leaders have to distinguish between reducing barriers and removing the work learning requires."*

It supports productive struggle when it gives a student a nudge, a prompt, a model after an attempt, or feedback that sends the student back into the work. It erases productive struggle when it generates the summary, writes the answer, polishes the essay, or gives the student language that sounds like understanding before the student has actually built understanding.

Students do not become stronger readers and writers simply by seeing or observing strong reading and writing. They become stronger by doing the work, with the right instruction, the right practice, the right feedback, and the right support over time. AI should make that work more possible. It should not make that work unnecessary.

The feedback opportunity is real, but it needs guardrails

One of the strongest learning opportunities for AI is feedback. Teachers cannot possibly give immediate, detailed, individualized feedback on every student response, every draft, every annotation, every vocabulary attempt, and every comprehension breakdown at scale. So there is something very compelling about tools that can provide students more frequent feedback and give teachers more insight into learning patterns. The risk is that feedback is not automatically useful.

Hattie and Timperley's (2007) work is important here. They describe feedback as one of the most powerful influences on learning, but they also make clear that feedback can be positive or negative depending on the type of feedback, the timing, the level, and the way students use it. Effective feedback helps learners understand where they are going, how they are doing, and what they need to do next (Hattie & Timperley, 2007). Feedback quality depends on the task, the criteria, the instructional purpose, and whether the feedback keeps the student responsible for the next move.

A useful leadership test is this: does the feedback return responsibility to the student, or does it quietly take responsibility away? If AI fixes the writing, the product may improve while the learner does not. If AI gives generic praise, the student may feel encouraged without knowing what to do next. But if AI points the student back to the task, the evidence, the sentence, the word choice, or the revision decision, then the tool may be supporting learning rather than replacing it.

© 2026 AIM Institute for Learning & Research

> *"Good feedback does not rescue students from the work. It points them back into the work with a clearer next move."*

That is the same standard we should apply with AI feedback. If the tool simply fixes the student's writing, we should be cautious. If it gives generic praise, we should be cautious. If it rewrites the student's thinking so thoroughly that the teacher can no longer see what the student understands, we should be cautious. But if the tool points the student back to the task, that is different. Feedback such as, "Your claim is clear, but your evidence does not yet support it. Go back to paragraph three and find the sentence that best proves your point," keeps the student responsible for the next move. So does feedback such as, "You used the word 'things' three times. Replace each one with a more precise noun." In those cases, AI is not simply improving the product. It is helping the student re-enter the work with a clearer next step. The challenge is that most AI tools do not do this automatically or consistently. That kind of use requires instructional design, teacher judgment, and clear guardrails.

The market is ahead of the research

This is the part of the AI in education conversation that everyone, including leaders, needs to acknowledge openly. AI in education did not begin with ChatGPT in November 2022. Intelligent tutoring systems, adaptive learning environments, automated feedback tools, and learning analytics have been researched for decades (Kulik & Fletcher, 2016; Ouyang & Jiao, 2021; Zawacki-Richter et al., 2019). And that research tells us that technology can support learning under the right conditions, but effects are not automatic (Kulik & Fletcher, 2016; Ouyang & Jiao, 2021). They depend on the instructional design, the learning goal, the student, the teacher, the implementation, and the way outcomes are measured.

When it comes to generative AI-enabled tools, the market is ahead of the research (Kasneci et al., 2023; Bond et al., 2024). It's also ahead of many school systems' governance structures. The behavioral data bear this out: even as student AI use climbed sharply through 2025, most students reported that their schools had no schoolwide rule governing it, with that ambiguity greatest in middle school (Schwartz & Diliberti, 2026). That does not mean generative AI-enabled tools have no value. And it doesn't mean we should do nothing. It means the claims around them need to be treated as hypotheses until there is evidence behind them.

Since the release of ChatGPT, schools have been flooded with tools that promise personalization, faster feedback, easier lesson planning, adaptive support, and greater access for students. Some of those tools may prove incredibly valuable. Some may save teachers time. Some may strengthen practice, feedback, writing, vocabulary development, or access to complex text. But an impressive demo at a

conference or during a Zoom meeting is not the same as evidence of improved literacy outcomes in the classroom.

> *"An impressive AI demo is not the same as evidence of improved learning."*

This is why the emerging work of groups like Boston University's Evidence-Based AI in Learning (EVAL) Collaborative matters. BU's work on independent evaluation of AI tools for K–12 education reflects a larger need in the field - that schools should not have to rely only on vendor claims, marketing materials, or conference demonstrations when making decisions about tools that affect children's learning. Educators need better ways to examine what a tool actually does, what evidence supports it, and whether it improves outcomes for the students it claims to serve.

That distinction is especially important in literacy. Quickly creating AI-generated passages, questions, summaries, writing prompts, feedback, or practice activities that are polished, is not the same as instructional quality. A tool may produce materials that look useful while still misaligning with how students learn to read, write, and build language. A tool may appear to personalize learning while simply adjusting task difficulty without addressing the underlying skill need. A tool may support access without strengthening the instructional journey that helps a student become more independent.

The right approach should be neither blanket resistance nor blind adoption. The right approach should be disciplined optimism.

Implementation is the work

This is where implementation science and change management need to be brought into the AI conversation. Any tool can be promising, research-informed, and well-designed, and still fail if it is dropped into a school system without the conditions that make it usable. Implementation science reminds us that adoption is not the same as implementation. Buying the tool, announcing the tool, or training people how to apply the tool correctly is not the same as changing practice in a way that improves outcomes (Fixsen et al., 2005; Proctor et al., 2011).

And this matters for AI because AI is not just another digital resource on the shelf. It has the potential to change workflows, decision-making, student behavior, teacher judgment, feedback cycles, assessment signals, and sometimes even what people think counts as learning. That means leaders have to think beyond access and compliance. They have to think about readiness, fit, coaching, pedagogy, shared language, guardrails, data routines, communication, and the everyday moments

where teachers and students decide how the tool will actually be used in the classroom.

> *"Adoption is not implementation. Access to AI does not mean a system knows how to use it well."*

Change management is the human side of this. People need to know why the change matters, what is expected, what is not expected, what support they will receive, and how the system will learn from what is happening. If we use it effectively, AI can actually support that work. It can help leaders see patterns in teacher feedback, summarize implementation barriers, draft clearer guidance, compare student work samples, surface common misconceptions, and make the invisible parts of change more visible. But again, the same rule applies: AI should support the thinking, not replace the judgment. It should help a system learn faster without pretending that change is just a technical rollout.

In practical terms, this means leaders should not only ask, "Do teachers have access?" They should also ask: Do teachers know when AI use is instructionally appropriate? Do students know what work still belongs to them? Do coaches and school leaders know what strong implementation looks like? Do teams have a way to look at student work, not just usage data? Does the system know when to stop, revise, or scale?

The resulting implementation question is not, "How do we get everyone using AI?" The better question is, "What learning problem are we trying to solve, what practice has to change, what support will make that change possible, and how will we know whether students are actually better off?" That is where AI moves from a conversation piece to disciplined action.

A leadership test before adopting AI

Before adopting or scaling AI, leaders need a practical way to test whether the use case is instructionally sound. The goal is not to slow innovation for the sake of caution. The goal is to make sure AI is being used in service of learning, not simply in service of access, efficiency, or output.

A useful leadership test begins with five questions:

1. What literacy outcome is this supposed to improve? Is the goal decoding accuracy, fluency, vocabulary, comprehension, written organization, sentence-level control, evidence use, revision, oral language, background knowledge, access, teacher efficiency, or something else?

© 2026 AIM Institute for Learning & Research

2. How will we know whether that outcome improved? Usage, logins, engagement, satisfaction, and time saved may tell us something. But they do not tell us whether students are learning more, performing more independently, or transferring learning when the tool is removed.
3. What work is the AI doing, and what work is the student still doing? If AI is generating the core response, the student may not be practicing the intended skill. If AI is prompting, questioning, giving feedback, or helping the student revise after an attempt, the student may still be doing the cognitive work.
4. What does the teacher see more clearly because of this tool? Strong tools should make student thinking more visible. Weak tools may make student work look better while making student thinking harder to see.
5. What evidence would cause us to stop, revise, or scale? A disciplined pilot should not be designed only to prove that a tool works. It should help leaders learn under what conditions the tool is useful, for whom, and why.

These questions shift the leadership task. It moves AI from a tool conversation to an instructional decision-making conversation. The leadership test helps teams evaluate whether a particular AI use case is instructionally sound. The five moves that follow help leaders turn that judgment into disciplined adoption across a school or system.

> ***The Golden Rule Question:*** *What thinking is this student meant to do, and does this AI use support it or replace it?*

Five moves for disciplined AI adoption

If the tool is not the pedagogy, then leadership cannot simply be about access, purchasing, permission, or prohibition. Leaders have to help their systems make better instructional decisions. And that means moving the conversation from "Are we using AI?" to "Are we using AI in a way that protects and strengthens learning?"

- **Name the learning problem before naming the tool.** Start with the instructional need. Are we trying to improve decoding accuracy, fluency, vocabulary, comprehension, written organization, evidence use, revision, access to complex text, teacher feedback, or intervention planning? The tool should enter the conversation after the learning need is clear, not before.
- **Decide whether AI is acting as a substitute, scaffold, or mirror.** If AI is producing the core response, it may be replacing the work. If it is prompting, questioning, giving feedback, or helping a student revise after an attempt, it may be supporting the work. If it helps teachers see patterns in student thinking more clearly, it may be strengthening instructional judgment.

- **Protect the student's cognitive work.** Faster completion is not the same as stronger learning. Cleaner work is not the same as deeper understanding. Students still need to read, decode, reread, infer, explain, draft, revise, select evidence, monitor meaning, try again, and transfer the skill when the tool is removed.
- **Preserve teacher visibility.** A good AI use should help teachers see more, not less. What did the student do independently? What did the tool suggest? What decision did the student make next? What misconception remains? If a tool makes student work look stronger while making student thinking harder to see, leaders should slow down.
- **Pilot with evidence before scaling.** Usage, engagement, satisfaction, and time saved may tell part of the story, but they are not enough. Leaders should look at student work, teacher feedback, classroom routines, and transfer to unaided performance. Then they should make the next decision based on evidence, not excitement.

These moves do not require leaders to have all of the answers before they act. They require leaders to act with discipline. Start with the learning problem. Define the role of the tool. Protect the student's work. Preserve what teachers can see. And scale only when there is evidence that students are better off.

> *"Disciplined AI adoption starts with the learning problem, defines the role of the tool, protects the student's work, preserves teacher visibility, and scales only when evidence says students are better off."*

The tool is not the pedagogy

The future of AI in literacy instruction will not be determined by whether educators declare AI good or bad. It will be determined by whether leaders develop the judgment, discipline, evidence standards, and instructional guardrails to use AI in the right context and use it well. That means moving the conversation from the acronym to the application.

What is the tool doing? What is the student doing? What is the teacher able to see? What learning outcome is being supported? What evidence shows that learning transfers beyond the tool? And does the technology reduce barriers, or does it remove the work students need to do?

© 2026 AIM Institute for Learning & Research

AI will become a meaningful part of the literacy ecosystem. It may extend the reach of teachers, specialists, and interventionists. It may help students get more feedback, more access, more practice, and more opportunities to revise and reflect. But it cannot become a shortcut around the learning journey itself.

Literacy develops through language, attention, practice, knowledge, error, feedback, revision, and meaning-making. AI can support those conditions. It can help create more of them. It can make some of them more accessible. But it should not be allowed to stand in for them.

Leaders should not ask only whether their systems are using AI. They should ask whether AI is preserving the work that learning requires.

The promise is not in the acronym of AI. It is in the application. The tool is not the pedagogy.

REFERENCES

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). *A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour*. *International Journal of Educational Technology in Higher Education, 21*, Article 4.

Castles, A., Rastle, K., & Nation, K. (2018). *Ending the reading wars: Reading acquisition from novice to expert*. *Psychological Science in the Public Interest, 19*(1), 5–51.

Crompton, H., & Burke, D. (2023). *Artificial intelligence in higher education: The state of the field*. *International Journal of Educational Technology in Higher Education, 20*, Article 22.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). *Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology*. *Psychological Science in the Public Interest, 14*(1), 4–58.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Hayes, L., Justice, L., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). Foundational skills to support reading for understanding in kindergarten through 3rd grade. Institute of Education Sciences, What Works Clearinghouse.

Hattie, J., & Timperley, H. (2007). *The power of feedback*. *Review of Educational Research, 77*(1), 81–112.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., et al. (2023). *ChatGPT for good? On opportunities and challenges of large language models for education*. *Learning and Individual Differences, 103*, Article 102274.

Kulik, J. A., & Fletcher, J. D. (2016). *Effectiveness of intelligent tutoring systems: A meta-analytic review*. *Review of Educational Research, 86*(1), 42–78.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press.

Ouyang, F., & Jiao, P. (2021). *Artificial intelligence in education: The three paradigms*. *Computers and Education:* Artificial Intelligence, 2, Article 100020.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. Administration and Policy in Mental Health and Mental Health Services Research, 38, 65–76.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? International Journal of Educational Technology in Higher Education, 16, Article 39.

© 2026 AIM Institute for Learning & Research