Beyond Adoption

A National R&D Agenda for Generative AI in K–12 Mathematics and Science Education

AmplifyGAIN Center | 2026

Authors:
Authors: Alex Liu[2], Min Sun[1,2], Zewei (Victor) Tian[1], Michael Xiao[1], Yulia Lapicus[2], Zachary Zhang[2], Kevin He[2]

[1] University of Washington, [2] Colleague AI

The AmplifyGAIN Center is supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C240012, and by awards from the National Science Foundation (NSF #2043613, 2300291, 2405110) to the University of Washington and an NSF SBIR/STTR award to Hensun Innovation LLC (#2423365). The opinions expressed are those of the authors and do not represent the views of the funders.

Why This Agenda, Why Now

Generative AI (GenAI) is reshaping the economy students will enter. That places K–12 schools, and math and science classrooms in particular, near the center of the nation's workforce readiness challenge.

Two years of nationally representative survey data now give the field its first reliable picture of how GenAI is settling into teachers' daily work. Both waves were collected with the RAND Corporation, in spring 2025 and spring 2026 (Esbenshade et al., 2025; Liu et al., 2026). The second year's finding is that the technology has been adopted while the structure around it has not followed. More teachers use GenAI, and users use it more intensively, yet what they use it for has barely moved. GenAI remains largely teacher-facing. It is concentrated in planning and the production of instructional materials, and it reaches students and the assessment of student work far less often.

> The technology has been adopted while the structure around it has not followed.

The questions that matter most now are ones the survey cannot answer. Surveys can document adoption. They cannot tell us whether GenAI-assisted planning produces better teaching, what guided student use should look like, or which forms of professional learning and policy actually change practice. Answering those questions requires research and development.

The 2026 report describes a workforce that is adopting the technology while thinking critically about it, and sets the field's task as empowering responsible use by addressing the concerns teachers name, among them over-reliance, the risk to human creativity, and the need for clearer training and district guidance. This agenda takes that reading as its starting point. It names the priorities where we believe the field's effort should concentrate over the next three to five years, and it invites researchers, developers, districts, and funders to take up that work.

What Two Years of National Data Establish

1. **Adoption is growing and deepening among users.** In 2026, 58 percent of math and science teachers reported using GenAI for instructional tasks, up from 50 percent in 2025. The gain came at the intensive end. Frequent users, meaning weekly or more, rose from 21 percent to 31 percent, while occasional users fell from 29 percent to 26 percent and non-users from 50 percent to 42 percent. Among users, 65 percent reported using the tools more frequently over the school year, and 44 percent reported growth in both frequency and range of tasks.

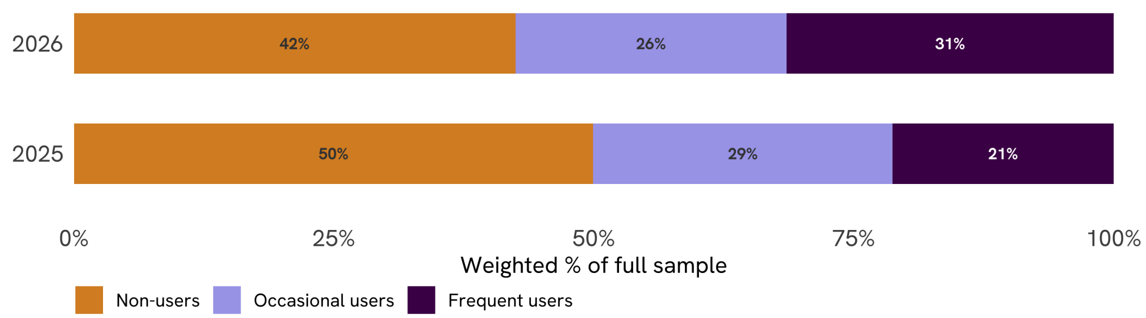

Figure A. Usage intensity groups, 2026 and 2025. Intensity groups collapse the frequency scale from Q1C, with frequent use meaning daily or weekly and occasional use meaning monthly or rarely. Responses are mutually exclusive and sum to 100 percent within each year. Full sample; 2025 N = 979, 2026 N = 1,011.

Source. Figure 1 in Liu et al. (2026), 2026 Survey on Teachers' Use of Generative AI in Math and Science Instruction, AmplifyGAIN Center and RAND American Teacher Panel.

2. **The purposes of use have barely moved.** Instructional planning remains the dominant use, reported by 76 percent of users in both years. In 2026 it was followed by creating assignments and assessments (58 percent) and adapting activities for different learners (57 percent), though the adapting-activities item was reworded for 2026 and is not comparable with 2025. Year-over-year movement was limited to modest shifts in the materials teachers turn to GenAI to produce.

3. **Use remains almost entirely teacher-facing.** In 2026, 74 percent of teachers did not include student GenAI use in their classes, fewer than a quarter held conversations with students about responsible use, and 10 percent monitored or reviewed students' AI use on assignments. Academic integrity was the area where teachers rated GenAI's impact on students most negatively (60 percent).

4. **A substantial and durable population of non-users persists.** In 2026, 42 percent of math and science teachers did not use GenAI for instructional tasks. They include teachers who have not heard of the tools (11 percent of all teachers), deliberate decliners (14 percent), and a growing group who tried GenAI and walked away (9 percent, up from 5 percent).

5. **Perceptions are diverging along lines of use.** Overall judgments of GenAI's impact on student learning remain split (29 percent positive, 32 percent negative, 39 percent neither positive nor negative), but the split increasingly tracks teachers' own use. Positive ratings among non-users fell from 16 percent to 8 percent and among occasional users from 37 percent to 28 percent, while 56 percent of frequent users continued to rate the impact positively. Ratings at both extremes grew, from 2 percent to 6 percent very positive and from 6 percent to 9 percent very negative. The share of teachers reporting each concern the survey asked about also rose. The survey is descriptive and does not identify what drives these movements.
6. **Support infrastructure is improving from a low base and still lags demand.** In 2026, 58 percent of teachers reported no formal GenAI training (down from 66 percent) and 42 percent reported no district guidelines (down from 49 percent). Formal guidelines had been established for 11 percent of teachers, with a policy under development for a further 20 percent, so 31 percent reported one or the other, roughly double the 18 percent of 2025. Training presence and policy presence both accompany heavier use, though the survey cannot establish the direction of that relationship.
7. **Frequent users report more planning time, not less.** Teachers who use GenAI frequently reported spending 7.3 hours in the past week planning and preparing lessons, 1.4 hours more than non-users, with total working hours roughly equal across groups. The comparison is between groups at one point in time rather than a change within teachers, and the survey cannot say why the difference exists. It is nonetheless hard to square with the common efficiency narrative.

Six Priority Areas for Research and Development

Priority 1. From GenAI-Assisted Planning to Instructional Quality and Student Learning

Planning is where GenAI already lives in teachers' practice, yet the field knows almost nothing about what GenAI-assisted planning produces. Teachers report creating better materials (45 percent of users), but these are perceptions, not measured outcomes. The finding that frequent users report more planning time than non-users suggests that teachers may be reinvesting time in expanded or deepened preparation, or may be absorbing new verification and revision work that the tools create. The field needs direct evidence along the whole chain, from the instructional quality of GenAI-assisted artifacts, to how those artifacts are enacted in classrooms, to what students learn as a result, and on how planning work is redistributed along the way.

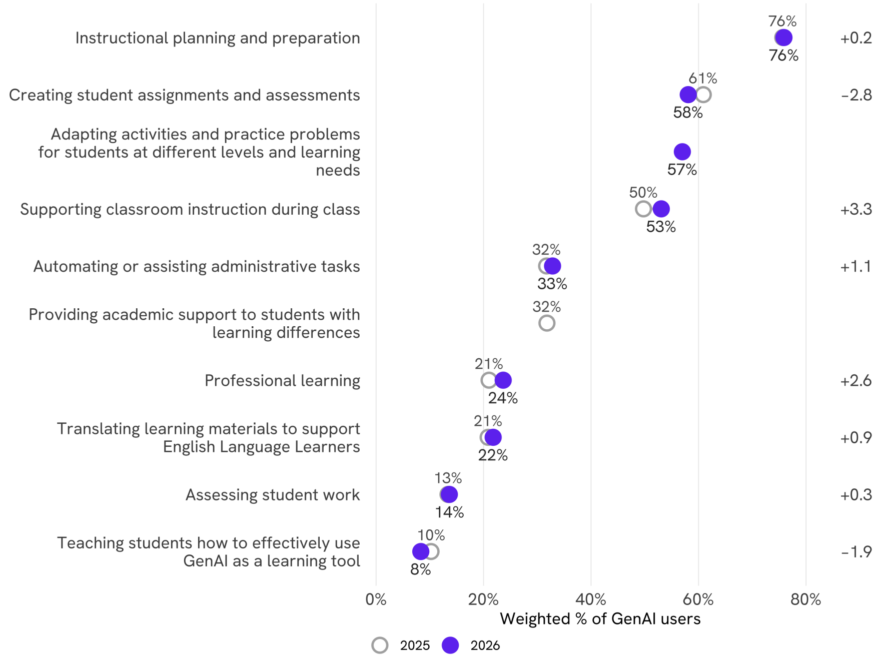

Figure B. Instructional tasks for which teachers have used GenAI tools, 2025 (open circles) and 2026 (filled circles), weighted percent of GenAI users. Multiple responses allowed. Planning dominates in both years, while assessing student work remains rare. The adapting-activities and academic-support rows are the two wordings of a single item, fielded differently in each year, and are not compared. GenAI users; 2025 N = 486, 2026 N = 588.

Source. Figure 6 in Liu et al. (2026), 2026 Survey on Teachers' Use of Generative AI in Math and Science Instruction, AmplifyGAIN Center and RAND American Teacher Panel.

Questions for the field include the following.

1. How does the quality of GenAI-assisted lesson plans, tasks, and assessments compare with teacher-authored materials when judged against established frameworks for cognitive demand, coherence, and standards alignment in mathematics and science?
2. What do teachers actually do during GenAI-assisted planning, and where does the additional planning time of frequent users go?
3. How faithfully are GenAI-assisted materials enacted, and what do teachers change, drop, or add when they teach from them?
4. Does teaching from GenAI-assisted materials change observable instructional quality, including the cognitive demand of the tasks students actually work on, the questions teachers ask, and the classroom discourse that follows?
5. What effect does GenAI-assisted planning have on student learning in mathematics and science, and how does that effect compare with other uses of the same planning time?
6. Through what mechanism would any such effect run, whether the quality of the materials, the planning time freed or absorbed, teacher content knowledge, or the responsiveness of instruction to the students in the room?
7. What prompting, curation, and revision practices distinguish teachers whose GenAI-assisted materials meet high quality bars, and are those practices teachable?

Priority 2. Designing and Testing Guided Student Use

Student-facing use is the largest gap between GenAI's claimed potential and observed practice. Three quarters of teachers do not include student GenAI use in their classes, and their strongest reservation concerns academic integrity. At the same time, one in four teachers now includes student use in some structured form, most often conversations about responsible use or explicit guidelines. This is where design work is most needed, where the risks teachers name are most concentrated, and, if students are to enter an AI-transformed workforce prepared to use these tools well, where schools will need the most help. The field should treat guided, supervised student use as a design and evaluation problem, not an adoption problem.

Questions for the field include the following.

1. What instructional designs allow students to use GenAI in math and science in ways that preserve, or increase, the intellectual work students do themselves?
2. How do structured classroom routines for student GenAI use affect academic integrity behaviors, student self-monitoring, and teachers' trust?
3. What does developmentally appropriate GenAI use look like at the elementary level, and how does it differ from what is appropriate in secondary mathematics and science?

4. What AI literacy competencies do math and science students need, and how can those competencies be taught and measured within subject-matter instruction rather than alongside it?
5. How do students themselves experience and interpret sanctioned classroom GenAI use, and how does this differ from their unsanctioned use outside class?

Priority 3. The Assessment Asymmetry

Teachers readily use GenAI to create assessments (58 percent of users) but rarely to evaluate student work (14 percent), and the share of teachers reporting any GenAI impact on their assessment practices fell by 7.3 percentage points between 2025 and 2026. This asymmetry held across both survey years. It likely reflects a combination of trust, data privacy exposure, and the professional judgment teachers reserve for themselves, but the survey cannot separate these. Because feedback and formative assessment are among the highest-leverage instructional practices, understanding whether and how GenAI can responsibly support them is a priority, as is understanding when teachers are right to keep it out.

Questions for the field include the following.

1. What accounts for teachers' reluctance to use GenAI in evaluating student work, and which concerns reflect accurate appraisals of current tool performance versus addressable design or policy gaps?
2. Under what conditions can GenAI-assisted feedback on student mathematical and scientific thinking reach acceptable standards of accuracy, specificity, and fairness?
3. How do data protection arrangements shape teachers' willingness to bring student work into GenAI tools, and what agreement structures make responsible use feasible at district scale?
4. What hybrid workflows, in which GenAI drafts and teachers adjudicate, preserve teacher judgment while reducing feedback latency for students, and what are their effects on student revision behavior?

Priority 4. System Implementation, Professional Learning, and Policy as Interventions

Both surveys show that training and policy presence accompany heavier use. Both are descriptive, so neither training nor policy can be assumed to cause it. District-provided professional development rose by 8.3 percentage points, to 30 percent of teachers, and established formal policy rose from 5 percent to 11 percent, yet time to learn remains the single largest barrier at

61 percent in both years, and lesson planning remains the most requested training topic at 73 percent. The report reads that pattern as a constraint of structured time rather than of willingness, which places the design of professional learning, not teacher persuasion, at the center of the problem. Related Center research describes how policy, professional learning, and messaging shape teachers' GenAI use (Nucci et al., 2026), but the field has correlations where it needs causal evidence. The unit of study should be the implementation system a district actually runs, in which professional learning, policy, and tool provisioning arrive together or not at all. Professional learning models and district policies should now be studied as designed interventions with specified mechanisms, not as background conditions.

Questions for the field include the following.

1. Which features of GenAI professional learning (application-grounded practice, collaborative structures, sustained duration, subject specificity) produce durable changes in the quality of teachers' GenAI-supported practice, and at what cost?
2. What happens to teacher practice when districts convert informal guidance into formal policy, and which policy provisions matter most for teachers' willingness to move beyond private, teacher-facing use?
3. How do policy, professional learning, and tool provisioning interact, and is there a minimum coherent package below which each component underperforms?
4. What formats of professional learning address the time barrier directly, for example embedded planning-time models, rather than adding training hours to full workloads?

Priority 5. Non-Use and Divergence

The 42 percent of math and science teachers who do not use GenAI for instructional tasks are not a residual category. They include teachers who have not heard of the tools (11 percent of all teachers), deliberate decliners (14 percent), and a growing group who tried GenAI and walked away (9 percent, up from 5 percent). Meanwhile, perceptions are diverging along lines of use, and policy, training, and use cluster together, raising the possibility of a widening divide in which some teachers and their students accumulate GenAI-related capacity while others do not. The survey documents the pattern but cannot tell us whether it is settling into place. The field should study non-use with the same seriousness it studies adoption, without presuming that adoption is the correct endpoint for every teacher or context.

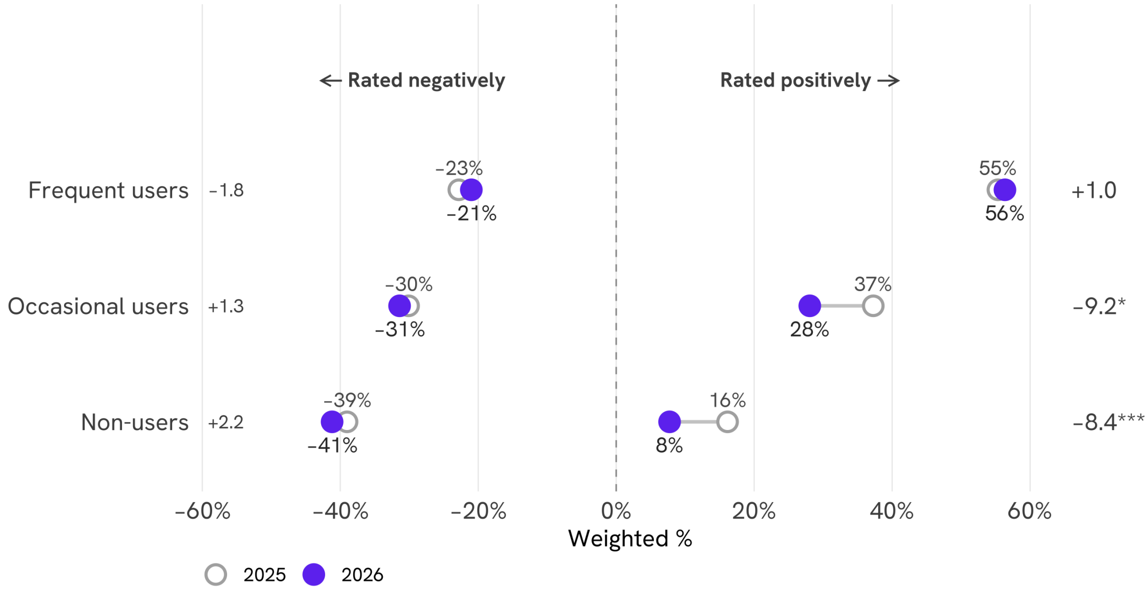

Figure C. Positive and negative ratings of GenAI's impact on students' learning by usage intensity, 2025 (open circles) and 2026 (filled circles). Positive ratings fell among non-users and occasional users while holding steady among frequent users. Descriptive association, not causal, since teachers are not randomly assigned to usage-intensity groups. Full sample; 2025 N = 979, 2026 N = 1,011.

Source. Figure 13 in Liu et al. (2026), 2026 Survey on Teachers' Use of Generative AI in Math and Science Instruction, AmplifyGAIN Center and RAND American Teacher Panel.

Questions for the field include the following.

1. What distinguishes principled, informed non-use from non-use rooted in lack of access, training, or institutional support, and how should policy respond differently to each?
2. Why do teachers who tried GenAI abandon it, and what do their experiences reveal about tool quality, fit with math and science teaching, and the credibility of current claims?
3. How do school-level social dynamics, including colleague networks and leadership signaling, shape teachers' movement between non-use, occasional use, and frequent use over time?

Priority 6. Feature Development and Tool Design

Teachers work mainly with general-purpose tools. ChatGPT remains the most widely used at 87 percent of users, and the survey cannot tell us how a tool built specifically for education, or for mathematics and science instruction in particular, would perform differently. Answering that question requires development work. Because the effects observed in classrooms depend in part on how a tool was designed, design decisions deserve the same scrutiny as classroom outcomes. Development is most likely to serve teaching well when it is vertically integrated, with the underlying model, subject-matter content, pedagogical supports, interface, and conditions of district use designed as one system, and when it is guided by domain knowledge, with quality judged by the standards of mathematics and science teaching rather than by general usefulness. Educators should help shape these tools throughout the design process, not only evaluate finished products. The field should invest in building education-specific tools and hold them to the same standards of evidence as any other intervention.

Questions for the field include the following.

1. What do education-specific GenAI tools do that general-purpose assistants do not, and where does that difference appear in the quality of what teachers produce and in what students learn?
2. Where do current tools fail in mathematics and science specifically, on symbolic notation, representation, modeling, and disciplinary argumentation, and how should subject-matter knowledge be built into a tool rather than prompted on top of a general model?
3. What does vertical integration add, and at what cost, when the model, content, pedagogical supports, interface, and conditions of district use are designed together rather than separately?
4. Which design choices, including defaults, guardrails, and transparency about limitations, produce the effects observed in use, and can those choices be varied experimentally to identify what does the work?
5. How can educators shape design decisions from the outset, and which participatory methods change what gets built rather than validating what was built already?
6. While tools are designed with transformative aims, how can development and design also serve classrooms as they run today, within existing curricula, routines, and constraints, and where do the two goals pull against each other?

Cross-Cutting Commitments

Work under this agenda, whatever its priority area, should reflect five commitments.

Measurement development.

The field lacks validated instruments for the quality of GenAI-supported practice, for AI literacy in math and science contexts, and for guided student use. Descriptive and causal work alike will stall without them, and instrument development is itself a contribution this agenda calls for.

Evidence beyond self-report.

Two years of survey data have taken perception-based evidence close to its limit. The next phase requires artifacts, observation, log data, student work, and learning outcomes.

Subject specificity.

Mathematics and science make distinctive demands, including symbolic notation, representation, modeling, and disciplinary argumentation, and GenAI's performance and pedagogical fit must be studied within them rather than inferred from general findings.

Practice partnership.

Research should be conducted with teachers, students, families, and districts in real settings, and its questions should be answerable in forms that practitioners and policymakers can act on.

Responsible implementation.

Data protection, academic integrity, bias, and the preservation of student intellectual work are constraints on every priority above, not a separate topic to be studied in isolation.

An Invitation to the Field

This agenda is a call, not a conclusion. We invite researchers at every career stage to take up these questions at scales ranging from single-classroom design studies to multi-district trials. We invite developers to treat the assessment asymmetry, guided student use, and the planning-time finding as product problems worth solving with evidence. We invite district and state leaders to partner in studying their own policy and professional learning investments. And we invite funders to recognize that the window for building an evidence base ahead of entrenched practice is open now and will not remain so.

The AmplifyGAIN Center will pursue this agenda through its own research, its partnerships, and its support for early-career scholars, and will revisit these priorities as each subsequent wave of national data arrives.

> The window for building an evidence base ahead of entrenched practice is open now and will not remain so.

References

Esbenshade, L., Sarkar, S., Nucci, D., Edwards, A., Nielsen, S., Rosenberg, J. M., Liu, A., Tian, Z., Sun, M., Zhang, Z., Han, T., Lapicus, Y., & He, K. (2025). Emerging patterns of GenAI use in K–12 science and mathematics education. AmplifyLearn.AI, University of Washington. https://www.amplifylearn.ai/wp-content/uploads/2025/08/Emerging-Patterns-of-GenAI-Use-in-K–12-Science-and-Mathematics-Education.pdf

Liu, A., Tian, V., Xiao, M., Nucci, D., Esbenshade, L., Rainier, S., Zhang, Z., Lapicus, Y., He, K., & Sun, M. (2026). Adopted but unstructured: Findings of GenAI use, training, and policy in K-12 math and science education from a 2026 national teacher survey. AmplifyGAIN Research & Development Center, University of Washington. https://www.amplifylearn.ai/wp-content/uploads/2026/08/rand_yr2.pdf

Nucci, D., Rainier, S., & Edwards, A. (2026). How policy and pedagogy influence STEM teachers' GenAI use: Findings from a 2025 exploratory study from the AmplifyGAIN Center. WestEd. https://www.wested.org/resource/how-policy-and-pedagogy-influence-stem-teachers-genai-use/

For questions, collaboration, or further information, contact

amplifylearn@uw.edu.