When AI Scaffolding Becomes Substitution: Retained Cognitive Responsibility and Dependence Risk in ESL/EFL Higher Education

Aiman Zulaikha Mohd Fadzli[1]*, Saripah Sinas[2], Nur Nadhirah Mijan[3]

Academy of Language Studies, Universiti Teknologi MARA (UiTM) Cawangan Sarawak

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.102100220

Received: 18 October 2026; Accepted: 23 October 2026; Published: 31 October 2026

ABSTRACT

Artificial intelligence (AI) offers university ESL/EFL learners immediate feedback and practice, but assistance may also displace the work needed for independent learning. Existing research seldom distinguishes perceived autonomy, regulation during AI use, and unassisted capability. This systematic review aimed to identify how AI supports self-regulation and autonomy, when assistance risks becoming cognitive substitution, the factors associated with that transition, and its theoretical and instructional implications. A Scopus search covering January 2023 to August 2026 yielded 232 records; staged screening and full-text assessment produced 25 empirical journal articles. Methodological appraisal rated five studies as high confidence, 13 as moderate, and seven as low. Findings were synthesised across self-regulated learning phases and an autonomy–dependence continuum. AI supported feedback, rehearsal, and revision, with clearer evidence of active regulation in guided workflows requiring initial learner effort, evaluation of suggestions, and reflection. Dependence risks included premature consultation, substantial task delegation, and uncritical acceptance of output, although lasting loss of unaided capability was rarely tested. Patterns varied by learner proficiency, task demands, pedagogy, tool features, and context, but formal evidence of moderation was limited. Existing frameworks explained adoption, motivation, or mediated support without adequately connecting AI use to later independent performance. The resulting provisional Retained Cognitive Responsibility framework proposes that AI acts as a scaffold when learners retain responsibility for planning, evaluation, and learning. Instruction should make these decisions visible while future studies should test their effects through behavioural evidence and delayed unaided assessments.

Keywords: Artificial intelligence; Self-regulated learning; Learner autonomy; Scaffolding; Dependence risk

INTRODUCTION

Recent studies show that university students learning English as a second or foreign language (ESL/EFL) use artificial intelligence (AI) tools to generate ideas, revise writing, translate text, obtain explanations, and practise language beyond the classroom (MohammedAli et al., 2025; Tanharat & Phootirat, 2025; Tran et al., 2025). Automated feedback can help learners identify areas for improvement and monitor their progress (Li & Kim, 2024). These opportunities are relevant to self-regulated learning (SRL), which involves planning learning activities, monitoring performance, and reflecting on outcomes (Zimmerman, 2002). Access to assistance, however, does not by itself show that learners carry out these processes independently.

Studies of structured AI use provide evidence of its potential benefits. In a cluster-randomised trial, students who completed an initial draft before receiving AI feedback and then evaluated that feedback showed greater gains in cognitive, metacognitive, and resource-management strategies and English achievement than students in a control condition (Chen & Alibakhshi, 2026). A project-based study also reported greater gains in achievement and learner autonomy when AI use was combined with collaboration and instructor guidance (Abdelhalim &

Almaneea, 2026). In a qualitative study, learners used automated feedback more critically within a course that included teacher support, portfolios, and reflection (Li & Kim, 2024). These findings suggest that the way AI support is organised matters for its contribution to learning.

Other studies raise concerns about what learners may delegate to AI. In a survey of 376 Jordanian EFL students, 61% reported consulting AI before attempting a task themselves and 73% reported that they did not critically evaluate AI-generated output. More frequent use was associated with lower self-reported self-directed learning, error-correction confidence, and vocabulary retention; the study's correlational design does not establish that AI use caused these outcomes (Rababah et al., 2026). Vietnamese students likewise reported relying on AI feedback instead of developing their own editing skills and finding it difficult to write without assistance (Nhan et al., 2025). Analysis of AI-assisted academic writing found improvements in some textual features alongside concerns about homogenised voice and reduced learner control (Tomchakovskyi et al., 2026). Conversely, a Saudi study found no statistically significant association between students' perceived usefulness of Grammarly and their reported overreliance on it (Abdullah et al., 2026). Together, these findings indicate that perceived usefulness or frequency of use alone cannot establish whether AI is supporting learning or replacing work the learner needs to practise.

This distinction also presents a measurement problem. Surveys can show how students perceive their autonomy and dependence, but they cannot on their own demonstrate how students evaluate AI output or perform without it (Al-Atrash et al., 2026a; Nhan et al., 2025). Chat histories, portfolios, and comparisons of student texts provide closer evidence of learners' actions during AI-assisted work (Hai et al., 2026; Li & Kim, 2024; Tomchakovskyi et al., 2026). A review of the evidence therefore needs to consider when learners seek assistance, which parts of a task they delegate, how they assess the response, and what evidence exists of independent performance.

The present systematic review examines empirical studies of AI use in ESL/EFL higher education published between January 2023 and August 2026. It aims to clarify when AI supports self-regulated learning and learner autonomy, and when its use is associated with cognitive substitution or dependence. Five research questions guided the review:

RQ1: How do artificial intelligence tools scaffold self-regulated learning and learner autonomy in higher education ESL/EFL settings?

RQ2: Under what conditions does machine scaffolding transition into cognitive offloading, overreliance, or diminished agency?

RQ3: Which learner, task, pedagogical, technological, and contextual variables moderate this transition?

RQ4: What theoretical frameworks dominate the empirical base, and what conceptual limitations persist?

RQ5: What instructional design principles and research directions can preserve autonomy while mitigating reliance risks?

This synthesis offers three contributions. First, it establishes clear analytical distinctions among perceived autonomy, enacted self-regulation with AI assistance, and their performance without said assistance. Second, it proposes retained cognitive responsibility as a conceptual lens for examining when AI functions as a scaffold and when it may substitute for learners' own work. Third, it brings these findings together in a provisional process model and identifies implications for instructional design and future research. Together, these contributions focus attention on which learning activities remain under the learner's control during AI-assisted work.

LITERATURE REVIEW

Self-Regulated Learning and Learner Autonomy

Self-regulated learning (SRL) involves a recurring cycle of forethought, performance, and self-reflection (Zimmerman, 2002). Learners set goals and plan their approach before a task, use and monitor strategies while working, and evaluate the outcome to guide later learning. AI tools can provide explanations, examples, and feedback during each phase (Chen & Alibakhshi, 2026; Li & Kim, 2024). Receiving such support, however, does not itself show that learners have planned, monitored, or evaluated their own work. The relevant question is how they use the support to make and revise decisions.

In this review, learner autonomy refers to learners' capacity to make informed decisions about their goals, learning strategies, and evaluation of their work. It can be exercised with support from teachers, peers, and tools (Li & Kim, 2024). Self-determination theory adds a motivational perspective by identifying autonomy, competence, and relatedness as basic psychological needs (Deci & Ryan, 2000). In a study of students using machine translation and generative AI for academic English, perceptions of these needs varied with instructional and disciplinary context (Tran et al., 2025). A learner's feeling of autonomy is therefore relevant, but it should be distinguished from evidence that the learner regulates a task or can complete it without assistance.

Scaffolding, Cognitive Offloading, and Dependence

Scaffolding, as used in this review, refers to assistance that supports task completion while keeping learners engaged in the decisions needed for independent work. Studies of guided AI use illustrate this distinction. A draft-first, AI-feedback, and human-evaluation sequence was associated with improved learning strategies and achievement (Chen & Alibakhshi, 2026). Learners in a course combining automated feedback with teacher guidance and reflection also described more critical use of tool suggestions (Li & Kim, 2024). In both cases, the instructional activity required learners to do more than receive an automated answer.

Cognitive offloading occurs when a learner delegates part of a task's mental work to an external resource. Cognitive load theory draws attention to limited processing capacity and the importance of using it for learning and schema development (Sweller, 1988). Applied to AI-assisted language learning, this suggests that the value of offloading depends on the task objective. Automated correction of surface errors may allow attention to remain on argument and organisation; generating a complete passage may instead reduce opportunities to practise language formulation and evaluative judgement. The quality of the resulting text alone cannot show which learning processes the student carried out (Tomchakovskyi et al., 2026).

Dependence risk is more specific than the use of external help. It concerns patterns such as consulting AI before making an attempt, accepting output without evaluation, or reporting difficulty working without assistance. These patterns appear in survey and interview findings from Jordan and Vietnam (Nhan et al., 2025; Rababah et al., 2026). Because much of this evidence is self-reported or correlational, it indicates a risk rather than proving that AI use caused a lasting loss of ability. Assessing that claim requires evidence of learners' performance when support is reduced or removed.

METHODOLOGY

Research Design

This study used a systematic review design to synthesise empirical evidence on AI use, self-regulated learning, learner autonomy, and dependence risk in higher education ESL/EFL settings. The published empirical study report was the unit of analysis. Before data extraction, a review protocol documented the research questions,

search strategy, eligibility criteria, extraction fields, and quality appraisal approach. PRISMA 2020 guided the reporting of study identification, screening, selection, and synthesis (Page et al., 2021).

The included studies differed in research design, AI tools, learning tasks, and outcome measures. Their findings were therefore synthesised thematically rather than combined in a meta-analysis. The theoretical foundations outlined in Section 1.1 informed interpretation of the evidence. The retained cognitive responsibility framework was developed from patterns identified across the included studies and is presented after the synthesis in Section 4.3.

Study Corpus and Learner Populations

The review included 25 empirical journal articles published between 2024 and 2026. The articles reported research with postsecondary learners of English as a second or foreign language, or English for Academic Purposes. Learners were participants in the original studies; the published study reports were the texts examined in this review. The characteristics of the included studies, including their contexts, learner populations, sample sizes, research designs, and intervention durations, are presented in Appendix B, Table B.1.

To be eligible, a study had to examine learners' direct use of an AI tool during an English language learning task and report empirical evidence relevant to self-regulation, autonomy, or potential dependence. Studies involving learners at several educational levels were eligible only when the postsecondary findings could be identified separately. Reviews, theoretical papers, theses, preprints, conference abstracts, instructor-only studies, and studies without relevant learner evidence were excluded. The complete inclusion and exclusion criteria are presented in Table 1.

Table 1. Inclusion and Exclusion Criteria for Study Selection

Criterion	Inclusion criteria	Exclusion criteria
Publication period	Published from January 2023 to August 2026	Published before January 2023 or after the search date
Language	Full article published in English	Article not available in English
Publication type	Peer-reviewed empirical journal article	Review, theoretical or conceptual paper, editorial, book chapter, dissertation, thesis, preprint, conference abstract, or other non-peer-reviewed publication
Research design	Quantitative, qualitative, or mixed-methods empirical study	Non-empirical discussion, system-development paper, algorithm-development study, or study without learner data
Educational setting	University, college, polytechnic, or other postsecondary institution	Primary or secondary school, workplace-only, or non-educational setting
Participants	ESL, EFL, L2 English, or EAP learners in higher education; mixed samples were eligible only when postsecondary findings could be extracted separately	Instructor-only studies; non-language learners; or mixed-level samples where higher-education findings could not be separated
AI involvement	Direct learner interaction with an AI-supported tool during English-language learning activities	Non-AI technology, automated detection systems, or studies without direct learner–AI interaction
Relevant outcomes	Evidence concerning self-regulation, metacognition, learner autonomy, scaffolding, strategic help-seeking, cognitive	Studies reporting only technology acceptance, general attitudes, academic integrity, or language performance without evidence relevant to regulation, autonomy, or dependence

	offloading, uncritical acceptance, overreliance, or unassisted performance	
Focal-triage requirement	Title or abstract explicitly indicated dependence risk or related regulatory dynamics relevant to the scaffolding–dependence boundary	Broadly eligible studies whose titles and abstracts did not signal relevant dependence-risk or regulatory evidence
Full-text availability	Complete, legitimate full text retrievable from a publisher, institutional repository, journal repository, ERIC, or verified author-hosted source	Full text unavailable after retrieval attempts or obtainable only through unverified or unauthorised sources

Review Instruments

Two instruments were used to organise and evaluate the evidence. The first was a reviewer-developed data-extraction matrix. It recorded each study's publication details, geographical and institutional context, learner characteristics, ESL/EFL or EAP setting, research design, intervention duration, AI tool and version, language-learning task, theoretical framework, and principal findings. It also captured evidence relating to self-regulation, scaffolding, autonomy, cognitive offloading, dependence behaviours, moderating factors, and methodological limitations. Evidence was classified according to whether it originated from behavioural or performance data, learner self-reports, instructor observations, or the primary authors' interpretations. The characteristics extracted using this matrix are summarised in Appendix B, Table B.1.

The second instrument was a methodological quality-appraisal checklist based on the Mixed Methods Appraisal Tool, version 2018 (MMAT; Hong et al., 2018). The MMAT criteria appropriate to each study design were supplemented with ten AI-specific reporting indicators: identification of the AI platform and version, duration of use, interaction or prompting protocols, permitted tool functions, learner AI-literacy preparation, instructor mediation, objective usage or process data, output-verification safeguards, data-ethics procedures, and differentiation between tool-assisted and unassisted outcomes. The combined checklist supported the classification of studies as providing high, moderate, or low confidence evidence. Study-level extraction and appraisal records are provided in Supplementary Material S1 (S1.2 and S1.3).

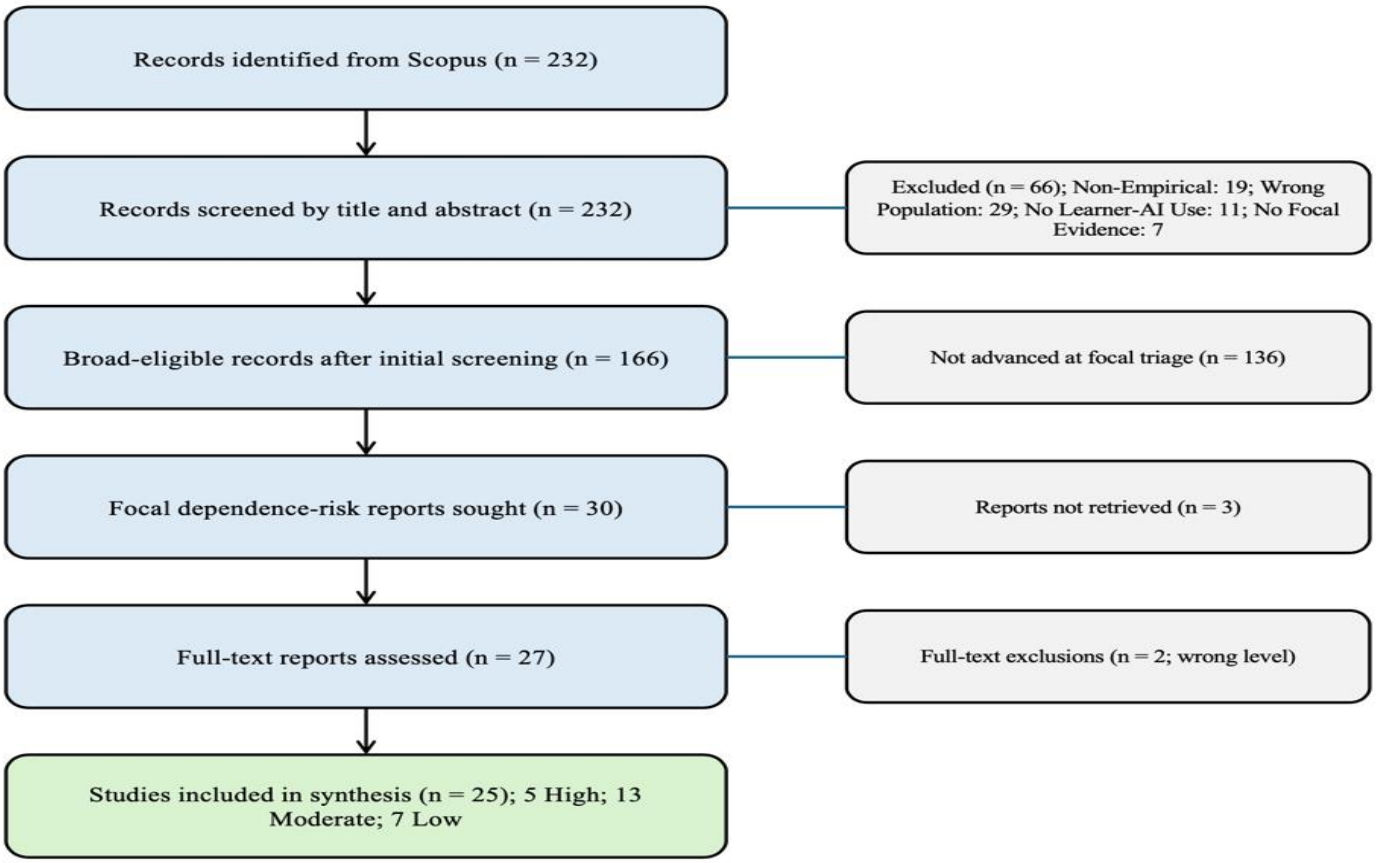
Data Collection Procedures

The literature search was conducted in Scopus and closed on 18 August 2026. It covered records published from January 2023 to the search date. The search combined four concept blocks: (a) artificial intelligence tools and automated systems; (b) ESL, EFL, L2 English, or English for Academic Purposes; (c) university and other postsecondary settings; and (d) self-regulation, metacognition, autonomy, agency, scaffolding, help-seeking, cognitive offloading, overreliance, or uncritical acceptance. No subject-area filters were applied. Results were restricted to English-language journal articles. The complete search string is provided in Appendix A.

The search produced 232 records. Title-and-abstract screening against the eligibility criteria in Table 1 excluded 66 records: 19 were non-empirical, 29 involved ineligible populations or educational levels, 11 did not examine direct learner–AI interaction, and seven did not report relevant outcomes. The remaining 166 records met the broad eligibility criteria. A prespecified second-stage focal screening retained 30 reports whose abstracts explicitly indicated dependence risk or related regulatory processes. The other 136 broadly eligible records were not advanced because their abstracts did not indicate evidence concerning the autonomy–dependence relationship. This stage was introduced because many broadly eligible records reported AI-assisted performance or attitudes without examining how responsibility for learning was shared between learner and tool.

Full-text retrieval was attempted for the 30 focal reports through publisher platforms, journal repositories, institutional databases, and verified author-hosted manuscripts. Twenty-seven full texts were obtained. Three reports could not be retrieved because no complete, authorised version was available through these sources by the final search date. Unverified copies were not used. Full-text assessment excluded two studies because their postsecondary findings could not be separated from data collected at lower educational levels. The final corpus therefore comprised 25 studies. Data from these reports were entered into the extraction matrix described in Section 2.3. All searching, screening, selection, and extraction decisions were completed and documented by one reviewer. Figure 1 presents the selection process in accordance with PRISMA 2020 reporting guidance (Page et al., 2021).

Figure 1. PRISMA-Style Flow of Study Selection



Data Analysis Procedures

The extracted evidence was analysed using a framework-based thematic synthesis. Before coding, a coding dictionary defined the principal constructs and the evidence required for each category. Findings were first distinguished by evidence source: behavioural or performance data, learner self-reports, instructor observations, or the primary authors' interpretations. Claims of autonomy based solely on learners' perceptions were not treated as evidence of demonstrated independence unless they were supported by behavioural records, performance measures, or unassisted transfer outcomes.

The synthesis proceeded through three iterative analytical passes. First, findings were mapped onto Zimmerman's (2002) self-regulated learning phases of forethought, performance, and self-reflection. Second, the allocation of cognitive work between the learner and the AI tool was examined along a continuum from resource and scaffold to collaborator, substitute, and potential dependency. Third, factors influencing these relationships were coded at

the learner, task, pedagogical, technological, and contextual levels. Themes were compared across studies and repeatedly checked against the extraction matrix and original study reports.

Methodological confidence ratings were considered when interpreting the strength of each finding. Behavioural and performance evidence from stronger study designs was given greater interpretive weight than unsupported causal claims or self-reported perceptions. Contradictory, negative, and null findings were retained rather than resolved through vote counting. The cross-study patterns generated through this process informed the Retained Cognitive Responsibility framework presented in Section 4.3. Study-level synthesis records are provided in Supplementary Material S1 (S1.3).

Quality Appraisal Procedure

Following data extraction, each study was appraised using the MMAT criteria appropriate to its research design (Hong et al., 2018). The design-specific assessment was supplemented by the ten AI-reporting indicators described in Section 2.3. The MMAT judgments and AI-specific indicators informed an overall confidence classification of high, moderate, or low. These classifications represent review-specific confidence judgments rather than numerical MMAT scores. No study was excluded solely because of its appraisal outcome; instead, the ratings were used to determine the interpretive weight given to its findings and to identify recurring weaknesses in the evidence base. Study-level appraisal ratings and their supporting rationales are provided in Supplementary Material S1 (S1.2 and S1.3).

RESULTS

Study Characteristics

The review included 25 studies published in 2024 (n = 1), 2025 (n = 11), and 2026 (n = 13), indicating that the evidence base developed rapidly during the final two years of the review period. Research was concentrated in Asian and Middle Eastern higher-education settings. Saudi Arabia contributed five studies, while Vietnam contributed four single-country studies and one cross-national study conducted with Australian and Vietnamese learners. Other research contexts included Australia, China, India, Iran, Jordan, Oman, Pakistan, Palestine, South Korea, Taiwan, Thailand, and Ukraine. This geographical distribution should be considered when assessing the wider applicability of the findings.

Participants were primarily university undergraduates, English majors, foundation-programme students, and international students enrolled in ESL, EFL, or EAP courses. Reported analytical samples ranged from seven qualitative participants to 416 survey respondents. Academic and general English writing represented the most frequently studied learning activities. Other tasks included speaking practice, translation, vocabulary development, grammar correction, project work, and general English learning. ChatGPT and related generative AI platforms were the most frequently examined tools, either independently or in combination with Grammarly, QuillBot, Copilot, Bard, Poe, Canva AI, Narakeet, machine-translation applications, automated feedback systems, and ELSA Speak.

The corpus included cross-sectional surveys, qualitative interview studies, mixed-method investigations, case studies, Q methodology, text and interaction-trace analyses, single-group interventions, quasi-experiments, and one cluster-randomised trial. Intervention periods ranged from brief two-week implementations to a full academic semester, although several cross-sectional studies examined existing patterns of use rather than a defined intervention. The characteristics of individual studies are reported in Appendix B, Table B.1.

Quality appraisal classified five studies as providing high-confidence evidence, 13 as moderate-confidence evidence, and seven as low-confidence evidence. The stronger studies incorporated comparison groups, structured interventions, behavioural or performance measures, and clearer accounts of instructional procedures. Recurring

limitations included reliance on cross-sectional self-reports, short or unspecified exposure periods, absence of comparison groups, limited reporting of implementation fidelity, unvalidated dependence measures, and causal interpretations unsupported by the study design. Table 2 summarises the principal evidence clusters and their confidence patterns.

Table 2. Summary of Included Evidence Clusters

Evidence cluster	Representative studies	Main contribution	Confidence pattern
Structured interventions	Hwang et al. (2025); Abdelhalim and Almaneea (2026); Chen and Alibakhshi (2026); Vo and Ho (2026)	Tests of draft-first protocols, guided feedback, project work, AI literacy, reflection, and human evaluation	Two high, one moderate, one low; strongest evidence favored structured integration
Feedback and writing performance	Mujeeb et al. (2026); Aljohani (2025); Lai (2025); Tomchakovskyi et al. (2026)	Performance gains in accuracy, cohesion, efficacy, text complexity, and voice	Short-term gains common; transfer and component isolation weak
Autonomy and perceptions	Al-Atrash et al. (2026a, 2026b); Al Zakwani and Binu (2025); MohammedAli et al. (2025); Tran et al. (2025)	Perceived planning, flexibility, competence, motivation, and relational context	Mostly moderate confidence; autonomy heavily operationalized as self-report
Direct dependence evidence	Rababah et al. (2026); Nhan et al. (2025); Mubarak (2026); Abdullah et al. (2026)	Evidence of pre-emptive use, non-evaluation, unassisted difficulty, and a null usefulness–reliance association	Direct empirical focus; limited by cross-sectional self-report designs
Process and mechanism evidence	Zheng and Wang (2026); Hai et al. (2026); Li and Kim (2024); Tanharat and Phootirat (2025); Tomchakovskyi et al. (2026)	Dissonance regulation, chat histories, critical feedback literacy, selective use, and revision traces	Detailed mechanism accounts; small or single-site samples

Research Question 1: AI Support for Self-Regulation and Learner Autonomy

Across the included studies, AI increased access to feedback, explanations, examples, speaking practice, and writing assistance. These functions supported task planning and revision, but improved AI-assisted performance did not necessarily demonstrate greater learner autonomy. Evidence of active regulation was clearer when learners produced initial responses, evaluated suggestions, and retained final decisions.

Structured interventions provided the strongest evidence of benefits. Chen and Alibakhshi (2026) randomised ten intact university classes to an AI-supported or control condition. A six-hour AI-literacy workshop preceded a 12-week workflow involving initial drafting, AI critique, human evaluation, and reflection. The AI-supported group achieved higher scores in cognitive, metacognitive, and resource-management strategies and English achievement (partial η^2 = .16–.22). Reflective journals documented planning, monitoring, and selective help-seeking. Abdelhalim and Almaneea (2026) likewise reported greater achievement and learner-autonomy growth among 53 Saudi undergraduates in a 12-week AI-supported collaborative project. Because both interventions combined AI with structured instruction, their effects cannot be attributed to AI access alone.

Other interventions reported performance gains with weaker causal evidence. Hwang et al. (2025) found that placement-test scores rose from 22.34 to 24.87 during a semester-long, draft-first writing programme, but the study lacked a control group and reported inconsistent analytical sample sizes. Mujeeb et al. (2026) reported larger short-term writing gains with automated than teacher feedback, although assignment was non-random and more than one third of learners accepted corrections without adequate evaluation. Vo and Ho (2026) found higher grammar post-test scores in an AI-supported class, but group assignment and statistical reporting were unclear. Lai (2025) reported stronger language performance, self-efficacy, and resilience with AI assistance, whereas traditional instruction better supported self-regulation and metacognitive growth.

Process evidence showed what active regulation involved. Hwang et al. (2025) used initial drafts, sentence-level queries, revision rationales, and final verification. Vo and Ho (2026) asked learners to compare their own passages with AI-generated alternatives. In Li and Kim's (2024) study, students described learning to compare feedback sources and reject unsuitable suggestions within a course that also provided teacher conferences, peer review, portfolios, and reflection. These studies linked autonomy to learners' evaluative actions rather than access to a tool.

Text and interaction analyses further complicated the performance picture. Among 60 EAP law students, AI-assisted texts were more complex, cohesive, and accurate, but also more homogeneous and less individually voiced (Tomchakovskyi et al., 2026). Revision traces distinguished learners who compared and selectively edited output from those who accepted templates. Similar improvements in final texts could therefore reflect different levels of learner involvement.

Survey and interview studies reported perceived gains in flexibility, confidence, idea generation, and access to help (Al-Atrash et al., 2026a, 2026b; Khan et al., 2025; Tanharat & Phootirat, 2025). These findings relied largely on self-report. Chuanpipatpong (2025), for example, found extensive tool use alongside recurring errors in students' unaided writing. Overall, AI supported immediate performance and regulation within guided workflows, but few studies tested whether these gains persisted when assistance was withdrawn.

Research Question 2: Transitioning from Scaffolding to Cognitive Substitution

Evidence of cognitive substitution appeared most frequently in studies examining when learners consulted AI, which parts of a task they delegated, how they evaluated generated output, and whether they could complete related work without assistance. Most evidence was based on cross-sectional self-reports and interviews; behavioural observations, interaction records, and delayed unassisted assessments were comparatively uncommon.

The most direct quantitative evidence came from studies that measured reliance alongside autonomy-related outcomes. Among 376 Jordanian undergraduates, 61% reported consulting generative AI before attempting tasks independently, while 73% reported accepting generated output without critical evaluation. AI-use frequency was negatively correlated with self-directed learning ($r = -.48$), error-correction confidence ($r = -.42$), and self-reported vocabulary retention ($r = -.39$) (Rababah et al., 2026). Because the study was cross-sectional and relied largely on self-reports, these associations do not establish that AI use caused the lower outcomes.

Comparable tensions appeared among 150 Vietnamese English majors. Participants reported that generative AI supported their writing and autonomy, while also reporting excessive reliance ($M = 3.95$), difficulty completing tasks without AI ($M = 3.90$), hindered self-correction ($M = 3.88$), reduced confidence ($M = 3.78$), and diminished critical engagement ($M = 3.85$) (Nhan et al., 2025). These findings demonstrate that perceived support and perceived dependence can coexist, although the study did not objectively assess unassisted performance.

Perceived usefulness was not consistently associated with lower dependence risk. In a study of 61 male Saudi foundation students, Grammarly received a high usefulness rating ($M = 4.03$), while the mean overreliance score was 3.31 and 37.7% of respondents were classified as highly reliant. Usefulness was not significantly correlated

with overreliance (ρ = .105, p = .419) (Abdullah et al., 2026). In a separate study involving 120 Saudi undergraduates, 70% strongly agreed that excessive reliance on AI translation and generative tools could reduce independent language practice, with lower-proficiency learners reporting greater reliance (Mubarak, 2026). These findings were perceptual and did not include objective tests of capability after tool removal.

Studies of feedback practices identified uncritical acceptance as another indicator of substitution. More than one third of the 400 participants in Mujeeb et al. (2026) reported accepting automated corrections without sufficient evaluation, despite short-term improvements in writing performance. Qualitative accounts similarly indicated that some learners used AI for basic sentence construction, vocabulary selection, or complete task generation rather than attempting these operations independently (Hai et al., 2026; Khan et al., 2025). Thai English majors also reported that ready-made text could reduce effort and vocabulary internalisation, even when they valued AI for brainstorming and organisation (Tanharat & Phootirat, 2025).

The limited process evidence supported these reported patterns. Analysis of chat histories in Hai et al. (2026) showed that some lower-proficiency learners submitted complete assignment prompts and performed little subsequent verification. Tomchakovskyi et al. (2026) distinguished learners who iteratively prompted, compared, and revised output from those who accepted ready-made templates. Template-based use was associated with reduced authorship and more homogeneous, less individually voiced texts. These findings provide evidence of responsibility being transferred during task completion, although neither study measured whether the behaviour produced long-term deterioration in independent ability.

Across the corpus, cognitive substitution was therefore associated with four recurring indicators: a) consultation before an independent attempt, b) delegation of core learning operations, c) acceptance of output without evaluation, and d) reported or observed difficulty performing without assistance. Interaction frequency alone did not consistently distinguish scaffolding from substitution. However, the predominantly cross-sectional evidence does not demonstrate how repeated substitution develops into enduring dependence over time. Longitudinal studies incorporating delayed, unassisted performance measures were largely absent.

Research Question 3: Moderating Factors

Five groups of factors were associated with whether AI functioned as a scaffold or a substitute: learner, task, pedagogical, technological, and institutional or contextual factors. Most included studies did not test these variables as statistical interaction effects. They were more commonly identified through subgroup comparisons, correlations, interviews, observations, or cross-study interpretation. They should therefore be understood as reported or inferred influences rather than confirmed causal moderators.

Learner Factors

Language proficiency was the most consistently reported learner-level factor. Lower-proficiency learners benefited from translation, explanations, and corrective feedback but possessed fewer linguistic resources with which to evaluate output. In Hai et al. (2026), lower-proficiency students were more likely to submit complete assignment prompts and perform limited verification. Mubarak (2026) similarly found greater reported reliance among lower-proficiency learners. In contrast, beginners in an Omani foundation programme reported greater perceived autonomy than advanced learners (Al Zakwani & Binu, 2025). This result concerned perceived autonomy and was not supported by an objective assessment of independent performance.

AI literacy, previous tool-use habits, self-efficacy, and trust also shaped reported interaction practices. The structured AI-literacy training examined by Chen and Alibakhshi (2026) was accompanied by stronger self-regulation outcomes within a guided workflow. Zheng and Wang's (2026) Q-methodological study identified three learner profiles: efficiency–capacity dissonance, instrumental–traditional dissonance, and trust–reliance dissonance. Participants managed these tensions through selective neglect, task sequencing, cognitive reframing,

contextual practice, verification, or conformity-based rationalisation. These findings indicated that scepticism about AI accuracy did not necessarily prevent continued use.

Task Factors

The function delegated to AI varied according to the task. Bounded operations such as grammar checking, vocabulary retrieval, and pronunciation feedback generally left learners responsible for broader communicative decisions. Generative operations involving idea development, argument construction, organisation, and rhetorical voice transferred a larger proportion of the target task to the system. However, few studies compared task types experimentally.

The writing studies reviewed in Sections 3.2 and 3.3 illustrate this pattern. Template-based users in Tomchakovskyi et al. (2026) transferred more compositional responsibility than learners who iteratively prompted and revised, while short-term gains from automated feedback in Mujeeb et al. (2026) coexisted with uncritical acceptance of corrections. Across these studies, final-product improvement did not consistently indicate how much cognitive work learners had performed.

Pedagogical Factors

Teacher mediation and task structure were the most frequently reported pedagogical influences. Studies associated draft-first requirements, bounded queries, written revision rationales, peer review, reflective logs, and final human evaluation with more selective AI use (Abdelhalim & Almaneea, 2026; Chen & Alibakhshi, 2026; Hwang et al., 2025; Li & Kim, 2024; Vo & Ho, 2026). These procedures required learners to make their reasoning and revision decisions visible.

By comparison, complete task outsourcing and uncritical acceptance were reported more frequently in settings where tool use was largely unstructured (Hai et al., 2026; Rababah et al., 2026). In Mujeeb et al. (2026), 59% of participants preferred a combination of teacher and AI feedback. The evidence therefore consistently associated human guidance and structured learning procedures with greater learner evaluation, although the bundled nature of most interventions prevented the independent effect of teacher mediation from being isolated.

Technological Factors

Tool design and output characteristics also shaped learner behaviour. Sentence-level feedback restricted opportunities for wholesale text replacement in Hwang et al. (2025), although learners reported unnatural expressions, misinterpretations, and confusion when feedback did not account for broader discourse. Li and Kim (2024) likewise found that automated systems supplied useful corrections alongside unsuitable recommendations, requiring learners to compare feedback sources.

Multimodal systems enabled iterative prompting, visual feedback, and re-expression across modes, but they also made complete templates readily available (Tomchakovskyi et al., 2026). Zheng and Wang (2026) further showed that learners could distrust output accuracy while continuing to use AI because of its efficiency. Evidence concerning interface transparency, uncertainty indicators, and specific model features remained limited because many studies did not report the platform version or permitted functions in sufficient detail.

Institutional and Contextual Factors

Institutional setting, disciplinary expectations, peer practices, access to human feedback, and local learning conditions influenced how learners described AI use. In a cross-national study of 416 survey respondents and 17 interviewees, autonomy, competence, and relatedness varied across Australian ESL and Vietnamese EFL settings, disciplines, instructional approaches, and peer–teacher relationships. Some EFL participants also described social pressure to use AI because their peers were doing so (Tran et al., 2025).

AI access supported continuity of learning during political disruption in Palestine (Al-Atrash et al., 2026b) and supplemented limited instructional resources in a rural Vietnamese setting (Vo & Ho, 2026). In Saudi Arabia, students and instructors valued AI for language learning and research assistance, while instructors expressed concerns about critical thinking and called for clearer institutional guidance (Alqahtani et al., 2026). These findings were largely perceptual but demonstrated that the meaning and consequences of AI assistance varied across educational environments.

Overall, the evidence indicated that the transition from scaffolding to substitution was conditional rather than uniform. Language proficiency, AI literacy, task demands, teacher mediation, tool characteristics, and institutional expectations repeatedly appeared to shape learner behaviour. Nevertheless, the limited use of longitudinal designs and formal moderation analyses means that the magnitude and causal role of these factors remain uncertain.

Research Question 4: Theoretical Frameworks and Conceptual Gaps

The included studies used overlapping frameworks, so the following categories are not mutually exclusive. Thirteen of the 25 studies used learner autonomy, self-directed learning, self-regulated learning, or metacognitive regulation as an organising construct (Abdelhalim & Almaneea, 2026; Al-Atrash et al., 2026a, 2026b; Chen & Alibakhshi, 2026; Li & Kim, 2024). Autonomy was operationalised inconsistently: some studies examined planning, monitoring, and strategic decisions, while others inferred it from confidence, satisfaction, or access to AI assistance. Few measured learners' regulatory behaviour or their ability to work unaided.

Six studies used the Technology Acceptance Model or a related approach to examine usefulness, ease of use, attitudes, and intentions (Aljohani, 2025; Al Zakwani & Binu, 2025; Hai et al., 2026; Khan et al., 2025; MohammedAli et al., 2025; Vo & Ho, 2026). These constructs helped explain AI adoption but offered limited insight into which cognitive tasks learners delegated or whether they could perform independently afterwards.

Three studies invoked sociocultural theory or the zone of proximal development to explain mediated learning (Rababah et al., 2026; Tanharat & Phootirat, 2025; Vo & Ho, 2026). They rarely assessed whether support was reduced over time or whether learners transferred skills to unaided tasks. Two studies drew on cognitive load theory (Khan et al., 2025; Lai, 2025) but did not establish whether AI freed capacity for higher-order learning or bypassed the processing learners needed to practise.

Tran et al. (2025) used self-determination theory to examine autonomy, competence, and relatedness across Australian ESL and Vietnamese EFL settings. Zheng and Wang (2026) applied cognitive dissonance theory to tensions between AI's efficiency, trustworthiness, and learners' own capability. Two studies did not specify a sufficiently clear framework for interpreting self-regulation or dependence (Chuanpipatpong, 2025; Giang et al., 2025).

Across this corpus, no single framework adequately connected the timing and scope of AI assistance with learner monitoring, knowledge internalisation, and later unaided performance. This gap informed the proposed integrative framework presented in Section 4.3.

DISCUSSION

Summary of Findings

This review identified a conditional relationship between AI assistance and learner autonomy in higher-education ESL/EFL settings. AI expanded access to feedback, explanations, rehearsal, and revision support, but its contribution to autonomous learning depended on how responsibility was divided between learner and system. Stronger comparative studies associated improved performance and self-regulation with structured workflows incorporating AI literacy, initial independent drafting, bounded consultation, reflection, and human evaluation

(Abdelhalim & Almaneea, 2026; Chen & Alibakhshi, 2026). Acknowledging that these interventions combined AI with several pedagogical components, their effects cannot be attributed to the technology alone.

Dependence-related evidence converged around premature consultation, complete task delegation, uncritical acceptance, reduced monitoring, and reported difficulty working without assistance (Hai et al., 2026; Nhan et al., 2025; Rababah et al., 2026). Process evidence also associated template acceptance with reduced authorship and stylistic individuality (Tomchakovskyi et al., 2026). However, most dependence findings came from cross-sectional self-reports. Causal direction and long-term effects therefore remain uncertain. Learners experiencing existing language difficulties may seek AI more frequently, while repeated delegation may simultaneously reduce opportunities for independent practice.

The synthesis distinguishes perceived autonomy, enacted regulation, and capability autonomy. Perceived autonomy concerns confidence and access to support; enacted regulation concerns planning, monitoring, verification, and decision-making during tool use; capability autonomy concerns performance after assistance is withdrawn. Most studies measured perceptions, fewer documented regulatory behaviour, and very few assessed delayed or unassisted performance. Tool-assisted product quality therefore provides incomplete evidence of durable language development.

Language proficiency, AI literacy, task demands, teacher mediation, tool characteristics, and institutional expectations shaped these relationships. Existing theories explained individual dimensions of adoption, motivation, scaffolding, and cognitive processing but did not fully specify how responsibility for learning operations moved between learner and system. Retained cognitive responsibility emerged from the synthesis as the mechanism connecting AI-supported activity with either developmental scaffolding or cognitive substitution. Sections 4.2 and 4.3 develop this interpretation.

Theoretical Implications

The findings refine the meaning of learner autonomy in AI-supported language learning by separating the three levels identified in Section 4.1. Access to immediate assistance may increase learners' confidence and sense of control without demonstrating that they can plan, monitor, or perform independently. The coexistence of perceived support and reported difficulty working without AI illustrates why autonomy cannot be represented by a single self-report score (Nhan et al., 2025). Future theoretical applications should therefore distinguish learners' perceptions from observable regulatory behaviour and subsequent unassisted performance.

The findings also position cognitive offloading as task-dependent. Cognitive load theory explains how external support can release processing capacity (Sweller, 1988), but its educational value depends on which operation is delegated. Assistance with mechanical correction may preserve attention for argument development, whereas automated generation of the argument itself may replace the intended learning process. Structured workflows requiring independent drafting, output evaluation, and reflective revision retained greater learner involvement (Chen & Alibakhshi, 2026).

By comparison, acceptance of complete templates produced polished texts while reducing evidence of individual authorship (Tomchakovskyi et al., 2026). Scaffolding should therefore be evaluated through the timing, scope, monitoring, internalisation, and transfer of assistance rather than through final-product quality alone.

Finally, dependence should be conceptualised as a dynamic relationship among learner characteristics, task demands, pedagogy, technology, and institutional context. It cannot be treated solely as a stable learner deficit or a direct property of an AI system. Proficiency affects the capacity to evaluate output; assessment requirements shape incentives to delegate; tool design affects the availability of complete answers; and peer or institutional expectations influence patterns of adoption (Tran et al., 2025; Zheng & Wang, 2026).

A comprehensive explanation must therefore trace how responsibility is repeatedly allocated across these interacting conditions. This requirement provides the theoretical basis for the Retained Cognitive Responsibility framework presented in Section 4.3.

Proposed Retained Cognitive Responsibility Framework

The synthesis generated the Retained Cognitive Responsibility (RCR) framework as a provisional explanation of how AI assistance can function as developmental scaffolding or cognitive substitution. RCR refers to the extent to which learners remain responsible for defining a task, producing an initial response, monitoring performance, evaluating automated suggestions, and internalising knowledge for later independent use. The framework evaluates the allocation of cognitive work during a specific learning task rather than the frequency of AI use.

As illustrated in Figure 2, the framework contains four connected components:

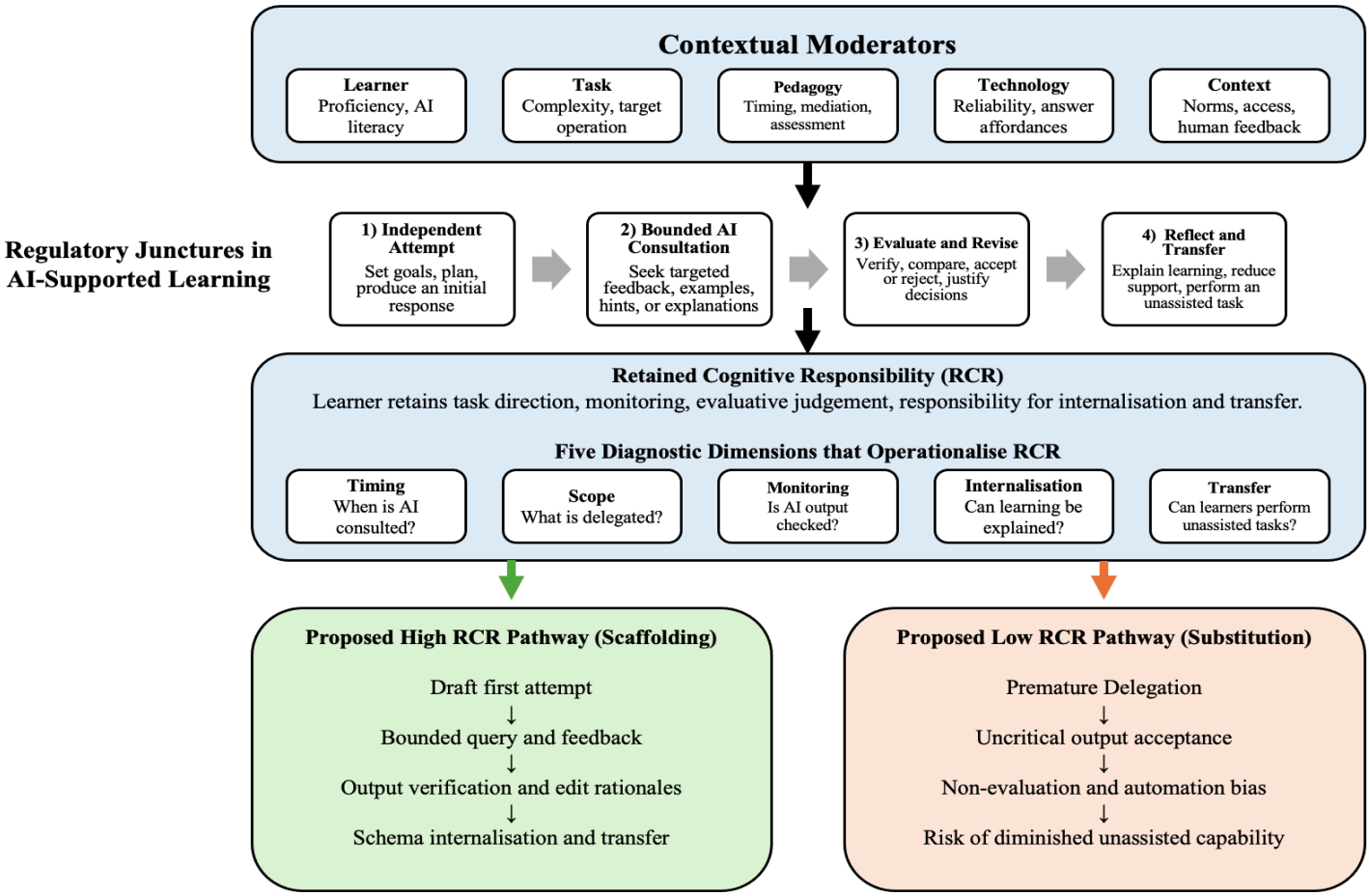
a) Contextual Moderators: Learner proficiency and AI literacy, task demands, pedagogical structure, tool characteristics, and institutional expectations shape how assistance is used.
b) Regulatory Junctures: Responsibility is allocated during task definition, initial production, AI consultation, output evaluation, revision, and subsequent transfer.
c) Cognitive allocation: Responsibility is retained when learners attempt the target operation, restrict assistance to a defined need, verify output, justify revisions, and reconstruct the underlying knowledge. Responsibility is transferred when AI performs the central learning operation before meaningful learner engagement or evaluation.
d) Developmental pathway: Repeated high-RCR interactions are expected to support internalisation and capability autonomy. Repeated low-RCR interactions are expected to increase the risk of automation bias, reduced monitoring, diminished authorship, and difficulty performing without assistance.

Five diagnostic dimensions operationalise RCR: timing, scope, monitoring, internalisation, and transfer. Timing identifies whether assistance occurs before or after an independent attempt. Scope establishes whether AI supports a bounded subproblem or completes the central learning operation. Monitoring concerns whether learners compare and verify generated output. Internalisation considers whether learners can explain and reuse the relevant principles. Transfer examines whether performance continues when assistance is reduced or removed.

The framework proposes two developmental pathways. In the high-RCR pathway, learners attempt the task, consult AI selectively, evaluate its output, justify their decisions, and demonstrate subsequent independent performance. In the low-RCR pathway, learners consult AI prematurely, delegate the core task, accept output with limited evaluation, and obtain fewer opportunities to develop the targeted capability. These pathways explain why intensive use can coexist with active regulation, while brief use can become substitutive when it replaces the central cognitive operation.

The RCR framework remains provisional because few included studies measured delayed transfer or changes in unassisted capability over time. It generates testable expectations concerning the timing and scope of assistance, the mediating role of verification, the effect of scaffold fading, and variation across learner and instructional contexts. Longitudinal and experimental research is required to determine whether repeated patterns of retained or transferred responsibility produce the proposed developmental outcomes.

Figure 2. Proposed Retained Cognitive Responsibility Framework for AI-Supported Self-Regulated Learning



Research Question 5: Instructional Design Principles

The proposed RCR framework suggests four linked practices for preserving learner responsibility during AI-supported tasks. First, learners produce a plan, draft, or oral attempt before consulting AI. This makes their initial thinking visible and gives subsequent assistance a defined purpose. Second, they request feedback or explanations for specific difficulties rather than delegating the complete task. Structured approaches incorporating initial drafting and bounded consultation were associated with improved performance or regulation in several studies (Chen & Alibakhshi, 2026; Hwang et al., 2025; Vo & Ho, 2026).

Third, evaluation becomes an explicit part of the assignment. Learners compare AI suggestions with a rubric, reference text, peer response, or teacher feedback; identify what they accept or reject; and justify consequential revisions. Li and Kim (2024) documented more selective feedback use within a course that combined automated assistance with human discussion and reflection. Fourth, learners explain what they learned and complete a related task with reduced or no AI support. This final step tests whether improved assisted performance extends to independent capability, an outcome rarely measured in the included studies.

Support should be adjusted to learner proficiency and task demands. Students who have difficulty evaluating AI output may need clearer prompts, worked examples, and teacher guidance while retaining opportunities to formulate and check their own responses (Hai et al., 2026; Mubarak, 2026). Assessment can include initial drafts, revision explanations, and unassisted follow-up work alongside the final product. These practices are evidence-informed design principles; because existing interventions combined several components, the independent effect of each practice remains to be tested.

Implications for AI System Design

The findings suggest that educational AI systems could make selective assistance and critical evaluation easier for learners. Writing tools, for example, could offer graduated hints or feedback on a learner's draft before offering complete alternative passages. Interfaces could show original and suggested text side by side and invite learners to accept, reject, or explain substantial revisions. These proposals reflect the bounded feedback and comparison practices associated with active learner involvement (Hwang et al., 2025; Li & Kim, 2024; Vo & Ho, 2026).

Systems could also distinguish mechanical corrections from suggestions that substantially alter an argument, organisation, or authorial voice. This distinction matters because more polished AI-assisted texts did not necessarily preserve individual expression or reveal how learners reached their decisions (Tomchakovskyi et al., 2026). Clear identification of generated content and accessible ways to check questionable suggestions may support evaluation, particularly where learners cannot readily judge output accuracy.

It should be noted that these are design proposals, not features proven to prevent dependence. Their value should be assessed through learners' verification and revision behaviour and, ultimately, their performance on related tasks without AI. Time spent using a tool or the quality of an assisted final product would be insufficient outcomes on their own.

Research Directions

Future research should test the RCR framework using measures that distinguish when learners seek assistance, what they delegate, how they evaluate output, and what they can subsequently do without AI. Surveys have documented perceived reliance (Nhan et al., 2025; Rababah et al., 2026), while text and interaction analyses have revealed differences in learner involvement (Tomchakovskyi et al., 2026). Combining these approaches with baseline tasks and delayed unassisted assessments would help determine whether patterns of AI use are associated with lasting changes in capability.

Experimental studies should also isolate the effects of individual instructional practices. Existing interventions combined AI-literacy training, initial drafting, teacher mediation, and reflection, making it difficult to identify which components contributed to improved outcomes (Chen & Alibakhshi, 2026). Multi-site studies could compare these components while examining differences by proficiency and access to human feedback. Longer-term research should investigate how assessment rules, peer norms, and institutional guidance influence learners' decisions across settings (Tran et al., 2025). Cross-context comparisons should also examine whether paid tool access, internet infrastructure, and institutional policies affect who benefits from AI-supported learning.

Studies should report the AI platform and version, permitted functions, duration and conditions of use, instructional support, and whether outcomes were measured with or without AI. These details are necessary to evaluate and replicate proposed mechanisms. Table 3 summarises the priority designs and outcomes. The proposed RCR pathways remain hypotheses until tested against behavioural evidence and unassisted performance over time.

Table 3. Priority Research Agenda

Research priority	Recommended design	Primary outcome	Dependence-sensitive indicator
Validate RCR and Dependence	Multi-site scale development and trace validation	Planning, monitoring, evaluation, and transfer	Premature help-seeking, non-evaluation, and unassisted skill decline
Test Scaffold Components	Factorial randomized classroom trial	Language achievement and self-regulation	Comparative impact of draft-first rules, edit rationales, and support fading
Separate	Assisted post-test with	Retention and	Assistance-performance gap and

Performance from Learning	delayed unassisted transfer	generalization capacity	recovery speed post-withdrawal
Model Proficiency Interactions	Adaptive intervention stratified by CEFR level	Skill growth and calibrated help-seeking	Frequency of whole-task delegation and verification quality by proficiency
Examine Institutional Effects	Multilevel longitudinal study	Autonomy, relatedness, and policy compliance	Assessment incentives, peer norms, and instructor workload effects

LIMITATIONS

Several considerations frame the interpretation of this review. First, Scopus served as the sole bibliographic database, providing a consistent and reproducible search base; extending future searches to other international and regional databases would further broaden coverage. Focusing on English-language journal articles similarly supported consistent appraisal, while including other languages and publication formats would capture a wider range of evidence. As the search closed on 18 August 2026, studies published after this date offer opportunities to extend the present synthesis.

Second, the focal screening stage concentrated full-text review on studies whose abstracts indicated dependence risk or related regulatory processes. This kept the synthesis aligned with the review's central question, and full-text screening of broadly eligible records in future reviews could identify further studies that address these issues beyond their abstracts. Three potentially relevant reports could not be retrieved through authorised sources and may be incorporated as they become available. Screening, extraction, appraisal, and coding were conducted by one reviewer. The prespecified protocol, decision records, extraction matrix, and coding dictionary provide a transparent audit trail, and independent verification by a second reviewer would further strengthen the reliability of these judgements.

Finally, the evidence base is still maturing. Most included studies were rated as moderate or low confidence (Section 3.1), and many used cross-sectional self-reports, brief interventions, single-group designs, or dependence measures yet to be validated, which points to clear opportunities for more rigorous designs. The review's conclusions are therefore best interpreted as a structured synthesis of an emerging evidence base that can guide future causal testing.

CONCLUSION

Across 25 empirical studies, AI expanded access to feedback and supported immediate language-task performance. Its contribution to learner autonomy depended on whether learners remained responsible for planning, producing, monitoring, and evaluating their work. Structured interventions reported stronger language and self-regulation outcomes when AI use was combined with initial drafting, reflection, and human evaluation (Abdelhalim & Almaneea, 2026; Chen & Alibakhshi, 2026). Studies also documented premature consultation, delegation of complete tasks, uncritical acceptance of output, and reported difficulty working without AI (Hai et al., 2026; Nhan et al., 2025; Rababah et al., 2026). These are indicators of dependence risk, although predominantly cross-sectional and self-reported evidence cannot establish lasting loss of independent capability.

The proposed Retained Cognitive Responsibility framework brings these findings together through five diagnostic dimensions: timing, scope, monitoring, internalisation, and transfer. It distinguishes feeling autonomous from regulating learning during AI use and performing independently afterwards. For instruction, the central priority is to make learners' initial efforts, evaluative decisions, and subsequent independent performance visible.

These conclusions reflect a single-database, single-reviewer synthesis of a developing evidence base, as discussed in Section 4.7. The RCR framework is therefore presented as a set of evidence-informed propositions, and the

design principles offered here are intended to reduce, rather than eliminate, the risk of dependence. Longitudinal and experimental studies with delayed unassisted assessments are well placed to test the proposed pathways and extend the framework.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Academy of Language Studies, Universiti Teknologi MARA (UiTM), Sarawak Branch, for its continued support and the academic environment that enabled this research. They also thank everyone who contributed valuable insights and constructive feedback during the preparation of the manuscript.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Aiman Zulaikha Mohd Fadzli contributed to conceptualisation, methodology, investigation, data curation, formal analysis, project administration, and writing the original draft. Saripah Sinas contributed to validation, resources, supervision, and reviewing and editing the manuscript. Nur Nadhirah Mijan contributed to methodology, formal analysis, visualisation, and reviewing and editing the manuscript. All three authors contributed substantially to the work, approved the final manuscript, and agree to be accountable for all aspects of the work.

AI-Use Declaration

During the preparation of this manuscript, the authors used OpenAI's ChatGPT and Codex to assist with refining the review questions and search strategy, organising extracted information, supporting thematic synthesis, improving language and structural coherence, and formatting the manuscript according to the journal template. The authors critically reviewed, verified, and revised all AI-assisted outputs, including the interpretation of findings, in-text citations, references, tables, and figures. The AI tools were not treated as authors and did not independently determine the study's conclusions. The authors take full responsibility for the accuracy, originality, integrity, and final content of the manuscript.

Supplementary Material

Supplementary Material S1 is a single workbook containing the search string and record counts at each selection stage (S1.1), the characteristics, quality ratings, and DOIs of the 25 included studies (S1.2), and the study-level evidence matrix covering scaffolding mechanisms, autonomy and dependence evidence, RCR coding, outcomes, and appraisal rationales (S1.3). The underlying publisher PDFs are not redistributed.

REFERENCES

1. Abdelhalim, S. M., & Almaneea, M. O. (2026). GenAI-supported collaborative project-based language learning: Effects on EFL undergraduates' achievement and autonomy. Asian-Pacific Journal of Second and Foreign Language Education, 11, Article 50. https://doi.org/10.1186/s40862-026-00430-8

2. Abdullah, A. M. A., Nasim, S. M., & Alqahtani, M. A. (2026). The role of Grammarly in Saudi EFL writing: Students' perceptions of its effectiveness, efficiency, and overreliance. FWU Journal of Social Sciences, 20(1), 133–148. https://doi.org/10.51709/19951272/Spring2026/10
3. Al-Atrash, A., Hamdan, M., & Dweikat, K. (2026a). Exploring EFL students' perspectives on ChatGPT-assisted autonomous practices in enhancing self-directed learning at Palestinian universities. Cogent Education, 13(1), Article 2689737. https://doi.org/10.1080/2331186X.2026.2689737
4. Al-Atrash, A., Hamdan, M., & Dweikat, K. (2026b). Qualitative insights into EFL students' use of ChatGPT-assisted autonomous learning for enhancing self-directed learning in Palestinian higher education. Discover Global Society, 4, Article 16. https://doi.org/10.1007/s44282-026-00350-5
5. Aljohani, N. J. (2025). Enhancing writing accuracy and empowering students: The transformative influence of ChatGPT's informative feedback on students' writing. Forum for Linguistic Studies, 7(7), 1109–1128. https://doi.org/10.30564/fls.v7i7.9849
6. Alqahtani, N., Alshammari, K. K., Jamshed, M., & Karim, M. R. (2026). From pedagogy to research: Exploring ChatGPT's transformative role in Saudi EFL educational innovation and scientific research. World Journal of English Language, 16(4), 234–245. https://doi.org/10.5430/WJEL.V16N4P234
7. Al Zakwani, M., & Binu, P. M. (2025). Exploring student insights on ChatGPT as a resource for learning English and its influence on learner independence. Forum for Linguistic Studies, 7(2), 813–824. https://doi.org/10.30564/fls.v7i2.7955
8. Chen, J., & Alibakhshi, G. (2026). AI-powered applications' effects on English language learners' cognitive, metacognitive, and resource management strategies, and language achievement. Journal of Computer Assisted Learning, 42(1), Article e70171. https://doi.org/10.1002/jcal.70171
9. Chuanpipatpong, A. (2025). Thai EFL university students' writing in the digital age: Error analysis revisited. PASAA, 70, 308–339. https://doi.org/10.58837/chula.pasa.70.10
10. Deci, E. L., & Ryan, R. M. (2000). The "what" and "wy" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
11. Giang, N. T., Loc, N., & Tuyen, T. (2025). Using AI-powered tools for enhancing speaking skills of EFL learners. North American Journal of Psychology, 27(4), 1015–1016.
12. Hai, N. T. M., Ho, P. V. P., & Huong, N. T. (2026). Generative AI in online English learning: A qualitative case study of non-English-major students in a Vietnamese blended-learning environment. CALL-EJ, 27(2), 134–153. https://doi.org/10.54855/callej.262727
13. Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). Mixed Methods Appraisal Tool (MMAT), version 2018: User guide. McGill University.
14. Hwang, E., Im, H., & Son, H. (2025). The effects of AI feedback learning activities in English writing on university students' English achievement and writing motivation. English Teaching, 80(5), 95–117. https://doi.org/10.15858/engtea.80.5.202512.95
15. Khan, R., Qamar, M. T., Ansari, M. S., & Yasmeen, J. (2025). Enhancing or impairing? Exploring Indian EFL learners' academic writing narratives with ChatGPT. Cogent Education, 12(1), Article 2514329. https://doi.org/10.1080/2331186X.2025.2514329
16. Lai, Z. C.-C. (2025). The impact of AI-assisted blended learning on writing efficacy and resilience. International Journal of Computer-Assisted Language Learning and Teaching, 15(1), 1–21. https://doi.org/10.4018/IJCALLT.377174
17. Li, L., & Kim, M. (2024). It is like a friend to me: Critical usage of automated feedback systems by self-regulating English learners in higher education. Australasian Journal of Educational Technology, 40(1), 1–18. https://doi.org/10.14742/ajet.8821
18. MohammedAli, E. A., Veettil, R. P., Al-Zaabi, M. A., Sharafudeen, M., & Qureshi, T. (2025). Exploring Omani EFL learners' perceptions of AI tools. Digital Technologies Research and Applications, 4(3), 206–216. https://doi.org/10.54963/dtra.v4i3.1922

19. Mubarak, R. (2026). AI-powered translation apps and their impact on English language learning in Saudi Arabia. Arab World English Journal, 2026(Special Issue 3), 346–360. https://doi.org/10.24093/awej/AI3.22
20. Mujeeb, N., Lodhi, M. A., Ali, A., Mashwani, W. K., Mostafa, S. M., & Ahmadini, A. A. H. (2026). Beyond red ink: AI feedback and ESL academic writing development in higher education. Language Testing in Asia, 16, Article 24. https://doi.org/10.1186/s40468-026-00442-8
21. Nhan, L. K., Hoa, N. T. M., & Quang, L. V. N. (2025). Leveraging AI for writing instruction in EFL classrooms: Opportunities and challenges. Educational Process: International Journal, 15, Article e2025158. https://doi.org/10.22521/edupij.2025.15.158
22. 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., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
23. Rababah, L. M., Al-Namarneh, S. I. M. S., Rababah, H. A., & Ismail, I. A. (2026). From self-learners to system-dependents: The negative effects of AI on EFL autonomy in Jordan. Digital Technologies Research and Applications, 5(2), 238–250. https://doi.org/10.54963/dtra.v5i2.2081
24. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
25. Tanharat, N., & Phootirat, P. (2025). Integrating generative AI in EFL academic writing: Thai English-major students' purposes, perceptions, and experiences with ChatGPT. rEFLections, 32(3), 1891–1916. https://doi.org/10.61508/refl.v32i3.285982
26. Tomchakovskyi, O., Tomchakovska, Y., Strochenko, L., Shvelidze, L., & Yeryomenko, S. (2026). Multimodal AI tools as mediators of linguistic development in EL academic writing. Arab World English Journal, 2026(Special Issue 3), 390–403. https://doi.org/1024093/awej/A3.25
27. Tran, H., Crosthwaite, P., & Pham, Q. H. P. (2025). Studnts' self-determination in using machine translation and generative AI tools for English for acadeicpurposes. Journal of English for Academic Purposes, 78, Article 101578. https://doi.org/10.1016/j.jea.2025.101578
28. Vo, T. L., & Ho, T. M. P. (2026). A quasi-experimenta study on enhancing EFL learners' grammar precision in writing skills under supported-ChatGPT approach at a university in a remote region of Vietnam. Arab World English Journal, 2026(Special Issue 3), 361–375. https://doi.org/10.24093/awej/A3.23
29. Zheng, C., & Wang, Y. (2026). Generative AI in EFL contexts: Q-methodological analysis of cognitive dissonance patterns and self-regulation strategies among Chinese tertiary learners. Acta Psychologica, 264, Article 106531. https://doi.org/10.1016/j.actpsy.2026.106531Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

APPENDIX A. SCOPUS SEARCH STRING

TITLE-ABS-KEY(("artificial intelligence" OR "generative artificial intelligence" OR "generative AI" OR GenAI OR "large language model*" OR LLM* OR ChatGPT OR chatbot* OR "conversational agent*" OR "intelligent tutor*" OR "intelligent tutoring system*" OR "adaptive learning" OR "automated feedback" OR "AI feedback" OR "automated writing evaluation" OR "automated writing feedback" OR "machine translation" OR "AI writing assistant*" OR "AI-assisted" OR "AI-supported" OR "AI-mediated" OR "AI-enhanced") AND (EFL OR ESL OR "English as a foreign language" OR "English as a second language" OR "English language learner*" OR "English language learning" OR "English language teaching" OR "second language English" OR "L2 English" OR "English for academic purposes" OR EAP) AND (universit* OR college* OR polytechnic* OR postsecondary OR "post-secondary" OR "higher education" OR undergraduate* OR postgraduate* OR "tertiary education") AND ("self-regulated learning" OR "self regulated learning" OR self-regulat* OR metacognit* OR "learner autonom*" OR "learning autonom*" OR "learner agency" OR "student agency" OR scaffold* OR "self-directed learning" OR "self directed learning" OR "help seeking" OR "help-seeking" OR overreliance OR "over-reliance" OR dependenc* OR "cognitive offloading" OR "automation bias" OR "critical thinking" OR "uncritical acceptance")) AND PUBYEAR > 2022 AND PUBYEAR < 2027 AND (LIMIT-TO(LANGUAGE, "English"))

APPENDIX B. INCLUDED-STUDY EVIDENCE MATRIX

This appendix summarises the 25 empirical studies retained in the final synthesis. Quality ratings are review-specific confidence judgements (Section 2.5.1). Study codes correspond to records in Supplementary Material S1, which also reports DOIs.

Table B.1. Characteristics and Methodological Profile of Included Studies

Study	Country/Context and Sample	Design and Duration	AI Tool(s) and Language Focus	Theoretical Framework	Quality Rating
Hwang et al. (2025) SCOPUS001	South Korea. 244 first-year university EFL students; intermediate placement	One-group pre/post mixed-method study; four tasks across one semester	ChatGPT-3.5; Writing	Self-directed learning orientation; motivation theory	Moderate
Lai (2025) SCOPUS008	Taiwan. Lower-intermediate university EFL learners; sample size not reported in the accessed version	Quasi-experimental AI-assisted blended versus traditional instruction; duration not reported in the accessed version	AI-assisted writing tools; Writing	Self-efficacy; resilience; cognitive load; blended learning	Low
Al-Atrash et al. (2026b) SCOPUS014	Palestine. 30 undergraduates at two universities	Qualitative semi-structured interviews	ChatGPT; General English/self-directed learning	Learner autonomy and self-directed learning; retrospective constructivist and experiential interpretation	Moderate
Al-Atrash et al. (2026a) SCOPUS031	Palestine. 348 adult undergraduates in EFL programs at two universities	Cross-sectional quantitative survey collected September 2024 to March 2025	ChatGPT; General English/self-directed learning	Learner autonomy and self-directed learning; relational and socially mediated autonomy	Moderate
Zheng and Wang	China. 25 Chinese	Q methodology	Generative AI tools;	Cognitive	High

(2026) SCOPUS062	tertiary EFL learners	with 45-statement sorts plus post-sort interviews	General EFL/self-regulation	Dissonance Theory	
Chuanpipatpong (2025) SCOPUS066	Thailand. 70 university EFL students	In-class writing/error analysis plus 14-item questionnaire	Google Translate; websites; generative AI; Writing	Error analysis; no explicit SRL framework	Low
Giang et al. (2025) SCOPUS071	Vietnam. 151 English majors at Van Hien University	Mixed-method pre/post quasi-experiment questionnaire and interviews; two weeks	ChatGPT; ELSA Speak; Speaking	Communicative competence orientation; explicit framework not clear	Low
Mujeeb et al. (2026) SCOPUS073	Pakistan. 400 computing undergraduates; AI n=200 teacher feedback n=200	Quasi-experimental mixed-method study; eight weeks with five-week performance phase	Grammarly; ChatGPT; Academic writing	Formative feedback and metacognitive regulation	Moderate
Khan et al. (2025) SCOPUS076	India. 43 adult EFL students in English-medium academic courses with at least one year of ChatGPT use	Qualitative semi-structured interviews and thematic analysis	ChatGPT; Academic writing	Cognitive Load Theory; Technology Acceptance Model; AI-IDLE; Ideal L2 Self; SRL discussed	Moderate
Aljohani (2025) SCOPUS079	Saudi Arabia. 50 undergraduates; 49 apparently retained in paired analyses; five raters	One-group mixed-method intervention; 16 weeks	ChatGPT; Writing	Technology Acceptance Model and Theory of Planned Behavior	Low
Abdelhalim and Almaneea (2026) SCOPUS094	Saudi Arabia. 53 female first-year computer-science undergraduates; experimental n=28 control n=25	Explanatory sequential mixed-method quasi-experiment; 12 weeks	ChatGPT; Copilot; Canva AI; Narakeet; Grammarly; QuillBot; translation tools; Integrated language/project work	Project-based language learning; socially mediated autonomy and SRL	High
MohammedAli et al. (2025) SCOPUS100	Oman. 198 students at two higher-education institutions	Cross-sectional questionnaire with open responses and correlations	ChatGPT and other AI tools; General English	Technology Acceptance Model	Moderate
Chen and Alibakhshi (2026) SCOPUS103	Iran. 310 undergraduates in ten intact classes; AI n=139 control n=171	Cluster-randomized mixed-method intervention; six-hour literacy workshop plus 12 weeks	ChatGPT; Poe; Bard; General English/learning strategies	Self-regulated learning and language-learning strategy framework	High
Vo and Ho (2026) SCOPUS113	Vietnam. 90 first-year non-English majors in a rural university; experimental n=45 control n=45	Quasi-experimental study; eight weeks	ChatGPT; Writing/grammar	Vygotskian sociocultural theory and ZPD; Krashen input; process writing; TAM	Low
Hai et al. (2026) SCOPUS137	Vietnam. 30 third-year non-English majors; eight teachers; 23 chat	Qualitative case study in a blended university course	Generative AI tools; Online English learning/writing	Technology Acceptance Model	Moderate

	histories				
Rababah et al. (2026) SCOPUS140	Jordan. 376 undergraduates across multiple universities; 22 interviews	Sequential explanatory mixed-method cross-sectional study	Generative AI tools; General EFL/autonomy	Holec learner autonomy; sociocultural theory; cognitive offloading; SDT in discussion	Low
Al Zakwani and Binu (2025) SCOPUS141	Oman. 190 foundation-program students; 18 interviews	Mixed-method cross-sectional survey and interviews	ChatGPT; General English	Learner independence; technology acceptance orientation	Moderate
Nhan et al. (2025) SCOPUS143	Vietnam. 150 fourth-year English majors; follow-up interviews	Mixed-method survey and interviews	Generative AI writing tools; Writing	Constructivist and learner-autonomy orientation	Moderate
Tanharat and Phootirat (2025) SCOPUS145	Thailand. 50 senior English majors; five scenario interviews	Mixed-method cross-sectional survey and interviews	ChatGPT; Academic writing	Sociocultural theory and ZPD; process writing; AI-mediated scaffolding	Moderate
Li and Kim (2024) SCOPUS165	Australia. Seven focus-group/e-portfolio participants and 32 survey respondents from a cohort of 93 international learners	Qualitative-dominant mixed-method case study of a personalized autonomous English course	ELSA; Grammarly; automated feedback systems; Writing/automated feedback	Personalized autonomy model; self-regulated learning; self-efficacy and SDT implications	High
Mubarak (2026) SCOPUS194	Saudi Arabia. 120 undergraduates at three universities and eight instructors	Mixed-method survey and interviews	AI translation apps; ChatGPT; Translation-supported English learning	Autonomy and digital-learning orientation	Moderate
Tran et al. (2025) SCOPUS198	Australia and Vietnam. 416 survey respondents and 17 interviewees across ESL and EFL universities	Mixed-method cross-sectional study with CFA mixed-effects regression and interviews	Machine translation and generative AI tools; English for academic purposes	Self-Determination Theory	High
Abdullah et al. (2026) SCOPUS208	Saudi Arabia. 61 male preparatory-year EFL students	Cross-sectional descriptive-correlational questionnaire	Grammarly; Writing/grammar feedback	Automated writing evaluation and learner autonomy orientation	Moderate
Alqahtani et al. (2026) SCOPUS210	Saudi Arabia. 283 undergraduates and 19 instructors	Mixed-method learner survey and instructor open responses	ChatGPT; General EFL/academic research	Educational innovation and learner autonomy	Low
Tomchakovskyi et al. (2026) SCOPUS224	Ukraine (Odesa). 60 second- and third-year EAP law students; 120 paired texts	Within-participant multimodal writing study with prompt and revision traces questionnaire and reflections	Multimodal generative AI tools; Academic writing	Sociocognitive writing; noticing; distributed cognition; scaffolding	Moderate

Note. Studies included in synthesis (n = 25): High 5 | Moderate 13 | Low 7. Study-level evidence on scaffolding mechanisms, autonomy and dependence, RCR coding, and appraisal rationales is provided in Supplementary Material S1.