The Sethi Ledger

PEOPLE / PRACTICE / POSSIBILITY

LEADERSHIP • WEEKEND RESEARCH SUPPLEMENT

Learning needs more than an AI tool

September 19-25, 2026

This week's research puts the focus on the design of help and the evidence needed to judge learning. The practical question is what students can understand and do after the tool is removed.

The week at a glance

01 Evaluate motivation and independent learning separately. A positive experience is a reason to investigate further, not a substitute for demonstrated understanding.

02 Specify the human role in tutoring pilots. Responsive help, teacher review and independent assessment belong in the program design and budget.

03 Read technical claims carefully. Prediction accuracy, simulated gains and expert review answer different questions from a classroom learning study.

04 Use staff and student perspectives to shape clear examples of acceptable help, while testing adult evidence cautiously in K-12 settings.

ABOUT THIS BRIEF

Selected research published September 19-25, 2026. Findings are assessed for relevance to K-12 leadership, evidence quality and practical implications. Each summary identifies important limitations.

01 / LEADERSHIP

Motivation gains need a local learning test

Peer-reviewed meta-analysis
Fang and colleagues / Frontiers in Psychology

THE FINDING

Does generative AI make students more motivated to learn? A new meta-analysis combines 42 controlled studies involving 6,059 participants and reports an average positive effect of 0.764 standard deviations. That average conceals substantial variation: the prediction interval stretches from a negative effect to a large positive one. The included studies span primary, secondary and university education; exploratory estimates were smaller in school settings than in universities. The finding supports testing thoughtfully designed uses, but does not establish that a particular district product will help. Motivation was the outcome under review. Better grades, durable understanding and independent performance cannot be inferred from the pooled motivation result.

THE EVIDENCE

The synthesis covers 56 effects from experimental and quasi-experimental studies published in 2023-2025. Heterogeneity was high: I-squared 93.7%.

LIMITS TO KEEP IN VIEW

Study-level subgroup differences are exploratory. A positive average does not guarantee local benefit; instruments, interventions and populations vary substantially.

WHAT LEADERS CAN DO

Editorial recommendation: Have curriculum leaders define success before expanding a pilot. Pair a brief motivation survey with an unaided task that uses new material and a delayed assessment. Compare with normal instruction and examine students needing additional support separately. A student who enjoys the tool but cannot explain the work needs a different instructional response from one who gains both confidence and understanding. Budget teacher time for reviewing those differences.

WHO / WHAT WAS STUDIED

Primary, secondary and university learners; mixed educational contexts.

READ THE ORIGINAL SOURCES

GenAI and learning motivation: meta-analysis
www.frontiersin.org · Published 2026-09-24 · Open source

02 / LEADERSHIP

A tutor should adjust help to the learner

Preprint; within-participant experiment
Hou and colleagues / arXiv

THE FINDING

How can an AI tutor help without taking over? CoMeT increases assistance when a learner struggles with a specific decision, then reduces it when the learner responds. In a study of 131 adults completing three Python tasks, researchers compared it with an unrestricted assistant and a question-only tutor. CoMeT prompted more reported effort than the assistant and less frustration than the question-only design. It surrendered the complete answer in 6.1% of sessions, compared with 16.0% for the question-only tutor. These are promising interaction results, not proof of stronger learning: the analysis included only participants who completed all tasks, in one sitting.

THE EVIDENCE

Each participant used all three tutors; tutor order was balanced. Shared model and task infrastructure helped isolate the support design.

LIMITS TO KEEP IN VIEW

Preprint, adult programming context and completer-only analysis. No delayed unaided assessment establishes retention; findings do not validate a school-age deployment.

WHAT LEADERS CAN DO

Editorial recommendation: Add a demonstration to tutor procurement. Ask a vendor to show what happens when a student makes a plausible mistake, repeats it, asks for the answer and then explains a correction. Require a teacher-readable record of the support given. In a pilot, observe whether students can transfer the idea without the tool; completion rates alone cannot answer that question.

WHO / WHAT WAS STUDIED

131 adult learners, 393 task sessions; indirect K-12 evidence.

READ THE ORIGINAL SOURCES

CoMeT: contingent AI tutoring, version 1
arxiv.org · Published 2026-09-19 · Open source

03 / LEADERSHIP

Make robotics learning visible beyond the robot

Preprint; curriculum design and expert review
Dong, Cao and Wang / arXiv

THE FINDING

A high-school robotics curriculum connects assembly, simulation and reinforcement learning through a shared humanoid robot project. Pairs of students work toward training a walking behavior and transferring it to hardware. Five experts reviewed the design, exposing tensions between authentic engineering and novice workload, system integration and early progress, and teamwork and individual accountability. The proposed revisions introduce supplied code, bounded choices and staged checkpoints. A walking robot can therefore be a useful project outcome while still leaving a separate assessment question: what did each learner understand? The paper reports expert feedback on curriculum design; student learning data are reserved for separate work.

THE EVIDENCE

The panel included robotics, engineering and secondary STEM expertise. The revised responses to its feedback had not been re-evaluated.

LIMITS TO KEEP IN VIEW

Preprint and small formative expert panel. No comparative learning gains or demonstrated district implementation cost are established by this study.

WHAT LEADERS CAN DO

Editorial recommendation: For an elective or summer pilot, define an individual explanation task alongside the team build. Have each learner predict how a change will affect behavior, inspect a result and explain the mismatch. Fund instructor preparation, safe testing space and replacement parts before scaling enrollment. Keep a simulation activity ready for hardware downtime, with its own learning objective and clear limits.

WHO / WHAT WAS STUDIED

Curriculum designed for grades 9-12; five expert reviewers.

READ THE ORIGINAL SOURCES

High-school reinforcement learning curriculum, version 1
arxiv.org · Published 2026-09-22 · Open source

04 / LEADERSHIP

Prediction accuracy is not a learning gain

Peer-reviewed computational benchmark
Zhang and Zhang / Discover Artificial Intelligence

THE FINDING

An adaptive-learning paper illustrates a distinction district buyers need to make. Its system combines predictions of engagement and achievement with recommendations, using existing OULAD and ASSISTments data. On OULAD, its engagement classification accuracy was 95.8%, but a conventional XGBoost model performed better overall. The ASSISTments component used retrospective simulation; its reported learning gains were simulated measures, not changes observed after teaching students. The study contributes a technical design and a benchmark comparison. It does not show that assigning the system's recommendations improves achievement. Even the engagement label represents behavior such as activity logs, rather than a direct measure of a learner's thinking or interest.

THE EVIDENCE

Historical datasets support prediction experiments. Five-fold comparisons and sensitivity analyses examine model behavior, rather than a new classroom intervention.

LIMITS TO KEEP IN VIEW

Behavioral proxies can be misread as psychological engagement. Retrospective and simulated results cannot establish causal classroom improvement or local student-level validity.

WHAT LEADERS CAN DO

Editorial recommendation: Require a vendor to separate three claims in its evidence packet: predicting an outcome, choosing an intervention and improving the outcome after that intervention. Ask for the comparison group and independently measured student outcome supporting the third claim. If only benchmark accuracy is available, treat the product as a limited evaluation candidate and preserve educator review of every consequential recommendation.

WHO / WHAT WAS STUDIED

OULAD higher-education records and historical ASSISTments learning data.

READ THE ORIGINAL SOURCES

Adaptive learning: predictions and simulated validation
link.springer.com · Published 2026-09-24 · Open source

05 / LEADERSHIP

Science AI research leaves governance gaps

Peer-reviewed review and research mapping
Molnar / Discover Education

THE FINDING

What does the science-education research base actually cover? A review maps 80 peer-reviewed studies across technology, curriculum integration and ethics. Personalization and adaptive learning dominate, accounting for 47 studies, while privacy and data protection appear as the coded ethical concern in only one. Higher education receives more attention than K-12 or teacher education. These patterns identify gaps in what the selected literature examines; they do not establish that a particular school tool is unsafe or ineffective. The review's proposed framework links the technology being used, its instructional role and its ethical implications. Those relationships are conceptual propositions, rather than demonstrated causes of better learning.

THE EVIDENCE

The author combined bibliometric and qualitative analysis of English-language studies indexed in Web of Science and Scopus, spanning 1990-2026.

LIMITS TO KEEP IN VIEW

Single-author coding and two-database coverage constrain the map. Publication patterns are not effect estimates, and omitted topics may be studied elsewhere.

WHAT LEADERS CAN DO

Editorial recommendation: Ask each science pilot team to complete a short evidence map: the concept students will learn, the activity AI changes, the student information required and the evidence the teacher will collect. Assign curriculum and technology leaders joint responsibility for unresolved questions. An attractive simulation should earn a place through its instructional purpose, with an equally clear account of the information it consumes.

WHO / WHAT WAS STUDIED

80 research publications; higher education predominates.

READ THE ORIGINAL SOURCES

AI in science education: curriculum and ethical gaps
link.springer.com · Published 2026-09-23 · Open source

06 / LEADERSHIP

Teachers and learners see AI differently

Peer-reviewed qualitative survey
Ang / Discover Education

THE FINDING

An exploratory Singapore study asked educators and students to complete statements about how AI had changed education, was changing it and might change it next. The 223 respondents included 71 educators and 152 university students. Educators emphasized assessment, standards and institutional responsibilities; students emphasized access to information, learning support, efficiency and organizing their work. The groups often recognized similar developments but understood their significance through different roles. These responses describe how people interpret change, rather than measuring what AI caused. The references to past, present and future also come from a single survey, not repeated observations tracking a population over time.

THE EVIDENCE

Inductive thematic analysis was followed by descriptive content analysis. Theme counts indicate prominence within this dataset, not statistical importance or population prevalence.

LIMITS TO KEEP IN VIEW

University perceptions in Singapore are indirect K-12 evidence. Short sentence responses offer limited depth and do not measure workload savings or learning outcomes.

WHAT LEADERS CAN DO

Editorial recommendation: Discuss one concrete assignment with staff and students. Ask what help is useful, what work remains the learner's and how understanding will be demonstrated. Reconcile the answers before revising guidance. Include families when explaining secondary-school expectations, using the same practical example so they can support consistent use.

WHO / WHAT WAS STUDIED

71 educators and 152 university students in Singapore.

READ THE ORIGINAL SOURCES

Educator and student interpretations of AI change
link.springer.com · Published 2026-09-21 · Open source

07 / LEADERSHIP

AI guidance should allow for differing views

Peer-reviewed cross-sectional survey
Zohourian and colleagues / SN Social Sciences

THE FINDING

A US university survey finds that students combine enthusiasm and concern in different ways. Researchers analyzed 237 responses and grouped participants into optimistic, neutral and skeptical profiles, comprising approximately 41%, 38% and 21% of this sample. Frequent AI use was associated with a more optimistic profile, but the study cannot tell whether use changed attitudes or favorable attitudes encouraged use. The analysis describes perceptions of benefits, challenges and effectiveness; it does not verify the learning improvements students believe they experience. For school leaders, the transferable question is how guidance can address different starting points without assuming that students share either confidence in AI or opposition to it.

THE EVIDENCE

An online US participant pool supplied the cross-sectional sample. The analysis used 23 perception items and a separate ethical-concern scale.

LIMITS TO KEEP IN VIEW

Self-reported adult attitudes and online recruitment limit generalization. The clusters are analytical descriptions, not stable student types or causal evidence of adoption benefits.

WHAT LEADERS CAN DO

Editorial recommendation: Offer a common core of examples and a choice of practice activities. Confident users can examine a convincing error and defend their correction; hesitant users can compare an approved workflow with an ordinary approach. Collect anonymous questions before and after the session. Use the responses to improve instruction, while avoiding permanent labels or assumptions about a student's capability based on enthusiasm.

WHO / WHAT WAS STUDIED

237 university students in the United States; indirect K-12 evidence.

READ THE ORIGINAL SOURCES

University AI perceptions across student profiles
link.springer.com · Published 2026-09-23 · Open source

FROM EVIDENCE TO ACTION

The week ahead

Monday / Define the learning claim

01 Curriculum and assessment leads: select one active AI pilot. Write down the intended learning outcome, normal-instruction comparison and an unaided transfer task. Identify who will review results and when.

This week / Specify the support

02 Instructional and student-support leads: document what the adult does before, during and after an AI session. Include referral and disengagement procedures; estimate the staffing commitment.

03 Procurement and instructional technology: classify each supplier claim as perception, prediction, simulation or measured learning. Request the missing evidence before expansion.

Before expansion / Make a decision

04 Principals and teacher teams: review student work, implementation burden and subgroup experience together. Choose to continue, revise or stop the pilot against the criteria set at the start.

05 Communications lead: share one practical assignment example with staff, students and families showing useful assistance and the understanding the student must demonstrate.

KEEP READING

Visit The Sethi Ledger · Explore the archive