ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

The Role of AI-Powered Tutoring Systems in Personalized Learning and Academic Performance in Modern Classrooms

Mahrukh Rehman Mirza
Mahrukh.rehman@arch.uol.edu.pk
Assistant Professor, School of Architecture, The University of Lahore

Sadaf Raja
sadaf.cad@pu.edu.pk
Assistant Professor, Architecture Department, College of Art & Design, University of the Punjab

Amal Noor Nizami
amal.noor@arch.uol.edu.pk
Assistant Professor, School of Architecture, The University of Lahore
Corresponding Author: * Mahrukh Rehman Mirza Mahrukh.rehman@arch.uol.edu.pk

Received: 17-07-2025 **Revised:** 24-08-2025 **Accepted:** 12-09-2025 **Published:** 25-09-2025

ABSTRACT

The given qualitative multiple-case study will focus on the way AI-based tutoring systems influence individual learning and perceived academic success in the classroom in the present-day. Over 6-10 weeks in four different classrooms (math, science, language arts), we observed lessons based on AI (2-3 lessons per classroom), talked to teachers and students, gathered de-identified learning artifacts, coded platform feedback traces descriptively, and analyzed teacher reflective journals. Reflexive thematic analysis with triangulation and cross-case synthesis provided us with the findings that AI can be most useful in cases where the technical affordances of adaptive pacing, step-by-step hints, localized explanations and mastery dashboards are realized through constrained autonomy and routine agency. In good classrooms, the use of cycle of commit- compare- revise and touch upon reveal after commit hints combined with teacher revoicing/withholding AI feedback and dashboard triaging to targeted conferencing.

Three processes are associated with AI use and perceived performance: (1) timely, specific feedback, making the next actionable; (2) cognitive scaffolding, which changes as one becomes more independent, and results in self-explanations; (3) metacognitive activation, and the transition to the AI says to I decided. The rewards (on-demand assistance, multilingual assistance) were based on reliability of the infrastructure, clear data procedures, and the advanced learner challenge paths. Limitations: There are context-boundedness, descriptive traces and short duration. We suggest a pragmatic logic model that relates features to practices to mechanisms to outcomes and provides implementation recommendations to the teachers, designers, and leaders.

Keywords: *AI tutoring; personalized learning; self-regulated learning; formative feedback; classroom orchestration; equity; qualitative case study.*

INTRODUCTION

Within the last five years, AI-driven tutoring has shifted its research prototype to the classroom reality. The speed is at least partially due to infrastructure (1:1 devices, cloud platforms, and learning management systems) rather than breakthroughs in at least 1:1 large language models (LLMs) and learning analytics. Unlike the previous generation of rule-based tutors, modern systems can be bent to the questions of students in natural language, offer step-by-step advice, and engage in a discussion of any confusion, in levels that can hardly be implemented in the classroom environment. There is preliminary data from real-world environments that well-designed AI tutors have the potential to improve learning and engagement compared to powerful alternatives such as active learning, even in a single class session (Kestin et al., 2025). Simultaneously, systematic reviews warn that the effect size differs according to contexts and designs, and that the usefulness of AI tutoring does not necessarily come on a silver platter but depends on the pedagogy, activity, and equity (Létourneau et al., 2025; Lin et al., 2023). Concisely, there is a promise of personalization but the classroom is a complex ecosystem where new tools have to be orchestrated, trusted and workable.

Why AI tutors now? A decade later following a consistent flow of intelligent tutoring systems (ITS), the paradigm was once again reset by the emergence of conversational LLMs and multimodal analytics. AI tutors can now adjust pace (when to proceed), direction (what issue or hint follows), and feedback (how to clarify, prompt or motivate) within close real-time. Overall, the syntheses of classroom ITS deployments describe the positive outcomes on learning and also point to unbalanced gains as compared to non-intelligent systems and the necessity of a longer duration of study (Wang et al., 2023; Létourneau et al., 2025). Adjunctive meta-analytic studies of feedback in technology-intensive settings demonstrate medium-reliable effects on performance, and explanatory feedback is more effective than less complex right/wrong feedback (Cai et al., 2023). These strands combined explain a baseline: the most educatively useful use of personalization mechanisms is to provide timely and elaborated feedback that is able to respond to state of the student: exactly the niche that AI tutors can occupy when designed well.

Classroom realities. Real classrooms do not personalize only to student-AI dyad, but a teacher-classroom-curriculum system. Educators have to be able to observe numerous learners, make decisions when to intervene, and combine AI-based practice with a whole-class discussion and evaluation. New teacher-facing analytics and orchestration tools, which summarize who has gotten stuck, which misconceptions still exist, and what hints have worked, are meant to bring AI tutoring to a legible form and a form of actionable activity on the part of an instructor (Aleven et al., 2022). These tools re-position the teacher as the orchestrator of activity, but not a supervisor of automation, but also come with new requirements (interpreting signals, aligning to learning goals, and planning exit tickets) that affect whether AI tutoring can enhance instruction or simply create cognitive load.

Associating AI tutoring with perceived educational achievement and self-belief. The literature has three recurrent pathways, which are mechanistic. First, the level of cognitive load, extraneous thoughts, can be minimized with scaffolding worked examples, step-by-step hints, and contingent prompts, which boost mastery experiences, which contribute to self-efficacy (Kestin et al., 2025). Second, perceived learning and persistence are mediated by both delivery of formative feedback at the appropriate grain size, which promotes the diagnosis and revision of errors (Cai et al., 2023). Third, affective attunement, which is the detection or reaction to frustration, boredom, or confusion, may maintain involvement and persistence and is applicable in the variants of affect-aware tutoring (Fernández-Herrero, 2024). These moment-to-moment experiences of the quality of explanation, speed, and support are of interest to students; the judgments made by teachers are of interest to evaluate the evidence of using strategies and transforming concepts in student work; demonstrated mastery manifests itself in artifacts (solutions, reflections, drafts) that reflect evidence of transfer or understanding.

Context issues: infrastructure, policy and equity. Availability of devices, connectivity and technical support determine the dependability of AI tutors to be present and responsive at times of classes. The policy frameworks concerning the responsible AI use in teaching and learning (e.g., institutional policies on AI, the instructions on disclosure and acceptable use) determine how teachers use AI-generated explanations or feedback in teaching/evaluation (Chan, 2023). More significantly, previous studies on algorithmic bias in education caution that personalization is prone to encode and repeat inequities in case the models are trained on discriminatory data, or the levels of detection are biased by demographic (Baker and Hawn, 2022). Educators thus must have effective tools of orchestration and guardrails usable such as transparency, opt-outs, data minimization, and bias monitoring that AI tutoring can enhance and not diminish equity objectives. The recent practitioner research also presents the issues of privacy, teacher-student relationships, and workload, which necessitates the designing of the product, which prioritizes human agency and aligns with classroom rhythms (Delello et al., 2025).

Scope and definitions. Both classic ITS and more recent LLM-based conversational tutors embedded into the curriculum can be discussed as AI-powered tutoring, which (a) simulates the knowledge, strategies, or affect of the learner; and (b) responds through tasks, hints, or dialogue. Personalized learning refers to the processes of instruction that adhere to personal needs and objectives and align the pace (speed of content coverage), path (sequencing of the problems, representations, or supports), and feedback (content, timing, and tone of feedback) to meet the needs of the individual learner or their objectives. Operationalization of academic performance is perceived performance (self-reports with regard to learning and confidence by students) and teacher ratings (formative assessment of both understanding and strategy application), and mastery of artifacts (evidence in student work products). These definitions focus on the enactment of classrooms, and coincide with the current reviews of AI tutoring and sustainable AI-in-education practice (Lin et al., 2023).

Significance. Through the lived classroom enactment, this qualitative study cuts across three communities. To practice, it sheds light on the process of organizing AI tutoring by teachers in the context of common limitations, and how students feel when their lessons are personalized. In the case of product design, it previews mechanisms (feedback grain size, scaffolding, affect support) and teacher facing affordances (orchestration views, evidence of learning) that are important to adopt and make a difference. On policy, it emerges under circumstances where AI tutoring can facilitate equity and professional judgment, including in the policies of institutions (regarding AI procurement, data management, and professional learning), procurement, and data governance (Chan, 2023; Baker and Hawn, 2022).

Problem statement, research questions and objectives. Although the use of AI-powered tutors is spreading rapidly, we have limited information about how personalisation works in reality (how students experience it, how it is facilitated and directed by teachers, what mechanisms plausibly mediate the relationship between AI tutoring and perceived academic accomplishment and self-efficacy), nor is it yet well known what contextual factors (infrastructure, policy, equity) facilitate or impede meaningful use; to answer this question, the research takes four qualitative research questions: RQ1: How do students experience personalisation (pace, path, feedback) with AI-powered tutors RQ2: How teachers plan AI tutoring in common instruction and assessment? RQ3: How are AI tutoring and perceived academic performance and self-efficacy of the students interconnected? RQ4: What contextual forces (infrastructure, policy, equity) facilitate or impede meaningful use? They include mapping the lived processes of personalization and orchestration, theorizing mechanisms along the path of the interaction between tutors and perceived performance and self-efficacy, and making design and policy implications that support equitable and teacher-centered application of AI tutors in contemporary classrooms.

LITERATURE REVIEW

Individualized learning places the learner in the center of control of goals, strategies, and speed what most people refer to as learner agency. In the online world, agency is enhanced as systems expose options (e.g. paths to problems, difficulty, modalities) and allow students to control time-on-task and help-seeking. Competency Mastery-oriented feedback is immediate, goal-oriented, and goal-oriented, and aimed at goal attainment of criteria and does not cause ego-threat, which is particularly crucial in frequent feedback. Operationalisations of these concepts by AI-powered tutors include (a) dynamically increasing/decreasing the difficulty of problems; (b) giving step-by-step formative feedback; (c) detecting misconceptions and (d) allowing instruction to proceed at a variable pace, which means you can fast-forward and rewind of your learning (Lin et al., 2023; Zhang and Aslan, 2021).

Self-regulated learning (SRL) planning, monitoring, control and reflection and personalization interact in both ways. Systems scaffolding goal setting, immediate metacognitive verifications in the form of explain why or whats your plan, and visualizing progress toward mastery have shown an increase in the SRL behaviors of students; students with better SRL have been found to benefit more (Järvela et al., 2024; Sobocinski et al., 2024). Notably, personalization, also in the presence of non-SRL supporting systems, can degenerate into efficient floundering, in which students can move fast but learn superficially; therefore, mastery-oriented feedback and SRL prompts should be combined. Qualitative classroom research indicates that in pacing autonomy, the educator mentions that the engagement and autonomy grow, but the autonomy must be maintained through deadlines, checkpoint quizzes, and express reflection routines, which may hinder procrastination (Tripathi et al., 2025; Shi et al., 2024).

The AI tutors of today are a combination of the model-based student diagnosis, hint sequencing, and stepwise feedback and the natural-language dialogue. Such systems have the ability to recognize patterns of errors, suggest just-in-time hints, breakdown multi-step problems, and provide worked-example comparisons, and newer systems can produce explanations with citations and can ask reflective questions to test comprehension (Lin et al., 2023). Recently, in a physics course at a university, a randomized controlled trial with an AI tutor constructed on the best-practice pedagogy approach was found to produce significantly higher post-test performance in a shorter time than in-class active learning, and the students mentioned that they felt more engaged and motivated with the AI tutor (Kestin et al., 2025). But AI tutors possess documented drawbacks as well: hallucinations (reality that is untrue and assigns a meaningless action), inconsistency between sessions, and obscurity. The Explainable AI (XAI) in the education field asserts the presence of interfaces that reveal evidence chains, uncertainty, and model confidence to provide teachers and learners with the opportunity to audit and bootstrap trust (Khosravi et al., 2022).

There are other problems of integration than accuracy. The teachers explain issues in orchestration, which involves handling a combination of human and AI assistance, determining when to freeze AI assistance in assessment, and aligning group work with individual AI assistance. The new literature on work called AI co-orchestration explores the situation where teachers and systems share control over the pace, assignment of tasks, and timing of feedback; teachers prefer designs that provide them with powers to override and class-wide situational awareness (Ahn et al., 2024). In practice, classroom integration needs to be consistent with the curriculum and assessment policies and access to devices; qualitative case studies demonstrate that its adoption can be enhanced when teachers are provided with exemplars, timely libraries, and schedules of the phases of the activities involving AI-off and AI-on (Shi et al., 2024; Kim and Kim, 2022).

The major learning process of AI tutoring is formative feedback. Meta-analyses of automated writing evaluation (AWE) demonstrate small-to-moderate improvements in writing quality when AI provides immediate and criterion-based feedback especially where students iteratively revise their work and where the teacher presents AI feedback as being draft-level rather than as final-level feedback (Deng et al., 2023). Adaptive and stepwise hints have a greater effect on cognitive scaffold (worked examples → faded steps → independent problem solving), lessened cognitive load, and induced self-explanations than global comments do in quantitative and mixed-method research (Lin et al., 2023; Bauer et al., 2024).

The metacognition is essential because students should determine when they should request hints, when they should persevere and when they should contemplate. In the case of generative AI tools, there are increased requirements of metacognition: learners will be required to test AI output, weigh evidence, and provide justification as to why the suggestions should or should not be accepted or altered (Fan et al., 2025). Designs that include explain your reasoning, have students guess what will be given feedback and then report before one is given, or demand that a plan be of plan quality checklists are more likely to enhance transfer. Incentive systems are combined. Other studies have indicated more engagement and self-efficacy through quick and customized commentary; others have determined that AI commentary is less acknowledged than human remarks because of perceived social distance and trust (Hein et al., 2024). These results indicate that the social construction of AI feedback (or its tone, transparency, and teacher approval) are as significant as the technical correctness of the feedback (Venter et al., 2025; Bauer et al., 2024).

With the shortage of human tutoring, AI tutoring may expand access to high-quality and on-demand tutoring. The advantages however depend on the availability of the device, bandwidth and an support feature (e.g. text-to-speech, multilinguality). One of these is algorithmic bias: the errors of the model may systematically discriminate against certain groups of people, and the lack of transparency in scoring may undercut due process in evaluation (Baker and Hawn, 2022). In K-12 and higher education, a variety of issues related to privacy are relevant: telemetry based on fine-grained learning traces, student essay content, and behavioral analytics implies a precarious purpose, retention, and consent; cross-cultural evaluations of AI privacy coverage indicate the convergence of worries regarding proctoring, surveillance, and data governance, albeit at a different emphasis (Xue et al., 2025).

The workload of teachers is a two-sided coin. According to qualitative studies, AI was considered capable of decreasing routine marking and creating first-draft feedback, but according to the disclosures of the teachers, more invisible work was generated: curating prompts, vetting AI work, and controlling classroom orchestration (Ahn et al., 2024; Shi et al., 2024). The development of the profession and regulation (what is acceptable AI use, when to turn off, how to document work assisted by AI) are needed to provide equitable sustainable implementation.

In both K-12 and higher-ed, both qualitative and mixed-methods research come together around some themes. First, educators see AI tutors and AI supported scaffolds as having potential to differentiate and provide writing assistance but fear transparency and role changes of the teachers (Kim and Kim, 2022). Second, according to case study research on AI-enabled classroom tools, it has been indicated that the orchestration is improved when teachers maintain the steering control and are able to freeze or order AI supports (Ahn et al., 2024). Third, recent qualitative studies among K-12 teachers report unbalanced AI literacy, the necessity to use exemplars to meet the standards, and cultural/ethical issues (Tripathi et al., 2025; Filiz et al., 2025). Fourth, intelligent tutoring and automated feedback have generally positive effects on learning, but effects differ according to subject, duration, and presence of revision cycles (Lin et al., 2023; Deng et al., 2023).

However, gaps remain. There is a limited number of qualitative studies associating classroom practices that can be observed (e.g., teacher re-voicing AI hints, rules regarding peer-before-AI help, metacognitive routines of commit then compare) to the variation of SRL and perceived academic performance of students under one term. Equity-oriented implementation (how schools orchestrate device sharing, construct privacy-sensitive defaults, develop AI-feedback literacy in students, etc.) is scarcely described. Lastly, limited comparative work in the classroom on the various AI-tutor designs (open-ended dialogic vs constrained stepwise) and their downstream impact on motivation, agency and teacher workload exists. Recent RCTs demonstrate huge advantages of a tightly scaffolded AI tutor (Kestin et al., 2025), yet the literature on qualitative explanations of why such designs are effective on daily classroom settings (and to whom) is limited.

Gap. There are no qualitative and classroom-proximal reports that follow the influences of particular AI-tutor characteristics (diagnosis, hints, stepwise feedback; explainability affordances) on teacher orchestration moves and to the extent that the practices mediate perceived academic performance and equity outcomes in the long-term. Precisely, available work documents that AI tutors can assist; we require dense descriptions of how characteristics into practices to mechanisms result in perceived performance improvements under actual classroom conditions.

Framework (model).

- **Inputs (AI-tutor features):** Diagnosis granularity; hint types (strategic vs. procedural); stepwise scaffolding; mastery dashboards; pacing controls; explainability/uncertainty displays.

- **Teacher practices (orchestration):** Gate/sequence AI access; set "AI-off/AI-on" phases; revoice or withhold AI hints; monitor dashboards; calibrate workload by delegating routine feedback.
- **Learner practices (agency & SRL):** Goal setting; plan-then-seek feedback; self-explain; regulate pacing; evaluate/triangulate AI outputs; revise iteratively.
- **Mechanisms:** Feedback quality (specificity, timing, alignment to criteria); cognitive scaffolding (worked→faded); metacognitive activation (monitoring, control); motivational climate (competence signals, autonomy support).
- **Short-term outcomes:** Higher perceived clarity of next steps; greater engagement; fewer unproductive impasses.
- **Perceived academic performance:** Improved self-reported mastery and confidence; teacher judgments of readiness; assessment evidence where permitted.
- **Equity/ethics moderators:** Access, privacy safeguards, bias audits, and PD for AI literacy.

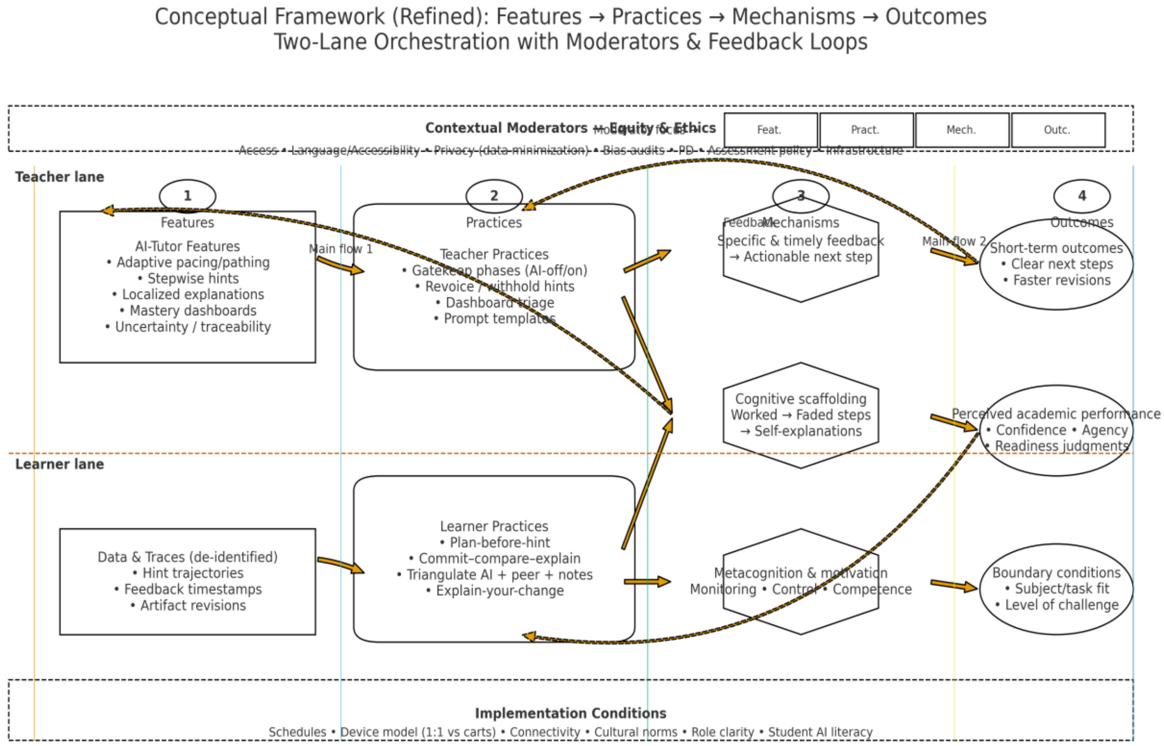

Figure 1 could visualize this as a left-to-right flow: AI-tutor features → teacher & learner practices → mechanisms → outcomes, with equity/ethics as contextual moderators.

METHODOLOGY

This study adopts a qualitative multiple-case study design from an interpretivist stance to generate rich, contextualized accounts of how AI-powered tutoring systems shape personalized learning and perceived academic performance in contemporary classrooms. A "case" is defined as a classroom in which an AI-tutoring tool is an integral component of routine instruction (at least two days per week) for ≥ 8 weeks. The unit of analysis is the classroom; embedded units include teachers, students, classroom activities, learning artifacts, and AI-platform interactions. Interpreting participants' meanings, practices, and constraints is prioritized over causal estimation; nevertheless, systematic triangulation across methods and sites supports credible, transferable insights.

Setting & Participants

Sites

Data will be collected in 2–4 schools representing diverse institutional types (e.g., public/low-resource, public/mid-resource, and private or charter), with subjects spanning mathematics, science, and language arts at upper-primary through lower-secondary grades (roughly ages 10–16). Candidate schools will already be piloting or fully deploying an AI tutor (e.g., for problem-solving, practice, or writing support). To enable cross-case contrasts, we will seek variation in (a) device access models (1:1 vs. cart-based), (b) connectivity quality, and (c) existing teacher professional development around AI.

Participants

Across sites, we will recruit approximately 24–36 students (6–9 per classroom), 6–8 teachers (one per classroom plus subject colleagues where relevant), and 2 school leaders (e.g., principal and instructional technology coordinator). Sampling will follow maximum variation principles to reflect diversity of grade level, subject specialization, teaching experience, and classroom connectivity. Where classes exceed 30 students, a stratified invitation process (by achievement band and gender) will prevent over-representation of any subgroup. All participants will be assigned pseudonyms.

Inclusion Criteria

- Classrooms using an AI-tutoring tool as a **regular** instructional resource for **≥ 8 consecutive weeks** prior to data collection;
- Teachers who consent to observations and interviews and agree to share de-identified artifacts;
- Students with parent/guardian consent and student assent;
- AI platforms capable of exporting **non-identifiable** traces (e.g., feedback messages, hint types, timestamps) or allowing on-screen capture with identifiers masked.

AI-tutoring Context

Although vendor-agnostic, the target tools share common **core features**:

1. **Personalized pacing and pathing** (adaptive item difficulty, mastery thresholds, optional enrichment/remediation);
2. **Hints and explanations** (stepwise scaffolds, strategic prompts, worked-example comparisons, and final solution rationales);
3. **Mastery dashboards** for students (progress toward goals, time-on-task, recent errors) and teachers (class-level heat maps, item analytics);
4. **Authoring or prompt templates** that let teachers constrain assistance (e.g., "nudge only," "no final answer," "point to misconception").

The inquiry attends to **how** these affordances are enacted: when hints are requested or withheld, how students interpret explanations, how teachers gate or sequence access, and how dashboards inform grouping or conferencing.

Sampling Strategy

We will use purposive sampling to select classrooms and participants exhibiting maximum variation along four axes: (1) grade level (upper-primary vs. lower-secondary), (2) subject (math/science vs. language arts/writing), (3) teacher experience with AI (novice vs. advanced), and (4) connectivity/device access (robust 1:1 vs. intermittent/shared). Within each classroom, students will be nominated via teacher lists to ensure inclusion of (a) frequent and infrequent AI-tool users, (b) a range of prior achievement or reading levels, and (c) accommodations (e.g., multilingual learners, IEP/504 where applicable). Replacement sampling will occur if consent rates fall below 60%.

Data Collection

Data collection spans **6–10 weeks** per site and integrates six sources. All instruments and protocols are included in Appendices A–D.

(a) Classroom observations

Each participating classroom will be visited **2–3 times** during AI-enabled lessons (40–60 minutes each). Observations are **non-participant** and guided by a structured protocol capturing:

- **Time-sampled activity maps** (5-minute intervals) noting task type, AI use phase (off/on/check), and grouping (whole-class, small group, individual);
- **Interaction traces** (vignettes of student–AI, teacher–student, and teacher–AI moments, with verbatim snippets when feasible);
- **Critical incidents** (breakdowns, breakthroughs, contested AI outputs, or equity-relevant events such as access barriers);
- **Orchestration moves** (teacher gating of AI, re-voicing or reframing of hints, conferencing prompts, peer-before-AI norms).
  Field notes will be expanded within 24 hours into narrative memos, with analytic comments flagged separately from description.

(b) Semi-structured interviews

We will conduct **30–45 minute** interviews with all participating teachers and a purposive subset of students.

- **Teacher interview foci:** goals for AI use; routines for personalization; criteria for acceptable AI help; dashboard interpretation; workload shifts; assessment policies; equity/ethics considerations; perceived impact on learning and confidence.
- **Student interview foci:** experiences of pacing and choice; how/when they request hints; interpreting explanations; revisions after AI feedback; feelings of competence; moments of confusion or mistrust; perceptions of fairness and privacy.
  Interviews will be audio-recorded (or securely video-recorded if remote), professionally transcribed, and returned to participants for **member checks** of factual accuracy.

(c) Focus groups (optional)

Where feasible and ethical, we will convene **student focus groups** (5–7 students; 45 minutes) to elicit social norms around AI use (e.g., peer-help vs. AI-help decisions) and to stimulate recall of classroom episodes. Focus groups will not disclose grades or sensitive personal data; participation is optional and separate from interviews.

(d) Learning artifacts

Teachers will provide **de-identified** samples of student work showing **revision cycles** (e.g., drafts with AI comments, problem sets with hint histories, teacher annotations). We will also collect **teacher feedback** artifacts (rubrics, comment banks) to understand alignment between human and AI advice. A sampling rubric (Appendix D) ensures representation of high/medium/low performance and different task types.

(e) Platform traces (for qualitative coding)

Subject to data-processing agreements and privacy review, we will export **textual feedback messages**, **hint trajectories** (types/sequence), and **timestamps** for sampled students. We will **not** collect names, emails, IPs, or keystroke logs. Traces will be used **descriptively**, e.g., to reconstruct the sequence of hints around critical incidents or to characterize the **form** of feedback (procedural vs. strategic) in episodes discussed in interviews. No inferential statistics will be reported.

(f) Reflective journals

Participating teachers will complete brief **weekly memos** (~10 minutes; 6–10 entries) responding to prompts about (a) what worked poorly/well with AI that week, (b) any orchestration adjustments, (c) notable equity issues, and (d) student confidence or engagement signals. These memos support longitudinal sense-making and enable timely **member-checking touchpoints**.

Procedures

1. **Approvals & onboarding.** Prior to recruitment, we will obtain institutional ethics approval and, where required, district/school permissions. We will meet with leaders to align on schedules, device policies, and data minimization.

2. **Recruitment & consent/assent.** Teachers will be invited via email and staff meetings; families receive plain-language consent forms describing voluntary participation, confidentiality, and the right to withdraw without penalty. Students provide assent in age appropriate language.

3. **Pilot & scheduling.** Instruments will be piloted in a non-study classroom; minor refinements will precede full rollout. Each class will be observed early, mid, and late in the window to capture developmental trajectories. Interviews occur after observations to leverage shared reference points.

4. **Member checks.** We will conduct a **midpoint** check (week 4–6) sharing brief analytic memos with teachers, and an end of study check presenting emergent themes to teachers and optional student advisory groups. Feedback will refine theme boundaries and clarify misinterpretations.

5. **Researcher safety & classroom continuity.** Observers will minimize disruption, refrain from providing content help, and avoid recording faces unless expressly permitted. Any classroom management incidents will be handled by teachers.

Data Management

All data will be stored in an encrypted cloud repository with role-based access. Audio/video files are stored separately from transcripts; a participant key mapping pseudonyms to identities is kept offline. Filenames follow a consistent convention (site class date source). Within 72 hours of collection, files are uploaded; within 7 days, audio is transcribed and spot-checked. De-identification includes redacting names, locations, and idiosyncratic details; transcripts preserve meaning while masking identifiers. An audit trail (collection logs, codebook versions, analytic memos, decision logs) will be maintained in a version-controlled workspace. Retention is capped at 5 years; raw video is destroyed within 12 months unless participants request deletion sooner. No data will be shared with platform vendors; only de-identified excerpts may appear in publications.

Data Analysis

We will conduct **reflexive thematic analysis** (RTA) complemented by **cross-case synthesis**.

Reflexive Thematic Analysis

1. **Familiarization.** Researchers read field notes, transcripts, artifacts, and traces holistically; analytic memos capture first impressions and questions, with attention to sensitizing concepts (personalization, feedback, agency, orchestration, equity).

2. **Initial coding.** Using qualitative software, we apply semantic and latent codes to meaningful units (utterances, episodes, artifact segments). Codes may be descriptive (e.g., “teacher freezes AI”), process-oriented (“commit-compare routine”), or evaluative (“feedback uncertainty flagged”).

3. **Theme development.** Codes are clustered into candidate themes (e.g., “feedback as hinge,” “bounded autonomy,” “co-orchestration”), with subthemes to reflect contextual contingencies (subject, connectivity). We iteratively write theme memos with within-case examples and counter-examples.

4. **Review & refinement.** Themes are tested against the full dataset; ambiguous codes are revisited; boundaries are sharpened by **negative case analysis**. We avoid forcing consensus RTA values researcher subjectivity and interpretive depth.

5. **Definition & naming.** Themes are defined by their **central organizing concept**, with explicit inclusion/exclusion criteria and illustrative quotes.

6. **Reporting.** The final write-up weaves **vivid excerpts** and **cross-case contrasts**, linking back to the study’s research questions and conceptual model.

Coding Frame

We begin with a provisional codebook anchored in the sensitizing concepts but explicitly invite inductive subthemes (e.g., “explanation calibration,” “AI-off assessment zones,” “equity workaround”). The codebook records code definitions, examples, and decision rules and evolves through analytic meetings. To support dependability, one researcher will code–recoded a 10–15% subsample after 2–3 weeks to check stability over time; discrepancies provoke discussion and clarification rather than a search for high “agreement” scores (consistent with RTA’s epistemology).

Triangulation & Cross-case Synthesis

- **Method triangulation:** observations, interviews, artifacts, traces, and journals are juxtaposed to corroborate or complicate claims (e.g., an observed hint episode is reconstructed with trace timestamps and discussed in interviews).
- **Source triangulation:** teacher and student accounts are compared; leader perspectives contextualize policy/infra constraints.
- **Site triangulation:** we analyze within-case first, then conduct a cross-case synthesis using a matrix that maps themes by site/subject/grade to identify recurring patterns and context-specific divergences.

Quality Checks

- **Peer debriefs:** monthly sessions with an external qualitative scholar to stress-test interpretations and expose blind spots; notes are appended to the audit trail.
- **Negative case analysis:** purposeful search for disconfirming evidence (e.g., cases where AI feedback reduced confidence).
- **Thick description:** detailed accounts of routines, language, and material conditions to support transferability judgments by readers.

Trustworthiness

- **Credibility:** iterative member checking (midpoint and end-of-study), triangulation across methods/sources/sites, and prolonged engagement (6–10 weeks per site) enhance plausibility.
- **Transferability:** we provide rich context (school type, subject, schedules, access models) so practitioners can judge fit.
- **Dependability:** a comprehensive audit trail documents decisions, codebook revisions, and analytic shifts; code–recode checks monitor stability.
- **Confirmability:** reflexive memos explicitly track researchers' assumptions and their influence on analysis; direct quotes anchor interpretations.

Researcher Positionality & Reflexivity

The core research team includes scholars with backgrounds in education technology, classroom teaching, and learning sciences. Prior positive experiences with adaptive platforms and skepticism about over-automation may influence sense-making. To mitigate bias, we will:

1. maintain reflexive journals logging expectations and reactions after each field visit;
2. seek peer debrief critique on theme naming, especially where interpretations might over-attribute agency to technology;
3. treat teachers and students as co-interpreters, privileging their explanations of routines and constraints;
4. explicitly attend to power dynamics in student interviews, using age-appropriate prompts and assuring no link to grading.

Ethics

Working with minors requires heightened protections. Key principles include **voluntariness, data minimization, confidentiality,** and **non-maleficence**.

- **Consent/assent:** Parents/guardians provide written consent; students provide assent. Teachers/leaders consent separately. Participation (or non-participation) will **not** affect grades, services, or employment.
- **Privacy & data minimization:** Only de-identified artifacts are collected. From platforms, we gather textual feedback messages, hint categories, and timestamps not names, emails, or keystrokes. If export requires an identifiable intermediary file, identifiers are stripped immediately and the original is deleted.
- **Recording:** Audio is default; video only where necessary (e.g., to capture screen interactions) and with explicit consent; faces are blurred or framed out.
- **Right to withdraw:** Participants may withdraw at any time; data linked to them will be deleted upon request (to the extent possible with aggregated analyses).
- **Risk management:** Potential risks include discomfort discussing challenges with AI or perceived surveillance. Mitigations include opt-out of any question, off-the-record clarifications, and teacher-present interviews for younger students if preferred.
- **AI-platform governance:** No raw data will be shared with vendors. Any platform access will be governed by a data-processing agreement specifying purpose limitation, retention, and security. The research team will not upload new student content to any external AI service.
- **Equity safeguards:** We will avoid scheduling burdens that differentially disadvantage students (e.g., missing lunch or essential services) and will provide translated consent materials.

LIMITATIONS & DELIMITATIONS

Limitations.

- **Time horizon:** A 6–10 week window may miss longer-term shifts in metacognition or identity and cannot fully capture sustainability of routines.
- **Novelty & context effects:** Early enthusiasm or resistance may inflate or mask patterns; findings are context-bound to participating schools' infrastructure and policies.
- **Trace incompleteness:** Platform logs vary by vendor; limited granularity may constrain reconstruction of certain episodes.
- **Researcher presence:** Even low-inference observations may subtly shape behavior (Hawthorne effects).

Delimitations.

- The study focuses on perceived academic performance (self-reports, teacher judgments, and artifact-based demonstrations), not standardized test score impacts.

- We sample classrooms already using AI tutoring rather than documenting adoption from scratch.
- We privilege classroom-proximal phenomena teacher orchestration, student agency, and feedback quality over comparative efficacy across vendors.

RESULTS

This section reports findings from a qualitative multiple–case study across four "modern classrooms" (two middle schools and one secondary school; mathematics, science, and language arts). Across the 6–10 week window, we conducted non-participant observations of AI-enabled lessons (2–3 per class), semi-structured interviews with students and teachers, optional student focus groups, collection of de-identified learning artifacts showing revision cycles, teacher weekly reflective memos, and limited platform traces (feedback messages, hint trajectories). Pseudonyms are used for all participants. We first provide a brief case sketch, then present six cross-case themes with contrasting evidence and negative cases. We close with a synthesis linking features → practices → mechanisms → perceived performance.

Case Sketches (brief)

- **Case A (Math 7, public/low-resource).** Cart-based devices, intermittent connectivity. AI tutor used for problem sets three times weekly. Teacher (Ms. Rivera) emphasized "commit first, then compare" routines.
- **Case B (Language Arts 8, public/mid-resource).** 1:1 devices. AI used for planning and revising analytical paragraphs. Teacher (Mr. Khan) restricted "final answer" generation and leaned on rubric-aligned hints.
- **Case C (Science 9, public/mid-resource).** 1:1 devices; blended labs. AI used to scaffold explanation writing and graph interpretation. Teacher (Ms. Okoye) relied on dashboards to triage conferences.
- **Case D (Math 10, private/charter).** Robust connectivity. AI used daily for mixed-practice and cumulative review; "AI-off" assessment windows enforced. Teacher (Mr. Lewis) customized hint levels ("nudge," "step," "worked example").

Theme 1 Bounded Autonomy: Personalization that "feels real" when routines scaffold choice

In all the sites, the students explained personalization by saying that they went at their own pace, or tried a new route, or did not wait until the class. Where educators had developed traditions that constrained autonomy, personalization was authentic and not anarchy. Beginning her sessions with a two-minute goal-setting slide and finishing with what changed was adopted by Ms. Rivera in Case A. micro-reflection. Students expressed the feeling of confidence due to the observable feedback: "When I see the bar turn green with 3/5 skills I know what to complete exactly (Aaliyah, Grade 7). In Case D, Mr. Lewis combined pacing autonomy and time-boxed checkpoint problems (no AI) which reset wandering.

In comparison, when there were no routines, personalization occasionally implied being more like the same. One Case C student observed, "It continued to offer me easy graphs, having failed at one of them, I wanted to do the hard one again (Diego, Grade 9). Teachers were taught how to adjust mastery thresholds and allow optional challenges to avoid adaptation breaking to remedial loops. The pragmatic implication: unstructured autonomy was either procrastination or efficient floundering; structured autonomy (goals, checkpoints and closure) was perceived as personalization.

Cross-case note. Options previewing (a) felt more real than any other personalization method, (b) AI-off intervals interspersed with AI-on work, and (c) help-seeking coaching by teachers (peer - AI - teacher) were all more dependable.

Negative case. A Case B student with strong writing skills perceived adaptation as “busy work”: “It keeps telling me to add evidence even when my paragraph already has two. It doesn’t know I’m done.” This prompted the teacher to enable a **“satisfaction” button** (“I’ve addressed this”) so students could signal closure.

Figure 1. AI-Orchestration Timeline

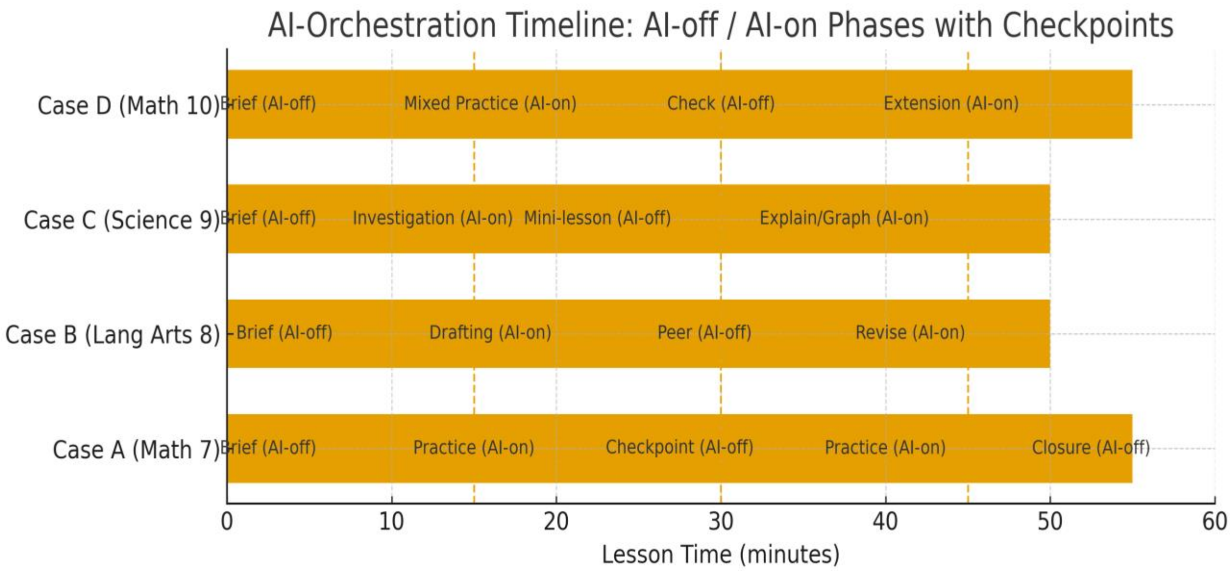

Figure 1. AI-Orchestration Timeline

Interpretive note: Gantt-style lanes showing AI-off/AI-on phases and assessment checkpoints across cases; illustrates bounded autonomy and timed gating.

Theme 2 Feedback as the Hinge: Specificity + timing outrank tone; uncertainty cues build trust

Students in both cases differentiated between useful (demonstrated precisely where my assertion went awry) and noisy (generic: 'be more specific') AI feedback. Localization (indicating the specific sentence/step), next-step clarity (a single, doable fix) and why-it-matters rationales were the three most appreciated characteristics of the messages. A Case C student said that the moment it pointed to the variable switch and stated that you switched units here, check line 3 I knew what to change (Maya, Grade 9).

The time was of the essence, not the content. Hints given too early negated perseverance--students would not go through with it. Hints given immediately after an attempt made. Ms. Rivera (Case A) explained how I switched the default to reveal after commit: As I made them make the try first, their second attempts got better. Teachers who paraphrased AI hints that turned a procedural hint into a strategic one (What changed between lines 2 and 3?) had a lower number of cycles of help seeking by clicking next.

Uncertainty cues (I may be mistaken check your unit conversion) were also favored by students and teachers, as opposed to false confidence. Students indicated that they were safer to disagree where the system identified low-confidence suggestions or where there were two plausible interpretations. The reason was that it stated: possible misconception: slope vs. rate to look into it, which indeed I did, (Jaden, Grade 10). Conversely, assertive explanations with wrong beliefs had the effect of mistrust and teacher overload to rebuke misperceptions.

Cross-case note. Language arts classes demanded **tone** sensitivity ("sounds formal/robotic"); math and science emphasized **specificity and localization**.

Negative case. In Case B, one student learned to "farm" hints to reconstruct a full paragraph. When the teacher switched to **"nudge-only"** mode (no sentence templates), the student initially stalled but later reported using the rubric more actively.

Figure 2. Feedback Quality vs. Timing Map

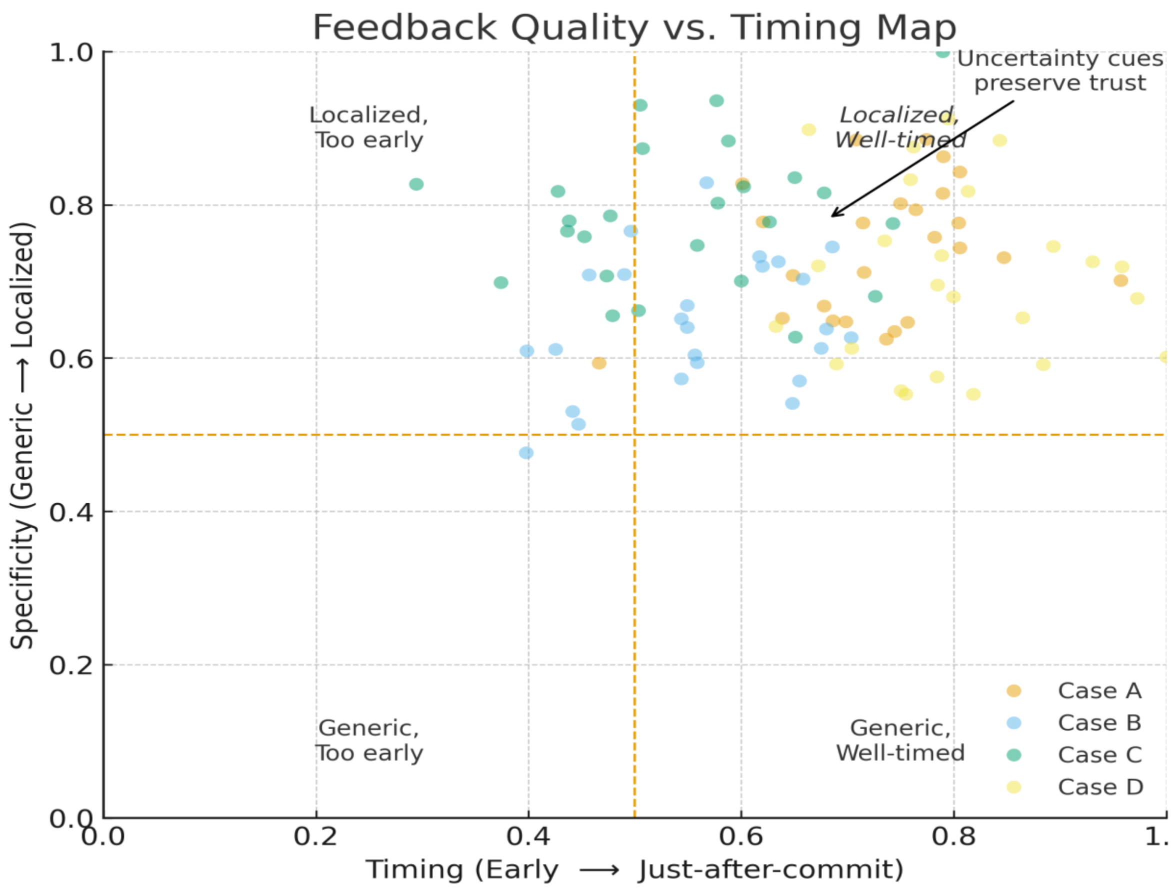

Figure 2. Feedback Quality vs. Timing Map

Interpretive note: Conceptual scatter by case demonstrating that localized, just-after-commit feedback clusters in the upper-right quadrant; dashed lines indicate thresholds.

Theme 3 SRL in the Loop: Metacognitive prompts convert help-seeking into learning

Where AI prompts were embedded in **SRL routines**, students articulated clearer strategies. Common routines included **"plan before hint"**, **prediction prompts** ("What will the AI say?"), and **"explain your change"** checkouts after revision. A Case D student reflected: "Before, I'd just click hints. Now he (teacher) makes us write what we expect the hint to be. If I'm wrong, I read it more carefully" (Rosa, Grade 10). Students described learning to **triangulate** advice ("AI + notes + partner"), especially after episodes of low-quality feedback.

Teachers reported that **visual progress** (dashboards showing clusters mastered) helped students self-pace, but **too many metrics** distracted novices ("completion %, streaks, time-on-task"). Simplified views (skills mastered + one next skill) were easiest for students to act on. Weekly **reflection prompts** in teacher journals documented a gradual shift from "the AI said to…" to "I decided to…" phrasing interpretive evidence of growing agency.

Cross-case note. SRL routines **traveled** across subjects when explicitly taught. Case A's "commit compare explain" showed up in Case C after teachers co-planned.

Negative case. Without explicit SRL framing, some students equated "green bar full" with "I understand." Exit tickets revealed **performance without understanding** ("I memorized steps but can't explain why"). Teachers responded by pairing dashboard progress with **explain-why** oral checks.

Theme 4 Teacher Orchestration Patterns: Gatekeeping, revoicing, and triage

Teachers developed **orchestration repertoires** for co-working with AI:

1. **Gatekeeping access** (*AI-off/AI-on phases*). In every case, teachers time-boxed AI to specific phases: **briefing** (AI-off), independent practice (AI-on with guardrails), **checking** (AI-off or nudge-only). Gatekeeping reduced over-reliance and simplified academic integrity decisions.

2. **Revoicing & withholding.** Teachers often withheld AI hints for students who clicked rapidly, nudging them to articulate a plan first. They **revoiced** AI feedback to align with classroom language. Mr. Khan (Case B): "It said 'expand evidence'; I ask, 'Which claim is under-warranted? Show me the verb that weakens it.' That re-aims the fix."

3. **Triage with dashboards.** Heat maps enabled targeted conferencing. Ms. Okoye (Case C) sorted students into three groups: ready to extend, needs one fix, stuck on misconception. She prioritized the third group for a five-minute mini-lesson while AI handled routine feedback for the second. Students in the first group got **challenge prompts** from a library.

4. **Prompt templating.** Teachers created guardrail prompts ("no final solutions," "ask a why question before any hint"). Over time, templates stabilized classroom norms and reduced teacher cognitive load.

5. **Assessment alignment.** All teachers froze AI during summative tasks; two kept **AI-on for drafting** but required AI-off explanations to demonstrate understanding. This hybrid approach reduced plagiarism concerns and surfaced students' reasoning.

Cross-case note. Orchestration moved from reactive ("put your screens down") to **proactive** (clearly signaled phases, visible timers, posted rules). Teachers reported **less fatigue** once routines matured.

Negative case. In Case A, early attempts at triage left a small group unvisited; students interpreted this as "teacher trusts AI more than me." The teacher corrected by using **two-cycle rotations** and inviting students to "flag" for human help.

Figure 4. Dashboard-Guided Triage Heatmap

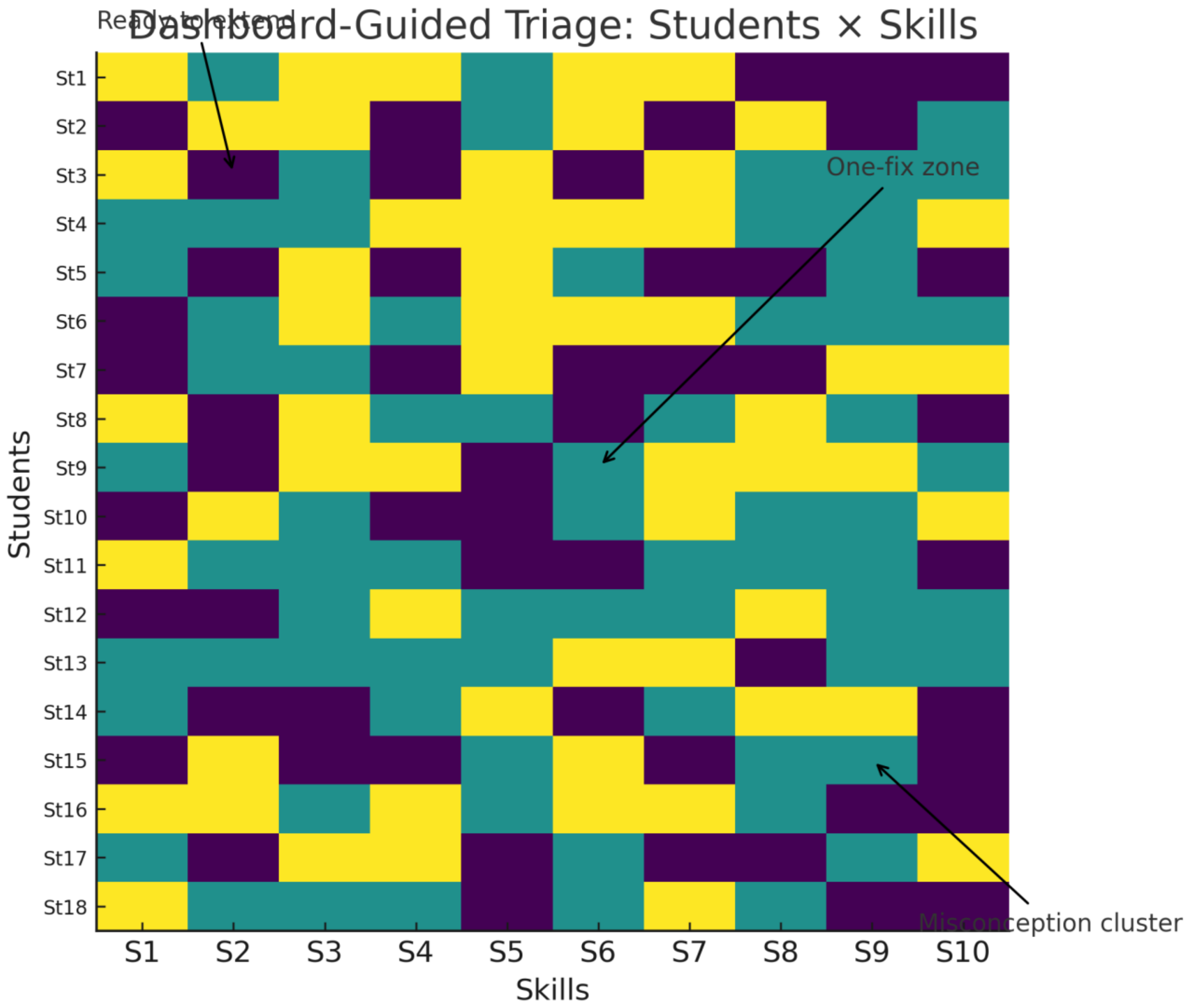

Figure 4. Dashboard-Guided Triage Heatmap

Interpretive note: Heatmap mock-up used for conferencing triage: cluster of misconceptions, one-fix opportunities, and ready-to-extend students across skills.

Theme 5 Equity Frictions: Access, language, and invisible work

Access. Cart-based classrooms (Case A) lost momentum during device swaps and log-ins; a five-minute delay compressed practice time and increased help-seeking pressure. Teachers mitigated with warm-up tasks (paper) and offline prompts mirroring AI nudge types.

Language & accessibility. In Case B, multilingual learners benefited when the AI **translated** rubrics and provided bilingual examples; however, literal translations sometimes distorted genre expectations (e.g., informal tone). Teachers coped by offering **curated exemplars** and instructing students to cross-check with rubric descriptors.

Privacy & agency. A minority of students expressed discomfort with **fine-grained data** ("It knows how long I stared at a sentence"). Trust improved when teachers **named** what data were visible to them and disabled time on task displays for peers.

ACADEMIA International Journal for Social Sciences
Volume 4, Issue 3, 2025 ISSN-L (Online): 3006-6638

Invisible work. Teachers' reflective journals logged **unseen labor**: curating prompt templates, auditing AI outputs, and troubleshooting log-ins. Ms. Rivera wrote, "When it worked, I graded less; when it glitched, I did three jobs." After weeks 3–4, invisible work decreased as **templates stabilized** and students internalized routines.

Cross-case note. Equity gains (more immediate help, multilingual support) were **conditional** on teacher mediation and infrastructure reliability.

Negative case. A student with IEP accommodations (Case C) found the AI's suggested "concise" rewrites **removed** scaffolded sentence frames; the teacher turned on **accessibility-preserving** settings and taught the student to **reject** unhelpful rewrites.

Figure 5. Logic Model

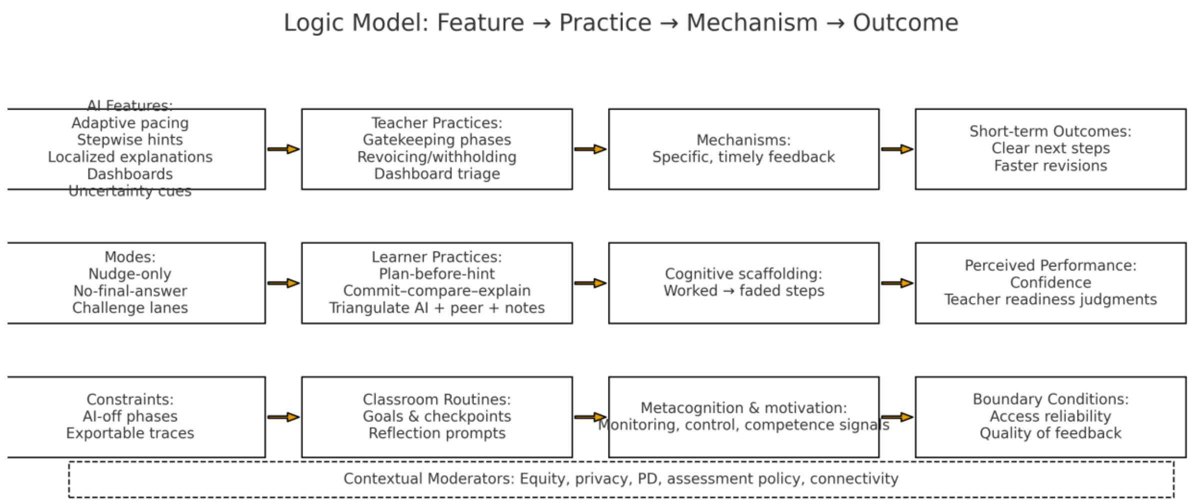

Figure 5. Logic Model

Interpretive note: Layered flow from AI features through teacher/learner practices to mechanisms and outcomes, with contextual moderators at the base.

Theme 6 Perceived Academic Performance: Narratives of progress, with caveats

Students and teachers narrated **performance improvements** tied to clearer next steps, more practice, and **faster revision cycles**.

- **Mathematics (Cases A & D).** Students reported fewer "dead ends." "Before, I'd get stuck on step 2 and quit. Now the hint says, 'Check where you changed the sign,' so I go right there" (Omar, Grade 10). Teachers noticed cleaner algebra and unit discipline after repeated, localized feedback. Short AI-off checks showed more students reaching procedural fluency, though conceptual language still required teacher coaching.

- **Language arts (Case B).** Students described stronger claim-evidence-reasoning chains and quicker drafting. "It's like having a checklist in my head now; I don't just write until it looks long" (Nadia, Grade 8). Teachers emphasized that rubric alignment not eloquence was the primary gain. Overreliance on AI phrasing sometimes flattened voice; the "nudge-only" mode countered that effect.

- **Science (Case C).** Students reported better graph reading and explanation clarity. "The AI asked me to define the system before the claim that made the rest make sense" (Arman, Grade 9). Teachers saw more students **revise** after first feedback rather than waiting for human conferences.

Students frequently used **confidence language** in exit tickets ("I can fix it if I know why"). Teachers' end-of-study judgments cited **readiness** ("more students are ready to try the challenge set") rather than test scores per se. Where AI feedback quality dipped or connectivity failed, perceived gains **stalled** and frustration spiked—highlighting the **contingent** nature of performance narratives.

Cross-case note. Gains clustered where three conditions co-occurred: **bounded autonomy**, **high-quality localized feedback**, and **SRL routines**.

Negative case. In Case D, two high-achieving students reported **plateauing**: "The hints are too cautious; I want pushback." The teacher created a challenge lane (proofs, non-routine problems) and permitted **"why-not?"** dialogues with the AI (students argue against AI's conservative steps), restoring perceived growth.

Figure 6. Equity Frictions & Enablers Map

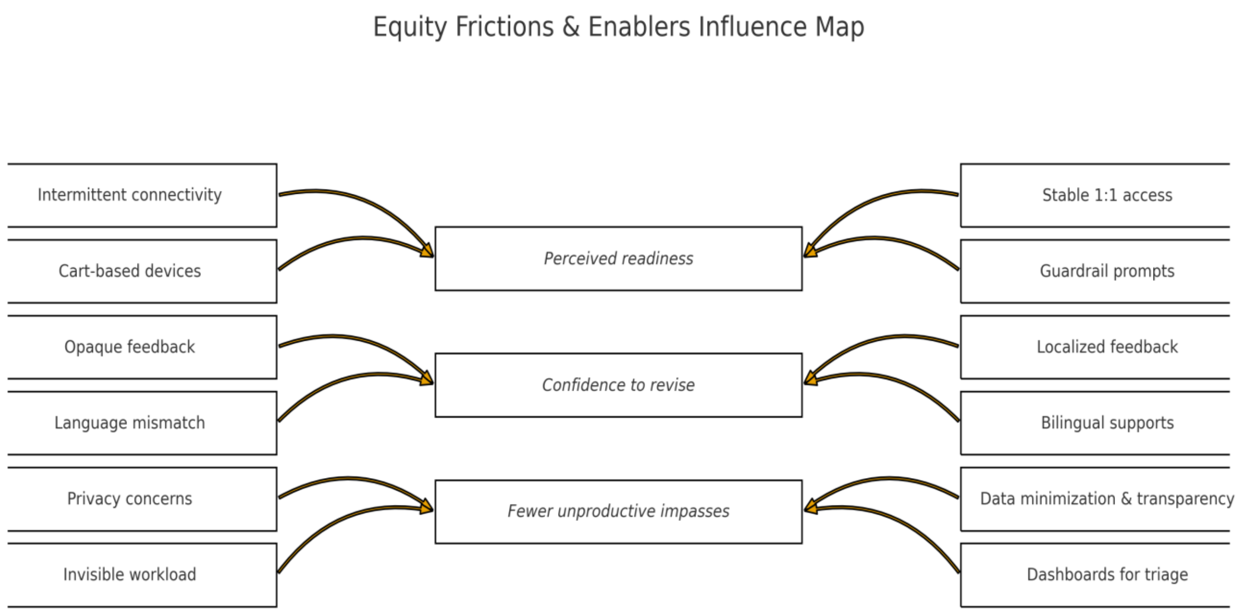

Figure 6. Equity Frictions & Enablers Map

Interpretive note: Influence map connecting frictions and enablers to outcomes via curved arrows to emphasize conditional equity gains.

Cross-case Synthesis: From features to practices to mechanisms to outcomes

Findings cohere as a **feature → practice → mechanism** chain that explains perceived performance differences:

1. AI features (adaptive pacing, stepwise hints, localized explanations, mastery dashboards, uncertainty cues) became educationally meaningful only when translated into teacher/learner

practices (gatekeeping phases, revoicing/withholding, commit–compare, plan before hint, reflective closure).

2. These practices activated **mechanisms**:
    - **Feedback quality:** specificity + timing increased actionable revisions; uncertainty cues preserved trust.
    - **Cognitive scaffolding:** worked-example and faded hints reduced cognitive load and encouraged self-explanations.
    - **Metacognitive activation:** prediction and justification prompts moved students from passive receipt to active regulation.
    - **Motivational climate:** bounded autonomy signaled competence and control, raising willingness to persist.
3. **Perceived outcomes** reflected these mechanisms: clearer next steps, faster revision cycles, increased confidence, and teacher judgments of class-wide readiness. Where infrastructure was fragile or orchestration was immature, mechanisms broke, and outcomes weakened.

Credibility Checks and Deviant Cases

Member-checking with teachers confirmed the scope of themes (especially the centrality of timing in feedback and workload dynamics as an invisible work being neutralized as routines become established). Two teacher respondents disagreed with us that uncertainty signals are consistently useful; in higher math, they favored brief over hedged hints so that they would not second-guess. We observe this as a contextual dependent boundary condition. Deviant case: even in socially complex writing tasks (where the voice problem is important) where AI feedback was correct, a subset of students favored peer-before-AI; the focus group study highlights the social aspect of learning.

DISCUSSION

This paper demonstrates that AI-based tutoring is a part of personalized learning where classroom practices can convert technical affordance into limited autonomy and quality feedback. In different cases, students were personalized in a way that they did not have unlimited choice but rather were systematically litated in terms of goal setting in the beginning, commit-compare in the middle of the work, and brief closure in the end. Such habits helped to explain when it was necessary to ask for assistance and avoided efficient floundering. Therefore, the pattern according to which students view personalization in the most favorable way when pacing and pathing are supported through rituals that render autonomy applicable figures as the response to RQ1.

In terms of RQ2, the orchestration moves of the teachers, including not only gatekeeping AI-on/AI-off but also revoicing or withholding hints and triaging using dashboards, were determining. Orchestration redistributed work: daily feedback was no longer in human hands, but the AI paid attention to it, teachers focused on misunderstandings and conferences. The workload in early weeks was more of invisible work (prompt curation, quality checks), but at the stabilization of templates load decreased. The implication is practical: schools must not only teach orchestration but they need to offer examples.

RQ3 focuses on the mechanisms between the use of AI and perceived academic performance. Three mechanisms surfaced. The specificity and promptness of feedback First, the specificity of feedback and timeliness: local feedback messages (just-after-commit) produced actionable revisions and maintained productive struggle. Second, stepwise hints led to self-explanations, particularly failure of supports.

Third, metacognitive activation--prediction--explain your change, shift talk: the AI says was replaced by I decided, and there was agency development. A combination of these mechanisms led to short term results (clear next steps, faster revisions) and stories of upward progress- our operationalization of perceived performance.

RQ4 brings forth moderators of context. Helps found on-demand (equity), and multilingual (equity) depended on reliability of the infrastructure, transparency of data practices, and mediation of tone and genre by teachers. Where connections were lost or interpretations were assuredly false, trust was destroyed and interests were held in abeyance. Notably, the students who excelled had to have challenge lanes and be allowed to challenge the conservative hints to prevent plateauing.

These are findings that are useful because they narrow in on a feature - practice - mechanism - outcome chain that describes why AI tutors are significant in daily instruction. Claims were also tied: results are context-dependent, trace data were descriptive, and the time window (6-10 weeks) will not be able to deal with long-term transfer or test-score effects. The future research must compare dialogic versus stepwise tutor designs under the classroom setting, test the longitudinal evolution of self-regulated learning, and experiment with tangible equity protection mechanisms (e.g., default to privacy, low-connectivity day offline routine).

In practice, schools must (1) institutionalize the routines of limited autonomy, (2) default on hint-timing-reveal, (3) provide teachers with guardrail prompt library and dashboard triage playbooks, and (4) incorporate explicit moderator policy on access, privacy, and assessment. In such circumstances, AI-tutoring will act as an agent of agency, clarity, and confidence-instead of noise.

CONCLUSION

This is a qualitative multiple-case study that examined the influence of AI-based tutoring systems in creating personalized learning and perceived academic achievement in contemporary courses. The technology had the most effect in four differentiated classrooms in which its own affordances, such as adaptive pacing, step-by-step hints, localized clarifications, and mastery dashboards, were incorporated in teacher-led rituals, which established limited autonomy. The students explained that personalization was guided latitude: devoted to an attempt, compared with direction, and concluded to a short look-back, and not unrestricted selection.

We had three mechanisms which explained the patterns. The next steps were clear without stifling the efforts through discouragement, which was first and specific feedback. Second, there was a cognitive scaffolding transition based on worked examples to independence that induced self-explanations. Third, metacognitive activation - planning, feedback prediction, and explanations - shifted talk to the AI says to I decided. These mechanisms, under good access and open data policies, provided more visible improvements and more expeditiousness in revision, and increased confidence- our operationalization of perceived performance. Gifts were conditional: fragile infrastructure, incomplete messages or too conservative suggestions dampened profits and added to the workload of the teachers.

It was context-bound and had a limited duration of study; platform traces were not inferential. Nevertheless, it adds a useful logic model that warrants features to practices, mechanisms, and outcomes, and moderators that schools can control. We suggest making reveal-after-commit timing institutional, curating guardrail prompt libraries, matching dashboards with brief conferences, and using privacy-first defaults. When put in this manner, AI tutoring will not be so much of a shortcut but rather a triggering agent in forming agency, clarity, and confidence, which allow classrooms to provide the many forms of personalization that sound both natural, human, and teach-back.

REFERENCES

Ahn, J., Jeon, M., & Holmes, W. (2024). How do K-12 teachers conceptualise shared control with an AI co-orchestrator? *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13372

Aleven, V., Blankestijn, J., Lawrence, L. E. M., Nagashima, T., & Taatgen, N. (2022). A dashboard to support teachers during students' self-paced AI-supported problem-solving practice. *In* I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, & T. Farrell (Eds.), *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption (EC-TEL 2022)* (pp. 16–30). Springer. https://doi.org/10.1007/978-3-031-16290-9_2

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*, 901–910. https://doi.org/10.1007/s40593-021-00285-9

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Bauer, E., ta Vilarinho, T., & Ifenthaler, D. (2024). Effects of AI-generated adaptive feedback on statistical skills. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13609

Cai, Z., Gui, Y., Mao, P., Wang, Z., Hao, X., Fan, X., & Tai, R. H. (2023). The effect of feedback on academic achievement in technology-rich learning environments (TREs): A meta-analytic review. *Educational Research Review, 39*, 100521. https://doi.org/10.1016/j.edurev.2023.100521

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*, 57. https://doi.org/10.1186/s41239-023-00408-3

Delello, J. A., Sung, W., Mokhtari, K., Hebert, J., Bronson, A., & De Giuseppe, T. (2025). AI in the classroom: Insights from educators on usage, challenges, and mental health. *Education Sciences, 15*(2), 113. https://doi.org/10.3390/educsci15020113

Deng, L., He, F., Li, Y., & Zhan, Z. (2023). Automated writing evaluation and students' writing performance: A meta-analysis. *Frontiers in Artificial Intelligence, 6*, 1162454. https://doi.org/10.3389/frai.2023.1162454

Fan, Y., Yang, H., & Zhao, X. (2025). Effects of generative artificial intelligence on learning outcomes: A systematic review. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13544

Fernández-Herrero, J. (2024). Evaluating recent advances in affective intelligent tutoring systems: A scoping review of educational impacts and future prospects. *Education Sciences, 14*(8), 839. https://doi.org/10.3390/educsci14080839

Filiz, O., Kaya, M. H., & Adiguzel, T. (2025). Teachers and AI: Understanding the factors influencing AI integration in K–12 education. *Education and Information Technologies, 30*, 17931–17967. https://doi.org/10.1007/s10639-025-13463-2

Giannakos, M., Krogstie, J., Divitini, M., & Jaccheri, L. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology*. Advance online publication. https://doi.org/10.1080/0144929X.2024.2394886

Hein, I., Kluge, A., & Zacher, H. (2024). Acceptance and motivational effect of AI-driven feedback in the workplace. *Frontiers in Organizational Psychology, 3,* 1468907. https://doi.org/10.3389/forgp.2024.1468907

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103,* 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*. https://doi.org/10.1038/s41598-025-97652-6

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*, 17458. https://doi.org/10.1038/s41598-025-97652-6

Khosravi, H., Sadiq, S., & Gasevic, D. (2022). Explainable artificial intelligence in education. *Computers & Education: Artificial Intelligence, 3,* 100074. https://doi.org/10.1016/j.caeai.2022.100074

Kim, N. J., & Kim, M. K. (2022). Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing. *Frontiers in Education, 7,* 755914. https://doi.org/10.3389/feduc.2022.755914

Létourneau, A., Deslandes Martineau, M., Charland, P., & Karran, J. A. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *NPJ Science of Learning, 10*, 29. https://doi.org/10.1038/s41539-025-00320-7

Létourneau, A., Martineau, M. D., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K–12 education. *npj Science of Learning, 10*, 29. https://doi.org/10.1038/s41539-025-00320-7

Lin, C.-C., Chen, Y.-T., & Lin, T.-J. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learning Environments, 10,* 26. https://doi.org/10.1186/s40561-023-00260-y

Lin, C.-C., Huang, A. Y. Q., & Lu, O. H. T. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learning Environments, 10,* 41. https://doi.org/10.1186/s40561-023-00260-y

Shi, L., Ding, A.-C., & Choi, I. (2024). Investigating teachers' use of an AI-enabled system and their perceptions of AI integration in science classrooms: A case study. *Education Sciences, 14*(11), 1187. https://doi.org/10.3390/educsci14111187

Tripathi, T., Sharma, S. R., Singh, V., Bhargava, P., & Raj, C. (2025). Teaching and learning with AI: A qualitative study on K–12 teachers' use and engagement with artificial intelligence. *Frontiers in Education, 10*, 1651217. https://doi.org/10.3389/feduc.2025.1651217

Venter, J., & Grussendorff, S. (2025). Exploring the use of AI in the delivery of effective assessment feedback. *Assessment & Evaluation in Higher Education.* https://doi.org/10.1080/02602938.2024.2415649

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research and future directions. *Computers & Education: Artificial Intelligence, 2,* 100039. https://doi.org/10.1016/j.caeai.2021.100039