Trends and Reviews of Technological Engineering in Education
Vol. 2, No. 2 (2026) | ISSN 3071-2211 | Open Access

REVIEW ARTICLE

AI-Supported Self-Regulated Learning in EFL Higher Education: A Review of Learner Literacy, Technology Perceptions, and Academic Outcomes

Dan Dan Ji[1,*], Juliana Rosmidah Jaafar[1]
[1] Faculty of Business, Information and Human Sciences, Kuala Lumpur University of Science & Technology, Kajang, 43000, Malaysia
Correspondence: 243925326@s.klust.edu.my

Received 18 December 2025 · Accepted 25 April 2026 · Published 30 June 2026

Abstract

Artificial intelligence (AI) applications — from automated writing evaluation and intelligent tutoring platforms to chatbots built on large language models — have moved into the everyday study routines of English as a Foreign Language (EFL) students in higher education. How these tools relate to learning may depend on both their capabilities and learners' self-regulatory behaviour. This article reports a PRISMA 2020-guided systematic review of 52 peer-reviewed empirical studies (January 2019 to June 2025) that examined AI-supported self-regulated learning (SRL) among EFL learners in higher education, organised around three questions: how learner AI literacy is conceptualised and measured; how learners perceive AI tools; and what academic and self-regulatory outcomes are associated with AI-supported study. Annual output rose sevenfold from 2019 to 2024; 2025 covers January–June only. East and Southeast Asian settings predominate; 24 studies examine LLM-based chatbots. Narrative synthesis indicates that AI literacy is unevenly conceptualised: functional operation is widely assumed, whereas the critical-evaluative capacities needed to appraise machine feedback are rarely measured and appear least developed. Perception evidence shows that usefulness and ease-of-use judgments are broadly favourable, but trust calibration and fear of over-reliance form a persistent counter-current that shapes strategy choices. Aggregation of 23 studies reports a pooled advantage for AI-supported conditions (Hedges' $g = 0.56$, 95% CI [0.44, 0.68]), with point estimates of $g = 0.68$ for writing performance and $g = 0.62$ for SRL strategy use. Contrasts by instructor scaffolding and duration are exploratory and do not establish causal moderation. We propose a framework linking learner readiness, technology perceptions, AI-mediated SRL engagement, and outcomes. These paths are hypotheses requiring direct testing. Critical AI literacy and strategy instruction are candidate curriculum priorities, but the reported associations do not establish how much either would improve achievement. The synthesis supports further evaluation of structured AI use while leaving the contribution of individual instructional components unresolved.

Keywords: Self-regulated learning; Artificial intelligence; EFL higher education; AI literacy; Technology acceptance; Learning outcomes; Systematic review

How to cite: Dan Dan Ji & Juliana Rosmidah Jaafar (2026). AI-Supported Self-Regulated Learning in EFL Higher Education: A Review of Learner Literacy, Technology Perceptions, and Academic Outcomes. TRTEE, 2(2), 29–51. https://doi.org/10.63646/YQ0Z5667

1. Introduction

Few developments have reshaped the daily practice of English as a Foreign Language (EFL) study as quickly as the arrival of accessible artificial intelligence. For two decades, computer-assisted language learning research treated intelligent systems as promising but peripheral; within a remarkably short period, automated feedback engines, speech-enabled assistants, and above all chatbots built on large language models have become tools that university language learners reach for without being told to [1], [2]. The release of conversational generative systems in late 2022 marked a visible inflection point: students discovered that a single free interface could draft, correct, explain, translate, and rehearse — functions that previously required separate tools, tutors, or considerable patience [3].
AI use can extend beyond direct supervision, making learner self-regulation relevant. Self-regulated learning (SRL) involves setting goals, selecting strategies, monitoring progress, and adapting. Evidence links these processes to achievement, but differs in population and outcome: a meta-analysis concerns children and adolescents [4], an online higher-education review examines strategy-achievement associations [5], and a MOOC study examines learner behaviour and goal attainment [6]. These sources do not establish SRL as one of the strongest modifiable causes of university EFL achievement. Language-learning research has increasingly situated learning strategies within a broader, cyclical regulatory system [7], [8]. AI tools may alter the activities learners regulate: generated text and feedback create decisions about delegation, verification, and evaluation. Their contribution to learning remains a question for investigation. This review examines outcomes and learner engagement with AI.
Research has responded energetically but unevenly. Three literatures now run largely in parallel. One asks what learners need to know about AI — the emerging construct of AI literacy and how that knowledge distributes across student populations. A second examines how learners perceive AI tools, typically through the vocabulary of technology acceptance: usefulness, ease of use, trust, and anxiety. A third estimates effects of AI-supported interventions on language proficiency and, less often, on self-regulatory behaviour itself. Each strand is informative alone and incomplete alone: outcome studies rarely explain why effects vary; perception studies rarely follow through to achievement; literacy studies rarely connect knowledge to actual study behaviour. Because AI capabilities change faster than curricula, the field needs a learner-centred synthesis that reads these strands against each other rather than another tool-centred inventory [9], [10].
This review provides that synthesis. It makes four contributions. First, it maps the empirical landscape of AI-supported SRL in EFL higher education through a PRISMA 2020-guided review of 52 studies published between January 2019 and June 2025, documenting the field’s growth, geography, designs, and tool families. Second, it synthesises how learner AI literacy has been conceptualised and measured, identifying a systematic gap between functional and critical dimensions of the construct. Third, it consolidates perception evidence through an acceptance-theoretic lens and examines links with strategic behaviour. Fourth, it summarises quantitative outcomes and exploratory contrasts by study characteristics. The integrated framework proposes testable relationships and a research agenda, without establishing a causal sequence.

© 2026 TRTEE. Open Access under CC BY 4.0

The article proceeds as follows. Section 2 situates the review within the SRL, educational AI, AI literacy, and technology acceptance literatures and states the research questions. Section 3 details the review methodology. Section 4 reports findings for each research question, including the effect-size aggregation. Section 5 develops the integrative framework. Section 6 discusses theoretical, pedagogical, and ethical implications together with the limitations of the evidence base, and Section 7 concludes with directions for future research.

2. Background and Literature Review

2.1 Self-Regulated Learning in Language Education

Self-regulated learning describes the processes through which learners deliberately activate and sustain cognition, motivation, and behaviour in pursuit of learning goals. Zimmerman's cyclical account organises these processes into forethought, performance, and self-reflection phases [11], while Pintrich's framework extends regulation across cognitive, motivational, behavioural, and contextual areas [12]; comparative analyses show that the major models converge on this cyclical, multi-componential core despite differences in emphasis [13]. A meta-analysis finds small achievement correlations in children and adolescents [4]. Other sources examine online university learning [5] and MOOC goal attainment [6]. These distinct settings and outcomes do not establish a common causal effect.
Language education arrived at SRL through research on learning strategies. Rich taxonomies and persistent measurement concerns prompted a turn toward self-regulation as an organising construct [8]. Models validated with EFL writers distinguish cognitive, metacognitive, social-behavioural, and motivational regulation [7]. Learners exercise these capacities in technology-rich settings beyond the classroom [14]. A cross-sectional study of 525 undergraduates found associations among technology-assisted SRL, self-efficacy, enjoyment, and English outcomes [15]. It did not establish causal direction or isolate AI-specific effects. This study provides background on technology-assisted learning. Scaffolding research addresses how systems and instructors can support planning, monitoring, and reflection [16], with meta-analytic evidence for embedded SRL scaffolds [17].

2.2 Artificial Intelligence Technologies in EFL Learning

Language-learning AI draws on several research traditions. Intelligent tutoring systems have been compared with human tutoring [18] and conventional instruction [19], although modelling open-ended communication remains difficult [20]. Chatbot-partner studies report situational interest and novelty decline relative to human partners [21], [22]. Mobile applications [23] and personal assistants extend practice opportunities. Listening and speaking evidence for university EFL learners is discussed in [24]. In contrast, Tai and Chen studied 112 eighth-grade learners using Google Assistant for two weeks and reported greater willingness to communicate [25]. That school-age study is contextual background, outside this review's higher-education eligibility criterion.

© 2026 TRTEE. Open Access under CC BY 4.0

Automated writing evaluation (AWE) has a substantial research history. An early review found limited evidence for better AWE-assisted texts and little evidence of transfer to general writing proficiency [26]. Learners differ in feedback uptake [27], engagement with automated and teacher feedback [28], and revision processes within teacher-AWE arrangements [29]. Large language models offer educational opportunities and risks [1], including language practice and feedback alongside accuracy and integrity concerns [2], [30]. Su, Lin, and Lai [31] propose ways to integrate ChatGPT into argumentative writing; their article is a pedagogical proposal, not a classroom effectiveness experiment. Yan [32] reports exploratory learner experiences, while Mahapatra [33] examines an intervention with undergraduate ESL students in India. The latter supplies related context, not automatically evidence eligible for this review's EFL corpus. Conceptual accounts describe partner-oriented AI [34] and hybrid human-AI regulation [35]. These sources motivate instructional questions while differing in design, population, and evidential role.

2.3 AI Literacy and Feedback Literacy

If AI tools are study partners, learners need partner-appraisal skills. AI literacy has been defined as the set of competencies that enables individuals to evaluate, communicate, and collaborate with AI critically and effectively [36], and reviews organise the construct into knowing and understanding, using and applying, evaluating and creating, and attending to ethical issues [37]. Two features of these frameworks matter here. First, they distinguish functional operation from critical evaluation — knowing how to prompt a system is not knowing when to believe it. Second, they overlap substantially with feedback literacy, the capacity to appreciate feedback, make judgments, manage affect, and take action [38]; acting productively on machine feedback is a joint test of both literacies. Parallel work on the instructor side, such as the Intelligent-TPACK framework, underscores that ethically integrating AI requires professional knowledge that cannot be presumed [39] — an assumption this review examines on the learner side.

2.4 Technology Perceptions and Acceptance

Technology acceptance research relates perceptions to adoption. TAM foregrounds usefulness and ease of use [40]; UTAUT includes performance and effort expectancy, social influence, and facilitating conditions [41]. In a survey of 503 Polish university students, habit, performance expectancy, and hedonic motivation were associated with intention to use ChatGPT [42]. This was general university use, not specifically EFL learning. Liu and Ma surveyed 405 EFL learners: ease of use related to attitude indirectly through usefulness, while its direct path was not significant [43]. Thus, the findings do not confirm every proposed acceptance path. Broader student accounts combine favourable appraisals with accuracy, privacy, and integrity concerns [44]. Questionnaire-based intention and reported use should be distinguished from observed learning behaviour over time. How perceptions relate to strategic engagement remains a question for this review.

© 2026 TRTEE. Open Access under CC BY 4.0

2.5 Prior Reviews and the Present Contribution

Several syntheses frame the broader territory. Reviews of AI in higher education document rapid growth alongside thin theoretical grounding and limited attention to ethics [45], [9], a diagnosis reiterated by meta-review evidence calling for greater rigour and collaboration [46] and by field-level mappings of opportunities and challenges [47]. Language-specific syntheses chart the roles and research foci of AI in language education [10], [48], the interactional affordances of dialogue systems for university EFL learners [49], and the pedagogical potential of educational chatbots [50]. Meta-analyses report positive average effects of chatbots [51] and of AI-based instruction more broadly [52] on learning outcomes. None of these, however, centres self-regulated learning in EFL higher education while jointly examining learner literacy, perceptions, and outcomes. The present review addresses that gap through three questions. RQ1: How is learner AI literacy conceptualised, measured, and distributed in EFL higher education research? RQ2: How do EFL learners perceive AI tools, and how do these perceptions relate to self-regulated use? RQ3: What academic and self-regulatory outcomes are associated with AI-supported learning, and under what conditions are they strongest?

3. Methodology

3.1 Review Design

The review followed the PRISMA 2020 statement for reporting systematic reviews [53]. A protocol specifying the questions, eligibility criteria, search strings, screening procedure, coding frame, and synthesis plan was agreed by the author team before searching began. The design is deliberately hybrid: a configurative, narrative synthesis addresses the conceptual questions (RQ1 and RQ2), while an aggregative component pools standardised effect sizes from the subset of studies reporting sufficient statistics (RQ3).

3.2 Search Strategy and Eligibility Criteria

Six databases Scopus, Web of Science, ERIC, IEEE Xplore, ScienceDirect, and ProQuest were searched for records published between 1 January 2019 and 30 June 2025 using the Boolean string (“artificial intelligence” OR “chatbot” OR “large language model” OR “ChatGPT” OR “automated writing evaluation” OR “intelligent tutoring”) AND (“EFL” OR “English as a foreign language” OR “L2 English”) AND (“self-regulated learning” OR “self-regulation” OR “learner autonomy” OR “perception” OR “literacy” OR “achievement”). The 2019 starting point captures the field’s modern phase after the first experimental chatbot-partner studies and before the generative turn so that pre- and post-2022 research can be compared within one corpus. Backward and forward citation searching of included studies supplemented the database search. Table 1 summarises the eligibility criteria applied at every screening stage.

Table 1. Inclusion and exclusion criteria applied during screening

Dimension	Inclusion requirement	Exclusion condition
Population	Students enrolled in higher education EFL programmes	School-age learners; second-language immersion contexts; instructor-only samples

© 2026 TRTEE. Open Access under CC BY 4.0

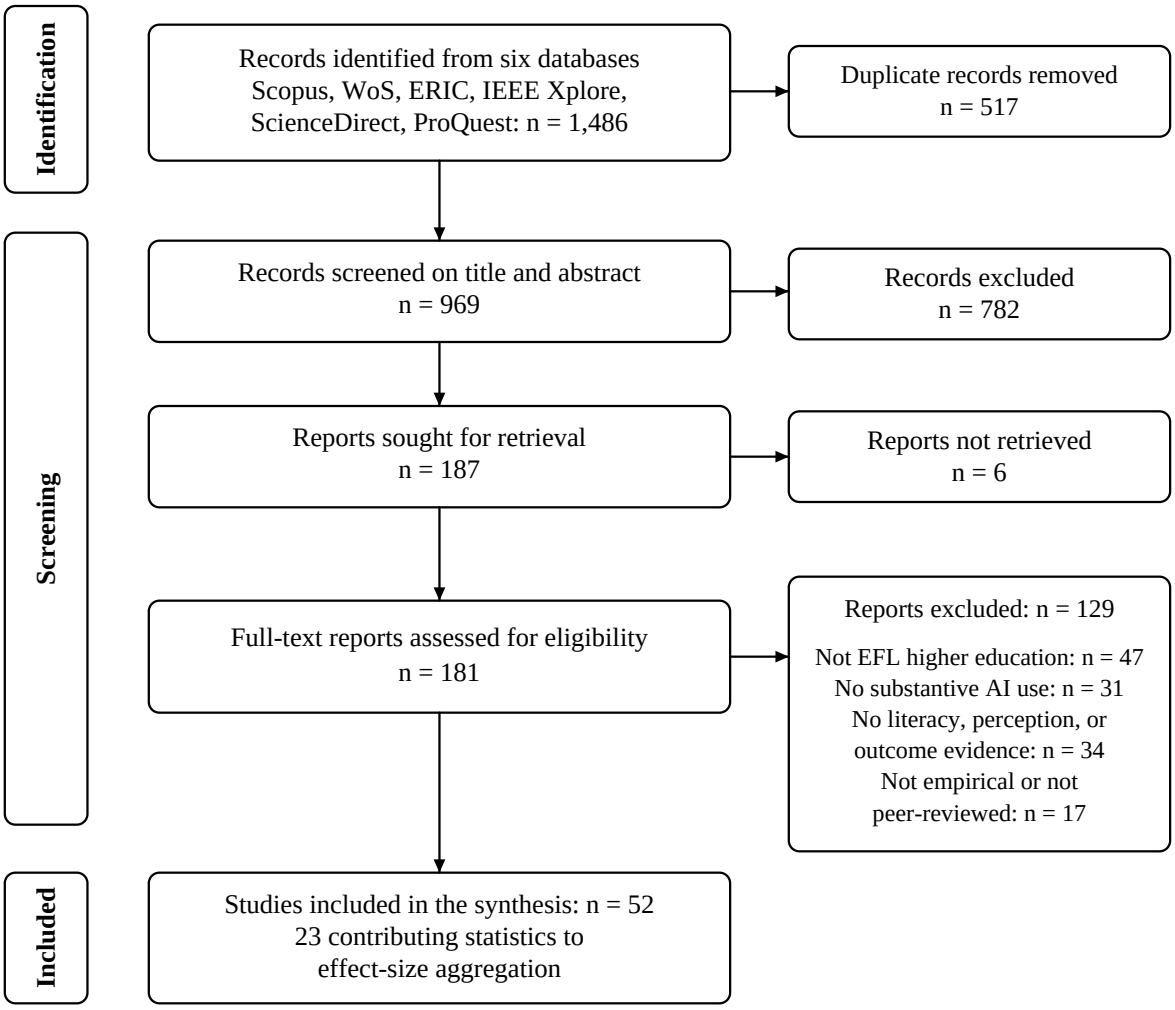
Dimension	Inclusion requirement	Exclusion condition
Technology	AI-driven tools: chatbots and LLM assistants, AWE, intelligent tutoring, speech agents, adaptive platforms	Static courseware or non-adaptive CALL without a substantive AI component
Constructs	Empirical evidence on AI or feedback literacy, tool perceptions, SRL processes, or learning outcomes	Purely technical system evaluations without learner data
Design and source	Peer-reviewed empirical journal articles or full conference papers	Editorials, position papers, theses, and other grey literature
Language and period	Published in English, January 2019 to June 2025	Other publication languages; records before 2019

The criteria were interpreted conservatively at the margins. Studies of English-medium instruction were included only when English learning itself was an explicit object of self-regulation; studies of pre-service English teachers were included only when participants were positioned as language learners rather than as future instructors.

3.3 Screening and Selection

Two reviewers independently screened all titles and abstracts against the criteria, achieving strong inter-rater agreement (Cohen’s kappa = 0.86, interpreted following established guidance [54]); disagreements were resolved by a third reviewer. The same procedure governed full-text assessment. Figure 1 presents the selection flow.
Database searching returned 1,486 records, reduced to 969 after de-duplication. Title and abstract screening excluded 782 records; of 187 reports sought, six could not be retrieved and 181 were assessed in full text. Full-text assessment excluded 129 reports, most commonly because the setting was not EFL higher education (n = 47) or because no construct of interest was measured (n = 34). The final corpus comprises 52 studies, of which 23 reported group contrasts or pre-post statistics sufficient for effect-size computation.

© 2026 TRTEE. Open Access under CC BY 4.0



Reported database flow; citation-search counts are not separately reported.

Figure 1. Reported Study Selection Flow Based on PRISMA 2020

3.4 Coding and Quality Appraisal

Each study was coded along five dimensions: context and sample, technology family, theoretical framing, constructs and instruments, and findings. The coding frame was piloted on eight studies, refined, and applied to the full set by two coders working independently on a 30 percent overlap sample (agreement 92 percent) and dividing the remainder. Methodological quality was appraised with an eight-item checklist adapted from established appraisal practice, covering sampling transparency, instrument validity evidence, design-inference fit, and reporting completeness. Quality scores informed the synthesis narratively and through sensitivity checks; no study was excluded on quality grounds alone, in keeping with the review's mapping purpose.

© 2026 TRTEE. Open Access under CC BY 4.0

3.5 Synthesis Approach

Narrative synthesis proceeded thematically within each research question, with themes generated inductively from the coded findings and audited against the source texts by a second coder. For RQ3, standardised mean differences were computed as Hedges' g with small-sample correction, combined under a random-effects model, and accompanied by 95 percent confidence intervals and the I-squared heterogeneity statistic; analyses were conducted in R using the metafor package. Given the heterogeneity of designs, instruments, and comparison conditions in a young literature, the aggregation is reported as an evidence summary with explicit caveats rather than as a definitive meta-analytic estimate, and moderator contrasts are treated as exploratory. Three sensitivity analyses are reported: exclusion of studies below the quality median, exclusion of single-session designs, and trim-and-fill for funnel asymmetry. The latter assesses a particular missing-study scenario; it does not establish an unbiased effect estimate.

4. Findings

4.1 Profile of the Evidence Base

Figure 2 reports two studies in 2019 and fourteen in 2024, a sevenfold ratio between complete calendar years. Its nine studies for 2025 cover January to June and are not a full-year total. The annual counts sum to 52. This growth can be considered alongside broader research on AI in higher education [9]. Table 2 classifies 24 of the 52 studies (46.2 percent) as LLM-based or generative chatbots, the largest technology category across the review window. That aggregate share does not establish a year-specific proportion or growth rate for any tool family.

Figure 2 reports 19 studies from East Asia and 11 from Southeast Asia, together 30 of 52 (57.7 percent). It also lists eight for the Middle East and North Africa, six for Europe, five as multi-region or other, and three for Latin America. These categories total 52. The combined Middle East/North Africa category and the multi-region/other category do not identify a separate total for Africa. The reported distribution limits geographic generalisation; it does not establish that African EFL settings are absent.

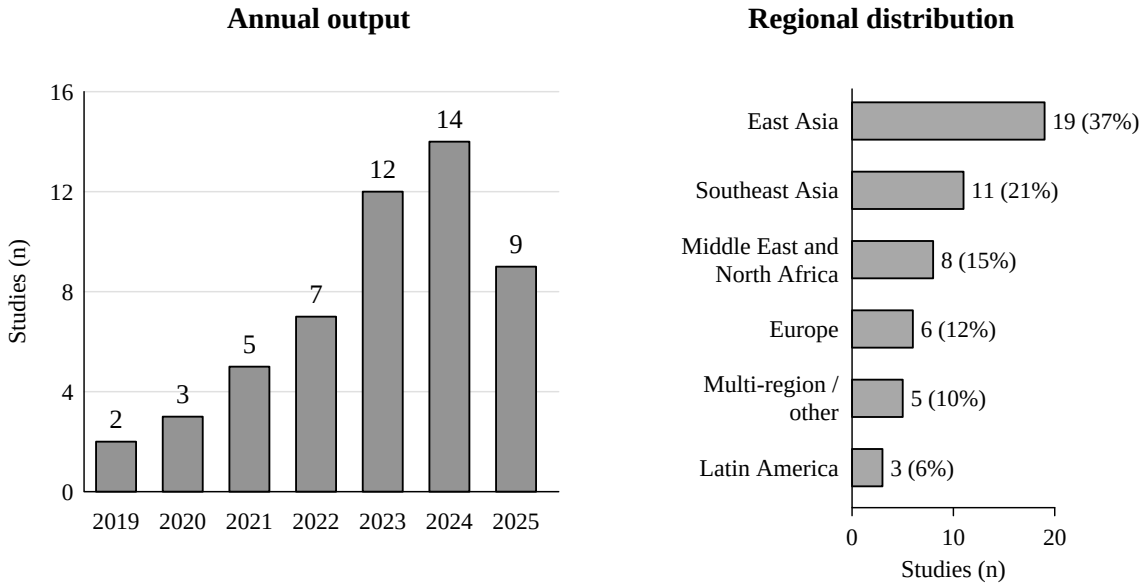
Table 2 lists 21 experimental or quasi-experimental studies (40.4 percent), 14 survey or correlational studies, 11 mixed-methods studies, and six qualitative or case studies. Its duration categories contain 15, 23, and 14 studies. The first two categories therefore contain 38 of 52 studies (73.1 percent). These duration counts cover the whole corpus, but their application to non-intervention designs is unspecified. Without a design-by-duration cross-tabulation, they cannot identify the number of short interventions or the combined share of surveys and short interventions.

Table 2. Characteristics of the 52 included studies

Dimension	Category	Studies (n)	Share (%)
Research design	Experimental or quasi-experimental	21	40.4
	Survey or correlational	14	26.9
	Mixed methods	11	21.2

© 2026 TRTEE. Open Access under CC BY 4.0

Dimension	Category	Studies (n)	Share (%)
	Qualitative or case study	6	11.5
Technology family	LLM-based or generative chatbots	24	46.2
	Automated writing evaluation	12	23.1
	Intelligent tutoring or adaptive platforms	7	13.5
	Speech-based assistants	6	11.5
	Other AI applications	3	5.8
Sample size	Fewer than 50 participants	12	23.1
	50–150 participants	24	46.2
	151–400 participants	11	21.2
	More than 400 participants	5	9.6
Reported duration	Single session to 4 weeks	15	28.8
	5–12 weeks	23	44.2
	One semester or longer	14	26.9



2025: January-June only. Percentages are rounded.

Figure 2. Reported Annual Output and Regional Distribution

© 2026 TRTEE. Open Access under CC BY 4.0

The median quality score was six of eight. Aims and instruments were more clearly reported than sampling justification and attrition; fewer than one in five quantitative studies reported an a priori power analysis, and nine shared instruments or data openly. Lower-scoring studies reportedly had larger effects on average, without a consistent difference in direction. This is a quality-related pattern, not evidence of small-study effects: quality scores do not measure sample size or precision. Section 4.4 reports sensitivity analyses.

4.2 Learner AI Literacy (RQ1)

The review reports that 17 of 52 studies explicitly measured AI literacy, using varied instruments informed by frameworks such as [36], [37]. Functional literacy concerns operating tools and obtaining output; critical literacy concerns evaluating accuracy, bias, and suitability. Assumed familiarity is not a measurement of functional competence. Six studies reportedly assessed both dimensions and found lower critical scores, with gaps described as about one standard deviation. Study identities, scale properties, and the standardisation method are needed to verify that comparison. This suggests a possible gap; it does not establish uniformly high functional skills or comparable differences across instruments. Generative-era accounts describe prompt competence: specifying tasks, constraining responses, and revising requests. These behaviours may indicate aspects of functional use, but do not by themselves validate a measure of broader AI literacy. The reported qualitative patterns include single-shot prompts, acceptance of initial responses, and limited checking. Earlier work on automated-feedback uptake [27] provides relevant background; it does not establish the prevalence of these patterns in the present corpus. Reported associations link critical literacy with selective uptake, verification, and task-tool matching, consistent with questions raised by feedback literacy [38]. Their causal direction remains unresolved. Much of the evidence is described as self-report, so associations may also reflect shared measurement methods. Performance tasks and longitudinal observations could test these interpretations. The reported absence of literacy-growth studies needs confirmation from the study-level coding record. The synthesis reports associations of digital coursework and informal AI experience with literacy, mostly non-significant gender comparisons, and higher functional scores among technology-adjacent majors in three studies. These descriptions do not rank causal determinants. Weak or non-significant associations with English proficiency do not show that language study can never contribute to critical literacy; estimates, precision, and measurement quality matter. Four studies reportedly embedded explicit AI literacy instruction and had high critical-literacy scores. Without comparable baseline measures, control groups, and study-level estimates, that pattern does not isolate an instructional effect or establish that institutional exposure is more influential than other factors. It motivates direct evaluation of literacy teaching. The original coding and results are needed to verify the reported counts and comparisons.

4.3 Technology Perceptions (RQ2)

Perception evidence is the most abundant strand of the corpus: 44 of the 52 studies report perception data in some form, most commonly framed by the technology acceptance model

© 2026 TRTEE. Open Access under CC BY 4.0

or its extensions [40], [41]. Table 3 synthesises the recurrent themes and their overall valence across studies.

Table 3. Synthesis of learner perception themes across the corpus

Theme	Recurrent pattern	Studies (n)	Overall valence
Perceived usefulness	Immediate, individualised, always-available feedback valued above all other affordances	41	Positive
Perceived ease of use	Low entry barrier; conversational interfaces reduce tool-learning cost	33	Positive
Trust and accuracy	Confidence in mechanics; doubt about content accuracy, idiomaticity, and cultural fit	27	Mixed
Over-reliance concern	Worry that delegation erodes effort, memory, and independent skill	19	Cautionary
Affective experience	Enjoyment and reduced anxiety, especially for speaking practice with non-judgmental agents	22	Positive
Academic integrity	Ambivalence about legitimate versus illegitimate assistance in graded work	16	Mixed

The perception themes suggest favourable judgments of usefulness and ease of use, including feedback immediacy and accessible interfaces [44], [43]. University-focused dialogue-system research provides context for interactional learning [49]. The anxiety reduction reported by Tai and Chen [25], however, concerns 112 eighth-grade learners; it is background evidence and cannot establish the same result for university EFL students. Trust also involves selective uptake: accepting mechanical corrections while checking content resembles the differentiated engagement described in automated-feedback research [28].

In Table 3, 19 of 44 studies report over-reliance concerns and 16 report integrity ambivalence. Themes can overlap; counts are studies, not learners.

© 2026 TRTEE. Open Access under CC BY 4.0

Related concerns appear in writing-practicum and education discussions [32], [3]. Reported self-imposed usage rules suggest questions about regulation. Strzelecki [42] links habit, performance expectancy, and hedonic motivation to intention in a university survey; it does not demonstrate sustained use over time or improved learning. Reported perceptions differ in emphasis across tool families: revision efficiency and error awareness for AWE, rehearsal and affective safety for speech agents, and usefulness alongside integrity concerns for generative chatbots. The summary does not supply comparable rating scales or formal contrasts that establish which family attracts the strongest responses. Likewise, accounts of more verification among higher-proficiency learners and more acceptance among lower-proficiency learners describe a possible pattern, not a tested moderation effect. Earlier automated-feedback research [27] offers a contextual comparison; it does not demonstrate that the same gradient recurs at scale across the present corpus. Study-level estimates and comparable measures are needed to separate proficiency from task, prior experience, and tool differences. Verification routines are a candidate instructional response to evaluate. Support can be informed by observed learner needs, while claims about its effectiveness require direct comparisons rather than an inference from perception ratings alone.

4.4 Academic and Self-Regulatory Outcomes (RQ3)

The quantitative subset contains 23 studies contributing 53 effects across seven domains. The reported random-effects estimate is Hedges' $g = 0.56$, 95% CI [0.44, 0.68], with $I^2 = 61\%$. Figure 3 shows domain summaries. The 53 effects are not 53 independent studies; study-level inputs and the treatment of dependent estimates are needed to verify the pooled result and its uncertainty. Writing has the largest reported domain estimate ($g = 0.68$, $k = 14$). AWE research [26], [29], a pedagogical proposal [31], and an ESL intervention [33] provide context; these citations do not identify the EFL studies contributing to that estimate. SRL strategy use is reported at $g = 0.62$, $k = 9$, alongside earlier scaffolding evidence [17]. Other estimates are vocabulary $g = 0.55$, speaking $g = 0.51$, grammatical accuracy $g = 0.47$, listening $g = 0.36$ ($k = 4$), and motivation $g = 0.45$ ($k = 8$). The domain counts total 53 effects, without establishing independence. Their ordering is descriptive; no formal between-domain test is reported. Social-cognitive theory [55] supplies context, but the ordering does not show that efficacy beliefs change more slowly than performance. Exploratory means are $g = 0.67$ with instructor scaffolding versus 0.41 without, and 0.63 for eight weeks or more versus 0.44 for shorter interventions. No consistent proficiency pattern is reported. Subgroup counts, intervals, and comparison tests are needed to assess these contrasts; study-level differences do not isolate causal effects. Related syntheses [51], [52] provide context. Reported funnel asymmetry has several possible explanations and does not make the pooled estimate an upper bound. Comparators range from no treatment to strong conventional feedback, a possible source of heterogeneity whose contribution has not been quantified.

© 2026 TRTEE. Open Access under CC BY 4.0

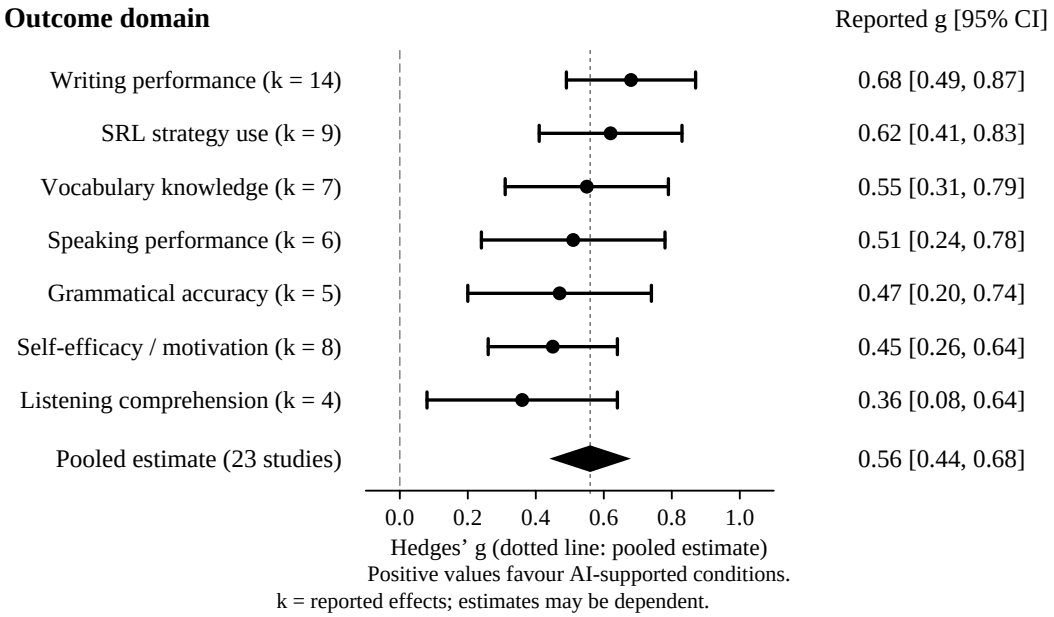


Figure 3. Reported Effects by Outcome Domain

Reported sensitivity estimates are g = 0.51 [0.38, 0.64] after excluding studies below the quality median, g = 0.53 after excluding single-session designs, and g = 0.49 with trim-and-fill. The latter two lack reported intervals here; their uncertainty cannot be compared fully. Trim-and-fill is a sensitivity scenario, not proof that bias has been removed. Similar estimates in chatbot [51] and broader AI-instruction syntheses [52] are contextual comparisons; without a study-overlap check they are not independent confirmation. The reported large effects (g > 1.2), all from interventions of four weeks or fewer, should not be used as programme planning parameters.

4.5 Interrelations Among Literacy, Perceptions, and Outcomes

Eleven studies reportedly measured at least two strands, with associations between critical literacy, differentiated perceptions, and engagement. The cited interaction analysis [56], however, followed 16 primary-school pupils for about three months. Its deep and passive learning profiles provide background, not direct evidence for the eligible university population. This distinction requires a transparent map of included studies and supporting citations. Four studies reportedly tested mediation, with weak or non-significant direct paths from access to achievement after engagement was included. Such fitted paths neither establish causal mediation nor demonstrate that access has no effect. Study identities, timing, confounding control, and estimates with uncertainty are needed to evaluate those reports. The proposed framework therefore organises hypotheses about readiness, perceptions, engagement, and outcomes; the contextual school-age study cannot validate that sequence in higher education.
The strength of these links is worth quantifying where the corpus allows. Across correlational studies, associations between critical AI literacy and the reported depth of strategy use ranged from r = .31 to r = .54, and associations between perceived usefulness

© 2026 TRTEE. Open Access under CC BY 4.0

and sustained voluntary use from r = .38 to r = .61. These are associations between measured constructs. The mediation reports describe indirect paths through engagement quality, but comparable estimates and uncertainty are not supplied here. Neither correlation ranges nor modelled indirect paths quantify the achievement gain from increasing literacy or engagement by half a standard deviation. That intervention prediction requires direct evidence and cannot be calculated from the reported associations alone.

5. An Integrative Framework

Figure 4 presents a proposed framework for AI-supported self-regulated learning in EFL higher education. Learner readiness comprises functional and critical AI literacy, feedback literacy, and a prior SRL profile. Technology perceptions concern usefulness, ease, trust, and threat. The engagement component follows cyclical SRL phases [11]: forethought includes decisions about which tasks to delegate or retain; performance involves prompting, monitoring, and verification; reflection concerns evaluating the interaction and adapting strategy. Language achievement, SRL growth, self-efficacy, and motivation are candidate outcomes. Arrows express hypothesised relationships, not pooled path coefficients or demonstrated causal effects. The framework guides study design; the outcome summaries in Section 4.4 do not test the full sequence.

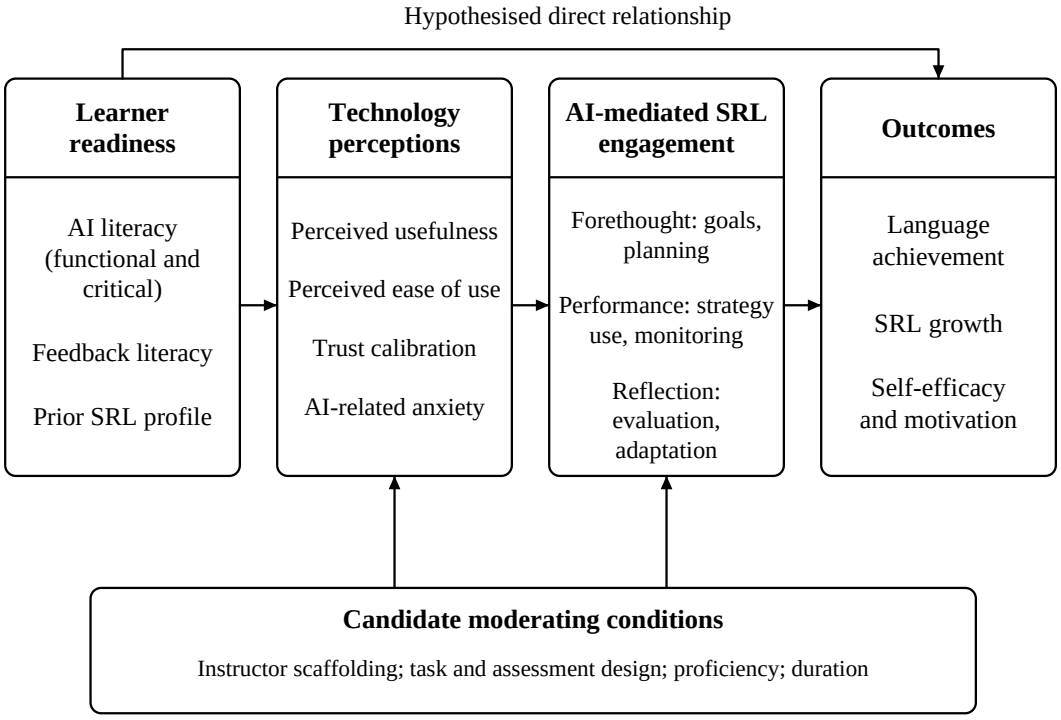


Figure 4. Proposed Model of AI-Supported Self-Regulated Learning

Four propositions guide future tests. P1: readiness may relate to engagement partly through perceptions. P2: instruction addressing perceptions may alter engagement quality, but its effects relative to other supports remain untested. P3: the upper solid arrow connects learner readiness to outcomes; it does not depict tool access. The strength of this direct relationship

© 2026 TRTEE. Open Access under CC BY 4.0

relative to paths through perceptions and engagement requires testing. P4: instructor scaffolding, task and assessment design, proficiency, and duration are candidate moderators; the exploratory contrasts in Section 4.4 do not establish their separate effects. Consistent with hybrid human-AI regulation accounts [35], the model places the tool inside a regulatory process whose contributions learners monitor. Longitudinal and experimental studies could test the proposed ordering, compare alternative pathways, and report direct and indirect estimates with uncertainty. Tool access and learner readiness should be measured as distinct constructs.

The framework suggests questions for curriculum, instruction, task design, and programme evaluation. Curricula can address AI and feedback literacy; instruction can examine usefulness and trust judgments; tasks can prompt planning, verification, and reflection. These are candidate design choices whose effects and resource requirements need evaluation. The current synthesis does not rank them or establish how quickly perceptions can change. Its proposed endpoint includes academic and self-regulatory outcomes, with engagement quality as a possible connecting process. In relation to earlier computer-based learning environments [16], the framework emphasises that AI contributions themselves can become objects of planning, monitoring, and evaluation. This conceptual distinction should guide measurement rather than substitute for evidence that any particular pathway improves learning.

6. Discussion

6.1 Theoretical Implications

Several implications can be considered for theory. First, SRL accounts may need to describe delegation alongside direct task execution. Generative tools allow learners to outsource parts of a task, potentially shifting regulatory activity toward specification, monitoring, and verification. Prompt competence and trust calibration are therefore useful constructs to examine within hybrid-regulation accounts [35] and partner-oriented approaches to educational AI [34]. Second, technology acceptance theory [40], [41] can be extended by studying strategic behaviour after adoption.

The present associations suggest questions about perceptions and engagement, without establishing which variable precedes or changes the other. Third, a possible role for AI literacy is to shape perceptions and engagement, but this remains a proposed mechanism. Studies measuring more than one strand supply relevant associations; they do not alone validate the sequence. Measurement also requires care: prompts and dialogue traces are not automatically valid indicators of planning, monitoring, or evaluation. The interaction analysis in [56] concerns 16 primary-school pupils, so its methods are contextual examples requiring validation for university learners. Future studies should assess whether trace-based measures agree with independent observations, distinguish alternative explanations, and establish temporal ordering. Such work could complement self-report without assuming that every prompt demonstrates self-regulation.

6.2 Pedagogical Implications

For programme leaders and instructors, these findings suggest priorities for evaluation. Critical AI literacy instruction can include verification

© 2026 TRTEE. Open Access under CC BY 4.0

routines, machine-output errors, and purposeful prompting [37], [38]. Structured planning, monitoring, and reflection are also candidates for course design, drawing on earlier scaffolding research [16], [17]. The reported means of g = 0.67 with scaffolding and 0.41 without come from different study groups. They do not isolate the added effect of scaffolding or show that it equals the value of the technology. Direct comparisons with comparable learners, tasks, and outcomes are needed. Prompt logs, drafts, and reflection memos can document use and inform discussions of academic integrity [44], while instructor preparation may support implementation [39]. Writing courses might evaluate draft-feedback-revision cycles with human judgment [29]; speaking courses might examine agent rehearsal with reflection [49]. Verification support for learners at different proficiency levels is another testable option. Programme evaluation should assess learning, workload, accessibility, and unintended reliance, rather than assume that any one design choice guarantees gains. These proposals follow from the synthesis as practical hypotheses, with their relative benefits still to be established.

6.3 Equity, Ethics, and Academic Integrity

The evidence base raises questions about unequal opportunities for AI-supported learning. Prior digital experience may matter for prompt competence and critical literacy, but the regional categories alone do not measure institutional resources or disadvantage. Figure 2 includes three Latin American studies and combines the Middle East with North Africa; it does not establish an absence of African EFL settings. Generalisation should respect this limited geographic detail. Concerns about privacy, opaque models, and the cultural scope of model English also appear in broader AI-in-education research [45], [46], [47]. Integrity ambivalence is reported in sixteen studies. Clear norms can distinguish assistance from substitution [3], while process-oriented assessment can make learners' use of AI more visible.

Material access deserves evaluation alongside perceptions. Subscription costs, device availability, connectivity, and interface accessibility may affect participation, but this review does not quantify their relationships with family income or learning outcomes. Paid access should not be treated as a verified proxy for quality or advantage. The synthesis also claims an absence of accessibility assessments and mentions cohort licences and loan-device schemes without identifying the relevant study-level records here; those claims require verification before they can support corpus-wide conclusions. The framework suggests examining access, literacy support, and scaffolding together. Evaluations can document actual costs, accessibility barriers, participation, and learning across learner groups. Comparisons should consider baseline differences and alternative explanations. The separate and combined contributions of provision and instructional support remain empirical questions rather than established effects of any single institutional response.

6.4 Limitations

Several limitations qualify the synthesis. English-language and peer-reviewed eligibility can restrict coverage; funnel asymmetry alone does not identify bias or its magnitude. The quantitative subset comprises

© 2026 TRTEE. Open Access under CC BY 4.0

23 studies with varied comparators and potentially dependent effects. The 52-study list, screening and coding records, and 23-study effect inputs are needed to verify eligibility, theme counts, and estimates; the contextual references are not a substitute. Self-reported constructs limit inference, and missing design-by-duration coding prevents an intervention-specific duration profile. Novelty effects in early chatbot work [21], [22] and rapid tool change also constrain interpretation. These gaps prevent independent reproduction of the synthesis from the information reported here.

7. Conclusion and Future Directions

This review examines AI-supported SRL among EFL students in higher education. Across the reported 52-study corpus, perceptions are generally favourable, while critical AI literacy is less consistently measured than functional use. The quantitative subset reports a moderate average advantage, with writing and strategy use having the largest domain point estimates. These summaries require caution about heterogeneous comparisons, dependent effects, and missing study-level records. Higher means in scaffolded and longer studies identify questions for testing; they do not show that adding scaffolding or extending duration causes a particular gain. The proposed readiness-perception-engagement framework offers an organising hypothesis, not an established mechanism or a ranking of institutional priorities. Curricula addressing critical literacy and structured AI use merit evaluation alongside access and assessment policies. The contribution of each component, and the conditions under which benefits persist, remain unresolved.

Four research priorities follow. First, longitudinal designs of a semester or longer, using interaction traces alongside self-report, are needed to establish whether early gains and regulatory behaviours persist beyond novelty. Second, the field needs validated, performance-based instruments for critical AI literacy in language learning contexts, so that the construct can be assessed rather than assumed. Third, experimental work should manipulate scaffolding directly — comparing tool-only, strategy-scaffolded, and literacy-plus-strategy conditions — to test the framework's central propositions about mediation and leverage. Fourth, the geographic base must widen to under-represented EFL regions, where infrastructure, policy, and linguistic ecology differ from the settings that dominate current evidence. Together, these studies could establish whether, how, and for whom AI supports language learning, and which conditions contribute to durable gains. They should test alternative explanations and make study-level information available for verification.

Acknowledgements

The authors thank the subject librarians who advised on the database search strategy, the second screening reviewer for their diligence during study selection, and the anonymous reviewers whose comments strengthened the manuscript.

© 2026 TRTEE. Open Access under CC BY 4.0

Author Contributions

D. D. J.: Conceptualization, Methodology, Investigation, Formal Analysis, Writing — Original Draft, Project Administration. J. R. J.: Methodology, Data Curation, Validation, Visualization, Writing — Review & Editing.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The search records, screening decisions, coding sheets, and extracted effect-size data supporting this review are available from the corresponding author on reasonable request.

Ethical Approval

Not applicable. This study is a systematic review of previously published literature and involved no new data collection from human participants.

Generative AI Statement

The original manuscript reports using generative tools only for minor language polishing. Post-publication editorial corrections used AI assistance for wording and statistical interpretation. No research data were generated and no statistical analyses were rerun.

© 2026 TRTEE. Open Access under CC BY 4.0

References

[1] E. Kasneci et al., “ChatGPT for good? On opportunities and challenges of large language models for education,” Learning and Individual Differences, vol. 103, art. 102274, 2023. https://doi.org/10.1016/j.lindif.2023.102274

[2] L. Kohnke, B. L. Moorhouse, and D. Zou, “ChatGPT for language teaching and learning,” RELC Journal, vol. 54, no. 2, pp. 537–550, 2023. https://doi.org/10.1177/00336882231162868

[3] A. Tlili et al., “What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education,” Smart Learning Environments, vol. 10, art. 15, 2023. https://doi.org/10.1186/s40561-023-00237-x

[4] A. L. Dent and A. C. Koenka, “The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis,” Educational Psychology Review, vol. 28, no. 3, pp. 425–474, 2016. https://doi.org/10.1007/s10648-015-9320-8

[5] J. Broadbent and W. L. Poon, “Self-regulated learning strategies and academic achievement in online higher education learning environments: A systematic review,” The Internet and Higher Education, vol. 27, pp. 1–13, 2015. https://doi.org/10.1016/j.iheduc.2015.04.007

[6] R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, “Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses,” Computers & Education, vol. 104, pp. 18–33, 2017. https://doi.org/10.1016/j.compedu.2016.10.001

[7] L. S. Teng and L. J. Zhang, “A questionnaire-based validation of multidimensional models of self-regulated learning strategies,” The Modern Language Journal, vol. 100, no. 3, pp. 674–701, 2016. https://doi.org/10.1111/modl.12339

[8] H. Rose, J. G. Briggs, J. A. Boggs, L. Sergio, and N. Ivanova-Slavianskaia, “A systematic review of language learner strategy research in the face of self-regulation,” System, vol. 72, pp. 151–163, 2018. https://doi.org/10.1016/j.system.2017.12.002

[9] H. Crompton and D. Burke, “Artificial intelligence in higher education: The state of the field,” International Journal of Educational Technology in Higher Education, vol. 20, art. 22, 2023. https://doi.org/10.1186/s41239-023-00392-8

[10] J.-C. Liang, G.-J. Hwang, M.-R. A. Chen, and D. Darmawansah, “Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach,” Interactive Learning Environments, vol. 31, no. 7, pp. 4270–4296, 2023. https://doi.org/10.1080/10494820.2021.1958348

[11] B. J. Zimmerman, “Becoming a self-regulated learner: An overview,” Theory Into Practice, vol. 41, no. 2, pp. 64–70, 2002. https://doi.org/10.1207/s15430421tip4102_2

© 2026 TRTEE. Open Access under CC BY 4.0

[12] P. R. Pintrich, “A conceptual framework for assessing motivation and self-regulated learning in college students,” Educational Psychology Review, vol. 16, no. 4, pp. 385–407, 2004. https://doi.org/10.1007/s10648-004-0006-x

[13] E. Panadero, “A review of self-regulated learning: Six models and four directions for research,” Frontiers in Psychology, vol. 8, art. 422, 2017. https://doi.org/10.3389/fpsyg.2017.00422

[14] C. Lai and M. Gu, “Self-regulated out-of-class language learning with technology,” Computer Assisted Language Learning, vol. 24, no. 4, pp. 317–335, 2011. https://doi.org/10.1080/09588221.2011.568417

[15] Z. An, C. Wang, S. Li, Z. Gan, and H. Li, “Technology-assisted self-regulated English language learning: Associations with English language self-efficacy, English enjoyment, and learning outcomes,” Frontiers in Psychology, vol. 11, art. 558466, 2021. https://doi.org/10.3389/fpsyg.2020.558466

[16] R. Azevedo and A. F. Hadwin, “Scaffolding self-regulated learning and metacognition — Implications for the design of computer-based scaffolds,” Instructional Science, vol. 33, no. 5–6, pp. 367–379, 2005. https://doi.org/10.1007/s11251-005-1272-9

[17] L. Zheng, “The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis,” Asia Pacific Education Review, vol. 17, no. 2, pp. 187–202, 2016. https://doi.org/10.1007/s12564-016-9426-9

[18] K. VanLehn, “The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems,” Educational Psychologist, vol. 46, no. 4, pp. 197–221, 2011. https://doi.org/10.1080/00461520.2011.611369

[19] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, “Intelligent tutoring systems and learning outcomes: A meta-analysis,” Journal of Educational Psychology, vol. 106, no. 4, pp. 901–918, 2014. https://doi.org/10.1037/a0037123

[20] E. M. Golonka, A. R. Bowles, V. M. Frank, D. L. Richardson, and S. Freynik, “Technologies for foreign language learning: A review of technology types and their effectiveness,” Computer Assisted Language Learning, vol. 27, no. 1, pp. 70–105, 2014. https://doi.org/10.1080/09588221.2012.700315

[21] L. K. Fryer, M. Ainley, A. Thompson, A. Gibson, and Z. Sherlock, “Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners,” Computers in Human Behavior, vol. 75, pp. 461–468, 2017. https://doi.org/10.1016/j.chb.2017.05.045

[22] L. K. Fryer, K. Nakao, and A. Thompson, “Chatbot learning partners: Connecting learning experiences, interest and competence,” Computers in Human Behavior, vol. 93, pp. 279–289, 2019. https://doi.org/10.1016/j.chb.2018.12.023

[23] S. Loewen, D. Crowther, D. R. Isbell, K. M. Kim, J. Maloney, Z. F. Miller, and H. Rawal, “Mobile-assisted language learning: A Duolingo case study,” ReCALL, vol. 31, no. 3, pp. 293–311, 2019. https://doi.org/10.1017/S0958344019000065

© 2026 TRTEE. Open Access under CC BY 4.0

[24] G. Dizon, “Evaluating intelligent personal assistants for L2 listening and speaking development,” Language Learning & Technology, vol. 24, no. 1, pp. 16–26, 2020. https://hdl.handle.net/10125/44705

[25] T.-Y. Tai and H. H.-J. Chen, “The impact of Google Assistant on adolescent EFL learners’ willingness to communicate,” Interactive Learning Environments, vol. 31, no. 3, pp. 1485–1502, 2023. https://doi.org/10.1080/10494820.2020.1841801

[26] M. Stevenson and A. Phakiti, “The effects of computer-generated feedback on the quality of writing,” Assessing Writing, vol. 19, pp. 51–65, 2014. https://doi.org/10.1016/j.asw.2013.11.007

[27] J. Ranalli, “Automated written corrective feedback: How well can students make use of it?” Computer Assisted Language Learning, vol. 31, no. 7, pp. 653–674, 2018. https://doi.org/10.1080/09588221.2018.1428994

[28] Z. Zhang and K. Hyland, “Student engagement with teacher and automated feedback on L2 writing,” Assessing Writing, vol. 36, pp. 90–102, 2018. https://doi.org/10.1016/j.asw.2018.02.004

[29] S. Link, M. Mehrzad, and M. Rahimi, “Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement,” Computer Assisted Language Learning, vol. 35, no. 4, pp. 605–634, 2022. https://doi.org/10.1080/09588221.2020.1743323

[30] J. S. Barrot, “Using ChatGPT for second language writing: Pitfalls and potentials,” Assessing Writing, vol. 57, art. 100745, 2023. https://doi.org/10.1016/j.asw.2023.100745

[31] Y. Su, Y. Lin, and C. Lai, “Collaborating with ChatGPT in argumentative writing classrooms,” Assessing Writing, vol. 57, art. 100752, 2023. https://doi.org/10.1016/j.asw.2023.100752

[32] D. Yan, “Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation,” Education and Information Technologies, vol. 28, pp. 13943–13967, 2023. https://doi.org/10.1007/s10639-023-11742-4

[33] S. Mahapatra, “Impact of ChatGPT on ESL students’ academic writing skills: A mixed methods intervention study,” Smart Learning Environments, vol. 11, art. 9, 2024. https://doi.org/10.1186/s40561-024-00295-9

[34] F. Ouyang and P. Jiao, “Artificial intelligence in education: The three paradigms,” Computers and Education: Artificial Intelligence, vol. 2, art. 100020, 2021. https://doi.org/10.1016/j.caeai.2021.100020

[35] I. Molenaar, “Towards hybrid human-AI learning technologies,” European Journal of Education, vol. 57, no. 4, pp. 632–645, 2022. https://doi.org/10.1111/ejed.12527

[36] D. Long and B. Magerko, “What is AI literacy? Competencies and design considerations,” in Proc. 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–16. https://doi.org/10.1145/3313831.3376727

© 2026 TRTEE. Open Access under CC BY 4.0

[37] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. S. Qiao, “Conceptualizing AI literacy: An exploratory review,” Computers and Education: Artificial Intelligence, vol. 2, art. 100041, 2021. https://doi.org/10.1016/j.caeai.2021.100041

[38] D. Carless and D. Boud, “The development of student feedback literacy: Enabling uptake of feedback,” Assessment & Evaluation in Higher Education, vol. 43, no. 8, pp. 1315–1325, 2018. https://doi.org/10.1080/02602938.2018.1463354

[39] I. Celik, “Towards Intelligent-TPACK: An empirical study on teachers’ professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education,” Computers in Human Behavior, vol. 138, art. 107468, 2023. https://doi.org/10.1016/j.chb.2022.107468

[40] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” MIS Quarterly, vol. 13, no. 3, pp. 319–340, 1989. https://doi.org/10.2307/249008

[41] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, “User acceptance of information technology: Toward a unified view,” MIS Quarterly, vol. 27, no. 3, pp. 425–478, 2003. https://doi.org/10.2307/30036540

[42] A. Strzelecki, “Students’ acceptance of ChatGPT in higher education: An extended unified theory of acceptance and use of technology,” Innovative Higher Education, vol. 49, no. 2, pp. 223–245, 2024. https://doi.org/10.1007/s10755-023-09686-1

[43] G. Liu and C. Ma, “Measuring EFL learners’ use of ChatGPT in informal digital learning of English based on the technology acceptance model,” Innovation in Language Learning and Teaching, vol. 18, no. 2, pp. 125–138, 2024. https://doi.org/10.1080/17501229.2023.2240316

[44] C. K. Y. Chan and W. Hu, “Students’ voices on generative AI: Perceptions, benefits, and challenges in higher education,” International Journal of Educational Technology in Higher Education, vol. 20, art. 43, 2023. https://doi.org/10.1186/s41239-023-00411-8

[45] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, “Systematic review of research on artificial intelligence applications in higher education — where are the educators?” International Journal of Educational Technology in Higher Education, vol. 16, art. 39, 2019. https://doi.org/10.1186/s41239-019-0171-0

[46] M. Bond et al., “A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour,” International Journal of Educational Technology in Higher Education, vol. 21, art. 4, 2024. https://doi.org/10.1186/s41239-023-00436-z

[47] T. K. F. Chiu, Q. Xia, X. Zhou, C. S. Chai, and M. Cheng, “Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education,” Computers and Education: Artificial Intelligence, vol. 4, art. 100118, 2023. https://doi.org/10.1016/j.caeai.2022.100118

© 2026 TRTEE. Open Access under CC BY 4.0

[48] X. Huang, D. Zou, G. Cheng, X. Chen, and H. Xie, “Trends, research issues and applications of artificial intelligence in language education,” Educational Technology & Society, vol. 26, no. 1, pp. 112–131, 2023. https://doi.org/10.30191/ETS.202301_26(1).0009

[49] C. Zhai and S. Wibowo, “A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students’ interactional competence in the university,” Computers and Education: Artificial Intelligence, vol. 4, art. 100134, 2023. https://doi.org/10.1016/j.caeai.2023.100134

[50] G.-J. Hwang and C.-Y. Chang, “A review of opportunities and challenges of chatbots in education,” Interactive Learning Environments, vol. 31, no. 7, pp. 4099–4112, 2023. https://doi.org/10.1080/10494820.2021.1952615

[51] R. Wu and Z. Yu, “Do AI chatbots improve students’ learning outcomes? Evidence from a meta-analysis,” British Journal of Educational Technology, vol. 55, no. 1, pp. 10–33, 2024. https://doi.org/10.1111/bjet.13334

[52] L. Zheng, J. Niu, L. Zhong, and J. F. Gyasi, “The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis,” Interactive Learning Environments, vol. 31, no. 9, pp. 5650–5664, 2023. https://doi.org/10.1080/10494820.2021.2015693

[53] M. J. Page et al., “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” BMJ, vol. 372, art. n71, 2021. https://doi.org/10.1136/bmj.n71

[54] M. L. McHugh, “Interrater reliability: The kappa statistic,” Biochemia Medica, vol. 22, no. 3, pp. 276–282, 2012. https://doi.org/10.11613/BM.2012.031

[55] D. H. Schunk and M. K. DiBenedetto, “Motivation and social cognitive theory,” Contemporary Educational Psychology, vol. 60, art. 101832, 2020. https://doi.org/10.1016/j.cedpsych.2019.101832

[56] X. Wang, Q. Liu, H. Pang, S. C. Tan, J. Lei, M. P. Wallace, and L. Li, “What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis,” Computers & Education, vol. 194, art. 104703, 2023. https://doi.org/10.1016/j.compedu.2022.104703

© 2026 TRTEE. Open Access under CC BY 4.0