

Modality Principle in Learning L2: A Systematic Review of How Proficiency Moderates the Efficacy of Reading-While-Listening on Receptive Skills

Padma Lamo

Indian Institute of Technology, Bhubaneswar, India

<A22hs09004@iitbbs.ac.in>

Saleem Mohd Nasim

Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

<s.nasim@psau.edu.sa>

Rajakumar Guduru

Indian Institute of Technology, Bhubaneswar, India

<rajakumarguduru@iitbbs.ac.in>

Abstract

Despite growing interest in reading-while-listening (RWL) for second language (L2) learning, findings remain inconsistent and difficult to translate into classroom decisions. This systematic review synthesizes empirical evidence on how learner proficiency moderates the effects of reading-only (RO), listening-only (LO), and RWL on L2 receptive outcomes. Following PRISMA 2020 guidance, five databases were searched (1970–2024), yielding 24 eligible empirical studies. Study quality was appraised using Joanna Briggs Institute critical appraisal checklists, and findings were synthesized narratively due to substantial heterogeneity in designs, measures, and interventions. Across studies, proficiency emerged as the most consistent moderator. For lower-proficiency learners, RWL often functioned as scaffolding, with large gains reported for reading comprehension in extended interventions and strong benefits for struggling readers. Listening outcomes frequently favored supported input, with several studies reporting large improvements. Process evidence further suggests that RWL can alter online comprehension behavior by increasing integrative eye movements. In contrast, for higher-proficiency learners, RO sometimes outperformed RWL on text comprehension, consistent with redundancy and divided-attention accounts. A prior meta-analysis reported only a small overall RWL advantage over RO, reinforcing that modality effects are conditional rather than uniform. Building on these patterns, the review proposes a Proficiency-Scaffolding Model and offers a proficiency-based decision guide to help educators align modality choice with learner level, task demands, and cognitive load.

Keywords: reading-while-listening; input modalities; receptive skills; L2 proficiency; systematic review; cognitive load

A Critical Analysis of Prior Syntheses and the Present Systematic Review's Contribution

Receptive skills—listening and reading—are fundamental to language development, enabling learners to establish form–meaning links from auditory and visual input. In instructional settings, input is commonly delivered through unimodal channels, listening-only (LO) or reading-only (RO), or through the bimodal channel of reading-while-listening (RWL), which combines auditory and visual information in real time (Cheetham, 2019). Evidence from unimodal-focused work suggests that listening and reading outcomes reflect modality-specific processes, such as strategic regulation during listening and targeted comprehension-building during reading, making unimodal conditions a necessary baseline for interpreting the added value or cost of bimodal input (Nasim, 2022; Nasim et al., 2024). Understanding how unimodal and bimodal input affects receptive performance is therefore central to evidence-based ESL instruction.

Although systematic reviews and meta-analyses have examined these input modes, their applicability to classroom practice is often limited. Syntheses vary widely in scope, learner populations, outcome measures, and methodological rigor. As text–audio formats become more common in technology-mediated ESL/EFL classrooms, teachers require clearer guidance on when RO, LO, or RWL is most appropriate, particularly for learners at different proficiency levels. The present systematic review addresses this need by synthesizing empirical studies that directly compare RO, LO, and RWL for receptive outcomes in ESL/EFL learners, with particular attention to proficiency as a moderator. The review follows PRISMA reporting guidance (Page et al., 2021) and appraises study quality using Joanna Briggs Institute (JBI, 2017) critical appraisal checklists.

A Landscape of Fragmented Comparisons and Heterogeneous Focus

Existing reviews often present fragmented comparisons, preventing a unified understanding of the relative efficacy of RO, LO, and RWL within a single analytical framework. Many syntheses rely on pairwise contrasts, typically RO versus LO followed by RWL versus RO, rather than integrating all three modalities together. When pooled, RO has shown a modest advantage in some contexts, whereas RWL often shows little overall benefit over RO, with any advantage becoming more visible under experimenter-paced conditions (Clinton-Lisell, 2022, 2023). This structure makes it difficult to draw coherent, classroom-relevant conclusions about how the three modes compare under comparable conditions.

Interpretation is further complicated when syntheses include all three modes without categorizing studies, mixing populations and task types in ways that distance the evidence from typical ESL acquisition. Some reviews combine ESL learners with native speakers and include general information-acquisition tasks, shifting the focus from second language development to broader information processing (Adesope & Nesbit, 2012; Neuman & Koskinen, 1992). Similarly, audiobook-focused syntheses often identify gaps for older learners and nonfiction texts but also include studies involving learners with disabilities, which introduce distinct instructional aims and constraints (Singh & Alexander, 2022; Stepien-Bernabe et al., 2019). As a result, these reviews blend outcomes and populations, reducing comparability across studies and limiting direct transfer to mainstream ESL contexts focused on receptive-skill development.

Transferability is further constrained by heterogeneous learner groups and research aims that pull evidence away from typical ESL learners and core listening and reading development. Reviews on text-to-speech, for example, often focus on struggling readers or include learners studying languages other than English, such as Spanish or French, shifting both context and outcomes

beyond mainstream ESL acquisition (Mohsen, 2016; Wood et al., 2018; Shaojie et al., 2022). In that broader literature, audiovisual input is frequently discussed in terms of engagement, or media "help options" are treated mainly as possible sources of cognitive overload, rather than isolating modality effects within RO, LO, and RWL comparisons (Mohsen, 2016; Shaojie et al., 2022). Because these syntheses combine different populations, languages, and outcome priorities, their conclusions do not translate cleanly to unimodal and bimodal text processing in ESL classrooms. This leaves a clear gap for a review focused specifically on ESL/EFL learners and reading and listening comprehension gains.

Limitations of Prior Syntheses

Beyond scope and population issues, limitations in how prior reviews appraised and synthesized evidence reduce the clarity and usefulness of their conclusions. This section evaluates limitations in prior syntheses rather than reviewing primary studies; it explains why a PRISMA-guided, quality-appraised three-mode synthesis is still needed.

Methodological rigor is the most persistent weakness across prior syntheses, limiting the confidence with which their conclusions can be applied. Many reviews summarize findings without formally appraising the quality of the primary studies they include. Without a transparent risk-of-bias appraisal using established tools, readers cannot judge whether reported effects reflect strong evidence or methodological artifacts, which undermines the credibility of the synthesis.

A second limitation is that many reviews describe study selection and data extraction with insufficient precision to support clean interpretation. Loosely defined inclusion and exclusion criteria can unintentionally blend conceptually different outcomes. Some syntheses, for example, combine discrete vocabulary gains with broader comprehension measures, making it harder to interpret what the evidence shows about receptive skills, specifically reading and listening comprehension. This definitional drift reduces comparability across studies and can produce conclusions that extend beyond what the underlying data can support.

A third limitation is that prior reviews often rely on descriptive summaries rather than structured, condition-based comparisons. Instead of comparing RO, LO, and RWL under specified conditions, they frequently report results study by study without integrating patterns across learner proficiency, task demands, or pacing conditions. Finally, learner perception evidence is commonly treated as supplementary rather than synthesized systematically. When perception data appear, they are rarely analyzed using a transparent qualitative method, leaving an incomplete account of not only what works but also how learners experience these modalities and why those experiences may matter for classroom use.

Addressing the Gaps: The Contribution of the Current Study

The critical analysis above reveals several interconnected limitations that this study is designed to overcome. Table 1 summarizes the key gaps in earlier syntheses and the corresponding contributions of this review.

Table 1. Gaps in Previous Literature and Specific Contributions of the Current Study

Identified Gap in Previous Literature	Specific Contribution of the Current Study
1. Fragmented comparisons: Prior reviews were limited to pairwise contrasts or lacked a unified framework for RO, LO, and RWL (Clinton-Lisell, 2022, 2023).	1. Three-mode synthesis: This review compares RO, LO, and RWL within a single analytical framework.
2. Heterogeneous focus: Inclusion of native speakers, learners with disabilities, or other languages reduces relevance for typical ESL/EFL acquisition.	2. ESL/EFL focus: We apply explicit inclusion criteria targeting typical ESL/EFL learners and reading and/or listening comprehension outcomes.
3. Limited methodological rigor: Prior syntheses often omit formal quality appraisal and report eligibility and extraction decisions imprecisely.	3. Methodological rigor: This review follows PRISMA (Page et al., 2021) and includes systematic quality appraisal using JBI critical appraisal checklists (JBI, 2017).
4. Limited qualitative synthesis: Prior syntheses emphasize quantitative outcomes and rarely synthesize learner perceptions transparently.	4. Qualitative synthesis: This review synthesizes learner perceptions thematically (RQ3) and integrates moderating factors (RQ2) throughout the analysis.

This review integrates RO, LO, and RWL within one analytic frame, directly addressing the fragmentation created by prior pairwise syntheses. Additionally, it appraises study quality systematically using JBI checklists, strengthening interpretability by making risk of bias explicit. Furthermore, it synthesizes learner perceptions thematically alongside outcome patterns, linking modality effects to learner experience and classroom feasibility. Together, these contributions help reconcile inconsistencies in the literature and provide more actionable guidance for matching RO, LO, and RWL to learner proficiency, task demands, and learning goals.

Methodology

This systematic review aimed to synthesize empirical research on the effects of unimodal (RO or LO) versus bimodal (RWL) input modalities on the receptive skills of EFL/ESL learners. To ensure transparency and methodological rigor, the synthesis procedure followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) updated by Page et al. (2021). The procedure encompassed the following stages: (1) identification of studies via a systematic search strategy, (2) screening of records, (3) eligibility assessment, (4) data extraction, (5) quality appraisal, (6) data synthesis, and (7) evidence integration.

This review is guided by the following research questions:

1. How do different input modes influence reading and listening performance?
2. What potential moderating factors, apart from input modes, affect listening and reading performance?
3. How do English language learners perceive different modes of input in reading and listening tasks?
4. What are the potential research directions for the use of different modes of input in the ESL domain?

Identification and Search Strategy

A systematic search of the literature was performed to identify relevant empirical studies published between 1970 and 2024. The search was executed across five electronic databases: Web of Science, Scopus, ERIC (Education Resources Information Center), Google Scholar, and ProQuest

Dissertations and Theses. This approach ensured comprehensive coverage of high-impact, peer-reviewed journal articles as well as relevant grey literature, such as unpublished dissertations, to mitigate publication bias.

The search strategy employed a Boolean query that combined key terms related to the interventions (input modalities) and outcomes (receptive skills). To maintain the focus on the core modalities of interest and isolate the effect of synchronous audio-text input, studies involving learning disabilities or subtitles were explicitly excluded from the search string. While subtitles represent a relevant form of bimodal input, they introduce additional visual and cognitive components that fall outside the scope of this review. The final search string was: ("reading while listening" OR "bimodal input" OR "audiobook") AND ("reading comprehension" OR "listening comprehension" OR "receptive skills") AND ("ESL" OR "EFL" OR "English as a second language" OR "English language learners") NOT ("subtitle" OR "video" OR "animation" OR "dyslexia" OR "learning disabilities").

To supplement the database search, backward and forward citation tracking was performed on all included primary studies and relevant prior systematic reviews to identify any additional pertinent publications that may have been missed.

Screening and Study Selection

The study selection process is outlined in the PRISMA flow diagram (Figure 1; Page et al., 2021). The database search identified 3,668 records, and a further 15 records were identified through manual searching and citation tracking. After removing 1,443 duplicates, 2,240 records were screened by title and abstract against the predefined inclusion and exclusion criteria. This process yielded 76 reports for full-text assessment. Of these, 52 reports were excluded due to the inclusion of subtitles or additional modalities (e.g., video), a focus on unrelated variables (e.g., strategy instruction only), or an emphasis on discrete vocabulary outcomes rather than global reading or listening comprehension. The final corpus comprised 24 empirical studies.

Two reviewers independently screened titles/abstracts and full texts; disagreements were resolved through discussion.

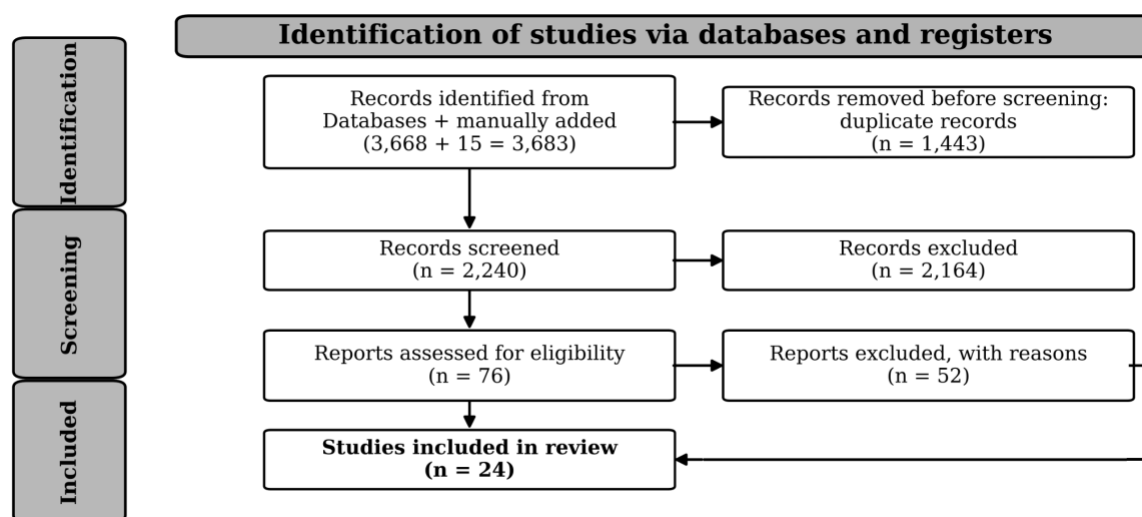


Figure 1. PRISMA 2020 flow diagram depicting the literature search and study selection process.

Eligibility—Inclusion and Exclusion Criteria

Studies were screened for inclusion based on the following pre-defined criteria:

Inclusion Criteria:

- Empirical research employing an experimental or quasi-experimental design with measurable outcomes.
- Involved ESL or EFL learners.
- Participants possessed sufficient language proficiency to engage with reading and listening tasks.
- The dependent variable(s) included gains in either reading comprehension, listening comprehension, or both.
- The study included at least one experimental condition featuring bimodal input (RWL) and a comparative condition employing a different input modality (RO or LO).
- Assessments of receptive skills utilized standardized or comparable measurement instruments.
- Published in English between 1970 and 2024 in peer-reviewed sources and/or indexed theses/dissertations (e.g., ProQuest).

Exclusion Criteria:

- Non-empirical studies (e.g., theoretical papers, literature reviews, editorials).
- Studies incorporating animations, videos, or on-screen text (subtitles/captions).
- Published in a language other than English.
- Conducted with learners with diagnosed learning or physical disabilities, or with children below the age of seven.
- Focused on learners of languages other than English.
- Examined only a specific aspect of language (e.g., isolated vocabulary gains) rather than global reading or listening comprehension.
- Measured dependent variables beyond the scope of reading and listening performance (e.g., speaking fluency).

Data Extraction

We extracted study metadata, design features, outcomes, and moderators using a standardized form. For each included study, we recorded bibliographic details (author, year); participant characteristics (sample size, age, proficiency level, L1); design features (e.g., Btn- vs. within-subjects design, treatment duration); intervention details (modalities compared); outcome measures; key quantitative results (including reported or calculated effect sizes); qualitative findings (e.g., learner perceptions); and any moderators explicitly examined (e.g., proficiency, text type) to address RQ2. The extracted data are summarized in a simplified comparative synthesis table (Tables 2a, 2b & 2c), while detailed study-level quantitative findings are reported separately in Appendix Table S1.

Table 2a. Simplified Comparative Synthesis and Quality Assessment of the 12 Included Empirical Studies - Reading-dominant outcomes

No.	Study	Context / L1 profile	N	Duration	Design	Key measures	Risk of bias	Main finding
1	Aka (2024)	Japanese HS EFL learners	157	3 wks	Within	RC, P	Moderate	Within: no sig. overall; low-proficiency trend toward RWL.
2	Chang & Millett (2015)	Chinese secondary EFL learners	64	26 wks	Btn	RF, RC	Moderate	Btn: sig.; RWL improved RF and RC, large.
3	Diao & Sweller (2007)	Chinese university EFL learners	60	2 ses	Btn	RC, CL	Low	Btn: sig.; RO > RWL on RC.
4	Eppard et al. (2020)	Mixed-L1 university L2 learners	61	8 wks	Btn	RF, RC, P	Moderate-high	Btn: no sig.; perceived vocabulary benefit only.
5	Friedland et al. (2017)	Mixed-L1 Grade 3 learners	46	4 wks	Btn	RF	Low	Btn: borderline; trivial RF gain.
6	Holmes & Allison (1986)	Elementary readers	48	1 ses	Within	RC	Low	Within: small effect; no overall poor-reader advantage; good readers disadvantaged by RWL.
7	Liu et al. (2019)	Chinese undergraduate learners	42	35 min	Btn	RC	High	Btn: sig.; dual input > single mode on RC.
8	Nakashima et al. (2018)	Japanese university EFL learners	130	15 wks	Mixed	RC, P	High	Mixed: sig.; RWL and silent reading > LO on RC; preferences split.
9	Rasinski (1990)	Grade 3 readers	20	1 ses	Within	RF	Moderate-high	Within: time effect only; no treatment advantage.
10	Serrano & Pellicer-Sánchez (2019)	Cat./Span. bilingual children	36	10–24 min	Within	EM, RC	Low	Within: process shift under RWL; no sig. RC difference.
11	Tragant Mestres et al. (2019)	Cat./Span. primary learners	96	17 wks	Btn	RC, V, RF, P	Low-moderate	Btn: sig.; vocabulary gains > control; more enjoyment.
12	Verlaan & Ortlieb (2012)	10th-grade H S students (110 score pairs / 3 days)	110 score pairs	3 days	Within	RC	Moderate	Within: sig.; strongest benefit for struggling readers.

Table 2b. Simplified Comparative Synthesis and Quality Assessment of the 3 Included Empirical Studies - Mixed/process outcomes

No.	Study	Context / L1 profile	N	Duration	Design	Key measures	Risk of bias	Main finding
13	Pellicer-Sánchez et al. (2020)	Cat./Span. bilingual child EFL readers	30	50 min	Within	EM, RC	Low	Within: sig. process shift; no clear RC gain.
14	Pellicer-Sánchez et al. (2021)	Adult L1/L2 readers	47	1 ses	Within	EM, RC	High	Within: sig. process shift; effects differed by reader type.
15	Yang et al. (2022)	Mandarin-speaking Grade 3 EFL learners	41	1 ses	Btn	CL, EM, RC	Moderate-high	Btn: no main RC advantage; attention shifted under audio support.

Table 2c. Simplified Comparative Synthesis and Quality Assessment of the 9 Included Empirical Studies - Listening-dominant outcomes

No.	Study	Context / L1 profile	N	Duration	Design	Key measures	Risk of bias	Main finding
16	Chang (2009)	Chinese university EFL learners	84	1 Ses	Within	LC, P	Moderate	Within: sig.; RWL > LO on LC, medium.
17	Chang (2011)	Chinese university EFL learners	19	26 wks	Btn	LC, V	High	Btn: sig.; LC and vocabulary gains, large.
18	Chang & Millett (2014)	Chinese low-intermediate university EFL learners	113	13 wks	Btn	LC	High	Btn: sig.; RWL showed the strongest LC gains.
19	Chang et al. (2019)	Chinese college EFL learners	69	39 wks	Btn	LC	Moderate-high	Btn: sig.; RWL showed the largest LC gains.
20	Hartshorn & Stephens (2023)	Advanced mixed-L1 ESL learners	31	14 wks	Btn	LC	Low	Btn: sig.; transcript-supported listening > control.
21	Hui (2024)	University English learners	46	1 ses	Within	C (RO/LO /RWL)	Low	Within: sig.; RWL > LO; RWL $\approx$ RO.
22	Kartal & Simsek (2017)	Turkish university EFL learners	66	13 wks	Btn	LC, P, M	Low-moderate	Btn: sig.; audiobook group > control on LC.
23	Kim (2021)	Korean university EFL learners	75	5 wks	Btn	LC, RC	Moderate	Btn: sig.; LC highest in RWL; RWL/RO > LO on RC.
24	Milliner (2019)	Japanese beginner univ EFL learners	58	15 wks	Btn	LC, RC, V	Moderate-high	Btn: sig.; LC improved most under RWL; RC favored control/RO.

Note. RC = reading comprehension; LC = listening comprehension; RF = reading fluency; V = vocabulary; P = perceptions/preferences; EM = eye movement; CL = cognitive load; M = motivation; C = overall comprehension across modality conditions.

Study Quality and Risk of Bias Assessment

We assessed study quality to make the strength of the evidence explicit rather than assumed. Using the JBI critical appraisal checklists for experimental and quasi-experimental studies (JBI, 2017), two trained reviewers independently rated each included study across domains such as participant description, allocation/randomization procedures, blinding (where applicable), validity and reliability of outcome measures, control of confounding variables, and appropriateness of statistical analyses. We report on study-level ratings in Table 2 and summarize the overall distribution by risk of bias category in Table 3.

Table 3. JBI Critical Appraisal and Risk of Bias Assessments

Risk category	n (%)	Prototypical study characteristics	Analysis by design type
Low	7 (29.2)	Rigorous designs (e.g., counterbalanced within-subjects or randomized allocation); valid/reliable measures; appropriate control of confounds; appropriate analyses.	Within-subject (n = 9): 4 (44.4%) / Between-subject (n = 15): 3 (20.0%)
Moderate	7 (29.2)	One major limitation (e.g., no pretest, baseline imbalance, unclear allocation, non-validated instruments), but results remain interpretable.	Within-subject (n = 9): 3 (33.3%) /Between-subject (n = 15): 4 (26.7%)
High	10 (41.7)	Multiple critical limitations (e.g., non-equivalent groups, very small samples, high attrition, weak measurement reliability, inadequate control).	Within-subject (n = 9): 2 (22.2%) /Between-subject (n = 15): 8 (53.3%)
Total	24 (100)		

Note. Between-subject designs included one randomized controlled trial (RCT). Percentages in the “Analysis by design type” column are calculated within design type.

Data Synthesis and Evidence Integration

Given the significant heterogeneity observed in populations, interventions, and outcome measures, a quantitative meta-analysis was deemed unfeasible. Consequently, a systematic narrative synthesis approach was adopted, guided by the structured framework proposed by Popay et al. (2006). This process involved (1) developing a preliminary synthesis by tabulating findings, (2) exploring relationships between study characteristics and outcomes, and (3) assessing the robustness of the synthesis. The final, integrative step involves weaving these threads together to address the broader phenomenon and to support the discussion and conclusions of the review.

Synthesis Procedures by Research Question

To comprehensively address all research questions, the synthesis employed specific techniques:

- For RQ1 (Influence on Performance): Studies were grouped by the receptive skill (reading or listening comprehension) and the modality comparison (e.g., RWL vs. RO). The direction, magnitude, and consistency of effects were compared across these groupings. The methodological quality and risk of bias assessments were integral to this process, with findings from studies with a low risk of bias being accorded greater weight in interpretation.
- For RQ2 (Moderating Factors): We conducted a framework analysis. Factors cited by authors as influencing outcomes were recorded during data extraction and iteratively categorized into key domains: Learner Characteristics (e.g., proficiency, age), Task Features (e.g., text genre, complexity), and Methodological Factors (e.g., pacing). The influence of each moderator was traced across the corpus to identify consistent patterns and discrepancies.
- For RQ3 (Learner Perceptions): Qualitative data were synthesized using a thematic synthesis approach (Thomas & Harden, 2008), involving line-by-line coding, the development of descriptive themes, and the generation of analytical themes to interpret learner experiences.
- For RQ4 (Potential Research Directions): The gaps, inconsistencies, and limitations identified during the syntheses for RQ1–RQ3 were systematically collected and categorized to generate a coherent agenda for future research.

Analysis and Discussion

RQ1: The Influence of Input Modalities on Receptive Skills

Input modality effects are conditional rather than uniform. Across the 24 included studies, outcomes varied most consistently with (a) learner proficiency and (b) the target receptive skill (reading vs. listening). This context-dependence aligns with meta-analytic evidence showing a small overall effect of RWL relative to RO across a broader pool of bimodal-input studies, suggesting that RWL is not universally beneficial (Clinton-Lisell, 2023). The evidence supports a pattern in which RWL functions as a scaffold for specific learner groups and objectives, while its value diminishes or changes as learner capabilities develop.

Reported effects ranged from trivial differences in overall comprehension (Aka, 2024) to large gains in extended supported-reading interventions for reading comprehension (Chang & Millett, 2015) and substantial improvements in listening outcomes under RWL-supported conditions (Chang & Millett, 2014; Chang et al., 2019). Stronger positive effects were more common in

lower-proficiency groups and longer interventions, whereas findings for higher-proficiency learners were more mixed.

The Primacy of Proficiency: A Key Moderator of RWL Efficacy

Proficiency emerged as the most consistent moderator across the corpus. As shown in Table 4, reading-while-listening (RWL) was more likely to support lower-proficiency learners by easing decoding demands, whereas higher-proficiency learners benefited only under narrower task conditions. Across studies, the evidence suggests that the value of RWL is conditional rather than uniform.

Table 4. Representative Quantitative Signals for Proficiency as a Moderator (RQ1)

Study	Comparison	Learner group	Key quantitative signal
Verlaan and Ortlieb (2012)	RWL vs. silent reading	Overall and struggling readers	Overall $d = 0.36$; struggling readers $d = 0.99$
Aka (2024)	RWL vs. RO	Low-proficiency subgroup	RWL $M = 8.00$ vs. RO $M = 7.00$; $p = .07$
Diao and Sweller (2007)	RO vs. RWL	First-year EFL majors	RO > RWL on free recall; $\omega^2 = 0.07$; lexical $\omega^2 = 0.37$
Chang and Millett (2015)	Audio-assisted reading vs. silent reading	Secondary EFL learners over 26 weeks	Reading comprehension $d = 1.55$ at post-test
Hui (2024)	RWL vs. LO / RO	University EFL learners	RWL > LO; RWL $\approx$ RO

For lower-proficiency and struggling readers, RWL most often supported comprehension. This pattern was especially visible in adolescent and struggling-reader samples, where dual-channel input appeared to reduce decoding pressure and free attentional resources for meaning construction. The trend reported by Aka (2024) and the stronger effects observed by Verlaan and Ortlieb (2012) point in the same direction, suggesting that RWL is most helpful when learners still require support in coordinating written and spoken input. Evidence from Chang and Millett (2015) further suggests that sustained audio-assisted reading can produce substantial gains when this support is extended over time.

For more proficient learners, however, the advantage of RWL narrows and can even reverse. Once decoding becomes relatively automatic, concurrent audio and text may create redundancy or split attention rather than facilitate comprehension. This pattern is most clearly illustrated in Diao and Sweller (2007), where RO outperformed RWL on lexical comprehension and free recall. Hui (2024) also supports a conditional interpretation by showing that although RWL outperformed LO, it did not significantly differ from RO. Taken together, these findings indicate that proficiency shapes whether RWL functions as a useful scaffold or as an unnecessary layer of input.

Differential Effects on Reading Versus Listening Outcomes

RWL effects also vary by the specific receptive outcome targeted. As shown in Table 5, the evidence does not support a single, uniform advantage for RWL across reading fluency, reading comprehension, and listening comprehension. Instead, the pattern differs by outcome domain, with the clearest benefits appearing in fluency-oriented and listening outcomes.

Table 5. Representative Quantitative Patterns Across Reading and Listening Outcomes (RQ1)

Outcome domain	Studies included in this subsection	Compact quantitative signal	Overall pattern
Reading fluency / rate	Chang and Millett (2015); Rasinski (1990)	Large gains in extended practice ($d = 1.00-1.55$); Rasinski reported significant time effects only	Most consistent positive pattern
Reading comprehension	Verlaan and Ortlieb (2012); Chang and Millett (2015); Holmes and Allison (1986); Liu et al. (2019); Tragant Mestres et al. (2019); Eppard et al. (2020); Serrano and Pellicer-Sánchez (2019)	Mixed findings: subgroup benefit ($d = 0.99$), one significant dual-channel effect ($\eta^2 = .145$), several ns results	Conditional and task-dependent
Listening comprehension	Chang and Millett (2014); Chang et al. (2019); Kartal and Simsek (2017); Milliner (2019); Hartshorn and Stephens (2023); Kim (2021)	Positive effects were frequent and often substantial ($d = 0.71-2.27$; $\eta^2 = .186-.25$; $\omega^2 = .527$).	Clearest and most consistent RWL advantage

Note. ns = non-significant.

Reading fluency and rate show the most consistent positive pattern. RWL appears to support fluency development because audio provides a pacing and prosodic model that promotes automaticity. The strongest gains were observed in extended audio-assisted reading, particularly in Chang and Millett (2015), while Rasinski (1990) suggests that time and repeated practice also contribute meaningfully to fluency growth. Taken together, these findings indicate that RWL is especially well suited to fluency-oriented goals, where speed, rhythm, and processing ease are central.

Reading comprehension effects are more mixed and more condition-dependent. Some studies reported clear benefits, particularly for struggling readers and in sustained supported-reading programs, whereas others found little or no difference between RWL and RO. Verlaan and Ortlieb (2012) and Chang and Millett (2015) suggest that RWL can support comprehension under favorable conditions, but Holmes and Allison (1986) showed that the effect may differ by reader profile, with good readers sometimes disadvantaged by concurrent audio. Likewise, Tragant Mestres et al. (2019) and Eppard et al. (2020) suggest that RWL may support vocabulary, engagement, or reading experience without necessarily producing higher comprehension scores, while Liu et al. (2019) and Serrano and Pellicer-Sánchez (2019) show that any advantage may be task-specific. Overall, RWL appears to support the reading process more reliably than it guarantees higher reading-comprehension performance.

Listening comprehension shows the clearest and most consistent advantage for RWL. Across extended interventions and controlled comparisons, RWL more often outperformed RO and LO on listening outcomes than any other modality. This pattern appears across different proficiency levels and settings, from low-to-intermediate learners to some more advanced cohorts. The convergence of findings across Chang and Millett (2014), Chang et al. (2019), Kartal and Simsek (2017), Milliner (2019), Hartshorn and Stephens (2023), and Kim (2021) suggests that textual support helps learners stabilize sound-meaning mappings while listening, especially when phonological decoding and lexical access are still developing.

This pattern further suggests that RWL may function as a developmental scaffold rather than as a permanent endpoint. By reducing uncertainty during listening and reinforcing form-meaning

connections, textual support can strengthen comprehension for learners with less developed listening skills. As proficiency grows, however, the role of RWL may shift from primary support to selective assistance, helping learners transition toward more autonomous listening.

Insights From Cognitive Processing: Mechanisms Behind Modality Effects

Eye-tracking and cognitive-load findings help explain why comprehension outcomes diverge across studies. As shown in Table 6, reading-while-listening (RWL) changes how learners allocate attention during comprehension, but the effects of that redistribution depend on learner profile and task conditions. In some contexts, RWL appears to reduce inefficient decoding and support integrative processing; in others, it shifts attention without producing a clear comprehension advantage.

Table 6. Representative Quantitative Signals From Cognitive-Processing Studies (RQ1)

Study / context	Learner-task profile	Compact quantitative signal	Process pattern
Pellicer-Sánchez et al. (2020); Serrano and Pellicer-Sánchez (2019)	Young learners reading multimodal texts	Less text fixation under RWL ($d = -0.47$) and more integrative saccades ($d = 4.19$); longer text dwell time negatively related to comprehension	Integration support
Pellicer-Sánchez et al. (2021)	Adult L1/L2 readers	Greater text-processing time positively related to L2 comprehension; interaction effect $d = 0.58$	Learner-dependent
Tragant Mestres et al. (2019)	Young learners reading without complementary visuals	No clear differences in eye movements or comprehension across modes	Visual-support dependent
Yang et al. (2022)	Grade 3 EFL learners under audio assistance	Mode shifted visual attention, $F = 10.523$, $p < .001$, $\eta^2 = .086$, but no main comprehension advantage	Attention shift only

For young learners, process evidence suggests that RWL shifts attention away from prolonged text fixation and toward integrative processing across available information sources. In these studies, reduced fixation on text does not necessarily indicate weaker processing. Rather, it appears to reflect reduced decoding burden and greater coordination across text, audio, and visual input. This helps explain why RWL may function as a scaffold for younger or struggling readers, especially when multimodal information is available to support meaning construction.

For adults, the optimal attention pattern appears less uniform. Pellicer-Sánchez et al. (2021) suggests that greater text-focused processing may support L2 comprehension, indicating that more proficient or mature readers do not always benefit from reduced attention to print. At the same time, Tragant Mestres et al. (2019) found no clear differences in eye movements or comprehension when complementary visuals were absent, suggesting that bimodal advantages may depend less on audio-text synchrony alone and more on how the task environment distributes attention across sources of information.

Yang et al. (2022) further reinforces this boundary-condition account. Although audio support significantly shifted visual attention, it did not produce a main comprehension advantage. This suggests that auditory input does not automatically reduce cognitive load in a beneficial way.

Under demanding conditions, it may instead redirect attention without improving understanding. Taken together, these process findings support dual-channel accounts only when the additional input is complementary rather than redundant, consistent with Moreno and Mayer's (2002) framework.

RQ2: What Moderating Factors, Apart From Input Modes, Affect Listening and Reading Performance?

Four moderator domains consistently shaped receptive-skill outcomes across the included studies: learner characteristics, instructional design, cognitive-affective factors, and environmental context (Table 7). Cognitive Load Theory provides a useful lens for integrating these domains because the same modality can reduce extraneous load for one learner while increasing it for another (Sweller et al., 2011). In practical terms, input mode (RO, LO, RWL) sets the format of exposure, but these moderators largely determine how efficiently learners process that input and how reliably comprehension develops.

Table 7. Synthesis of Moderating Factors for Receptive-Skill Performance (RQ2)

Moderator domain	Key factor	Main impact	Representative studies
Learner characteristics	Proficiency level	Low proficiency: RWL can reduce decoding demands; high proficiency: RWL may add extraneous load, and RO may be equally effective or superior.	Aka (2024); Diao and Sweller (2007); Verlaan and Ortlieb (2012)
Learner characteristics	Cognitive strategies and aptitude	Strategic behaviors can help manage load; vocabulary knowledge and aptitude shape attention during RWL.	Nakashima et al. (2018); Serrano and Pellicer-Sánchez (2019)
Instructional design	Exposure duration and practice	Sustained practice supports automatization; short exposure may not overcome initial task demands.	Chang and Millett (2014, 2015); Chang et al. (2019); Kim (2021)
Instructional design	Material design and scaffolding	Graded materials, repetition, and pacing can support processing while limiting overload.	Chang (2009, 2011); Serrano and Pellicer-Sánchez (2019)
Cognitive-affective factors	Motivation and engagement	Higher motivation can increase persistence and productive effort allocation across modes	Kartal and Simsek (2017); Tragant Mestres et al. (2019)
Cognitive-affective factors	Anxiety and working-memory constraints	Anxiety and limited working memory increase extraneous load and may make RWL burdensome.	Hui (2024); Diao and Sweller (2007)
Environmental context	Pedagogical support	Teacher guidance can prevent over-reliance on scaffolds and support strategic shifting between modes.	Hartshorn and Stephens (2023)
Environmental context	Background knowledge and resources	Prior knowledge reduces intrinsic load; limited resources may constrain the benefits of modality support.	Eppard et al. (2020)

Learner Characteristics: Determining the Need for Scaffolding

Proficiency most strongly predicts whether RWL helps or hinders performance. As summarized in Table 8, lower-proficiency learners generally benefited from textual support, whereas more proficient learners showed smaller gains or no clear advantage. This pattern suggests that RWL is most useful when it reduces decoding demands and stabilizes comprehension.

Table 8. Representative Quantitative Signals for Learner Characteristics (RQ2)

Study	Learner group	Compact quantitative signal	Overall pattern
Aka (2024)	Low-proficiency subgroup	RWL $M = 8.00$ vs. RO $M = 7.00$; $p = .07$	Trend favored RWL.
Kartal and Simsek (2017)	Lower-proficiency university learners	RWL $M = 6.53$ vs. silent reading $M = 4.94$; $d = 0.71$	RWL advantage
Diao and Sweller (2007)	More proficient EFL learners	RO > RWL on recall and lexical tasks; $\omega^2 = 0.07-0.37$	RO advantage

For more proficient learners, the same audio can become redundant and induce extraneous load. Instead of assisting comprehension, concurrent audio-text input may compete for attention and weaken recall-oriented processing. The clearest comparisons therefore support the cognitive-overload hypothesis: scaffolding becomes less helpful once learners can decode and integrate text more independently.

Beyond proficiency, cognitive strategies and individual differences further shape how learners engage with input. Strategic behaviors such as note-taking and mental translation help learners manage cognitive load and direct attention during RWL (Nakashima et al., 2018). Age-related cognitive flexibility may also make younger learners more receptive to dual-channel input (Tragant Mestres et al., 2019). Crucially, eye-tracking research confirms these characteristics manifest in real-time processing: individual differences in vocabulary knowledge directly predict how learners allocate visual attention between text and images during RWL, which ultimately governs comprehension success (Serrano & Pellicer-Sánchez, 2019).

Instructional Design: Orchestrating Cognitive Load

The structure of the learning intervention plays a crucial role in managing cognitive load. Table 9 highlights the clearest quantitative signals showing that longer, structured interventions tend to produce stronger gains than brief exposure alone. Sustained practice appears to help learners automate lower-level processing and free working memory for higher-level comprehension.

Table 9. Representative Quantitative Signals for Instructional Design (RQ2)

Study	Design feature	Compact quantitative signal	Overall pattern
Chang and Millett (2015)	26-week audio-assisted reading	Reading comprehension $d = 1.55$ at the post-test; delayed rate gains remained strong.	Strong sustained gains
Chang and Millett (2014); Chang et al. (2019)	13-39 week supported listening	Large repeated listening gains under RWL across practiced and unpracticed texts	Longer exposure supported transfer.
Kim (2021)	5-week aligned multimodal program	$\eta^2 = .25$ for listening; $\eta^2 = .17$ for reading	Shorter aligned programs can still work.

Shorter interventions can still succeed when the design is coherent and the materials align closely with the target assessments, but they do not guarantee gains. Overall, the evidence suggests that duration, repetition, pacing, and material difficulty work together to determine whether RWL acts as helpful support or as an added burden.

Equally important is the design of instructional materials. The use of familiar, graded readers ensures learners are not overwhelmed by intrinsic load while promoting germane processing that supports language acquisition (Serrano & Pellicer-Sánchez, 2019). The strategic use of repetition (Chang, 2009) and scaffolding techniques (Chang, 2011) helps reduce extraneous cognitive load, particularly for lower-proficiency learners, guiding them toward independent processing.

Cognitive–Affective Factors: The Engine of Engagement

Affective factors, such as motivation, can enhance germane load. The engaging nature of audiobooks has been linked to increased motivation and willingness to persist, suggesting that enjoyment and interest may offset some of the perceived difficulty of RWL for certain learners.

However, cognitive and affective challenges can induce extraneous load. The core trade-off of RWL is clear: it can either reduce extraneous load as a supportive scaffold or increase it through overload. Eye-tracking studies show that successful comprehension in RWL correlates with increased attention to images, indicating efficient, meaning-focused processing (Pellicer-Sánchez et al., 2020). Conversely, RWL also led to higher subjective mental load and significantly poorer comprehension, quantifying its potential downside (Diao & Sweller, 2007). Factors such as anxiety (Hui, 2024) and limited working memory further constrain cognitive resources, highlighting the importance of effective cognitive load management.

Environmental and Contextual Influences: Setting the Stage

The broader learning context constrains and shapes how all other factors operate. Print-poor environments can create foundational deficits in reading skills that no input mode can instantly overcome (Eppard et al., 2020). The presence of background knowledge provides crucial schematic support that reduces the intrinsic load of a text, while unfamiliar content increases it.

Most importantly, the pedagogical environment determines whether technology is used effectively. The mere presence of a transcript, a form of RWL, is not enough; structured teacher guidance on how to use it prevents over-reliance and fosters balanced skill development (Hartshorn & Stephens, 2023). This highlights that the teacher’s role is to curate and mediate input modes based on the other moderating factors at play, rather than treating modality as a fixed setting.

RQ3: How Do English Language Learners Perceive Different Modes of Input in Reading and Listening Tasks?

Learner perceptions cluster into three consistent themes rather than a single preference pattern. Across surveys, interviews, and open-ended responses from the included studies, learners evaluated RO, LO, and reading-while-listening (RWL) in ways that reflect (a) the value of scaffolding, (b) perceived cognitive load, and (c) proficiency-linked needs. The thematic synthesis therefore moves beyond the question of which mode learners prefer to explain why particular modes are perceived as helpful or difficult (Table 10).

Table 10. Thematic Synthesis of Learner Perceptions (RQ3)

Analytical theme	Supporting descriptive themes	Illustrative learner quotes
Bimodal input as a strategic scaffold	Enhances comprehensibility and bridges text-sound gaps; builds confidence and foundational skills; increases engagement and reduces anxiety	“I can understand the text better when I listen and read at the same time” (Kartal and Simsek, 2017, p. 108); “The audio support helped me with decoding and understanding the content better... It increased my confidence in reading” (Eppard et al., 2020, p. 759).
The cognitive toll of multimodal processing	Causes feelings of overload and distraction; leads to strategic preference for simpler, unimodal input	“It was difficult to concentrate on both the text and the audio at the same time” (Diao and Sweller, 2007, p. 85).
The proficiency divide in unimodal preferences	Lower proficiency: anxiety without textual support in LO and reliance on RWL; higher proficiency: value for autonomy and control in RO and viewing LO as an authentic challenge.	“I cannot check the spelling of words I don’t know when I only listen” (Nakashima et al., 2018, p. 58); “I prefer reading only because I can re-read at my own pace without being rushed by an audio track” (Aka, 2024, p. 158).

Analytical Theme 1: Bimodal Input as a Strategic Scaffold for Comprehension and Confidence

RWL is consistently perceived as a valuable scaffold, particularly at lower and intermediate proficiency levels (Aka, 2024; Eppard et al., 2020; Kartal & Simsek, 2017). Learner feedback indicates that this preference is strategic, rooted in the mode’s ability to actively build skills and reduce learning-related anxiety.

Enhanced comprehensibility through synchronized written and spoken input emerged as the primary perceived benefit. Learners reported that dual-channel processing clarified meaning and reinforced the connection between orthography and phonology. As one participant explained, “It was helpful to see the words while hearing them because sometimes I mishear words” (Hui, 2024, p. 192). This perception was echoed across studies, with learners reporting that they understood texts better when listening and reading at the same time (Kartal & Simsek, 2017).

Beyond comprehension, learners viewed RWL as a direct support for building foundational skills and confidence. The audio component was frequently described as a pronunciation model and as a support that reduced the cognitive burden of decoding (Kartal & Simsek, 2017). This scaffolding appeared especially important for struggling readers and lower-proficiency learners, who described difficult passages as easier to follow when audio accompanied text (Aka, 2024). These perceived benefits also translated into stronger engagement, with learners reporting that RWL made stories more enjoyable (Tragant Mestres et al., 2019).

Analytical Theme 2: The Cognitive Toll of Multimodal Processing

Despite its benefits, a clear counter-pattern emerged regarding the cognitive demands of RWL. Learner perceptions explicitly described the challenge of processing concurrent information streams, providing support for the limited-capacity principle in multimedia learning (Diao & Sweller, 2007).

Reports of poorer comprehension alongside higher subjective mental load under RWL were noted by Diao and Sweller (2007). One learner's comment, "It was difficult to concentrate on both the text and the audio at the same time," captures the experience of overload that can offset the advantages of bimodal input. Quantitative process evidence also aligns with these perceptions, as eye-tracking findings suggest that audio-assisted conditions can alter attentional allocation in ways that are not always beneficial for comprehension (Yang et al., 2022).

Analytical Theme 3: The Proficiency Divide in Unimodal Preferences

Preferences for unimodal input, whether LO or RO, are closely tied to proficiency level and reflect learners' attempts to adapt modality choice to their own needs (Aka, 2024; Hartshorn & Stephens, 2023; Nakashima et al., 2018).

LO elicited contrasting perceptions depending on proficiency. Higher-proficiency learners tended to view LO as a more authentic and less stressful challenge, whereas lower-proficiency learners expressed anxiety when textual support was absent (Nakashima et al., 2018). Their concern centered on not being able to verify unfamiliar words or spellings while listening, which highlights continued dependence on print support at earlier stages of development.

Perceptions of RO were similarly proficiency-dependent. Although some learners reported little difference between RO and RWL in comprehension (Hui, 2024), more advanced learners especially valued the control and autonomy that RO provides. They emphasized being able to re-read difficult passages at their own pace without being constrained by an audio track (Aka, 2024). This preference for control also appeared in other studies in which learners valued the ability to return to difficult parts of the text without needing to rewind audio (Hartshorn & Stephens, 2023).

Learner perceptions reflect a trade-off between support and load, moderated by proficiency. RWL is valued as a scaffold for comprehension, confidence, and form-sound mapping, especially among lower-proficiency learners (Aka, 2024; Eppard et al., 2020; Kartal & Simsek, 2017). At the same time, it can feel distracting or overwhelming when attention is split across channels (Diao & Sweller, 2007).

Unimodal preferences follow a clear proficiency pattern: lower-proficiency learners tend to rely on RWL to reduce anxiety and support decoding, whereas higher-proficiency learners increasingly value RO for autonomy and may view LO as an authentic target mode (Aka, 2024; Nakashima et al., 2018). Pedagogically, these findings argue for adaptive modality use rather than fixed input policies. Instruction is likely to be most effective when modality choices are aligned with learner proficiency and task goals and when learners are guided to select a mode that balances scaffolding with manageable cognitive load.

RQ4: What Are the Potential Research Directions for the Use of Different Modes of Input in the ESL Domain?

Future research must move beyond documenting outcomes to strengthening causal evidence and explaining underlying mechanisms. Although findings across the 24 included studies are promising, they remain conditional. The evidence base is also constrained by study quality, with 68% of the corpus rated as having moderate-to-high risk of bias (Table 3). These methodological limitations, particularly in between-subject designs, point to three priorities for future research: improving methodological rigor, broadening the scope to underrepresented contexts and learners, and deepening explanatory power through process-oriented methods.

Prioritizing Methodological Rigor: A Response to Elevated Risk of Bias

The most urgent priority is the generation of higher-quality primary studies. To reduce bias and improve interpretability, future investigations should prioritize high-fidelity experimental designs, including more randomized controlled trials with adequate sample sizes, active control conditions, and blinded outcome assessment where feasible. The sole randomized controlled trial in the current corpus, Friedland et al. (2017), yielded only borderline or negligible evidence of reading-fluency improvement (ANCOVA $p = .051$; $d = 0.05$), illustrating both the value of such designs and the need for longer, better-powered trials.

Longitudinal and ecologically valid designs are also needed. Classroom-based studies that track development over a full term or academic year are essential for determining whether modality effects persist, fade, or shift with proficiency. Existing longitudinal work spanning 13 to 39 weeks has shown sustained gains for RWL groups on both practiced and unpracticed listening texts (Chang & Millett, 2014; Chang et al., 2019), but these findings require replication through stronger designs.

A further methodological priority is the use of standardized and comparable measures. Wider adoption of validated assessments for reading and listening comprehension would reduce measurement bias and facilitate more meaningful cross-study synthesis. At present, the heterogeneity of outcome measures across the corpus, including TOEIC, GLCSS, and researcher-designed instruments, continues to limit quantitative pooling and direct comparison.

Expanding the Investigative Landscape: Moving Beyond the Review's Parameters

A second priority is to extend the evidence base into contexts that more closely reflect the diversity and complexity of real-world ESL learning. Because this review deliberately bounded its scope to isolate the effect of RWL, its conclusions are strongest within those parameters but less informative for settings that fall outside them. Future research should therefore broaden the evidence base in ways that improve ecological relevance without losing conceptual clarity.

One important direction is greater attention to underrepresented learner populations and instructional settings. More research is needed with school-aged learners in public-school contexts, alongside clearer reporting of first-language backgrounds and more consistent proficiency benchmarks. Although studies involving primary students (Holmes & Allison, 1986; Rasinski, 1990; Tragant Mestres et al., 2019) suggest developmental differences in modality effects, systematic investigation across age groups remains limited. Without broader sampling, it remains difficult to determine whether current findings generalize beyond the relatively narrow populations represented in the existing literature.

A second direction is to reconceptualize listening development as a longitudinal instructional pathway rather than a fixed comparison condition. Future studies should test sequencing models in which RWL functions strategically as a scaffold toward listening-only proficiency, rather than treating modalities as stable alternatives. Large effects for listening comprehension ($d = 1.50$ - 1.73 in Chang & Millett, 2014; $d = 1.35$ - 2.27 in Chang et al., 2019), together with the pooled estimate reported by Clinton-Lisell (2023), suggest that RWL may serve as a preparatory phase. However, the field still lacks studies that explicitly examine when scaffolded support should be reduced, maintained, or withdrawn as learners become more capable listeners.

A third direction is to examine modality use in digital environments that are more representative of contemporary language learning. Future studies should move beyond the static RO/LO/RWL paradigm to reflect how learners now encounter input through digital platforms. This includes AI-driven personalization, in which adaptive systems dynamically tailor modality choice to learner profile and real-time task demands, as well as integrated multimedia environments involving video, subtitles, interactive transcripts, and educational games. In such settings, the instructional question is no longer simply whether RWL is more effective than RO or LO, but how multiple forms of support interact under authentic learning conditions. This shift would allow future research to move from isolated modality contrasts toward a more realistic account of how learners navigate multimodal input in practice.

Deepening Explanatory Power Through Process-Oriented Inquiry

A third priority is to explain why modality effects occur and when they reverse. Heavy reliance on comprehension tests limits insight into the cognitive mechanisms that produce these conditional outcomes. Process-oriented methods are therefore essential, and the current evidence already demonstrates their value.

Eye-tracking research has shown that RWL shifts attention allocation, with learners spending less time fixating on text and more time integrating information from images (Pellicer-Sánchez et al., 2020). In addition, longer dwell time on text has been negatively correlated with comprehension ($r \approx -.47$; Serrano & Pellicer-Sánchez, 2019), suggesting that struggling readers may over-invest in inefficient decoding at the expense of meaning integration.

Think-aloud protocols also offer valuable insight into strategy use during multimodal processing. Participants have reported using text to disambiguate homophones and resolve lexical uncertainty (Clinton-Lisell, 2023), indicating that textual support may reduce uncertainty during listening and comprehension in ways that conventional outcome measures cannot fully capture.

Computational modeling provides a further explanatory layer. Yang et al. (2022) showed that auditory input may not itself impose extraneous cognitive load, yet it can still compromise comprehension by diverting visual attention under high-load conditions. This finding helps clarify the boundary conditions under which RWL is beneficial and when it may become counterproductive.

Future research should therefore place greater emphasis on psychophysiological and process measures, including eye-tracking, pupillometry, and EEG, to capture attention allocation, mental effort, and cognitive load in real time. It should also expand the use of robust qualitative inquiry. More theoretically grounded qualitative studies using structured interviews and rigorous thematic analysis are needed to explain how learners make modality choices, experience cognitive load, and develop strategy use across proficiency levels. As one learner explained, “It was difficult to concentrate on both the text and the audio at the same time” (Diao & Sweller, 2007, p. 85). Such comments mirror cognitive-load theory, but they require deeper exploration across more diverse populations and contexts.

Limitations of the Review

The conclusions of this systematic review should be interpreted within its methodological boundaries. Several limitations constrain the generalizability and strength of the synthesized evidence.

First, the review was deliberately narrow in scope. The exclusion of studies involving subtitles, video, or on-screen text was necessary to isolate the effect of synchronous audio-text input in reading-while-listening (RWL), but it also means that the findings cannot be generalized to other prevalent multimodal formats in digital learning. In addition, the focus on global reading and listening comprehension excluded studies examining only discrete vocabulary gains, which represent an important but distinct outcome domain. Finally, restricting the review to English-language studies of ESL/EFL learners increased sample homogeneity but may also have excluded relevant insights from second language acquisition research conducted in other languages.

A second limitation concerns the quality of the primary studies. The strength of the synthesized evidence is inherently constrained by the methodological rigor of the studies included. Low-risk studies generally reported more moderate and conditional effects, whereas higher-risk designs often produced larger but less reliable outcomes. This pattern suggests that the findings should be interpreted with caution and reinforces the need for more robust primary research in this area.

A third limitation lies in the synthesis itself. The heterogeneity of populations, interventions, and outcome measures across the corpus reduced the feasibility of a comprehensive meta-analysis and necessitated a narrative synthesis approach. Although this approach preserved the nuance of individual study findings, it also limited the precision with which effects could be pooled and compared across studies.

Notwithstanding these limitations, the review provides a comprehensive map of the available evidence and offers a theoretically grounded framework, the Proficiency-Scaffolding Model, to guide future inquiry and instructional practice.

Conclusion

This systematic review demonstrates that the influence of RO, LO, and reading-while-listening on ESL/EFL receptive skills is not absolute but is strongly moderated by learner proficiency, target skill, and cognitive load. The integrated evidence supports a Proficiency-Scaffolding Model in which RWL functions as a valuable scaffold for particular learner groups and instructional purposes, while its advantages diminish as learner capabilities develop.

For lower-proficiency and struggling readers, RWL offers important compensatory support. Struggling readers showed substantial gains ($d = 0.99$; Verlaan & Ortlieb, 2012), and low-proficiency Japanese high-school learners also showed a trend toward better comprehension under RWL ($p = .07$; Aka, 2024). Across the broader literature, the pooled estimate for RWL relative to RO remains modest ($g = 0.18$; Clinton-Lisell, 2023), but listening comprehension shows robust effects across multiple contexts ($d = 1.50$ - 2.27 ; Chang & Millett, 2014; Chang et al., 2019). For reading fluency, RWL also appears consistently beneficial through audio pacing and prosodic modeling.

For advanced learners, however, the scaffolding function of RWL can become redundant. In such cases, RO may outperform RWL on text comprehension, with small-to-moderate effects ($\omega^2 = 0.07$ - 0.37 ; Diao & Sweller, 2007). This contrast reinforces the central claim of the review: RWL is not inherently superior, but conditionally effective.

Four moderator domains systematically shape these outcomes: learner characteristics, especially proficiency; instructional design, including duration, materials, and scaffolding; cognitive-affective factors, such as motivation, anxiety, and working memory; and environmental context, including pedagogical support and background knowledge. These factors operate together as a

dynamic system, determining whether RWL reduces cognitive load and supports comprehension or increases redundancy and processing burden.

Learner perceptions reflect the same trade-off. Across the qualitative evidence, learners consistently valued RWL as a scaffold for comprehension, confidence, and form-sound mapping, particularly at lower proficiency levels. At the same time, they also reported cognitive costs: RWL could feel distracting or overwhelming when attention was split across channels. Unimodal preferences followed a clear proficiency pattern, with lower-proficiency learners relying more on RWL for support and higher-proficiency learners increasingly valuing RO for autonomy and LO as an authentic challenge (Nakashima et al., 2018).

Pedagogically, these findings support adaptive, proficiency-calibrated modality use rather than fixed input policies. For beginning and lower-proficiency learners, RWL may serve as a primary scaffold that reduces decoding burden and builds confidence. For intermediate learners, a more balanced approach may be appropriate, with RWL used for unfamiliar or demanding content and RO for consolidation and independent practice. For advanced learners, LO may be emphasized for authentic auditory processing, with RWL reserved for strategically demanding texts. In listening development specifically, RWL may function best as a preparatory phase that helps learners build phonological and lexical foundations before transitioning toward more autonomous listening.

Theoretically, this review advances a Proficiency-Scaffolding Model that helps resolve longstanding contradictions in the literature by showing that the efficacy of RWL is conditional rather than inherent. The same modality that reduces decoding load for beginners may create redundancy for advanced learners. In this way, the review extends Cognitive Load Theory by clarifying the boundary conditions of redundancy in L2 contexts and refines Multimedia Learning Theory by showing when dual-channel input is complementary and when it becomes intrusive.

The central question, therefore, is not whether RWL is better, but for whom, for what skill, and under what conditions it is most beneficial. This more nuanced understanding is essential for moving beyond one-size-fits-all recommendations toward targeted, evidence-informed pedagogical practice that supports learners across the proficiency spectrum.

About the Authors

Padma Lamo is a PhD scholar in Applied Linguistics at the Indian Institute of Technology, Bhubaneswar. Her research interests include English language teaching, second language listening, global Englishes, accent perception, neurodivergent learners, and technology-mediated language learning. ORCID ID: 0009-0001-3223-8799

Saleem Mohd Nasim is an Assistant Professor of TESOL/ELT with over 15 years of experience in Saudi Arabia and India. He specializes in language pedagogy, assessment, EAP/ESP, and educational technology. His research includes Scopus/WoS publications, peer review, and editorial service in applied linguistics and digital learning. ORCID ID: 0000-0001-5110-1547

Rajakumar Guduru is an Assistant Professor of English Linguistics at the Indian Institute of Technology, Bhubaneswar. His research interests include ESL vocabulary development, cognitive reading skills, second language acquisition, teacher education, communication skills, and technology-mediated language learning. ORCID ID: 0000-0002-0928-1166

To Cite this Article

Lamo, P., Nasim, S. M. & Guduru, R. (2026). Modality principle in learning L2: A systematic review of how proficiency moderates the efficacy of reading-while-listening on receptive skills. *Teaching English as a Second Language Electronic (TESL-EJ)*, 30(2).
<https://doi.org/10.55593/ej.30118a2>

References

- Adesope, O. O., & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology*, 104(1), 250–263.
<https://doi.org/10.1037/a0026147>
- Aka, N. (2024). Effects of reading-while-listening and reading-only on reading comprehension of Japanese high school EFL learners. *The Journal of Asia TEFL*, 21(1), 144–162.
<https://doi.org/10.18823/asiatefl.2024.21.1.9.161>
- Chang, A. C.-S. (2009). Gains to L2 listeners from reading while listening vs. *listening only in comprehending short stories*. *System*, 37(4), 652–663.
<https://doi.org/10.1016/j.system.2009.09.009>
- Chang, A. C.-S. (2011). The effect of reading while listening to audiobooks: Listening fluency and vocabulary gain. *Asian Journal of English Language Teaching*, 21, 43–64.
<https://www.cuhk.edu.hk/ajelt/vol21/abstract/a03.pdf>
- Chang, A. C.-S., & Millett, S. (2014). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal*, 68(1), 31–40.
<https://doi.org/10.1093/elt/cct052>
- Chang, A. C.-S., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System*, 52, 91–102.
<https://doi.org/10.1016/j.system.2015.05.003>
- Chang, A. C.-S., Millett, S., & Renandya, W. A. (2019). Developing listening fluency through supported extensive listening practice. *RELC Journal*, 50(3), 422–438.
<https://doi.org/10.1177/0033688217751468>
- Cheetham, D. (2019). Multi-modal language input: A learned superadditive effect. *Applied Linguistics Review*, 10(2), 179–200. <https://doi.org/10.1515/applirev-2017-0036>
- Clinton-Lisell, V. (2022). Listening ears or reading eyes: A meta-analysis of reading and listening comprehension comparisons. *Review of Educational Research*, 92(4), 543–582.
<https://doi.org/10.3102/00346543211060871>
- Clinton-Lisell, V. (2023). Does reading while listening to text improve comprehension compared to reading-only? A systematic review and meta-analysis. *Educational Research: Theory and Practice*, 34(3), 133–155. <https://files.eric.ed.gov/fulltext/EJ1403866.pdf>
- Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and Instruction*, 17(1), 78–88. <https://doi.org/10.1016/j.learninstruc.2006.11.007>

- Eppard, J., Baroudi, S., & Rochdi, A. (2020). A case study on improving reading fluency at a university in the UAE. *International Journal of Instruction*, 13(1), 747–766.
<https://doi.org/10.29333/iji.2020.13148a>
- Friedland, A., Gilman, M., Johnson, M., & Demeke, A. (2017). Does reading-while-listening enhance students' reading fluency? Preliminary results from school experiments in rural Uganda. *Journal of Education and Practice*, 8(7), 82–95.
<https://iiste.org/Journals/index.php/JEP/article/view/36012/37005>
- Hartshorn, K. J., & Stephens, C. (2023). The effects of transcript use on advanced ESL listening comprehension. *International Journal of TESOL Studies*, 5(4), 45–62.
<https://doi.org/10.58304/ijts.20230404>
- Holmes, B. C., & Allison, R. W. (1986). The effect of four modes of reading on children's comprehension. *Reading Research and Instruction*, 25(1), 9–20.
<https://doi.org/10.1080/19388078509557854>
- Hui, B. (2024). Scaffolding comprehension with reading while listening and the role of reading speed and text complexity. *The Modern Language Journal*, 108(1), 183–200.
<https://doi.org/10.1111/modl.12905>
- Joanna Briggs Institute. (2017). *Checklist for quasi-experimental studies (non-randomized experimental studies) [Critical appraisal tool]*. https://jbi.global/sites/default/files/2019-05/JBI_Quasi-Experimental_Appraisal_Tool2017_0.pdf
- Kartal, G., & Simsek, H. (2017). The effects of audiobooks on EFL students' listening comprehension. *The Reading Matrix: An International Online Journal*, 17(1), 112–123.
<https://www.readingmatrix.com/files/16-7w4b733r.pdf>
- Kim, N.-Y. (2021). E-book, audiobook, or e-audiobook: The effects of multiple modalities on EFL comprehension. *English Teaching*, 76(4), 33–52.
<https://doi.org/10.15858/engtea.76.4.202112.33>
- Liu, H., Cao, S., & Wu, S. (2019). An experimental comparison on reading comprehension effect of visual, audio, and dual channels. *Proceedings of the Association for Information Science and Technology*, 56(1), 716–718. <https://doi.org/10.1002/pra2.148>
- Milliner, B. (2019). Comparing extensive reading to extensive reading-while-listening on smartphones: Impacts on listening and reading performance for beginning students. *The Reading Matrix: An International Online Journal*, 19(1), 1–19. <https://www.readingmatrix.com/files/20-81br6g10.pdf>
- Mohsen, M. A. (2016). The use of help options in multimedia listening environments to aid language learning: A review. *British Journal of Educational Technology*, 47(6), 1232–1242.
<https://doi.org/10.1111/bjet.12305>
- Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, 94(1), 156–163.
<https://doi.org/10.1037/0022-0663.94.1.156>
- Nakashima, K., Stephens, M., & Kamata, S. (2018). The interplay of silent reading, reading-while-listening, and LO. *The Reading Matrix: An International Online Journal*, 17(1), 51–63.
<http://www.readingmatrix.com/files/18-47992957.pdf>

- Nasim, S. M. (2022). Metacognitive listening comprehension strategies of Arab English language learners. *Education Research International*, 2022, Article 9916727. <https://doi.org/10.1155/2022/9916727>
- Nasim, S. M., Mohamed, S. M. S., Anwar, M. N., Ishtiaq, M., & Mujeeba, S. (2024). Assessing the pedagogical effectiveness of the web-based cooperative integrated reading composition (CIRC) technique to enhance EFL reading comprehension skills. *Cogent Education*, 11(1), Article 2401667. <https://doi.org/10.1080/2331186X.2024.2401667>
- Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, 27(1), 94–106. <https://doi.org/10.2307/747835>
- 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>
- Pellicer-Sánchez, A., Conklin, K., Rodgers, M. P. H., & Parente, F. (2021). The effect of auditory input on multimodal reading comprehension: An examination of adult readers' eye movements. *The Modern Language Journal*, 105(4), 936–956. <https://doi.org/10.1111/modl.12743>
- Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, 42(3), 577–598. <https://doi.org/10.1017/S0272263120000091>
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme (Final report)*. ESRC Methods Programme.
- Rasinski, T. V. (1990). Effects of repeated reading and listening-while-reading on reading fluency. *The Journal of Educational Research*, 83(3), 147–151. <https://doi.org/10.1080/00220671.1990.10885946>
- Serrano, R., & Pellicer-Sánchez, A. (2019). Young L2 learners' online processing of information in a graded reader during reading-only and reading-while-listening conditions: A study of eye-movements. *Applied Linguistics Review*, 10(1), 1–22. <https://doi.org/10.1515/applirev-2018-0102>
- Shaojie, T., Samad, A. A., & Ismail, L. (2022). Systematic literature review on audio-visual multimodal input in listening comprehension. *Frontiers in Psychology*, 13, 980133. <https://doi.org/10.3389/fpsyg.2022.980133>
- Singh, A., & Alexander, P. A. (2022). Audiobooks, print, and comprehension: What we know and what we need to know. *Educational Psychology Review*, 34(2), 677–715. <https://doi.org/10.1007/s10648-021-09653-2>
- Stepien-Bernabe, N. N., Lei, D., McKerracher, A., & Orel-Bixler, D. (2019). The impact of presentation mode and technology on reading comprehension among blind and sighted

individuals. *Optometry and Vision Science*, 96(5), 354–361.
<https://doi.org/10.1097/OPX.0000000000001373>

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Altering element interactivity and intrinsic cognitive load. In *Cognitive load theory* (Vol. 1, pp. 203–218). Springer.
https://doi.org/10.1007/978-1-4419-8126-4_16

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, Article 45.
<https://doi.org/10.1186/1471-2288-8-45>

Tragant Mestres, E., Llanes Baró, À., & Pinyana Garriga, À. (2019). Linguistic and non linguistic outcomes of a reading while listening program for young learners of English. *Reading and Writing*, 32(3), 819–838. <https://doi.org/10.1007/s11145-018-9886-x>

Verlaan, W., & Ortlieb, E. (2012). Reading while listening: Improving struggling adolescent readers' comprehension through the use of digital audio recordings. In J. Cassidy, S. Grote-Garcia, E. Martinez, & R. Garcia (Eds.), *What's hot in literacy 2012* (pp. 30–36). Texas Association for Literacy Education. <http://www.texasreaders.org/first-yearbook>

Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *Journal of Learning Disabilities*, 51(1), 73–84.
<https://doi.org/10.1177/0022219416688170>

Yang, J., Qi, X., Wang, L., Sun, B., & Zheng, M. (2022). A reading model of young EFL learners regarding attention, cognitive-load, and auditory-assistance. *The Journal of Educational Research*, 115(1), 51–63. <https://doi.org/10.1080/00220671.2022.2027327>

Appendix Table S1. Detailed Quantitative Findings by Study

This supplementary table preserves study-level quantitative detail that has been removed from the streamlined main-manuscript Table 2 to improve readability and keep the journal table within page limits.

Study	Outcome domain	Detailed quantitative findings	Dominant quantitative signal
Aka (2024)	Comprehension	Overall comprehension: no significant difference ($t \approx 0.37-0.72$, $p > .05$, $d \approx 0.07-0.09$). Low-proficiency subgroup: RWL M = 8.00 vs. RO M = 7.00; interaction $p = .07$.	Trivial overall effect; possible low-proficiency advantage
Chang (2009)	Listening/comprehension	Input mode effect: $F(1,163) = 10.43$, $p = .001$, $\eta^2 = .06$. RWL mean = 72%; LO mean = 62%. 93% perceived better comprehension with RWL.	Moderate RWL advantage
Chang (2011)	Listening/vocabulary	Listening dictation: $t(17) = 3.53$, $p < .005$, $d = 1.54$. Vocabulary gain: 566 words vs. 123 in control.	Large RWL advantage
Chang & Millett (2014)	Listening fluency	RWL $d = 1.50-1.73$; LO $d = 0.67-1.02$; RO $d = -0.20$ to 0.16.	Large and most consistent advantage for RWL
Chang & Millett (2015)	Reading rate/comprehension	Reading rate: $t(62) = 3.92$, $p < .001$, $d = 1.00$; delayed $t(62) = 5.39$, $p < .001$, $d = 1.35$. Reading comprehension: $t(62) = 6.13$, $p < .001$, $d = 1.55$.	Large assisted-reading advantage
Chang et al. (2019)	Listening fluency	Practised texts: RLL $d = 1.78-1.96$; LO $d = 0.75-1.35$; RO $d = -1.17$. Unpractised texts: RLL $d = 1.35-2.27$; LO $d = 0.00-2.16$; RO $d = -0.93$.	Large overall advantage for RWL/RLL
Diao & Sweller (2007)	Reading comprehension	Lexical comprehension: $t(58) = 2.05$, $p = .05$, $\omega^2 = .37$. Free recall: $t(58) = 2.34$, $p = .02$; $t(55) = 2.23$, $p = .03$; $\omega^2 = .07$.	Moderate-to-large RO advantage
Eppard et al. (2020)	Reading outcomes	No statistically significant post-test group differences (all $p > .05$).	No significant group difference
Friedland et al. (2017)	Reading fluency	ANCOVA $p = .051$; improvement-score t-test $p < .05$; calculated $d = .05$.	Negligible effect
Hartshorn & Stephens (2023)	Listening comprehension	Time $\times$ Group interaction: $F(1,29) = 6.63$, $p = .015$, $\eta^2 = .186$. Treatment improved from .436 to .593; control from .425 to .439.	Large treatment advantage
Holmes & Allison (1986)	Reading comprehension	Ability $R^2 = .54$; mode $R^2 = .05$. Good readers were negatively affected by silent reading while listening; poor readers showed a literal-question benefit only in oral reading to an audience.	Small overall mode effect; moderator effect by ability
Hui (2024)	Overall comprehension	RWL > LO: Estimate = -0.06 , $t(97.46) = -2.61$, $p = .01$, $d \approx .53$. RWL = RO: $p = .66$.	Moderate RWL advantage over LO; no RWL-RO difference

Study	Outcome domain	Detailed quantitative findings	Dominant quantitative signal
Kartal & Simsek (2017)	Listening comprehension	Post-test M = 6.53 vs. 4.94, $t = -2.861$, $p = .006$, $d = .71$.	Moderate-to-large treatment advantage
Kim (2021)	Listening/reading comprehension	Listening: $F(2,72) = 11.68$, $p < .001$, $\eta^2 = .25$; RWL M = 65.40 vs. RO 41.54 and LO 39.79. Reading: $F(2,72) = 7.20$, $p = .002$, $\eta^2 = .17$; LO lowest.	Moderate-to-large RWL advantage overall
Liu et al. (2019)	Reading comprehension	ANOVA: $F(2,39) = 3.307$, $p = .047$, $\eta^2 = .145$. Pairwise $d \approx 1.71, 1.22, 0.60$.	Large dual-input advantage
Milliner (2019)	Listening/reading	Listening group effect: $F(2,57) = 34.39$, $p < .001$, $\omega^2 = .527$; RWL gain = +20.1. Reading interaction: $F(2,57) = 5.756$, $p = .005$, $\omega^2 = .049$.	Large listening advantage for RWL; small reading interaction
Nakashima et al. (2018)	Reading comprehension	Kruskal–Wallis $\chi^2(2) = 20.47$, $p < .001$, $\epsilon^2 = .174$. Post hoc: SR = RWL > L-only. Preferences split 52.4% vs. 47.6%.	Moderate-to-large advantage for SR/RWL over LO
Pellicer-Sánchez et al. (2020)	Process/attention	Dwell time on images increased under RWL ($d = -0.47$); integrative saccades $d = 4.19$. Text time negatively related to comprehension; image time positively related.	Moderate attention shift; very large process effect
Pellicer-Sánchez et al. (2021)	Process/attention	Dwell time shift $d = 0.26$; integrative saccades $d = 4.19$; interaction $d = 0.58$.	Small-to-moderate attention effect; very large process effect
Rasinski (1990)	Reading fluency	Time effect for speed: $F(1,19) = 28.71$, $p < .0001$, $\eta^2 = .60$; word recognition: $F(1,19) = 10.83$, $p < .01$, $\eta^2 = .36$. No treatment effect.	Improvement over time only; no treatment effect
Serrano & Pellicer-Sánchez (2019)	Process/comprehension	Text dwell: $\beta = .348$, $p < .001$; image dwell: $\beta = -.488$, $p = .004$. Comprehension no significant difference: $t(34) = 0.422$, $p = .676$.	Process shift only; no comprehension advantage
Tragant Mestres et al. (2019)	Vocabulary/engagement	Time $\times$ Group interaction: $F(2,78) = 6.98$, $p = .01$, $\eta^2 = .15$. RWL and RO > control for vocabulary; enjoyment higher in RWL.	Moderate-to-large advantage over control
Verlaan & Ortlieb (2012)	Reading comprehension	Overall: $t(109) = 3.74$, $p < .001$, $d = .36$. Struggling readers: $t(52) = 7.206$, $p < .001$, $d = .99$.	Small overall RWL advantage; large benefit for struggling readers
Yang et al. (2022)	Comprehension/attention	Comprehension: no main effect of assistance mode, $F = .084$, $p = .919$, $\eta^2 = .001$; text length $F = 21.943$, $p < .001$, $\eta^2 = .164$. Attention/PDT: $F = 10.523$, $p < .001$, $\eta^2 = .086$; audio-assisted lowest PDT.	No main comprehension advantage; moderate mode effect on attention

Note. The final column is a descriptive comparison aid across heterogeneous studies; it does not replace the detailed statistics in the preceding column.

Copyright of articles rests with the authors. Please cite TESL-EJ appropriately.