REVIEW ARTICLE

Asian Research Association

ASIAN JOURNAL OF INTERDISCIPLINARY RESEARCH

DOI: 10.54392/ajir2645

A Systematic Review of AI-Enhanced Adaptive Flipped Classrooms in Undergraduate Engineering: Analytical Frameworks, Pedagogical Mechanisms, and Research Imperatives

Quanbo Geng [a,*], Aidah Abdul Karim [a], Hazrati Husnin [a]

[a] Faculty of Education (FPEND), Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia.

* Corresponding author Email: eda@ukm.edu.my

DOI: https://doi.org/10.54392/ajir2645

Received: 20-05-2026; Revised: 25-09-2026; Accepted: 17-10-2026; Published: 30-10-2026

Abstract: Flipped classrooms are increasingly being combined with adaptive and AI-enabled technologies, yet empirical evidence meeting a strict definition of AI-adaptive flipped learning in undergraduate engineering remains limited. This systematic review examined literature published between 2015 and 2025 and applied explicit operational criteria requiring learner-data-driven algorithmic adaptation, a structured flipped design comprising pre-class preparation and in-class application, an undergraduate engineering or computing population, and an eligible comparator or analytic contrast. Source-level verification identified five eligible peer-reviewed journal reports. These reports do not represent five independent research programmes; several arise from the same numerical-methods/adaptive-learning programme and include overlapping or sequential cohorts. Comparative reports generally found small or inconsistent differences in achievement, whereas perceived learning, classroom environment, and student experience were more consistently favourable. Kaw additionally showed that adaptive-learning-platform data could identify students at risk early, although the subsequent outcome comparison was confounded by voluntary uptake of human tutoring. No report isolated the contribution of AI adaptivity from all surrounding pedagogical and support components. The evidence therefore supports feasibility and several testable mechanisms, but not a pooled causal claim of superiority. Priorities for future research include independent multi-site trials, transparent reporting of algorithmic decisions and intervention fidelity, delayed engineering outcomes, prospective cost analysis, and predefined equity assessment.

Keywords: Flipped Classroom, Adaptive Flipped Classroom, Artificial Intelligence, Engineering Education, Personalized Learning, Pedagogical Mechanisms, Implementation Fidelity, Engineering Competencies.

1. Introduction

1.1 Conceptual Framing and Theoretical Underpinnings

Flipped learning is fundamentally a pedagogical reorganisation rather than a technological inversion. Constructivist theories of learning emphasise active participation, problem solving, and socially mediated learning within the learner's zone of proximal development (Piaget, 1954; Vygotsky, 1978). In a flipped model, lower-level content acquisition is typically shifted to asynchronous pre-class activities, while higher-order tasks such as analysis, design, troubleshooting, and teamwork are reserved for class time (Bonwell and Eison, 1991). This architecture addresses a long-standing tension in engineering education between breadth of content coverage and depth of problem-solving practice within limited contact hours (Ambrose *et al.,* 2010).

Conventional flipped designs nevertheless rest on an important assumption: a common pre-class sequence will prepare a heterogeneous cohort sufficiently for in-class application. Engineering students vary substantially in prerequisite knowledge, problem-solving fluency, and conceptual understanding (Ambrose *et al.,* 2010). Standard flipped resources, including video lectures, readings, and pre-class quizzes, often provide the same pathway regardless of initial preparedness. Two inefficiencies may follow: underprepared students can enter class cognitively overloaded and unable to contribute effectively to collaborative work, whereas well-prepared students may spend time revisiting material they have already mastered (Tomlinson, 2001). Both outcomes can reduce the effective use of face-to-face time and limit the potential of flipped pedagogy.

AI-enabled adaptive learning offers one approach to this readiness heterogeneity by adjusting content difficulty, pacing, sequencing, or feedback in response to learner-performance data (VanLehn, 2011). Learner-modelling techniques can update estimates of knowledge state as students interact with instructional material, enabling more timely identification of misconceptions and targeted support (Corbett and Anderson, 1994). This capability creates a testable possibility: pre-class preparation may become more responsive to individual readiness while preserving the constructivist, student-centred aims of flipped learning.

1.2 Literature Gap

The literature on flipped learning is extensive. Systematic reviews document widespread use in higher education, and overall differences in motivation and achievement relative to traditional lectures are often favourable, although effects vary by context and design (Lundin *et al.,* 2018). Parallel literatures on intelligent tutoring systems, adaptive learning platforms, and learning analytics likewise report potential benefits for performance and engagement (Wang *et al.,* 2023). A more specific gap remains, however: there is limited systematic synthesis of AI-enabled adaptive instruction embedded within flipped learning, particularly in engineering contexts that emphasise complex problem solving.

Existing reviews generally examine flipped learning and AI-enabled personalisation as separate domains rather than as an integrated instructional design. This fragmentation leaves three questions unresolved. First, which mechanisms might explain the effects of AI-adaptive flipped learning: personalised preparation, increased engagement, earlier intervention, or improved alignment among preparation, classroom activity, and assessment? Second, under what conditions do these mechanisms operate, and do effects vary by course level, engineering discipline, assessment design, or student characteristics? Third, what trade-offs arise, including potential reductions in learner agency, inequities in data-driven profiling, and additional instructor workload? Engineering education provides a particularly relevant setting in which to examine these questions because it combines heterogeneous prior preparation with substantial demands for ill-structured problem solving.

1.3 Significance and Contributions

This review contributes to engineering-education research by applying an auditable boundary to AI-adaptive flipped learning; distinguishing algorithmic adaptivity from instructor-created differentiation, non-adaptive AI tools, and blended delivery; reconstructing the eligible evidence base from verified source publications; and separating repeated reports from genuinely independent evidence. It also translates readiness-responsive support, assessment alignment, and instructor orchestration into testable propositions rather than treating them as established causal mechanisms.

The review is intended to inform instructional design while preserving evidential limits. Because most eligible reports originate from a closely related numerical-methods research programme, consistency across publications should not be interpreted as independent replication. The central question is therefore not only whether adaptive-flipped designs can work, but whether their effects persist across institutions, disciplines, instructors, algorithms, and student populations.

2. Literature Review and Conceptual Framework

The integration of artificial intelligence (AI) into flipped-classroom models represents a notable development in engineering pedagogy. In flipped learning, lower-order activities such as content acquisition through video lectures, readings, and pre-class practice are moved to asynchronous pre-class settings, while higher-order activities such as collaborative problem solving, design work, and applied learning are undertaken during synchronous class time (Bonwell and Eison, 1991). This architecture operationalises constructivist principles by increasing opportunities for active participation and socially mediated learning within learners' zones of proximal development (Vygotsky, 1978). However, conventional flipped models often assume that a common pre-class sequence is sufficient for heterogeneous groups. Variation in prerequisite knowledge, problem-solving fluency, and conceptual understanding can therefore leave some students underprepared for collaborative work while others receive insufficient challenge (Tomlinson, 2001).

AI-enabled adaptive learning addresses learner heterogeneity by adjusting difficulty, pacing, sequencing, or feedback in response to learner performance data (VanLehn, 2011). Intelligent tutoring systems can provide stepwise feedback and mastery-oriented pathways (Azevedo *et al.,* 2022), while learning analytics can support more timely instructional decision-making (Williamson and Kizilcec, 2022). In the verified engineering corpus, adaptive platforms are used mainly for personalised pre-class pathways and early identification of students who may need additional support (Kaw *et al.,* 2019; Clark and Kaw, 2020a, 2020b; Clark *et al.,* 2022; Kaw *et al.,* 2025). Broader reviews of adaptive learning and intelligent tutoring systems indicate potential benefits but do not by themselves establish the effectiveness of AI-adaptive flipped designs in engineering (du Plooy *et al.*, 2024; Wang *et al.,* 2023).

2.1 Constructivism as a Pedagogical Foundation

Constructivism, associated particularly with Piaget (1954) and Vygotsky (1978), holds that learners actively construct meaning by integrating new information with existing knowledge and through socially mediated interaction. Cognitive constructivism emphasises individual sense-making, whereas social constructivism emphasises the cultural and collaborative dimensions of learning. Vygotsky's zone of proximal development (ZPD) describes the gap between what a learner can accomplish independently and what can be achieved with appropriate guidance. Flipped classrooms can support these principles by reserving class time for collaborative problem solving and guided application. Their effectiveness, however, depends on whether pre-class preparation provides sufficient and appropriately differentiated scaffolding for each learner.

Within this framework, AI-enabled adaptation can support constructivist learning in several ways. First, learner modelling can identify emerging misconceptions from patterns of performance. Second, algorithms can adjust content difficulty or sequencing to keep tasks appropriately challenging within a learner's evolving ZPD (Corbett and Anderson, 1994). Third, intelligent systems can provide process-focused feedback that prompts explanation rather than merely marking responses as correct or incorrect. Fourth, adaptive platforms can support metacognitive regulation by helping learners monitor progress and identify areas requiring further study. These functions do not in themselves guarantee deeper learning; rather, they provide scalable mechanisms through which more responsive pre-class preparation may be implemented.

2.2 ICAP Framework—Cognitive Engagement Architecture

The ICAP framework (Chi and Wylie, 2014) classifies learning activities into four increasingly generative modes of cognitive engagement. Passive (P) engagement involves receiving information, such as watching a lecture or reading text. Active (A) engagement involves manipulating or rehearsing information, for example by taking notes. Constructive (C) engagement requires learners to generate outputs that extend beyond the material presented, such as self-explanations or inferential diagrams. Interactive (I) engagement involves co-constructive dialogue and collaborative problem solving. The ICAP hypothesis (I > C > A > P) predicts stronger learning as engagement becomes more generative, a pattern broadly consistent with evidence from active-learning research (Freeman *et al.,* 2014).

Traditional flipped learning often distributes ICAP modes temporally: pre-class work tends to be passive or active, whereas class time is intended for constructive or interactive activity. This arrangement can fail when students arrive underprepared and are unable to participate meaningfully in higher-level tasks. AI-adaptive systems may enrich pre-class work by prompting self-explanation, adjusting question difficulty, or providing metacognitive feedback (Chi and Wylie, 2014). In principle, such functions could help more learners reach class prepared for constructive and interactive work. Whether this shift actually occurs, and whether it improves learning equitably across readiness levels, remains an empirical question rather than an established effect.

2.3 Adaptive Scaffolding and Zone of Proximal Development

Vygotsky's (1978) zone of proximal development provides a useful rationale for adaptive scaffolding: instruction should be sufficiently challenging to extend current competence while remaining achievable with support. Effective scaffolding is contingent on learner performance, fades as competence develops, and progressively transfers responsibility to the learner (Wood *et al.,* 1976). Conventional teacher-delivered scaffolding is difficult to

individualise simultaneously for large classes, creating a risk that some students receive too little support while others are over-guided (Tomlinson and Moon, 2013).

AI-enabled scaffolding can operationalise several principles at scale. First, learner modelling can update estimates of concept mastery as performance data accumulate (Corbett and Anderson, 1994). Second, adaptive systems can vary sequencing, difficulty, feedback, or hints according to learner performance (VanLehn, 2011). Third, scaffolds can be reduced as mastery improves, supporting a gradual transfer of responsibility to the learner (Duffy & Azevedo, 2015). Fourth, learning-analytics dashboards can provide instructors with information that may inform grouping, remediation, or differentiated in-class activities (Williamson and Kizilcec, 2022). These functions are best understood as potential components of human-AI orchestration rather than as substitutes for instructional judgement.

2.4 Theoretical Integration and Research Gap

These perspectives combine into an author-constructed architecture in which constructivism defines pedagogical objectives, ICAP specifies forms of cognitive engagement, and adaptive scaffolding describes how learner support might be adjusted within the zone of proximal development. The eligible reports did not collectively test this architecture as a formal model. It is therefore used as an interpretive and hypothesis-generating framework rather than as a validated theory of AI-adaptive flipped learning.

2.5 Mechanism–Context–Outcome Framework and Theory-to-Evidence Mapping

To meet the requirement that theoretical claims be traceable to evidence, we specify the framework in mechanism-context-outcome terms and state what would falsify each proposition. The framework comprises four links. First, diagnostic adaptation: where a system infers a learner's readiness state from pre-class performance data and varies subsequent content or support accordingly, learners should arrive at class with more comparable preparation than under a fixed pre-class sequence. This proposition would be falsified if variance in pre-class mastery were no lower under adaptive than under conventional flipped conditions. Second, orchestration: where instructors receive and act upon readiness information, in-class tasks should be differentiated in ways observable in classroom practice. This proposition would be falsified if outcomes were equivalent whether or not instructors accessed the readiness data. Third, engagement: where difficulty is held near capability and feedback is immediate, perceived competence and task value should rise, and should predict subsequent effort. This proposition would be falsified if engagement gains did not predict achievement after controlling for prior attainment. Fourth, assessment correspondence: effects on summative outcomes should appear only where assessment targets the capabilities the adaptive system developed. This proposition would be falsified if achievement gains occurred equally under aligned and misaligned assessment conditions.

Mapping these propositions to the verified corpus shows partial, indirect support but no direct causal test. The reports describe adaptive pre-class pathways, comparative achievement and perception outcomes, and data-triggered intervention, but none manipulated readiness detection, instructor access to adaptive information, engagement mediation, or assessment alignment as isolated factors. The theoretical links developed in this review should therefore be treated as hypotheses for future trials.

A recurring identification problem constrains causal inference. The 2019–2022 reports compare instructional packages that differ in adaptive pre-class preparation and sometimes delivery context, while the 2025 report combines prediction with optional human tutoring. Consequently, any observed difference may reflect the adaptive platform, instructional redesign, cohort effects, instructor experience, online delivery, tutoring, or self-selection. The distinct contribution of AI adaptivity can be estimated only when these components are explicitly controlled.

3. Methodology

3.1 Systematic Review Design and Justification

This systematic review was conducted and reported with reference to PRISMA 2020 (Page *et al.*, 2021). The review was not prospectively registered, and no protocol was published before screening.

Eligibility criteria, search concepts, and extraction fields were documented internally; final operational definitions were applied during an integrity re-screening and source-verification stage to resolve inconsistencies in the initial study set. The study-selection process is shown in Figure 1.

Because several eligible reports include overlapping or sequential cohorts from the same research programme, the evidence was synthesised narratively and the reports were not treated as statistically independent observations.

Adapted from Page MJ *et al.* The PRISMA 2020 statement. BMJ 2021; 372:n71. (Page *et al.,* 2021)

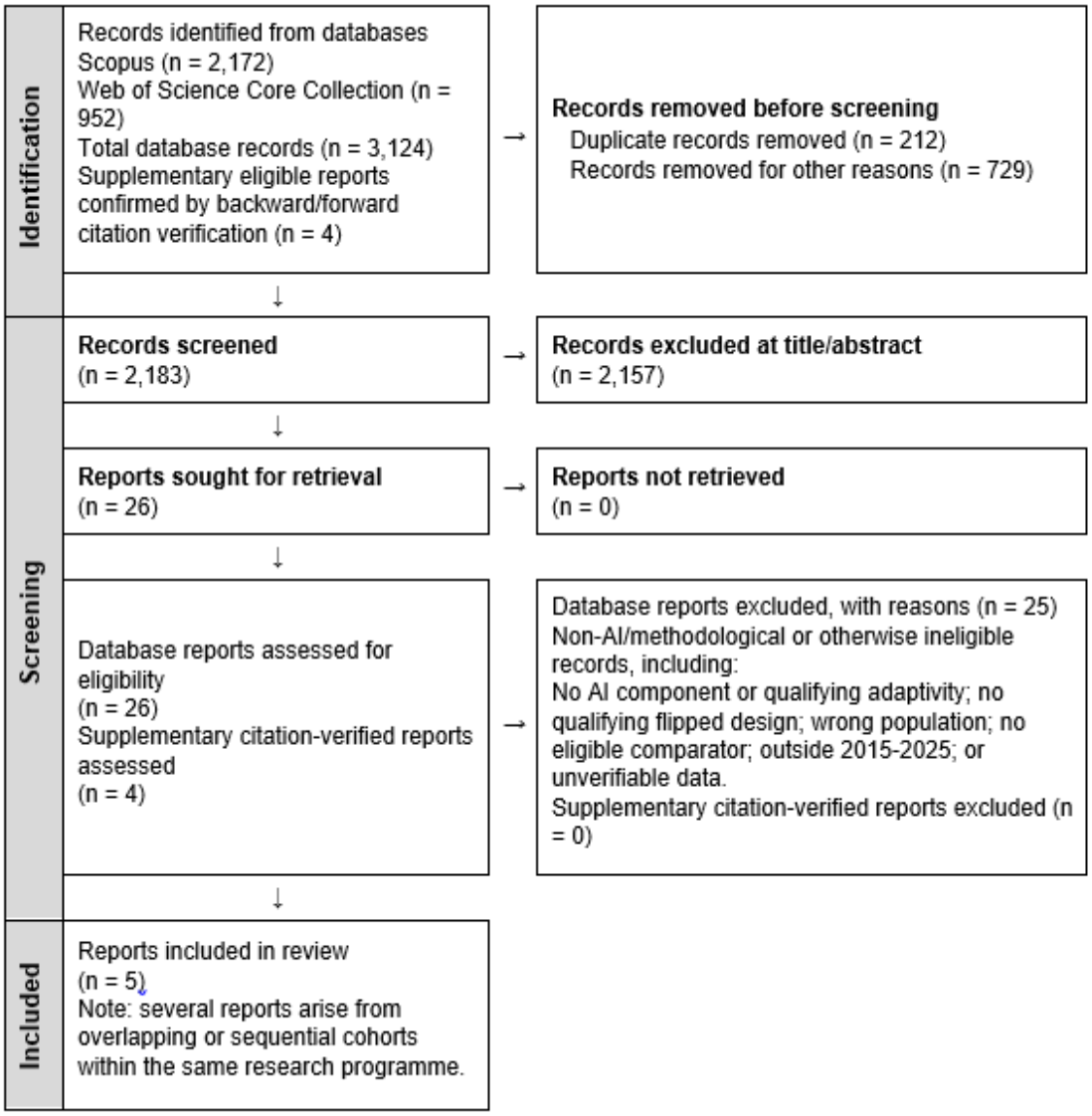


Figure 1. PRISMA 2020 Flow Diagram of Study Selection Process

Databases and coverage: Two multidisciplinary citation databases, Scopus and Web of Science Core Collection, were searched. Search strategies were developed with subject-specialist input and reported against the PRISMA-S extension to support reproducibility (Rethlefsen *et al.,* 2021). Records were deduplicated in EndNote and verified manually; 212 duplicates were removed. In addition to database searching, backward citation chasing of included studies and forward citation searching in Google Scholar were undertaken.

Search strategy: The search combined three concept blocks with Boolean operators: ("flipped classroom" OR "inverted classroom" OR "blended learning") AND ("artificial intelligence" OR "AI" OR "adaptive learning" OR "personalized learning" OR "intelligent tutoring") AND ("engineering education" OR "STEM education" OR "higher education"). The 2015–2025 window was selected because adaptive and AI-based educational platforms became increasingly prominent in higher education over this period. Because the search was completed before the end of 2025, that year is only partially represented and very recent evidence may be under-represented. The databases and core search terms are summarised in Table 1.

Retrieval and reporting of unretrieved records: The original flow diagram recorded 183 reports sought for retrieval and 157 reports not retrieved, implying an implausibly high full-text non-retrieval rate of approximately 86%. Re-inspection of the screening records showed that this discrepancy was a labelling error rather than a retrieval failure. The 157 records had been excluded during title-and-abstract screening and were incorrectly placed in the non-retrieval box. They are therefore included among the 2,157 records excluded at that stage. The corrected reconstruction leaves the full-text screening record intact: 26 reports were sought, all were retrieved and assessed for eligibility, and exclusions were assigned to documented reasons. Figure 1 has been redrawn accordingly; no report in this review was unavailable for full-text assessment.

Table 1. Core keywords applied and list of databases

List of Databases	Core keywords
Scopus and Web of Science	("flipped classroom" OR "inverted classroom" OR "blended learning") AND ("artificial intelligence" OR "AI" OR "adaptive learning" OR "personalized learning" OR "intelligent tutoring") AND ("engineering education" OR "STEM education" OR "higher education")

Inclusion criteria: peer-reviewed empirical research (quantitative, qualitative or mixed-method) reporting a defined research design; published in English between 2015 and 2025; conducted with undergraduate engineering or computing students; evaluating an instructional design that satisfies both the flipped-design definition and the AI-adaptivity definition set out below; and incorporating an eligible comparator, defined as a separate control or comparison group, a within-subject pre-post contrast, or an explicit analytic contrast between identified subgroups.

Exclusion criteria: descriptive case studies without any comparator or analytic contrast; grey and non-peer-reviewed literature; non-English publications; methodological or conceptual papers without primary empirical data; studies in which the technology performs no adaptive function as defined below; studies conducted outside undergraduate engineering or computing populations; and studies published before 2015. The inclusion and exclusion criteria are summarised in Table 2.

3.2 Operational Definitions Applied at Screening

Because the initial eligibility boundary had been applied inconsistently, each construct was explicitly operationalised and all candidate records were re-screened against the final definitions. AI adaptivity required the system to alter content, sequence, difficulty, feedback, or allocation of support for an individual learner on the basis of data generated by that learner. Fixed sequences, uniform automated feedback, descriptive dashboards that did not change what a learner received, and instructor-created differentiation without algorithmic input were considered ineligible. A flipped design required structured pre-class content acquisition and repurposed class time for application; blended delivery alone was insufficient. Eligible populations comprised undergraduate engineering or computing students, with mixed-discipline studies included only when engineering or computing results were reported separately. Eligible comparators were defined as described above.

3.3 Technology Taxonomy and Minimum Adaptive Functionality

Interventions meeting the AI-adaptivity definition were classified by mechanism to avoid pooling dissimilar technologies. Categories comprised: (1) adaptive sequencing platforms, which diagnose readiness and vary content order or difficulty; (2) intelligent tutoring systems, which model learner knowledge and generate targeted remediation; (3) recommender systems, which suggest resources on the basis of learner data; (4) predictive analytics and early-warning systems, which forecast risk and trigger differential allocation of human support; (5) generative-AI assistants, which provide responsive natural-language support; and (6) affect-sensing systems, which adapt on the basis of inferred engagement or emotional state. Gamification layers and non-adaptive digital tools were considered ineligible because they do not vary instruction on the basis of individual learner data. The minimum requirement for inclusion was that at least one instructional decision affecting an individual learner was determined by that learner's performance or behavioural data.

3.4 Outcome of Eligibility Re-Screening

The final source-level audit retained five peer-reviewed journal reports that satisfied the operational definition: Kaw *et al.* (2019), Clark and Kaw (2020a), Clark and Kaw (2020b), Clark *et al.* (2022), and Kaw *et al.* (2025). Four were recovered and confirmed through supplementary backward- and forward-citation verification of the numerical-methods adaptive-learning research programme. The audit also remove1d records retained in earlier screening rounds that did not satisfy the final criteria. Chen and Hsu (2022) used teacher-directed adaptive grouping rather than learner-data-driven algorithmic adaptation. González *et al.* (2022) evaluated AI virtual assistants in software-engineering capstone education but did not report the required pre-class/in-class flipped architecture. Hu *et al.* (2024) evaluated an AI-assisted engineering teaching system without a qualifying flipped design. Katona and Gyonyoru (2025), as resolved through its DOI, concerns adaptive programming education but not a flipped-classroom intervention. Previously identified medical-education, business-analytics, out-of-window, non-comparator, and unverifiable records remained excluded. The five retained reports were not assumed to represent five independent samples because several derive from the same NSF-supported numerical-methods programme.

Table 2. Criteria for inclusion and exclusion

Inclusion Criteria	Exclusion Criteria
Empirical peer-reviewed studies	Grey or non-peer-reviewed literature
Published in English	Published in languages other than English
Published between 2015 and 2025	Published before 2015 or outside the review window
Undergraduate engineering or computing population	No qualifying AI-adaptive function
Qualifying AI-adaptive function	No qualifying flipped design
Qualifying flipped design	Population outside undergraduate engineering or computing
Eligible comparator or analytic contrast	No eligible comparator or analytic contrast

3.5 Study Selection Process

Records were screened in Rayyan by two reviewers working independently at the title-and-abstract and full-text stages. Machine-learning prioritisation was used only to order records for human screening. During the final integrity audit, each retained record was checked against publisher or institutional source metadata to confirm bibliographic identity, population, adaptive functionality, flipped design, comparator, and publication status. Backward and forward citation verification was then used to identify directly related eligible reports missing from the provisional corpus. Overlap among reports was recorded explicitly to avoid treating repeated cohorts as independent evidence.

3.6 Quality Appraisal

The eligible literature is dominated by non-randomised comparisons, historical controls, correlational analyses, and self-selected uptake. A conservative ROBINS-I appraisal (Sterne *et al.,* 2016) therefore judged the comparative reports to be at serious risk of bias; Kaw *et al.* (2025) was judged at critical risk because students self-selected whether to accept algorithm-triggered tutoring. The 2019 report is primarily correlational and descriptive and therefore cannot establish an intervention effect. Table 3 summarises the principal bias concerns used to weight the narrative synthesis.

Quantitative pooling was not undertaken. The five reports are heterogeneous in design and outcome, and several arise from overlapping or sequential cohorts, violating the independence assumptions required for conventional meta-analysis. The synthesis therefore follows SWiM principles (Campbell *et al.,* 2020), reporting effect direction, measurement type, study limitations, and report overlap rather than combining effect estimates across publications.

Table 3. Risk-of-Bias Summary for the Five Eligible Reports

Study and design	Principal risk-of-bias concerns	Overall judgement
Kaw *et al.* (2019) Observational/correlational	Correlational association between adaptive-lesson metrics and exam performance; no independent non-adaptive comparator for the primary association; focus-group subgroup comparisons.	Serious
Clark and Kaw (2020a) Non-randomised comparative	Historical/semester comparisons across adaptive/non-adaptive and flipped/blended conditions; vulnerable to cohort, instructor, and temporal confounding.	Serious
Clark and Kaw (2020b) Non-randomised comparative	Adaptive flipped lessons compared with non-adaptive flipped and blended instruction; historical comparison and overlapping programme/cohorts limit independence.	Serious
Clark *et al.* (2022) Non-randomised comparative	Online flipped semesters with versus without adaptive lessons; pandemic-era temporal/cohort confounding and non-random assignment.	Serious
Kaw *et al.* (2025) Observational cohort	Prediction model followed by voluntary tutoring/advising uptake; strong self-selection and unadjusted confounding in the uptake comparison.	Critical

Evidence from repeated or overlapping reports is treated as one programme of evidence rather than multiple independent confirmations. Claims supported only by historical comparisons, correlational analyses, or self-selected uptake are described as exploratory or associative and are not used to establish causal mechanisms.

3.7 Data Extraction and Synthesis Framework

Data were extracted for author and year, engineering context, design, verifiable sample information, adaptive function, comparator, outcomes, and principal limitations. Table 4 presents the final evidence map. Kaw *et al.* (2019) reported 146 students across two semesters. The 2020 reports evaluated adaptive pre-class lessons against non-adaptive flipped and/or blended comparators and substantially overlap within the same research programme. Clark *et al.* (2022) compared online flipped instruction with and without adaptive pre-class lessons. Kaw *et al.* (2025) evaluated early-warning prediction and optional tutoring; its publisher abstract reports that only 23% of identified students accepted support, so this value is treated as an uptake proportion rather than as a sample size.

Implementation characteristics such as adaptive sequencing, pre-class preparation, assessment correspondence, instructor orchestration, online versus onsite delivery, and human tutoring were extracted descriptively. Because these characteristics co-vary with report, semester, institution, and comparator, they are not interpreted as statistically established moderators.

Table 4. Evidence Map of the Five Eligible Reports

Study	Aim and methodology	Sample	Key findings	Interpretation
1. Kaw *et al.* (2019)	Characterise adaptive pre-class learning in a flipped numerical-methods course. Observational analytics + focus groups	n = 146 across two semesters	Adaptive-lesson score correlated with final-exam score; focus groups generally identified benefits.	Feasibility and association; no causal adaptive-versus-non-adaptive estimate.
2. Clark and Kaw (2020a)	Evaluate adaptive learning in flipped and blended numerical-methods instruction. Non-randomised historical comparison	Sample size not independently verified from the abstract during this audit	Four instructional configurations compared; a notable open-ended-response effect of d = 0.34 was reported for the	Suggests possible modest benefit, but historical comparisons and programme overlap

			adaptive-flipped configuration.	limit causal inference.
3. Clark and Kaw (2020b)	Assess benefits of adaptive pre-class lessons in a flipped numerical-methods course. Non-randomised comparison	Sample size not independently verified from the abstract during this audit	Adaptive flipped instruction compared with flipped without adaptive lessons and blended instruction; direct and indirect outcomes generally favoured adaptive flipped.	Supports comparative evaluation but overlaps the same adaptive numerical-methods programme.
4. Clark *et al.* (2022)	Compare online flipped instruction with and without adaptive pre-class lessons. Non-randomised semester comparison	Sample size not independently verified from the abstract during this audit	Exam and concept-inventory scores were slightly higher but not significantly so; several perception and classroom-environment outcomes were more favourable with adaptive lessons.	Experience-related gains appear more consistent than direct achievement gains.
5. Kaw *et al.* (2025)	Use adaptive-platform data for early identification and targeted support of students at risk. Predictive modelling + observational uptake comparison	The publisher abstract reports that 23% of identified students accepted support; the total sample is not reported here without full-text verification	Best model achieved 85% predictive accuracy by week three; support acceptors scored about one-third of a letter grade higher than non-acceptors.	Feasible early-warning workflow; causal effect confounded by voluntary tutoring uptake.

4. Findings: Thematic Analysis

4.1 Theme 1: Adaptive Pre-Class Preparation and Readiness

4.1.1 Adaptive Pathways for Pre-Class Learning

Kaw *et al.* (2019) described adaptive pre-class lessons in an undergraduate numerical-methods course using personalised pathways through videos, text, simulations, and assessments. Across 146 students, final-examination performance was correlated with adaptive-lesson scores, and focus groups generally identified benefits of the adaptive platform. Because the report is correlational and does not isolate adaptive from non-adaptive flipped instruction, it supports feasibility and association rather than causal effectiveness.

Two 2020 reports by Clark and Kaw extended this programme through explicit comparisons. Clark and Kaw (2020a) compared adaptive and non-adaptive versions of flipped and blended numerical-methods instruction and reported an open-ended-response effect of Cohen's d = 0.34 in favour of the adaptive-flipped configuration, together with affective measures. Clark and Kaw (2020b) similarly compared adaptive-flipped instruction with flipped instruction without adaptive lessons and with blended instruction, reporting generally favourable direct and indirect outcomes. Because these publications are closely related, they should not be interpreted as independent replications.

Clark *et al.* (2022) provided a later comparison in an online, COVID-era flipped course conducted with and without adaptive pre-class lessons. Multiple-choice examination and concept-inventory outcomes were slightly higher with adaptive lessons but not statistically significant, whereas several classroom-environment and perceived-learning measures were more favourable. This pattern cautions against assuming that improvements in motivation or perception necessarily translate into substantial achievement gains.

4.1.2 Early-Warning Prediction and Triggered Support

Kaw *et al.* (2025) shifted the focus from adaptive sequencing towards predictive analytics. Adaptive-learning-platform data were used to build decision-tree models that identified students at risk of low performance; the best model achieved 85% predictive accuracy by week three. Identified students were offered weekly tutoring and advising. The publisher abstract reports that 23% of identified students accepted support and that these students scored approximately one-third of a letter grade higher than those who did not.

This result demonstrates the feasibility of coupling learner analytics with targeted human support, but it does not establish a causal effect of the algorithm. Uptake was voluntary, so students who accepted support may have differed systematically in motivation, help-seeking behaviour, prior preparation, or available time. The effect of prediction therefore cannot be separated from the effect of tutoring.

Taken together, the reports support two distinct adaptive functions: personalised pre-class pathways and early-warning allocation of support. The evidence does not yet establish which function contributes more to learning or whether either is superior to instructor-led differentiation implemented with equivalent resources.

4.2 Theme 2: Achievement, Engagement, and Implementation

4.2.1 Achievement Outcomes

Achievement findings are mixed rather than uniformly positive. The 2020 comparative reports suggest modest advantages for selected directly assessed outcomes, whereas Clark *et al.* (2022) found only small, non-significant differences in multiple-choice examinations and concept inventories. Kaw *et al.* (2025) reports a grade difference between students who did and did not accept triggered support, but that comparison is confounded by self-selection. The corpus therefore does not justify a pooled effect estimate.

The more defensible interpretation is that adaptive pre-class learning can be implemented without obvious loss of achievement and may produce modest gains under some conditions, but the magnitude and reproducibility of those gains remain uncertain.

4.2.2 Engagement and Classroom Experience

Perception and experience outcomes are more consistently favourable than direct achievement outcomes. Kaw *et al.* (2019) reported positive focus-group perspectives; the 2020 reports identified favourable classroom-environment and student-evaluation patterns; and Clark *et al.* (2022) reported improvements in perceived learning, responsibility, motivation for independent learning, and reduced reports of workload or stress when adaptive lessons were available. These outcomes are educationally relevant but are not equivalent to independently assessed competence.

No eligible report estimated a formal mediation model linking engagement or perception to achievement. Engagement should therefore be treated as an outcome and candidate pathway, not as an established mediator.

4.2.3 Instructor Orchestration and In-Class Activity

Across the programme, adaptive technology is embedded within a broader flipped architecture that preserves active or interactive class time and instructor support. Human orchestration is therefore an inseparable component of the intervention package. ICAP (Chi and Wylie, 2014) provides a useful vocabulary for describing in-class activity,

but the included reports do not experimentally isolate high- versus low-ICAP orchestration while holding the adaptive system constant.

Future trials should therefore report adaptive-system behaviour and instructor response separately: what the algorithm inferred, what instructional decision followed, who delivered the response, and whether that response differed from the comparator condition.

4.3 Theme 3: Evidence Limitations and Generalisability

4.3.1 Dependence and Concentration of the Evidence Base

The strongest limitation is not simply sample size but dependence within the evidence base. All five eligible reports involve overlapping research teams, and most concern numerical-methods engineering courses using related adaptive-learning platforms and pedagogical designs. Several draw on overlapping or sequential cohorts. Consequently, five publications do not constitute five independent replications, and publication-level consistency may overstate the breadth of the evidence.

4.3.2 Study Duration and Longitudinal Evidence

The eligible reports evaluate course-level or semester-level implementations. None establishes delayed retention, transfer to later engineering courses, capstone performance, or programme-level competency development. The evidence therefore concerns near-term course outcomes and student experience rather than durable engineering competence.

4.3.3 Outcome Measurement

Outcomes include final-examination items, concept inventories, course grades, classroom-environment scales, student surveys, platform analytics, and focus groups. These measures address different constructs and cannot be treated as interchangeable. Direct achievement effects are generally modest or inconsistent, whereas affective and perception outcomes are more often favourable.

4.3.4 Fidelity, Transparency, and Replicability

Replication requires reporting of the adaptive-platform version, model or rule set, learner inputs, pathway logic, instructor training, intervention dosage, adherence, and any human override. Studies of early-warning systems additionally require reporting of model calibration, decision thresholds, fairness metrics, and the exact tutoring or support process triggered by a flag. Incomplete reporting of these details limits reproducibility and transportability across institutions.

4.3.5 Risk of Bias and Causal Attribution

None of the eligible reports provides a clean randomised estimate of the incremental causal effect of AI adaptivity within an otherwise identical flipped course. Historical comparisons are vulnerable to cohort and instructor effects; correlational analyses cannot establish direction of causation; online-versus-onsite comparisons can be confounded by pandemic conditions; and the 2025 uptake comparison is self-selected. Causal claims must therefore remain qualified.

4.4 Evidence Boundary

The source-verification audit also showed that terms such as adaptive, AI-assisted, personalised, and flipped are used inconsistently in the wider literature. Several previously retained records failed the review's operational definition because adaptation was instructor-created rather than algorithmic, the intervention lacked a structured flipped architecture, or the population fell outside undergraduate engineering or computing. Maintaining a strict eligibility boundary reduces the apparent size of the corpus but improves construct validity.

No robust minimum duration, effect-size range, preferred algorithm, or implementation threshold can be inferred from this evidence base. The corpus is small, several reports are dependent, and intervention components vary across studies and time.

The most defensible summary is that AI-adaptive pre-class learning and analytics are feasible and can produce favourable student experiences, whereas evidence for reliable improvement in achievement remains limited and context dependent.

5. Discussion: Interpretations, Implications, and Unanswered Questions

5.1 Theoretical Synthesis: Readiness-Responsive Pedagogy as a Testable Framework

The verified literature supports a practical sequence rather than a proven causal model: learner performance is observed before or during preparation; the system personalises a pathway or identifies risk; and instructors or tutors use that information within a flipped course. This sequence is compatible with readiness-responsive scaffolding, but the reports do not experimentally isolate each link.

A rigorous future design should distinguish at least three conditions: a high-fidelity conventional flipped course, the same course with instructor-led differentiation using non-algorithmic readiness information, and the same course with algorithmically generated adaptation. Tutoring intensity, assessment, instructor contact, and in-class active-learning opportunities should be held constant wherever possible.

Within such designs, readiness, engagement, assessment alignment, and instructor orchestration should be measured prospectively and analysed as predefined mediators or moderators only when sample size and design support those inferences.

5.1 Engagement as an Outcome and Candidate Pathway

Across the eligible reports, classroom-environment and perception outcomes are often more favourable than direct achievement outcomes. This makes engagement, perceived responsibility, autonomy, workload, and acceptance important outcomes in their own right. However, no report establishes that these variables mediate achievement gains.

This distinction matters because students may prefer an adaptive environment without achieving substantially higher independently assessed performance, while performance may improve because of additional tutoring rather than increased motivation. Future studies should measure these mechanisms at multiple time points and avoid inferring mediation from simple co-occurrence.

Engineering outcomes should extend beyond course grades to independently scored problem solving, debugging, design tasks, transfer to novel problems, retention, teamwork, and subsequent-course performance. These outcomes are better aligned with programme-level engineering competencies.

5.2 Equity and Algorithmic Bias

The reports provide some subgroup analyses, including demographic focus groups, underrepresented-minority analyses, Pell-grant comparisons, and student reactions to targeted support. These analyses are informative but do not constitute a comprehensive fairness evaluation. In particular, predictive-model performance in Kaw *et al.* (2025) is not reported using subgroup sensitivity, specificity, false-positive rates, calibration, or predictive parity.

Concerns about algorithmic bias remain theoretically plausible and are supported by the wider fairness literature (Barocas and Selbst, 2016), but the current corpus cannot determine whether adaptive-flipped systems reduce or reproduce inequity. Future studies should predefine subgroup analyses, report model performance across relevant demographic and access groups, document consent and opt-out arrangements, and provide human review of consequential algorithmic flags.

Equity measurement should also include differential uptake of support, accessibility, connectivity constraints, perceived stigma, autonomy, and whether interventions change attainment gaps. Such analyses are especially important when students are labelled or contacted because an algorithm predicts low performance.

Accordingly, neither universal benefit nor systematic inequity can be inferred. The appropriate conclusion is that equity effects remain inadequately measured and should be incorporated prospectively into future evaluations.

5.3 Implementation Costs and Scalability

The verified reports do not provide a complete prospective cost analysis covering platform licensing, content development, model maintenance, instructor or tutor time, training, technical support, and data governance. This omission prevents estimation of cost-effectiveness and complicates comparison of algorithmic adaptation with well-organised human differentiation.

Future evaluations should collect resource inputs alongside educational outcomes. This is particularly important for early-warning systems in which an algorithm triggers labour-intensive tutoring, because both the principal cost and a substantial proportion of the educational benefit may arise from human support rather than from prediction itself.

5.4 Priority Research Questions

First, what incremental benefit does algorithmic adaptation provide over a high-fidelity flipped course with equivalent human support and instructional attention?

Second, which adaptive functions are educationally consequential: personalised sequencing, difficulty adjustment, automated feedback, resource recommendation, early-warning prediction, or allocation of human support? Studies should identify the adaptive decision explicitly and report the learner data, model or rule, threshold, and resulting instructional action.

Third, are the reported effects transferable beyond numerical methods and beyond the closely connected research teams represented in the present corpus? Independent, multi-site replication across engineering disciplines is required.

6. Conclusions

This systematic review identified five eligible peer-reviewed journal reports of AI-adaptive flipped learning in undergraduate engineering published between 2015 and 2025. The reports show that adaptive platforms can personalise pre-class preparation and that learning analytics can support early identification and targeted intervention. Student-experience and classroom-environment outcomes are generally favourable, whereas direct achievement differences are smaller, inconsistent, or susceptible to confounding.

The evidence should not be interpreted as five independent confirmations. The reports are concentrated in numerical-methods engineering, involve overlapping investigators and, in several cases, overlapping or sequential cohorts, and rely largely on historical comparisons, correlational analyses, or self-selected uptake of support. No report cleanly isolates the incremental causal contribution of AI adaptivity from high-quality flipped pedagogy, instructor orchestration, online-delivery conditions, or human tutoring.

The immediate research need is independent replication with stronger causal identification. Future studies should use transparent definitions of adaptivity and flipped design; include equivalent non-adaptive flipped comparators; preregister outcomes; document model versions, decision rules, intervention fidelity, and human support; assess delayed and independently scored engineering performance; collect cost and workload data; and specify equity and fairness analyses in advance. Until such evidence accumulates, AI-adaptive flipped learning is best described as a promising and feasible instructional approach with encouraging experience-related outcomes but an incompletely established advantage in achievement.

References

Ambrose, S.A., Bridges, M.W., DiPietro, M., Lovett, M.C., Norman, M.K. (2010). How learning works: Seven research-based principles for smart teaching. *Jossey-Bass*.

Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., Cloude, E., Dever, D., Wiedbusch, M., Wortha, F., Cerezo, R. (2022). Lessons learned and future directions of MetaTutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology*, 13, 813632. https://doi.org/10.3389/fpsyg.2022.813632

Barocas, S., Selbst, A.D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732. https://dx.doi.org/10.2139/ssrn.2477899

Bonwell, C.C., Eison, J.A. (1991). Active learning: Creating excitement in the classroom (ASHE-ERIC Higher Education Report. *ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183*.

Campbell, M., McKenzie, J.E., Sowden, A., Katikireddi, S.V., Brennan, S.E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, 368, l6890. https://doi.org/10.1136/bmj.l6890

Chen, H.R., Hsu, W.C. (2022). Do flipped learning and adaptive instruction improve student learning outcome? A case study of a computer programming course in Taiwan. *Frontiers in Psychology*, 12, 768183. https://doi.org/10.3389/fpsyg.2021.768183

Chi, M.T.H., Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243. https://doi.org/10.1080/00461520.2014.965823

Clark, R.M., Kaw, A.K. (2020a). Adaptive learning in a numerical methods course for engineers: Evaluation in blended and flipped classrooms. *Computer Applications in Engineering Education*, 28(1), 62-79. https://doi.org/10.1002/cae.22175

Clark, R.M., Kaw, A.K. (2020b). Benefits of adaptive lessons for pre-class preparation in a flipped numerical methods course. *International Journal of Mathematical Education in Science and Technology*, 51(5), 713-729. https://doi.org/10.1080/0020739X.2019.1617439

Clark, R.M., Kaw, A.K., Braga Gomes, R. (2022). Adaptive learning: Helpful to the flipped classroom in the online environment of COVID. *Computer Applications in Engineering Education*, 30(2), 517-531. https://doi.org/10.1002/cae.22470

Corbett, A.T., Anderson, J.R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253-278. https://doi.org/10.1007/BF01099821

du Plooy, E., Casteleijn, D., Franzsen, D. (2024). Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement. *Heliyon*, 10(21), e39630. https://doi.org/10.1016/j.heliyon.2024.e39630

Duffy, M.C., Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior*, 52, 338-348. https://doi.org/10.1016/j.chb.2015.05.041

Freeman, S., Eddy, S.L., McDonough, M.K., Smith, M.K., Okoroafor, N., Jordt, H., Wenderoth, M.P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

González, L.A., Neyem, A., Contreras-McKay, I., Molina, D. (2022). Improving learning experiences in software engineering capstone courses using artificial intelligence virtual assistants. *Computer Applications in Engineering Education*, 30(5), 1370-1389. https://doi.org/10.1002/cae.22526

Hu, J., Huang, Z., Li, J., Xu, L., Zou, Y. (2024). Real-time classroom behavior analysis for enhanced engineering education: An AI-assisted approach. *International Journal of Computational Intelligence Systems*, 17(1), 167. https://doi.org/10.1007/s44196-024-00572-y

Katona, J., Gyonyoru, K.I.K. (2025). AI-based adaptive programming education for socially disadvantaged students: Bridging the digital divide. *TechTrends*, 69, 925-942. https://doi.org/10.1007/s11528-025-01088-8

Kaw, A., Clark, R., Delgado, E., Abate, N. (2019). Analyzing the use of adaptive learning in a flipped classroom for preclass learning. *Computer Applications in Engineering Education*, 27(3), 663-678. https://doi.org/10.1002/cae.22106

Kaw, A., Yalcin, A., Clark, R. (2025). Using adaptive learning platform data in a flipped classroom for early detection and tutoring of low-performing students. *Computer Applications in Engineering Education*, 33(2), e70007. https://doi.org/10.1002/cae.70007

Lundin, M., Bergviken Rensfeldt, A., Hillman, T., Lantz-Andersson, A., Peterson, L. (2018). Higher education dominance and siloed knowledge: A systematic review of flipped classroom research. *International Journal of Educational Technology in Higher Education*, 15(1), 20. https://doi.org/10.1186/s41239-018-0101-6

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

Piaget, J. (1954). The construction of reality in the child. *Basic Books*. https://psycnet.apa.org/doi/10.1037/11168-000

Rethlefsen, M.L., Kirtley, S., Waffenschmidt, S., Ayala, A.P., Moher, D., Page, M.J., Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews*, 10, 39. https://doi.org/10.1186/s13643-020-01542-z

Sterne, J.A.C., Hernán, M.A., Reeves, B.C., Savović, J., Berkman, N.D., Viswanathan, M., Henry, D., Altman, D.G., Ansari, M.T., Boutron, I., Carpenter, J.R., Chan, A.W., Churchill, R., Deeks, J.J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y.K., Pigott, T.D., Craig R Ramsay, Regidor, D., Rothstein, H.R., Sandhu, L., Santaguida, P.L., Schünemann, H.J., Shea, B., Shrier, I., Tugwell, P., Turner, L., Valentine, J.C., Waddington, H., Waters, E., Wells, G.A., Whiting, P.F., Higgins, J.P.T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355, i4919. https://doi.org/10.1136/bmj.i4919

Tomlinson, C.A. (2001). How to differentiate instruction in mixed-ability classrooms (2nd ed.). *Association for Supervision and Curriculum Development*.

Tomlinson, C.A., Moon, T.R. (2013). Assessment and student success in a differentiated classroom. *Association for Supervision and Curriculum Development*.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychology Review*, 23(3), 309-342. https://doi.org/10.1080/00461520.2011.611369

Vygotsky, L.S. (1978). Mind in society: The development of higher psychological processes (M. Cole, V. John-Steiner, S. Scribner, E. Souberman, Eds.). *Harvard University Press*.

Wang, H., Tlili, A., Huang, R., Cai, Z., Li, M., Cheng, Z., Yang, D., Li, M., Zhu, X., Fei, C. (2023). Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective. *Education and Information Technologies*. https://doi.org/10.1007/s10639-022-11555-x

Williamson, K., Kizilcec, R.F. (2022). A review of learning analytics dashboard research in higher education: Implications for justice, equity, diversity, and inclusion. Proceedings of the *12th International Conference on Learning Analytics and Knowledge*, 260-270. https://doi.org/10.1145/3506860.3506900

Wood, D., Bruner, J.S., Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Author Contribution Statement

Quanbo Geng: Conceptualization, Data Curation, Methodology, Visualization, Writing - Original Draft. Aidah Abdul Karim: Methodology, Investigation, Writing - Review & Editing, Supervision. Hazrati Husnin: Formal Analysis, Writing - Review & Editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Data Availability

Data supporting the findings of this study are available from the corresponding author upon reasonable request.

Does this article screen for similarity?

Yes

Conflict of Interest

The authors have no conflicts of interest to declare. There is also no financial interest to report. The author certifies that the submission is original work and is not under review at any other publication.

About the License

© The Author(s) 2026. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International Licenses.

Cite this Article

Quanbo Geng, Aidah Abdul Karim, Hazrati Husnin, A Systematic Review of AI-Enhanced Adaptive Flipped Classrooms in Undergraduate Engineering: Analytical Frameworks, Pedagogical Mechanisms, and Research Imperatives, Asian Journal of Interdisciplinary Research, 9(4), (2026) 78-93. https://doi.org/10.54392/ajir2645