



Optimizing L2 English Reading Speed and Comprehension Through Artificial Intelligence: A Psycholinguistic and Pedagogical Framework

Saaïd Ali Omar 

English Department, Faculty of Arts, University of Zawia, Zawia, Libya

Corresponding Author's Email: s.belgasm@zu.edu.ly

Received: 03/8/2026 Revised: 20/8/2026 Accepted: 28/8/2026 Publication: 22/9/2026

Abstract

Artificial intelligence (AI) is proposed as a means of improving second-language (L2) English reading. The pedagogical problem, however, is not maximizing words per minute but helping readers allocate processing resources so that lexical access becomes more efficient without sacrificing syntactic integration, inference, monitoring, and transfer. This paper synthesizes research on L2 reading, evidence on fluency and comprehension interventions, studies of intelligent tutoring and generative AI, and scholarship on assessment and educational ethics. The synthesis supports a layered model in which reading performance emerges from interactions among lexical and orthographic quality, decoding, sentence and discourse integration, working memory, prior knowledge, strategic regulation, motivation, and text-purpose conditions. Speed and comprehension are related but dissociable outcomes; therefore, AI should optimize a conditional speed–accuracy profile rather than a reading-rate target. The paper proposes a teacher-governed AI architecture combining learner modeling, text selection, repeated and extensive reading, vocabulary recurrence, strategy coaching, uncertainty-aware feedback, and delayed transfer assessment. These mechanisms are recommendations derived from converging evidence, not claims that AI automatically produces durable L2 gains. Evaluation should include rate, accuracy, literal and inferential comprehension, vocabulary breadth and depth, immediate and delayed transfer, learner agency, fairness, and teacher workload. The paper concludes that AI is most defensible as a scaffold whose assistance fades as learners develop self-regulated reading, while teachers retain responsibility for interpretation, assessment, and equitable participation.

Keywords: L2 English Reading; Reading Fluency; Artificial Intelligence; Intelligent Tutoring Systems; Adaptive Learning.

1. Introduction

Reading in a second language requires a reader to coordinate processes that are partly automatic and partly deliberate. The reader must identify words, assemble syntactic relations, build a coherent discourse representation, use prior knowledge, monitor uncertainty, and adjust speed to purpose and text difficulty. L2 readers frequently devote more time and visual attention to words than L1 readers do. In natural reading, bilingual readers have been found to show longer sentence reading times, more fixations, shorter saccades, and less skipping in L2 than in L1 (Cop et al., 2015). Yet these differences do not imply a single deficit or a single instructional remedy. Exposure, vocabulary, spelling, decoding, working memory, topic knowledge, motivation, reasoning, literacy history, and the relation between the first and second writing systems all contribute to individual profiles.

The attraction of AI lies in its potential to make practice more differentiated. An AI system can select texts, reintroduce previously unfamiliar words, provide a prompt at a moment of difficulty, generate a passage-grounded question, or record how a learner responds to feedback. Intelligent tutoring research supports the general value of immediate and customizable

feedback, but its strongest evidence comes from K–12 and first-language contexts, and its effects are larger relative to conventional instruction than relative to human tutoring (Xu et al., 2019). Direct evidence for adult L2 English reading remains smaller. Computerized dynamic assessment and adaptive vocabulary recurrence have produced promising results, while recent generative-AI studies report short-term improvements in comprehension, inferencing, or self-regulated learning in particular settings (Wang, 2016; Yang & Qian, 2020; Çelik et al., 2024; Shafiee Rad, 2025). Such findings warrant design hypotheses, not automatic claims of efficacy. This paper asks how AI could be used responsibly to optimize L2 English reading speed and comprehension together. “Optimize” is defined here as improving the fit between processing efficiency and comprehension demands across purposes and texts. It does not mean maximizing rate regardless of meaning. The paper first establishes a conceptual background, then explains psycholinguistic mechanisms, presents a pedagogical framework and system architecture, and proposes an assessment model. It closes by examining teacher agency, equity, limitations, and a research agenda. Throughout, established findings are distinguished from proposed mechanisms. The central argument is that an AI reading system should be designed as a teacher-governed, passage-grounded, gradually fading scaffold that develops flexible fluency and comprehension rather than superficial speed.

2. Conceptual Background: Fluency Is Conditional, and Comprehension Is Distinct

Reading fluency is commonly conceptualized through automaticity, rate, accuracy, and, in oral reading, appropriate phrasing (Grabe & Yamashita, 2022). Passage fluency is often associated with comprehension in both L1 and L2 contexts, but the association should not be interpreted as identity. A reader can move quickly by guessing, skimming for a narrow purpose, or overlooking a relation that matters for a global inference. Conversely, a reader may slow down productively when integrating unfamiliar syntax or constructing a situation model. Reviews of speed reading conclude that dramatic increases in reading rate are unlikely to preserve thorough comprehension across ordinary learning tasks (Rayner et al., 2016). Rate must therefore be interpreted in relation to text difficulty, purpose, assistance, and accuracy.

The distinction is also visible in large-scale L2 evidence. In the Multilingual Eye-Movements Corpus, spelling, vocabulary, and decoding were associated with faster reading, more skipping, fewer and shorter fixations, and less rereading, whereas vocabulary and spelling were especially related to comprehension accuracy (Kuperman et al., 2023). This pattern suggests partly distinct fluency and comprehension pathways. Vocabulary breadth and depth are both relevant, but they need not contribute in the same way: breadth can support access to many words, while depth supports the semantic, morphological, and collocational knowledge needed to build meaning. Similarly, oral passage fluency may explain meaningful variance in silent comprehension but is not an adequate proxy when vocabulary and syntactic processing remain incomplete.

The proposed framework therefore treats reading as a conditional profile with at least four observable dimensions: rate, accuracy, comprehension, and regulation. Rate is words read per unit time under specified conditions. Accuracy includes word recognition and response accuracy. Comprehension includes literal understanding, local and global inference, evaluation, and integration across the passage. Regulation includes purposeful rereading, monitoring, help seeking, and adjustment of speed (Jeon, 2012). A system that raises rate while lowering inference or increasing answer guessing has not improved the target construct. A system that produces slower but more accurate reading of increasingly demanding texts may be supporting development even when a simple words-per-minute score falls.

This framework also reconciles apparently conflicting pedagogical evidence. Extensive reading produces small-to-medium gains across comprehension, vocabulary, decoding and fluency, motivation, writing, and broader proficiency, particularly when texts are accessible and participation is supported by light accountability (Sangers et al., 2025). A separate meta-

analysis found a positive between-group effect for extensive reading, with variation by age, context, medium, and curricular placement (Jeon & Day, 2016). Repeated reading can improve automatic word recognition and rate, and some studies report comprehension gains, but transfer to unfamiliar passages is less certain (Gorsuch & Taguchi, 2008). Explicit strategy instruction also shows a large mean effect, especially when learners are taught how to activate knowledge, attend to structure, predict, question, and connect information (Yapp et al., 2021). These approaches are not mutually exclusive. Extensive reading supplies volume and motivation; repeated reading supports automaticity; strategy instruction supports purposeful comprehension; vocabulary work increases the linguistic resources on which all three depend.

3. Psycholinguistic Mechanisms Relevant to AI Design

Understanding the psycholinguistic mechanisms underlying L2 reading is essential for designing AI-supported tools that align technological interventions with the cognitive processes involved in lexical access, syntactic processing, inference, and comprehension monitoring.

3.1 Lexical access, orthographic quality, and automaticity

L2 reading is often slowed by less efficient activation of word forms and meanings. Cop et al. (2015) found approximately 20% longer sentence reading times in L2 novel reading, along with more fixations and fewer skips. Whitford and Titone (2012) found larger L2 than L1 word-frequency effects in early and late eye-movement measures, with especially large L2 effects among bilinguals with less current L2 exposure. The results are consistent with experience-sensitive lexical activation: repeated encounters can make access faster and more stable, but the path is not simply a reduction in average response time.

Segalowitz et al. (1998) argue that extended L2 learning can restructure word-recognition processing toward automaticity. Variability in response time can therefore be informative alongside the mean. A proposed AI mechanism follows from this evidence: the system should identify words that are inconsistently recognized, not only words that are always slow, and should recycle them across meaningful contexts with spacing and retrieval (Zhang & Zhang, 2022). This mechanism is a design hypothesis, not a demonstrated causal effect of AI. It should be tested against ordinary teacher-led recurrence and matched practice time.

The same logic supports modality-aligned vocabulary instruction. Incidental learning through reading is real but partial; a meta-analysis estimated immediate learning of roughly 9–18% of target words and delayed learning of roughly 6–17%, with limited evidence on depth and long-term retention (Webb et al., 2023). Intentional instruction has positive effects, but the magnitude varies by context, proficiency, age, and method (Yousefi & Biria, 2018). An AI system should therefore distinguish a learner's recognition, recall, morphological, collocational, and contextual knowledge rather than treating a correct gloss selection as full vocabulary acquisition. Vocabulary knowledge is strongly associated with L2 reading comprehension, but correlational evidence does not by itself establish that an automated vocabulary intervention will cause comprehension growth.

3.2 Sentence integration, working memory, and prior knowledge

Reading requires the temporary maintenance and integration of information as a sentence unfolds. Working memory is central to this process, but competing accounts emphasize capacity, interference, activated long-term memory, and attention; different tasks can therefore produce different results (Cunnings, 2022). In L2 reading, working-memory capacity predicted comprehension especially when learners had sufficient topic knowledge. When topic knowledge was insufficient, L2 linguistic knowledge was more decisive (Joh & Plakans, 2017). This finding cautions against a generic “working-memory training” solution. A reader may appear slow because the text exceeds vocabulary and syntactic resources, not because a domain-general capacity is deficient.

A proposed AI mechanism is support matching. Before and during reading, the system should estimate whether a difficulty is primarily lexical, syntactic, discourse-level, or knowledge-based. It might offer a short word explanation for a lexical obstacle, a relation diagram for a dense sentence, or a concise background note for an unfamiliar topic. The support should be passage-grounded and optional. It should not silently replace the target text with a simplified version, because simplification can remove precisely the lexical, syntactic, cultural, or ambiguous features that learners need to learn to process (Çelik et al., 2024). The authentic passage should remain available, and any adaptation should be visible and reversible.

3.3 Eye movements and strategic regulation

Eye movements provide valuable but indirect evidence. Fixations, saccades, skips, regressions, and rereading are multiply determined by word frequency, predictability, layout, density, comprehension difficulty, and task. L2 readers with more efficient fixation patterns often comprehend better, yet individual studies show compensatory profiles. Tywoniw (2023) found that motivation and reasoning could offset less efficient reading behavior, and that different mixtures of proficiency, motivation, reasoning, and eye-movement measures predicted different comprehension tests. Thus, eye tracking may diagnose a specific task but should not be treated as a universal biomarker of comprehension.

For an ordinary AI platform without specialized eye-tracking hardware, interaction logs can still support limited hypotheses. Rereading, help requests, dictionary use, and time spent before a response can reveal where a learner may need support. They cannot reveal silent comprehension directly. A proposed mechanism is to use such data to select a low-cost prompt “Which sentence supports your inference?” or “What changed between these two events?” rather than to label the learner’s internal state. Learner explanations and confidence checks should remain part of the evidence.

3.4 Strategic comprehension and compensatory profiles

Comprehension involves constructing a coherent mental model. Explicit strategy instruction has shown substantial promise, including background-knowledge activation, attention to structure and signal words, semantic mapping, prediction, questioning, and links to prior knowledge (Yapp et al., 2021). However, knowing a strategy is not equivalent to applying it independently in authentic study. In academic reading, task prompts elicited reflection, while individualized feedback redirected attention to deeper processes and reminded learners of successful behavior (Mežek et al., 2022). A proposed AI mechanism is therefore faded metacognitive prompting: the system first makes decisions visible, then asks learners to choose a strategy, and finally reduces prompts while requiring justification.

This mechanism must accommodate compensatory profiles. One learner may compensate for slow lexical processing through strong reasoning; another may read quickly but lack sufficient vocabulary for global inference. A fixed algorithm that interprets all long fixations as weakness or all rapid reading as success risks misclassification. The learner model should represent uncertainty and use multiple evidence sources, including comprehension responses, explanations, confidence, vocabulary checks, and performance on novel passages.

4. A Pedagogical Framework for AI-Supported Fluency and Comprehension

The proposed framework combines four mutually reinforcing strands: accessible extensive reading, repeated reading for automaticity, explicit comprehension strategy instruction, and targeted vocabulary development. AI is positioned as a coordination layer rather than as a replacement for these practices.

First, the learner should encounter a sustained stream of engaging texts at an accessible level. Graded-text research suggests that beginning with highly accessible material can support comprehension, motivation, and self-efficacy, while greater challenge may increase motivation under some conditions (Yang et al., 2021). This does not justify a universal “one level below” rule. Instead, the system should offer a choice set within a tolerable difficulty range and observe

whether the learner can understand, persist, and later transfer learning to a more challenging passage. Text selection should also reflect interests, cultural relevance, genre diversity, and accessibility needs. Limited choice may improve matching, but learner agency should not be eliminated.

Second, repeated reading should be used selectively. Audio-assisted or silent repeated reading can focus attention on phrasing, word recognition, and efficient passage processing. A learner should not be rewarded for reciting a familiar passage faster without demonstrating comprehension. After one or more supported rereadings, the system should present a parallel unfamiliar passage with related language and a different structure. The contrast between trained-passage performance and transfer performance is essential because repeated-passage gains may reflect familiarity or instrumentation.

Third, the system should teach strategies explicitly and in context. Before reading, it may ask the learner to predict the topic and identify relevant knowledge. During reading, it may prompt attention to contrast markers, referents, causal links, or paragraph structure. After reading, it may request an evidence-based summary, an inference with a supporting sentence, or a question that distinguishes fact from interpretation. These prompts should be brief and connected to the passage. Generative AI can draft questions and explanations, but it is less reliable for nuanced inference and evaluation than for factual questions; detailed prompting improves alignment without eliminating oversimplification or content error (Wen & Chu, 2025). Teacher-reviewed question blueprints and passage evidence should constrain generation.

Fourth, vocabulary instruction should be integrated with reading rather than treated as a stream of disconnected definitions. The system can identify words for intentional attention, provide a short context-sensitive explanation, and schedule later recognition and recall. It should also test whether the learner can use the word in a new sentence, recognize a morphological family, or distinguish a near synonym. The goal is not to interrupt every unfamiliar word. Over-helping can weaken persistence and encourage answer hunting. Help should be graduated: reread, infer from context, consult a short clue, compare a gloss, and only then receive a fuller explanation. The strands can be organized into the following instructional logic:

Phase	Primary learner activity	AI function	Evidence of progress
Orient	Set purpose, preview structure, activate knowledge	Offer a prediction prompt and optional background note	Quality of prediction and stated purpose
Read	Read an accessible or appropriately challenging passage	Monitor interaction signals and provide graduated, passage-grounded help	Rate, accuracy, rereading, and confidence
Integrate	Answer literal, inferential, and evaluative questions	Generate or select human-validated prompts and request textual evidence	Comprehension by level and evidence quality
Recycle	Reencounter vocabulary and reread selectively	Schedule spaced retrieval and targeted recurrence	Recognition, recall, depth, and delayed retention
Transfer	Read a novel parallel passage without routine assistance	Reduce prompts and assess independent strategy use	Novel-text comprehension, flexible rate, and self-regulation

Table 1: Instructional Logic for Contextualized and Graduated Support

This table describes a proposed framework, not an established AI treatment effect. It also places learner explanations and teacher judgment alongside automated data. A system that cannot explain why it selected a text or prompt is difficult to audit and difficult for a teacher to improve.

5. AI System Architecture and Instructional Cycle

A defensible architecture would have five layers. The first is a content layer containing licensed or teacher-selected passages, graded and authentic texts, audio where appropriate, vocabulary metadata, genre labels, and human-validated questions. The second is a learner model representing proficiency, vocabulary knowledge, reading purpose, prior performance, confidence, preferred supports, accessibility requirements, and uncertainty. The model should avoid treating demographic or biometric data as necessary proxies for ability (Klella, 2024). The third is an adaptation engine that selects text difficulty, recurrence, prompt timing, and support modality. The fourth is a dialogue and feedback layer that uses retrieval from the assigned passage and constrained templates before open-ended generation. The fifth is an analytics and governance layer that stores minimal data, exposes logs, supports teacher override, and audits subgroup performance.

The instructional cycle begins with a diagnostic that is brief and multidimensional. A learner reads a passage under a specified purpose, answers literal and inferential questions, completes a small vocabulary task, and reports confidence. Rate is recorded, but not used as a sole placement criterion. The system then proposes a text and makes its rationale visible: for example, “This passage is similar in topic and syntax to your previous reading, but it contains a small number of new academic words.” The learner may accept or choose an alternative within the teacher’s range.

During reading, the system should preserve an authentic route. A learner can request a clue, a definition, a paraphrase, or a background note. The AI should quote or point to the relevant passage evidence and mark uncertainty. It should not present an unsupported interpretation as fact (Klella, 2026). For a dense sentence, a visual relation map may help, but the original sentence should remain central. If the system detects repeated rereading or a long pause, it should ask a diagnostic question rather than infer failure.

After reading, the learner answers a balanced set of questions. Factual questions are useful but insufficient. Local inference, global integration, evaluation, and summary tasks require the learner to build and justify a representation of the passage. A generative model may propose items, but a teacher or validated item bank should control consequential use. ChatGPT-generated reading items can have acceptable psychometric properties while still containing shortcomings identified by human reviewers (Lin & Chen, 2024). Competitive questioning and structured learner-generated questions are promising because they can improve both comprehension and the sophistication of questions, but evidence remains tied to particular short interventions.

The system then selects the next action. If rate is low but accuracy and inference are strong, it may prescribe more accessible volume or repeated reading without labeling the learner deficient. If rate is high but global comprehension is weak, it may slow the task, prompt evidence checking, and increase strategy instruction (Shang et al., 2025). If vocabulary recognition is adequate but recall and contextual use are weak, it may schedule spaced retrieval. If the learner performs well with prompts but poorly on a novel text, assistance should fade more slowly and transfer should become a central goal.

Teacher review occurs throughout the cycle. Teachers should be able to inspect the assigned passage, generated prompt, learner response, confidence, support history, and adaptation decision. They should be able to edit or reject feedback and override placement (Godwin-Jones, 2024). This is not merely a governance feature. Teacher interpretation is necessary because the same behavior can have different causes: long reading time may reflect difficulty, careful

inference, distraction, or a disability-related interaction barrier. Human–AI collaboration is therefore distributed agency, not autonomous personalization.

6. Measurement and Evaluation

Evaluation should begin by defining the target construct. A valid battery reports a joint speed–accuracy profile under specified text difficulty, purpose, display, and assistance conditions. It should include silent passage rate, word or proposition accuracy where appropriate, literal comprehension, local and global inference, evaluation, vocabulary breadth and depth, and immediate and delayed transfer (Klella & Al Garaghoolee, 2026). Immediate written recall is useful but mixes comprehension with memory, writing ability, and L2 output demands. Translation or recognition-supported tasks can reveal comprehension that recall suppresses, so no single response format should dominate.

A practical assessment model can include six components. Efficiency is rate combined with accuracy, not rate alone. Comprehension is reported by level, so an overall score does not conceal a collapse in inference. Vocabulary includes recognition, meaning recall, morphology, collocation, and contextual use (Chang, 2006). Transfer uses novel texts, genres, topics, and vocabulary after a delay. Regulation includes purposeful rereading, help selection, prediction, confidence calibration, and evidence checking. Equity and implementation include subgroup error rates, accessibility, device and bandwidth conditions, learner agency, and teacher workload.

Process data can enrich diagnosis, but it should not be overinterpreted. Eye tracking may explain more variance than reading-speed data for some comprehension tests, yet no single eye-movement predictor generalizes across tasks (Mézière et al., 2023). Platform logs are still less direct. Dwell time, rereading, and dictionary use may indicate productive processing or confusion. Their proper use is to generate a hypothesis that is checked with learner responses, not to function as an invisible score.

AI-specific evaluation should use active controls and ablations. A study might compare teacher-led reading, the same platform without adaptation, adaptive non-generative support, and generative support with teacher review. Time-on-task, passage exposure, prompt frequency, and teacher mediation must be reported (Klella & Mrghem, 2024). Human tutoring is a more demanding comparator than business-as-usual instruction; ITS effects are smaller against human tutoring than against conventional instruction (Xu et al., 2019). Studies should also measure whether a model’s feedback is correct, whether learners act on it, and whether it creates new errors.

The strongest outcome is delayed performance on unfamiliar texts. This requirement is especially important because only a small proportion of L2 strategy studies used delayed posttests (Yapp et al., 2021). A credible trial would pre-register primary outcomes, use parallel forms, report missing data and subgroup effects, and distinguish immediate engagement from learning (Mrghem & Klella, 2026). It would assess whether rate gains persist when help is removed and whether comprehension remains stable as text difficulty increases. A system should be considered successful only when it improves independent comprehension and flexible efficiency, not merely click-through, completion, or self-reported confidence.

7. Ethics, Equity, and the Teacher’s Role

AI-assisted reading should be bounded by privacy-by-default and meaningful human oversight. UNESCO recommends data protection, age-appropriate safeguards, and human-centered validation for generative AI in education (Miao & Holmes, 2023). The U.S. Department of Education similarly emphasizes human-in-the-loop use, inspectability, explainability, overridability, privacy, and fairness (U.S. Department of Education, 2023). In practice, a reading platform should collect only data needed for instruction, state retention and sharing rules clearly, offer a non-AI route, and prohibit biometric feedback unless a compelling, consented, and independently reviewed purpose exists.

Equity is not achieved by providing the same model to everyone. Digital inequality includes connectivity, devices, affordability, infrastructure, skills, and meaningful use (Organisation for Economic Co-operation and Development, n.d.). The system should support low-bandwidth and offline pathways, downloadable texts, teacher-mediated paper alternatives, and equivalent assessment conditions. Accessibility should follow the perceivable, operable, understandable, and robust principles of WCAG 2.2 (World Wide Web Consortium, 2024), but interface compliance does not guarantee that generated explanations are cognitively accessible or culturally appropriate. Learners with disabilities retain rights to reasonable accommodation and individualized support; AI must not replace those supports.

Bias can enter through passage selection, learner modeling, simplification, translation, question generation, scoring, and feedback. Research on language models in education has identified demographic bias in personalized educational content, including differences related to income and disability (Weissburg et al., 2025). L2 evaluation should therefore test explanations and examples across proficiency levels, languages, cultures, dialects, and disability needs. Aggregate accuracy is insufficient. Teachers and researchers should inspect harmful stereotypes, missing cultural perspectives, inappropriate difficulty, and systematic under-crediting of multilingual reasoning.

Academic integrity requires process-based design rather than unreliable AI-detector scores. Learners can be asked to make an independent prediction before requesting AI help, cite passage evidence, disclose interactions, and explain how a final answer changed. Formative AI assistance and high-stakes assessment should be separated. High-stakes decisions require validated tasks, human review, and due process (United Nations, 2006). Teacher expertise remains necessary for choosing culturally meaningful texts, interpreting unusual responses, checking generated explanations, and deciding when a learner needs linguistic, instructional, or wider institutional support (Ma et al., 2024; Miao & Cukurova, 2024). Teacher AI literacy should include evaluation of benefits and limits, prompting, assessment, ethics, and feedback.

8. Limitations and Research Agenda

The evidence base has four major limitations. First, much intelligent tutoring evidence concerns K–12 or first-language reading, not adult L2 English. Positive effects cannot simply be transferred across populations. Second, direct generative-AI reading studies are relatively few, often short, quasi-experimental, and concentrated in higher education and East Asian or Turkish EFL settings (Li et al., 2025; Lee et al., 2025). Third, compound interventions make causal attribution difficult. A reported gain may reflect extra time, easier text, vocabulary recurrence, novelty, teacher mediation, multimodal presentation, or the AI component itself. Fourth, many studies emphasize immediate comprehension, engagement, or self-report rather than delayed independent transfer.

Future research should first conduct well-powered, longitudinal trials with adult L2 learners across L1 writing systems, proficiency levels, educational contexts, and access conditions. Studies should compare bounded generative support with non-generative adaptive support and teacher-led equivalents. They should publish adaptation policies, learner-model variables, prompts, exposure time, error logs, and ablations so that “AI” is not treated as an irreducible treatment.

Second, researchers should test the proposed speed–accuracy frontier. Rather than asking whether AI raises words per minute, they should examine whether learners maintain or improve literal, inferential, and evaluative comprehension while reading more efficiently under new purposes and text difficulties. Eye movements and platform logs should be validated against task-specific outcomes, and learner explanations should be included when the system makes a diagnostic inference.

Third, studies should examine mechanism and transfer. Does spaced vocabulary recurrence improve contextual use and delayed comprehension? Does faded strategy prompting increase

independent regulation? Does visible simplification help access while weakening sensitivity to authentic form? Do repeated-reading gains transfer to unfamiliar passages? Mediation analyses and delayed novel-text tests can distinguish these possibilities.

Fourth, equity and labor must be primary outcomes. Researchers should report subgroup performance by language background, disability, socioeconomic access, device, and bandwidth. They should audit generated content and feedback, measure false positives in automated scoring, and quantify teacher review time. An AI intervention that produces modest learning gains by imposing unsustainable monitoring work is not pedagogically scalable.

Finally, the field needs design standards for uncertainty and failure. Systems should state when evidence is insufficient, retrieve from assigned passages, preserve authentic text, limit verbosity, and defer consequential judgments to teachers. Independent replications should verify recent single-study claims, including reported biometric-feedback and platform effects. The goal is not to inhibit innovation but to align technical capability with a construct-valid account of reading.

9. Conclusion

L2 English reading development is best understood as the coordination of efficiency and meaning. Psycholinguistic evidence shows that L2 readers often experience slower lexical access and more effortful visual processing, while also demonstrating heterogeneous compensatory profiles shaped by vocabulary, prior knowledge, working memory, motivation, reasoning, and purpose. Pedagogical evidence favors a coordinated model of accessible extensive reading, selective repeated reading, explicit strategy instruction, and systematic vocabulary development. AI can potentially coordinate these practices through adaptive text selection, passage-grounded prompts, spaced recurrence, graduated help, and formative feedback.

The evidence does not justify claiming that AI automatically improves L2 reading. Direct findings are promising but uneven, and the strongest intelligent-tutoring effects do not establish superiority to skilled teachers or durable transfer to adult L2 English. The defensible recommendation is narrower: use AI as a bounded and inspectable scaffold that supports flexible rate, accuracy, inference, vocabulary growth, and self-regulation; preserve an authentic-text route; measure novel-text performance after delay; and maintain teacher authority over interpretation and consequential decisions. If AI is implemented with privacy, accessibility, equity, and auditability built into the instructional cycle, it can become a useful instrument for developing readers who know not only how to read faster, but when to slow down, what evidence to seek, and how to understand more.

References

- Alrawashdeh, G. S., Fyffe, S., Azevedo, R. F. L., & Castillo, N. M. (2024). Exploring the impact of personalized and adaptive learning technologies on reading literacy: A global meta-analysis. *Educational Research Review*, 42, Article 100587. <https://doi.org/10.1016/j.edurev.2023.100587>
- Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix*, 1(1).
- Çelik, F., Yangın Ersanlı, C., & Arslanbay, G. (2024). Does AI simplification of authentic blog texts improve reading comprehension, inferencing, and anxiety? A one-shot intervention in Turkish EFL context. *The International Review of Research in Open and Distributed Learning*, 25(3), 287–303. <https://doi.org/10.19173/irrodl.v25i3.7779>
- Chang, Y.-F. (2006). On the use of the immediate recall task as a measure of second language reading comprehension. *Language Testing*, 23(4), 520–543. <https://doi.org/10.1191/0265532206lt340oa>
- Clahsen, H., & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, 40(3), 693–706. <https://doi.org/10.1017/S0272263117000250>
- Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLOS ONE*, 10(8), Article e0134008. <https://doi.org/10.1371/journal.pone.0134008>

- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Cunnings, I. (2022). Working memory and L2 sentence processing. In J. W. Schwieter & Z. Wen (Eds.), *The Cambridge handbook of working memory and language* (pp. 593–612). Cambridge University Press. <https://doi.org/10.1017/9781108955638.033>
- European Commission, High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*.
- Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5–31.
- Gorsuch, G., & Taguchi, E. (2008). Repeated reading for developing reading fluency and reading comprehension: The case of EFL learners in Vietnam. *System*, 36(2), 253–278. <https://doi.org/10.1016/j.system.2007.09.009>
- Grabe, W., & Yamashita, J. (2022). Reading assessment. In W. Grabe & J. Yamashita, *Reading in a second language: Moving from theory to practice* (pp. 460–486). Cambridge University Press. <https://doi.org/10.1017/9781108878944.021>
- Grabe, W., & Yamashita, J. (2022). Reading fluency, reading rate, and comprehension. In W. Grabe & J. Yamashita, *Reading in a second language: Moving from theory to practice* (pp. 403–418). Cambridge University Press. <https://doi.org/10.1017/9781108878944.018>
- Jeon, E. H. (2012). Oral reading fluency in second language reading. *Reading in a Foreign Language*, 24(2), 186–208. <https://doi.org/10.64152/10125/66860>
- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, 64(1), 160–212. <https://doi.org/10.1111/lang.12034>
- Jeon, E.-Y., & Day, R. R. (2016). The effectiveness of extensive reading on reading proficiency: A meta-analysis. *Reading in a Foreign Language*, 28(2), 246–265.
- Joh, J., & Plakans, L. (2017). Working memory in L2 reading comprehension: The influence of prior knowledge. *System*, 70, 107–120. <https://doi.org/10.1016/j.system.2017.07.007>
- Joo, D., & Belcher, D. (2025). Human-AI collaborative reading in academic contexts: An exploratory case study. *Journal of English for Academic Purposes*, 78, Article 101571. <https://doi.org/10.1016/j.jeap.2025.101571>
- Kaur, P., Kumar, H., & Kaushal, S. (2023). Technology-assisted language learning adaptive systems: A comprehensive review. *International Journal of Cognitive Computing in Engineering*, 4, 301–313. <https://doi.org/10.1016/j.ijcce.2023.09.002>
- Klella, A. S. (2024). Accessibility and inclusivity in digital language education. *University of Zawia Journal of Educational and Psychological Sciences*, 13(2). <https://journals.zu.edu.ly/index.php/UZJEPS/article/view/1063>
- Klella, A. S. (2026). Artificial intelligence in higher education and its transformative role in educational technologies and learning systems. *University of Zawia Journal of Educational and Psychological Sciences*, 15(1). <https://doi.org/10.26629/uzjeps.2026.05>
- Klella, A. S. A., & Al Garaghoolee, I. S. D. (2026). Cognitive and functional approaches to English language futurity in EFL contexts. *Al-Farooq Journal of Sciences*, 2(Supplement 3), 401–416.
- Klella, A. S. A., & Mrghem, Z. M. O. (2024). Artificial intelligence and human cognition: A systematic review of thought provocation through AI ChatGPT prompts. *ATRAS Journal*, 5(3), 432–444. <https://asjp.cerist.dz/en/article/253661>
- Kuperman, V., Siegelman, N., Schroeder, S., Acartürk, C., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Lõo, K., Marelli, M., Nisbet, K., Papadopoulos, T. C., Protopapas, A., Savo, S., Shalom, D. E., Slioussar, N., Stein, R., Sui, L., Taboh, A., Tønnesen, V., & Usal, K. A. (2023). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus. *Studies in Second Language Acquisition*, 45(1), 3–37. <https://doi.org/10.1017/S0272263121000954>
- Laufer, B., & Aviad-Levitzky, T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *The Modern Language Journal*, 101(4), 729–741. <https://doi.org/10.1111/modl.12431>
- Lee, J., Hicke, Y., Yu, R., Brooks, C., & Kizilcec, R. F. (2024). The life cycle of large language models in education: A framework for understanding sources of bias. *British Journal of Educational Technology*, 55(5), 1982–2002. <https://doi.org/10.1111/bjet.13505>

- Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, 335–359. <https://doi.org/10.1080/10494820.2025.2498537>
- Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, 7, Article 100266. <https://doi.org/10.1016/j.caeai.2024.100266>
- Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, 9, Article 100445. <https://doi.org/10.1016/j.caeai.2025.100445>
- Lin, Z., & Chen, H. (2024). Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items. *System*, 123, Article 103344. <https://doi.org/10.1016/j.system.2024.103344>
- Ma, Q., Crosthwaite, P., Sun, D., & Zou, D. (2024). Exploring ChatGPT literacy in language education: A global perspective and comprehensive approach. *Computers and Education: Artificial Intelligence*, 7, Article 100278. <https://doi.org/10.1016/j.caeai.2024.100278>
- Magliano, J. P., Millis, K. K., The R-SAT Development Team, & Levinstein, I. (2011). Assessing comprehension during reading with the Reading Strategy Assessment Tool (RSAT). *Metacognition and Learning*, 6(2), 131–154. <https://doi.org/10.1007/s11409-010-9064-2>
- Mežek, Š., McGrath, L., Negretti, R., & Berggren, J. (2022). Scaffolding L2 academic reading and self-regulation through task and feedback. *TESOL Quarterly*, 56(1), 41–67. <https://doi.org/10.1002/tesq.3018>
- Mézière, D. C., Yu, L., Reichle, E. D., von der Malsburg, T., & McArthur, G. (2023). Using eye-tracking measures to predict reading comprehension. *Reading Research Quarterly*, 58, 425–449. <https://doi.org/10.1002/rrq.498>
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104_eng
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693_eng
- Miao, F., Shiohira, K., & Lao, N. (2024). *AI competency framework for students*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391105_eng
- Mrghem, Z. M. O., & Klella, A. S. A. (2026). Cybersecurity Awareness in Higher Education: A Literature Review of Educational Practices and Digital Safety. *North African Journal of Scientific Publishing (NAJSP)*, 4(2), 264–274. <https://najsp.com/index.php/home/article/view/892>
- Nisbet, K., Bertram, R., Erlinghagen, C., Pieczykolan, A., & Kuperman, V. (2022). Quantifying the difference in reading fluency between L1 and L2 readers of English. *Studies in Second Language Acquisition*, 44(2), 407–434. <https://doi.org/10.1017/S0272263117000250>
- Organisation for Economic Co-operation and Development. (n.d.). *Digital divide in education*. Retrieved September 19, 2026, from <https://www.oecd.org/en/topics/sub-issues/digital-divide-in-education.html>
- Ozer, O. (2024). AI language models: A breach of academic integrity in online language learning? *Studies in Language Assessment*, 13(1), 237–260. https://www.altanz.org/uploads/5/9/0/8/5908292/7_sila_13_1_ozer.pdf
- Rayner, K., Schotter, E. R., Masson, M. E. J., Potter, M. C., & Treiman, R. (2016). So much to read, so little time: How do we read, and can speed reading help? *Psychological Science in the Public Interest*, 17(1), 4–34. <https://doi.org/10.1177/1529100615623267>
- Sangers, N. L., van der Sande, L., Welie, C., Dobber, M., & van Steensel, R. (2025). Learning a language through reading: A meta-analysis of studies on the effects of extensive reading on second and foreign language learning. *Educational Psychology Review*, 37(4), Article 96. <https://doi.org/10.1007/s10648-025-10068-6>
- Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, 19(1), 53–67. <https://doi.org/10.1017/S0142716400010572>

- Shafiee Rad, H. (2025). Reinforcing L2 reading comprehension through artificial intelligence intervention: Refining engagement to foster self-regulated learning. *Smart Learning Environments*, 12, Article 23. <https://doi.org/10.1186/s40561-025-00377-2>
- Shang, J., Huang, Y., Xu, M., Huang, Y., Shen, X., Wang, G., Wang, Y., & Zhang, L. (2025). Competition between human learners and ChatGPT: Enhancing university EFL students' reading comprehension and critical thinking through competitive questioning. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2025.2577356>
- Spichtig, A. N., Gehsmann, K. M., Pascoe, J. P., & Ferrara, J. D. (2019). The impact of adaptive, web-based, scaffolded silent reading instruction on the reading achievement of students in grades 4 and 5. *The Elementary School Journal*, 119(3), 443–467. <https://doi.org/10.1086/701705>
- Traxler, M. J., Long, D. L., Tooley, K. M., Johns, C. L., Zirnstein, M., & Jonathan, E. (2012). Individual differences in eye-movements during reading: Working memory and speed-of-processing effects. *Journal of Eye Movement Research*, 5(1), Article 5. <https://pmc.ncbi.nlm.nih.gov/articles/PMC4467465/>
- Tywniwi, R. (2023). Compensatory effects of individual differences, language proficiency, and reading behavior: An eye-tracking study of second language reading assessment. *Frontiers in Communication*, 8, Article 1176986. <https://doi.org/10.3389/fcomm.2023.1176986>
- U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. <https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report-core-messages.pdf>
- United Nations. (2006). *Convention on the Rights of Persons with Disabilities*. Office of the United Nations High Commissioner for Human Rights. <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-persons-disabilities>
- Wang, X., & Feng, Y. (2024). An experimental study of ChatGPT-assisted improvement of Chinese college students' English reading skills: A case study of *Dear Life*. In *Proceedings of the 15th International Conference on Education Technology and Computers* (pp. 21–26). Association for Computing Machinery. <https://doi.org/10.1145/3629296.3629300>
- Wang, Y.-H. (2016). Promoting contextual vocabulary learning through an adaptive computer-assisted EFL reading system. *Journal of Computer Assisted Learning*, 32(4), 291–303. <https://doi.org/10.1111/jcal.12132>
- Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, 56(2), 161–180. <https://doi.org/10.1017/S0261444822000507>
- Weissburg, I., Anand, S., Levy, S., & Jeong, H. (2025). LLMs are biased teachers: Evaluating LLM bias in personalized education. In *Findings of the Association for Computational Linguistics: NAACL 2025* (pp. 5665–5713). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.findings-naacl.314>
- Wen, Z., & Chu, S. K. W. (2025). Using generative AI for reading question creation based on PIRLS 2011 framework. *Cogent Education*, 12(1), Article 2458653. <https://doi.org/10.1080/2331186X.2025.2458653>
- Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, 19, 73–80. <https://doi.org/10.3758/s13423-011-0179-5>
- Wijekumar, K. K., Meyer, B. J. F., & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development*, 60, 987–1013. <https://doi.org/10.1007/s11423-012-9263-4>
- World Wide Web Consortium. (2024). *Web Content Accessibility Guidelines (WCAG) 2.2*. <https://www.w3.org/TR/WCAG22/>
- Xu, Z., Wijekumar, K. K., Ramirez, G., Hu, X., & Irey, R. (2019). The effectiveness of intelligent tutoring systems on K–12 students' reading comprehension: A meta-analysis. *British Journal of Educational Technology*, 50(6), 3119–3137. <https://doi.org/10.1111/bjet.12758>
- Xuan, Q., Cheung, A., & Sun, D. (2022). The effectiveness of formative assessment for enhancing reading achievement in K–12 classrooms: A meta-analysis. *Frontiers in Psychology*, 13, Article 990196. <https://doi.org/10.3389/fpsyg.2022.990196>

- Yang, Y., & Qian, D. D. (2020). Promoting L2 English learners' reading proficiency through computerized dynamic assessment. *Computer Assisted Language Learning*, 33(5–6), 628–652. <https://doi.org/10.1080/09588221.2019.1585882>
- Yang, Y.-H., Chu, H.-C., & Tseng, W.-T. (2021). Text difficulty in extensive reading: Reading comprehension and reading motivation. *Reading in a Foreign Language*, 33(1), 78–102. <https://doi.org/10.64152/10125/67394>
- Yapp, D. J., de Graaff, R., & van den Bergh, H. (2021). Improving second language reading comprehension through reading strategies: A meta-analysis of L2 reading strategy interventions. *Journal of Second Language Studies*, 4(1), 154–192. <https://doi.org/10.1075/jsls.19013.yap>
- Yousefi, M. H., & Biria, R. (2018). The effectiveness of L2 vocabulary instruction: A meta-analysis. *Asian-Pacific Journal of Second and Foreign Language Education*, 3, Article 21. <https://doi.org/10.1186/s40862-018-0062-2>
- Yuan, H. (2025). Artificial intelligence in language learning: Biometric feedback and adaptive reading for improved comprehension and reduced anxiety. *Humanities and Social Sciences Communications*, 12, Article 556. <https://www.nature.com/articles/s41599-025-04878-w>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, Article 28. <https://doi.org/10.1186/s40561-024-00316-7>
- Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, 26(4), 696–725. <https://doi.org/10.1177/1362168820913998>
- Zhang, S., Shan, C., Lee, J. S. Y., Che, S., & Kim, J. H. (2023). Effect of chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies*, 28, 15223–15243. <https://doi.org/10.1007/s10639-023-11805-6>