AI as a Teaching Partner

Executive Summary of the Full Report: Early Lessons from Classroom Codesign with Secondary Teachers

November 2025

Authors:

Alex Liu[1], Lief Esbenshade[1], Shawon Sarkar[1], Zewei (Victor) Tian[1], Min Sun[1], Zachary Zhang[2], Thomas Han[2], Yulia Lapicus[2], Kevin He[2]

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

For questions or further information about this report, please contact amplifylearn@uw.edu.

colleague.ai

AmplifyLearn.AI

IES Institute of Education Sciences

UNIVERSITY of WASHINGTON

NSF

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

Executive Summary

This report presents a comprehensive account of the Colleague AI Classroom pilot, a collaborative design (co-design) study that brought generative AI technology directly into real classrooms. In this study, AI functioned as a third agent, an active participant that mediated feedback, supported inquiry, and extended teachers' instructional reach while preserving human judgment and teacher authority.

Over seven weeks in spring 2025, 21 in-service teachers from four Washington State public school districts and one independent school integrated four AI-powered features of the Colleague AI Classroom into their instruction: **Teaching Aide, Assessment & AI Grading, AI Tutor, and Student Growth Insights.** More than 600 students in grades 6-12 used the platform in class at the direction of their teachers, who designed and facilitated the AI activities.

During the Classroom pilot, teachers were co-design partners: they planned activities, implemented them with students, and provided weekly reflections on AI's role in classroom settings. The teachers' feedback guided iterative improvements for Colleague AI. The research team captured rich data through surveys, planning and reflection forms, group meetings, one-on-one interviews, and platform usage logs to understand where AI adds instructional value and where it requires refinement.

Key Findings

- **Pedagogical judgment and AI competency jointly determine instructional impact.** Effective integration requires more than technical familiarity. Teachers needed both pedagogical judgement and tool fluency to structure student-AI interactions with clear goals, frame prompts, monitor progress, and model responsible use. Those with prior instructional experience using AI, and who perceived AI as instructionally useful prior to the pilot, were more likely to more actively experiment with students, iteratively refine AI-generated feedback, and report confidence in reuse.
- **AI functions as a "third agent" when teachers provide scaffolds and framing.** AI-supported features were most effective when teachers provided clear framing and instructions for both students and AI. Goal-oriented prompts (e.g., seminar preparation, proof sketches) led to deeper student-AI exchanges. In practice, AI acted as a "third agent" in the classroom, an additional instructional presence alongside teacher and student. AI proved especially useful for learning in guided exploratory preview and formative feedback cycles, assisting students with creating an initial draft, providing feedback, and prompting revision, while teachers retained final evaluative authority.
- **Implementation patterns vary by grade band and subject.** Teachers' use of AI reflected both the age group they served and their subject area. Middle-grade teachers (6-8) relied on clear workflows, shorter outputs, and strong monitoring, while high school teachers (9-12) made greater use of sustained dialogue with AI. In math and science, trust hinged on accurate, domain-specific feedback, whereas ELA and social studies teachers reported that off-topic or generic responses undermined learning value.

Main Features in Colleague AI Classroom

A demonstration of the Colleague AI platform's Classroom features is available on YouTube. This video incorporates learning from the codesign sessions to improve the tool.

Teaching Aide: The AI Teaching Aide functions as a classroom discussion partner for students, designed and guided by teachers. It supports student inquiry through structured, teacher-authored prompts that scaffold learning, reinforce key concepts, and sustain engagement during class activities. Students engage in one-on-one conversations with the AI; all students see the same initial message from the AI system and conversations automatically tailor themselves to the student's responses, while staying aligned to the teacher's instructions. Teachers can preview the tool as students, join live conversations, and review summaries that highlight common themes, misconceptions, and emerging needs. By mediating dialogue rather than directing it, the Teaching Aide extends teachers' instructional reach while preserving their professional oversight and judgment.

Assessment with AI Feedback: Colleague AI's assessment tools combine automated rubric generation with AI-assisted grading to make classroom evaluation faster, fairer, and more formative. Teachers can align assessments with national or state standards (e.g., CCSS, NGSS), select instructional frameworks such as Bloom's Taxonomy or Webb's Depth of Knowledge, and edit every criterion to ensure accuracy and contextual fit. The AI applies these rubrics to student work, producing instant narrative feedback that highlights strengths, identifies misconceptions, and suggests next steps, while teachers retain full authority over final grades.

AI Tutor: The AI Tutor provides students with on-demand academic support for homework, test preparation, and project-based learning under teacher supervision. Designed for guided independence, it helps students clarify concepts, practice reasoning, and receive formative feedback without giving away answers. All conversations are saved and summarized for teacher review, allowing educators to monitor progress, identify misconceptions, and ensure responsible use. By promoting self-advocacy and critical thinking while maintaining oversight, the AI Tutor extends learning beyond class time without diminishing the teacher's role.

Student Growth Insights: Student Growth Insights (SGI) transforms classroom data into actionable feedback for teachers. It aggregates patterns from student–AI interactions, highlighting common questions, misconceptions, and areas of progress at both the class and individual level. Teachers use these insights to differentiate instruction, plan reteaching, and communicate learning growth to parents and colleagues.

Colleague AI Classroom features demonstrated notable potential to reduce teacher burden while enhancing instructional feedback loops. Features like SGI automatically surfaced class-wide misconceptions and student growth patterns; AI-assisted grading streamlined feedback using teacher-authored rubrics; and the Teaching Aide enabled real-time personalized student support without requiring constant teacher intervention. Teachers reported that Colleague AI Classroom features augmented instructional practices through human-oversighted automation, differentiation, and targeted feedback.

The pilot suggests early-stage promise for both teacher- and student-level AI-amplified outcomes. At the teacher level, participants reported faster turnaround times for personalized feedback and growing confidence in leveraging AI for both in-class discussions and formative assessments. At the student level, observed outcomes included better scaffolded inquiry, visible self-improvement behavior in response to

feedback, and increased use of AI as a learning partner, particularly when teachers structured prompts or rubrics with clear goals and evaluative expectations. These outcomes, while exploratory, suggest measurable potential for AI-supported instructional improvement with sustained support and infrastructure.

Toward Effective Integration: Recommendations and Next Steps

AI adds value when it is structured, scaffolded, and goal-oriented. Clear prompts, short default replies, and visible end-products (exit tickets, notebook entries, seminar preparation) enabled students to engage thoughtfully with the AI. Without this framing, conversations devolved into aimless back-and-forth or became too overwhelming for younger learners.

Teacher AI competency drives adoption and impact. Teachers who entered the pilot with prior experience using AI for instruction, or who strongly believed in its potential, implemented more activities, framed the AI's role more effectively, and expressed higher confidence in reuse. Those without such experience typically conducted only one trial and reported challenges.

Formative feedback over summative scores. Across all assessments, teachers and students valued narrative feedback, specific comments that surfaced misconceptions and guided revision, far more than numeric scores, which were seen as inconsistent or opaque. Teachers used the AI as **a first-pass feedback engine** and retained final grading authority.

Developmental and subject differences matter. Middle school students benefited from concise outputs, sentence stems, and step-by-step guidance, whereas high school students tolerated deeper, more exploratory exchanges. Humanities and interdisciplinary STEM courses achieved the highest engagement, while math and earth science saw more mixed results due to the AI's difficulty with diagrams and technical reasoning.

SGI delivers high leverage with little effort. Student Growth Insights, which aggregate and display students' most common questions, saved teachers hours of scanning and enabled targeted reteaching. Teachers rated SGI as one of the most helpful elements for understanding student needs.

AI is a partner, not a replacement. Teachers consistently viewed AI as a third agent that augments teaching and learning rather than supplants it. Students appreciated quick, personalised feedback but wanted teachers to see their work and remain involved. The most successful implementations followed a collaborative model, with teachers framing the activity, AI providing initial support, and teachers interpreting and extending the AI's feedback.

Taken together, this pilot study demonstrates that generative AI can meaningfully support classroom instruction when it is thoughtfully designed, contextualised by teachers, and integrated into existing routines. This report details the study design and results for each feature, offers subject-specific insights, and concludes with practical implications for educators and school leaders.

Subject Insights: Math

- **Academic incentives boosted student engagement.** Math teachers tied AI conversations to participation grades, rubrics, or exit tickets, which increased effort and buy-in.
- **Scaffolding shaped success.** Teachers modeled math conversations, provided question stems, and set response-length expectations, while also prompting the AI to provide shorter, more focused replies.
- **Conceptual talk worked better than visuals.** Productive conversations centered on explanation, error analysis, and step-by-step reasoning, not diagrams or data tables where AI struggled.
- **Assessment feedback was useful, but scores were not.** Teachers valued AI-generated narrative comments to surface misconceptions and guide revision, but considered numeric scores unreliable or misaligned with expectations. They consistently reviewed and mediated AI output before using it with students.
- **Written reflections deepened understanding.** Having students restate math reasoning after AI exchanges reinforced learning and encouraged metacognition.

Subject Insights: Science

- **Verification was central to trust.** Teachers cross-checked AI outputs and modeled source checking, guiding students to validate claims and evidence, and encouraging them to think critically about the information provided rather than accepting it at face value.
- **Inquiry goals anchored productive use.** Science teachers saw strongest engagement when AI was tied to lab prep, Claims, Evidence and Reasoning writing, hypothesis generation, or debate preparation.
- **AI expanded representation in science.** Students encountered a broader range of scientists and perspectives, which helped expand their sense of who contributes to the field.
- **Middle schoolers needed concise scaffolds; high schoolers engaged with extended reasoning.** Middle school students responded best to short outputs and sentence starters that kept inquiry manageable, while high school students engaged productively with longer, Socratic-style explanations that supported deeper exploration.
- **Assessment feedback clarified reasoning, but scores were unreliable.** Teachers valued narrative feedback for helping students refine claims and evidence, but scoring inconsistencies required teacher mediation and final authority.
- **Structured prompts and clear products kept inquiry on track.** Focused entry questions and defined exit deliverables, such as lab notes or draft claims, prevented drift.

Download Full Report

For questions or further information about this report,
please contact amplifylearn@uw.edu.

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