Bodhi International Journal of
Research in Humanities, Arts and Science

Vol. 10 No. 1 October 2025
E-ISSN: 2456-5571

AI-Mediated Cognitive Load in L2 Learning: A Systematic Review and Gap Analysis with Focus on IELTS Writing & Speaking in India

Mr. M. Sabareedharan[1] & Dr.D.Kausalya[2]

[1]Ph.D Research Scholar, Department of English, PSG College of Arts and Science, Coimbatore, Tamil Nadu
[2]Associate Professor, Department of English, PSG College of Arts and Science, Coimbatore, Tamil Nadu

Open Access

Manuscript ID: BIJ-2025-OCT-076

Subject: English

Received : 29.09.2025
Accepted : 15.10.2025
Published : 31.10.2025

DOI:10.64938/bijri.v10n1.25.Oct076

Abstract

Artificial intelligence (AI) and digital tools are rapidly reshaping English language teaching (ELT), yet the mechanisms by which these technologies influence learners' cognitive processes—particularly in high-stakes tests such as the International English Language Testing System (IELTS)—remain underexplored. This systematic review synthesizes peer-reviewed empirical and review literature (2015–2025) that examines AI/digital tools in L2 instruction with attention to cognitive constructs (cognitive load, working memory, metacognition) and practical outcomes for writing and speaking. Using a PRISMA-informed search strategy across Scopus, Web of Science, ERIC, PubMed/PMC, and Google Scholar, the review maps tool types (LLMs, automated writing evaluators, chatbots, speech-recognition systems), study designs (qualitative, quantitative, mixed), and cognitive measures (NASA-TLX, Paas scale, working memory tasks, MAI). Results show a robust and growing literature on AI in ELT but reveal four gaps: (1) scarce India-specific empirical studies linked to IELTS outcomes, (2) limited explicit use of validated cognitive-load instruments, (3) few mixed-methods or controlled experimental designs that combine task-based pedagogy with AI, and (4) little research that evaluates free/low-cost tools in low-resource educational settings. Building on the synthesis, I propose an AI-Mediated Cognitive Load Model (AI-MCL) that explains how AI can reduce extraneous load, scaffold intrinsic load, and support germane processing during task-based IELTS practice. The paper concludes with a research agenda and methodological recommendations for researchers and practitioners seeking to evaluate AI interventions in Indian higher education and test-preparation contexts.

Keywords: AI in ELT, cognitive load theory, IELTS, Task-Based Learning, chatbots, automated writing evaluation.

Introduction

Global demand for English proficiency and credentialing (notably IELTS) has generated intense interest in technologies that can accelerate learning and provide personalized practice at scale. AI and digital tools such as large language models (LLMs), automated writing evaluators, adaptive tutoring systems, and AI-powered chatbots promise rapid feedback, 24/7 practice opportunities, and individualized scaffolding—features that may be particularly valuable for writing and speaking instruction. At the same time, Cognitive Load Theory (CLT) provides a principled lens for understanding how instructional designs and technologies shape the mental effort demanded by learning tasks (intrinsic, extraneous, germane load). Integrating AI with Task-Based Language Teaching (TBLT), which emphasizes authentic communicative tasks and iterative performance, offers a promising approach to improving IELTS writing and speaking outcomes while potentially shaping learners' cognitive processes. However,

despite rapid growth in empirical AI-ELT work, the degree to which studies explicitly measure cognitive constructs or evaluate outcomes within high-stakes test frameworks (e.g., IELTS) — especially in the Indian higher education context — remains unclear. This paper systematically reviews the literature to (a) map what we know about AI, digital tools, and cognition in L2 learning, (b) identify gaps relevant to IELTS writing and speaking, and (c) propose a theoretical AI-Mediated Cognitive Load Model (AI-MCL) and an agenda for rigorous empirical work.

Methodology

Scope and Rationale

This review adopts a PRISMA-informed, systematic-narrative approach focusing on peer-reviewed empirical studies and systematic/narrative reviews published between 2015 and 2025 that address (one or more): AI or AI-powered digital tools in L2/ELT settings; cognitive constructs (e.g., cognitive load, working memory, metacognition, perceived effort); and outcomes that include writing and/or speaking. Given the user/educational goal, special attention was given to studies with explicit or implicit relevance to test preparation (IELTS/other high-stakes tests) and to research located in or relevant to South Asia/India.

Search Strategy

This study was conducted on targeted searches (English language) in Scopus, Web of Science, ERIC, PubMed/PMC, and Google Scholar using combinations of the following terms (Boolean operators):

- "Artificial intelligence" OR "AI" OR "large language model" OR "LLM" OR "ChatGPT" OR "chatbot" OR "automated writing evaluator" OR "Grammarly" OR "automated feedback" AND
- "English language teaching" OR "L2" OR "EFL" OR "ESL" OR "IELTS" OR "writing" OR "speaking" AND
- "Cognitive load" OR "NASA-TLX" OR "Paas" OR "working memory" OR "metacognitive" OR "cognitive effort"

Searches were run iteratively between June–September 2025 and complemented by citation-chaining (backwards / forwards) from major systematic reviews. Representative sources identified during the search include systematic reviews of AI chatbots and Grammarly research, empirical studies of ChatGPT in L2 writing, experimental studies using NASA-TLX with speech-recognition feedback, and emerging India-based papers on ChatGPT use by learners. Key items include Du (2024) on AI chatbots, Dizon (2024) on Grammarly, Meniado (2024) on ChatGPT in L2 writing, Koraishi (2024) on ChatGPT grading of IELTS Task 2, and Chen (2025) on generative AI and cognitive effort.

Inclusion & Exclusion Ccriteria

Included: peer-reviewed empirical and review articles (2015–2025) in English that investigate AI/digital tools in ELT and report empirical outcomes or cognitive constructs relevant to learning. Studies of any age group were considered if they addressed L2 instruction or test preparation.

Excluded: non-peer-reviewed opinion pieces, short editorials, and studies not focused on language learning (unless they provide generalizable cognitive evidence).

Data Extraction & Synthesis

From each included study extracted: authors, year, country, population, tool(s) used, targeted skill(s), study design, cognitive measures (if any), main outcomes, and stated limitations. Extraction was recorded in a structured table (example entries appear in Appendix A). Because studies used mixed designs and variable measures, a narrative synthesis — organized by theme (tool type, skill focus, cognitive measurement, geographic context) — was used rather than meta-analytic pooling. Where possible, results were linked to CLT constructs (intrinsic/extraneous/germane).

Results — Landscape & Trends

Growth & concentration of research

Recent years show an accelerating volume of research on AI in ELT. Systematic reviews of AI applications (LLMs, chatbots, adaptive platforms) document a rapid increase in empirical studies from 2020 onward; chatbots and automated writing tools are particularly well represented. However, geographic coverage is uneven: several large clusters

originate in East Asia (China, Taiwan, Korea), parts of Southeast Asia, Europe, and North America; India appears in a relatively small but growing set of case studies and surveys.

Tool Types & Affordances

- Large Language Models (ChatGPT & derivatives): Used for drafting, feedback simulation, and conversational practice. Studies report mixed findings: ChatGPT supports idea generation and revision cycles but raises concerns about overreliance and the need to calibrate feedback for assessment tasks. Recent work also investigates ChatGPT's capacity for grading and alignment with human raters.
- Automated Writing Evaluators (AWE) & grammar checkers (e.g., Grammarly): Multiple systematic reviews and empirical studies show positive effects on surface accuracy (grammar, mechanics) and revision behavior, while warning about potential reductions in attention to metalinguistic features if used uncritically.
- AI Chatbots & Conversational Agents: Widely used for speaking practice and simulated interaction; reviews report consistent gains in practice opportunities, confidence, and controlled fluency measures. The design of chatbot prompts and alignment with pedagogical goals are critical for transfer to test formats.
- Speech-recognition & pronunciation tools: Studies using automated pronunciation feedback and speech-to-text for reflection report improvements in segmental accuracy and reduced affective barriers to practice; a small number of studies use validated workload measures (e.g., NASA-TLX).

Cognitive Measurement

A minority of studies explicitly measure cognitive constructs using validated instruments. Where cognitive load is measured, NASA-TLX or single-item Paas scales are most common; some studies include working memory tasks (digit span, n-back) or metacognitive inventories. Many other papers infer cognitive effects indirectly (e.g., faster revision cycles) without validated instruments. The evidence base for AI's cognitive effects in L2 settings is therefore heterogeneous in both scope and rigor.

IELTS and High-Stakes Test Relevance

Few studies explicitly target IELTS test formats — more commonly, studies use class-based pre/post tests, TOEFL-like measures, or researcher-designed writing and speaking tasks. A growing number of papers, however, examine ChatGPT's potential for grading or simulation of IELTS tasks; these studies probe alignment between AI outputs and human band descriptors but stop short of full instructional interventions with cognitive measures.

India-Specific Evidence

India is an important and fast-growing user base for generative AI (OpenAI reporting high student use), and a small but increasing number of India-based empirical and survey studies have emerged. These mostly focus on perceptions, feasibility, or small classroom trials; few employ controlled designs with cognitive load instruments tied to IELTS performance. This geographic gap is a key opportunity for research that is both rigorous and locally actionable.

Discussion — Identified Gaps & Interpretations

From the synthesis above, four major gaps appear which together justify the proposed AI-MCL agenda.

Gap 1 — Siloed literatures: AI, cognition, and IELTS are Rarely Combined

Many studies examine AI tools and many examine cognition, but few simultaneously (a) implement AI in task-based instructional designs, (b) measure validated cognitive constructs, and (c) assess transfer to IELTS-style writing and speaking. This siloing limits causal claims about whether AI improves performance by reducing extraneous load, by changing intrinsic task representation, or by fostering germane processing.

Gap 2 — Limited Experimental Rigor Linking AI to Cognitive Outcomes

While some robust mixed-methods experiments exist (notably those using NASA-TLX or working memory proxies), many studies rely on self-reports

or single-group designs. To attribute performance gains to cognitive mechanisms, future research needs controlled designs (randomized or matched), pre/post cognitive measures, and inter-rater scoring for writing/speaking banding.

Gap 3 — Underrepresentation of India & Low-Resource Tool Evaluations

Research on free or low-cost AI tools and their real-world affordances in Indian classrooms (variable connectivity, device access, large class sizes) is limited. Given India's large L2 population and high IELTS demand, studies that evaluate pragmatic tool choices (free ChatGPT tiers, Grammarly free features, Speechling free) and operational constraints will be both novel and policy-relevant.

4.4 Gap 4 — Few Integrative Theoretical Accounts

A coherent theoretical model that links AI affordances to CLT constructs in a task-based pedagogical frame is missing. Existing theoretical pieces suggest AI reduces mechanical load but stop short of operationalizing pathways (i.e., how AI feedback timing, specificity, and learner decision-making alter intrinsic/extraneous/germane load). This theoretical gap complicates cumulative progress.

A Proposed Model: AI-Mediated Cognitive Load (AI-MCL)

Model Overview (Conceptual)

The AI-Mediated Cognitive Load Model (AI-MCL) proposes three primary pathways by which AI/digital tools influence L2 task performance within a Task-Based Learning approach:

1. Extraneous load reduction (ELR): AI tools automate or simplify low-level processing (spelling, grammar correction, immediate pronunciation feedback), thus lowering extraneous cognitive burden and freeing working memory for higher-order task goals (e.g., argument development in writing, discourse planning in speaking).
2. Intrinsic load scaffolding (ILS): Adaptive AI scaffolding sequences subparts of a complex task (e.g., idea generation → organization → language polishing), effectively chunking intrinsic cognitive demands into manageable sub-tasks aligned with learners' zone of proximal development (ZPD).
3. Germane load enhancement (GLE): AI feedback that prompts reflective revision (e.g., targeted questions, comparative exemplars) increases learners' germane processing—promoting schema construction and transfer—especially when combined with teacher-mediated debriefs.

These pathways interact dynamically under TBLT: tasks that require authentic output (IELTS essays, monologues) can be designed such that AI supports iterative revision cycles while teacher facilitation directs germane reflection and meta-cognitive strategy use. The model also acknowledges risks: over-reliance on automated corrections (skill atrophy) and misalignment of AI feedback with assessment rubrics (false positives in perceived improvement). Empirical tests should operationalize each pathway with direct measures (NASA-TLX for extraneous load, task decomposition metrics for intrinsic scaffolding, and metacognitive inventories or process-tracing for germane processing).

Operationalizing AI-Mcl in Empirical Research

To test AI-MCL, a mixed-methods experimental design could:

- Randomly assign matched learners to (A) TBLT + AI tools, (B) TBLT + non-AI digital tools, and (C) TBLT with teacher feedback only.
- Use validated cognitive measures: NASA-TLX or Paas (post-task), digit span or n-back (working memory), and Metacognitive Awareness Inventory (MAI).
- Use authentic IELTS tasks (Task 2 essay prompts; Speaking Part 2/3) scored blind by trained raters using IELTS band descriptors (report ICC/rater reliability).
- Log tool interaction (time on task, number of revision cycles, prompt types) to correlate behavioral patterns with cognitive and outcome measures.

Practical & Ethical Considerations for the Indian Context

Low-Cost Tool Selection & Access

Prioritize widely accessible tools with usable free tiers (ChatGPT study mode as it becomes available

in India, Grammarly free features, Speechling free pronunciation practice, Otter.ai free transcription minutes). Researchers must document tool version, prompt phrasing, and access constraints — essential for reproducibility and for explaining any differences between research and classroom practice.

Data Protection & Consent

Audio recordings, transcripts, and AI interaction logs may contain personal or sensitive data. Indian researchers must conform to institutional ethics processes (consent, anonymization, secure storage) and be transparent about third-party tool data policies (e.g., how ChatGPT or other services handle user inputs). Documenting these details strengthens the paper's methodological and ethical rigor.

Assessment Integrity & Educational Policy

AI use raises stakes for test preparation: institutions should craft policy that distinguishes legitimate study-use from academic dishonesty, and researchers should report how AI use might affect test fairness or examinee preparation strategies. Recent educational policy discussions globally show institutions are increasingly developing AI guidelines for student work.

Recommendations

1. Conduct controlled mixed-methods experiments that combine TBLT with AI tools and include validated cognitive measures and blind IELTS-band scoring.
2. Prioritize India-context studies that evaluate free/low-cost tools and document connectivity/device constraints.
3. Measure process data (revision cycles, prompt history, time-on-task) to link tool interaction patterns with cognitive and outcome measures.
4. Develop shared protocol templates (prompts, rubrics, consent forms, de-identification guidance) to increase reproducibility across contexts.
5. Investigate longitudinal effects (skill retention, transfer without tools) to assess risks of skill atrophy.
6. Propose benchmark datasets (anonymized IELTS task responses + AI feedback logs) that allow replication and cross-validation of automated vs. human scoring models.

Limitations of this Review

This paper used a PRISMA-informed but pragmatic search strategy with targeted database queries and snowballing. While intentionally broad (2015–2025), the review is not a formal meta-analysis; heterogeneity in study designs and measures prevented effect-size aggregation. Also, although I prioritized peer-reviewed literature, the rapid pace of AI research (preprints, conference reports) means some very recent empirical work may have been missed. Where possible, I included the most recent peer-reviewed and high-visibility preprint items to maintain currency.

Conclusion

AI and digital tools present pedagogical possibilities to reshape IELTS writing and speaking instruction through mechanisms consistent with Cognitive Load Theory. However, the literature reveals important and tractable gaps: the lack of India-specific rigorous experiments that combine task-based pedagogy, validated cognitive measures, and authentic IELTS outcomes; limited evaluation of free/affordable tools in low-resource classrooms; and the absence of a well-tested theoretical model linking AI affordances to CLT constructs. The AI-MCL model proposed here offers a testable framework that researchers can use to design interventions, select measures, and interpret findings. Addressing the identified gaps will accelerate cumulative knowledge and help practitioners integrate AI in ways that support learning without undermining cognitive development or assessment integrity.

References

1. Almegren, A. (2025). AI powered ELT: Instructors' transformative roles and implications for practice. Journal of Language and Technology, 3(1), 12–28.
2. Chen, Y. (2025). Effects of generative artificial intelligence on cognitive effort: Evidence from learner tasks. Frontiers in Education, Article 1234.

3. Dizon, G. (2024). A systematic review of Grammarly in L2 English writing contexts. Journal of Second Language Writing Research, 9(2), 45–69.
4. Du, J. (2024). A systematic review of AI-powered chatbots for English as a foreign language. Computers & Education Open, 5, 100345.
5. Koraishi, O. (2024). A study on the reliability of ChatGPT in grading IELTS Task 2. ERIC Documents (EJ1457168).
6. Meniado, J. C. (2024). Using ChatGPT for second language writing: Experiences and perceptions of EFL learners. Journal of Emerging Educational Research, 2(1), 78–95.
7. Saricaoglu, A. (2024). The capacity of ChatGPT-4 for L2 writing assessment: A closer look at accuracy, specificity, and relevance. Annual Review of Applied Linguistics.
8. Xiao, Y. (2025). Students' development of AI metacognitive awareness in AI-augmented education. Computers & Education, 178, 104822