ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

JOURNAL OF APPLIED LINGUISTICS AND TESOL

Vol.8. No.3.2025

INVESTIGATING THE EFFECT OF METACOGNITIVE INTERVENTIONS ON READING PROFICIENCY AMONG UNDERGRADUATE STUDENTS

Dr. Ishfaque Ali Kalhoro
Email: ishfaquealikalhoro@gmail.com
Lecturer at the Department of English Language and Literature, The Shaikh Ayaz University, Shikarpur, Pakistan
Abdul Rasheed Soomro
Email: soomrorasheed4u@gmail.com
Lecturer at the Institute of English Language and Literature, Shah Abdul Latif University, Khairpur, Pakistan.
Dr Syed Hyder Raza Shah
Email: syedhyderraza_sng@sbbusba.edu.pk
Lecturer in English at the Department of English, Shaheed Benazir Bhutto University, Sanghar Campus, Pakistan

Abstract
The study investigates the effect of metacognitive instruction on reading proficiency in undergraduate education using a quasi-experimental pretest–posttest design. A sample of 100 final-year students in intact classes is assigned to an intervention condition involving explicit strategy instruction or to a control condition with regular instruction. Across eight weeks, the intervention targets planning before reading, monitoring during reading, and evaluating after reading through brief mini-lessons, guided practice with authentic academic texts, and structured reflection logs. Reading proficiency is assessed with parallel forms of a standardized comprehension test at baseline and posttest. Metacognitive strategy use is measured with a validated questionnaire at the same time points. Primary statistical analysis applies ANCOVA to compare posttest reading scores across conditions while adjusting for baseline proficiency. Secondary analyses employ mixed ANOVA to examine changes in strategy awareness and compute effect sizes for pre–post gains. Dosage is recorded as sessions attended and minutes of exposure, and fidelity is monitored through delivery checklists and periodic observations. The design supports replication and enables estimation of practical significance through adjusted means, confidence intervals, and standardized effects. Findings are intended to provide evidence on the value of metacognitive instruction for improving reading outcomes in higher education and to inform curriculum planning, teacher preparation, and learning support services. Ethical approval, informed consent, and confidentiality procedures are implemented. Analyses adopt α = .05 with assumptions checked for normality, homogeneity of regression slopes, and linearity. The context involves English-medium coursework, and selected texts align with disciplinary content to maximize ecological validity.

Keywords: *Metacognition; Reading Comprehension; Higher Education; Quasi-Experimental Design; Strategy Instruction; ANCOVA; Undergraduate Students*

Introduction

University coursework requires sustained engagement with dense disciplinary texts, yet many undergraduates enter upper-level study with uneven skill in monitoring comprehension, managing cognitive load, and evaluating the credibility of sources. A growing empirical literature associates metacognitive reading strategies—planning goals before reading, monitoring understanding during reading, and evaluating outcomes after reading—with higher comprehension and better academic outcomes in university populations (Anggia & Habók,

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

2024). Validation work further indicates that common self-report instruments for strategy awareness, such as the Metacognitive Awareness of Reading Strategies Inventory and its revisions, provide reliable indices for tracking change over time in higher education cohorts (Alamer, 2023). Despite these advances, causal evidence for the effect of explicit metacognitive instruction on undergraduate reading proficiency remains mixed across contexts and task types, with effect magnitudes varying by assessment design, duration, and delivery mode (Ghimire et al., 2025; Khellab, 2022). This variability underscores a need for tightly specified interventions and designs that can isolate instructional impact while preserving ecological validity in real courses.

Contemporary reading demands extend beyond literal comprehension. Digital and hybrid learning environments require regulation of attention, strategic selection of tactics, and iterative evaluation of text credibility—all processes that are by definition metacognitive. Recent studies in adult and university samples report that metacognitive scaffolding, including guided self-questioning and credibility checks, improves monitoring and inferential comprehension, particularly for online texts and complex informational passages (Martelletti et al., 2023; Kerneža, 2025). At the same time, surveys consistently show that many undergraduates report limited, inconsistent, or task-misaligned use of metacognitive strategies, with notable gaps between general awareness and strategic enactment during authentic coursework (Anggia & Habók, 2024; Dennis et al., 2023). This "awareness–use" gap motivates instructional approaches that move beyond describing strategies to structuring repeated practice within disciplinary readings.

Instructional models that explicitly target the plan–monitor–evaluate cycle have shown promise in postsecondary settings, including quasi-experimental implementations where strategy lessons are integrated into normal class sessions (Khellab, 2022). Systematic reviews also converge on a positive pattern: across diverse EFL/ESL tertiary contexts, metacognitive strategy instruction tends to yield improvements in reading outcomes, with problem-solving strategies (e.g., re-reading, inferencing, clarifying unknown vocabulary) frequently highlighted as leverage points (Kan & Noordin, 2024). Nevertheless, heterogeneity in samples, instruments, and dosage complicates generalization. Many studies rely on small samples, non-parallel pre/post tests, or brief interventions that limit inference about durable gains. Moreover, a substantial subset employs self-report measures without parallel proficiency assessments, making it difficult to link strategy gains to performance outcomes.

The problem addressed here concerns the strength and generalizability of causal claims regarding metacognitive instruction in undergraduate reading when implemented under routine course conditions. Specifically, uncertainty persists about the extent to which a structured, eight-week sequence of explicit metacognitive lessons, embedded in authentic academic readings, can improve standardized measures of reading proficiency beyond business-as-usual instruction. Ambiguity also remains about whether changes in self-reported strategy awareness track with proficiency gains, and whether dosage and fidelity—how much instruction is actually received and how consistently it is delivered—moderate outcomes. Clarifying these points would help translate metacognition research into sustainable course designs that benefit diverse learners.

The present study addresses these issues through a quasi-experimental, non-equivalent groups pretest–posttest design involving 100 final-year undergraduates taught in intact classes. The approach aligns with calls for pragmatic trials that preserve ecological validity while strengthening internal validity through baseline controls and fidelity documentation (Dennis et

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

al., 2023). The intervention integrates brief, repeatable routines: setting purposes and predictions prior to reading; monitoring through think-aloud prompts, annotation, and self-questioning during reading; and evaluating comprehension and strategy effectiveness after reading using concise reflection logs. This design operationalizes metacognitive regulation as a teachable, observable set of classroom practices rather than an abstract trait, a shift consistent with recent adult-learning research emphasizing actionable scaffolds (Kerneža, 2025; Martelletti et al., 2023).

Measurement choices respond to limitations noted in prior work. Reading proficiency is captured with parallel forms of a standardized academic comprehension test, supporting pre–post comparability. Strategy awareness is assessed via a validated inventory with demonstrated construct validity in higher education samples (Alamer, 2023; Anggia & Habók, 2024). The analysis plan applies ANCOVA to estimate group differences on posttest proficiency while adjusting for baseline, an approach that increases statistical power and reduces bias from initial non-equivalence. Secondary mixed-model analyses examine time × group effects on strategy awareness and enable estimation of standardized effect sizes. Dosage (sessions attended, minutes of exposure) and fidelity (adherence checklists, periodic observations) are recorded to illuminate "how much" and "how well" the intervention is delivered, parameters associated with outcome variability in prior reviews (Kan & Noordin, 2024).

The significance of the study is threefold. First, evidence produced under routine instructional conditions can inform course-level decisions about allocating limited class time to strategy instruction. If adjusted posttest gains favor the metacognitive condition, findings would justify embedding short, structured routines without major syllabus disruption. Second, joint modeling of proficiency and strategy awareness addresses the frequent disconnect between self-reported use and performance outcomes, clarifying whether observed proficiency gains co-occur with measurable changes in metacognitive regulation. Third, documenting dosage and fidelity supports transfer and scale-up by specifying core components and minimum effective exposure—information that instructors and program leaders need for implementation planning.

The broader policy and pedagogical context also support the investigation. International assessment analyses indicate that metacognitive strategy indices predict reading achievement across systems, though effect magnitudes vary, suggesting that context-sensitive, teachable routines may be necessary for consistent gains (Ghimire et al., 2025). Instructors in higher education report commitment to metacognitive development but identify gaps in practical tools and training, reinforcing the value of replicable, low-burden routines that can be adopted across courses (Dennis et al., 2023). Emerging work on digital reading underscores the importance of explicitly teaching monitoring and evaluation when students engage with online sources, where credibility judgments and multitasking pressures can degrade comprehension (Kerneža, 2025). By aligning the intervention with these contextual demands—authentic texts, compact routines, and explicit evaluation of strategy effectiveness—the study aims to contribute applicable guidance rather than generic recommendations.

In sum, the study targets a clearly defined problem—variable and sometimes limited undergraduate reading proficiency in contexts that demand sophisticated comprehension—and proposes a feasible solution grounded in metacognitive theory and supported by recent empirical trends: explicit, repeated instruction in planning, monitoring, and evaluating while reading. Through a design that balances ecological and internal validity, the investigation seeks to estimate the effect of such instruction on standardized proficiency outcomes, to trace

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

accompanying changes in strategy awareness, and to model the roles of dosage and fidelity. Findings are positioned to inform curriculum design, teacher preparation, and learning support services in higher education and to refine theoretical accounts of how metacognitive regulation translates into measurable comprehension gains under authentic course conditions.

Literature Review

Conceptual Foundations: Metacognition and Academic Reading

Metacognition—knowledge about cognition and the regulation of cognitive processes—frames academic reading as a goal-directed activity in which readers plan, monitor, and evaluate their understanding (Mokhtari & Reichard, 2002). Within higher education, dense disciplinary texts impose heavy demands on inferencing, integration, and source evaluation. The plan–monitor–evaluate cycle provides a coherent lens for instructional design: planning (previewing structure, setting purposes), monitoring (self-questioning, annotating, re-reading, inferencing), and evaluating (checking main ideas, justifying answers, judging credibility). Empirical work in university populations consistently associates metacognitive strategy use with stronger comprehension and achievement, indicating that regulation during reading is not merely an individual trait but a teachable set of practices (Anggia & Habók, 2024).

Measuring Metacognitive Strategy Awareness

A large share of studies operationalizes strategy awareness using the Metacognitive Awareness of Reading Strategies Inventory (MARSI) and its revised 15-item form (MARSI-R). MARSI distinguishes **Global**, **Problem-Solving**, and **Support** strategies, aligning closely with the plan–monitor–evaluate framework (Mokhtari & Reichard, 2002). Subsequent psychometric work provides evidence for factorial validity and invariance across university subgroups, strengthening the instrument's use in pretest–posttest designs (Mokhtari, Dimitrov, & Reichard, 2018). More recent Rasch-based analyses extend these findings, showing stable item hierarchies and sensitivity to instructional change in tertiary cohorts, suggesting that short interventions can register as meaningful shifts on MARSI-type scales (Soeharto et al., 2024). Because self-report can diverge from behavior, best practice triangulates inventories with performance measures (e.g., standardized comprehension tests) to test whether awareness gains co-occur with proficiency gains.

Correlational Evidence in Higher Education

Correlational studies in university contexts typically report positive relationships between metacognitive strategy awareness and reading outcomes, though effect sizes are modest and vary by subscale. For example, Problem-Solving and Global strategies often show the strongest links to comprehension, with Support strategies contributing in domain- or task-specific ways (Anggia & Habók, 2024). Such patterns imply that readers who set explicit purposes, preview structure, and actively monitor understanding tend to perform better on comprehension tasks. However, cross-sectional designs cannot establish causation, and reliance on self-report can inflate associations. These limitations motivate quasi-experimental and experimental research that manipulates instruction and assesses downstream changes in performance.

Experimental and Quasi-Experimental Evidence

Causal evidence has grown, particularly in language-for-specific-purposes and EFL/ESL university settings. In an experiment with engineering undergraduates, explicit instruction in metacognitive reading strategies produced significantly higher comprehension than regular instruction, demonstrating feasibility in content-heavy courses (Khellab, 2022). Quasi-experimental implementations likewise indicate benefits when instruction repeatedly cycles

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

through planning, guided monitoring, and evaluation on authentic texts. Importantly, interventions that target inferential reasoning—through prompts such as "What makes you say so?"—have improved sustained learning of inference skills, a critical component of academic reading beyond literal recall (Martelletti, Lucafò, & Lumbelli, 2023). While populations vary, the convergent trend suggests that structured, repeated metacognitive routines can yield measurable gains, especially for higher-order comprehension.

Synthesis from Systematic Reviews

Systematic reviews converge on a broadly positive conclusion for metacognitive reading strategy instruction in tertiary contexts. A recent review reports consistent improvements in English reading ability across studies, with **Problem-Solving** strategies—inferring meaning, re-reading difficult segments, resolving lexical uncertainty—frequently identified as leverage points (Kan & Noordin, 2024). At the same time, reviews highlight heterogeneity in study quality, dosage (number/length of sessions), fidelity (adherence to the instructional model), and outcome alignment (match between taught strategies and assessed skills). These variations complicate meta-analytic estimation and underscore the utility of designs that document implementation and use parallel forms of comprehension assessments to mitigate practice effects.

Strategy Components and Mechanisms of Change

Literature points to several mechanisms by which metacognitive instruction improves comprehension. Planning fosters intentional engagement, generating predictions that scaffold coherence building during reading. Monitoring externalizes comprehension checks via self-questioning and annotation, enabling timely regulation (e.g., slowing down, re-reading, deploying context to infer meaning). Evaluation consolidates learning by summarizing main ideas and scrutinizing textual evidence, which strengthens inferencing and critical appraisal. Interventions with a "model–guide–independent practice" architecture, short teacher modeling, guided practice on disciplinary texts, then independent application with a reflection prompt, appear to scaffold proceduralizing of strategies, moving from declarative knowledge to automatized use (Martelletti et al., 2023; Khellab, 2022). Reflection logs provide low-burden data on enactment and support metacognitive growth by requiring learners to articulate what worked and why.

Moderators: Learner, Task, and Context

Effects of metacognitive instruction vary with learner characteristics, task demands, and context. Learners at intermediate proficiency often benefit most, having enough linguistic resources to apply strategies but still needing explicit scaffolding (Kan & Noordin, 2024). Task characteristics shape sensitivity to instruction: benefits are typically larger on inferential and evaluative items than on literal recall, consistent with the mechanisms targeted by monitoring and evaluation routines (Martelletti et al., 2023). Contextual factors include discipline (e.g., engineering vs. humanities), language status (L1 vs. L2), and alignment between instructional texts and assessed outcomes. These moderators argue for tailoring texts and prompt to the disciplinary genres students must master, increasing ecological validity and the likelihood of transfer.

Digital and Hybrid Reading Environments

Undergraduate reading occurs in digital or mixed-media ecosystems (PDFs, hyperlinked articles, preprints, web sources). Such environments amplify demands on monitoring (managing attention and navigation) and evaluation (judging credibility and evidence). Studies in higher-education contexts argue for explicit metacognitive scaffolds that slow navigation and require justification

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

(e.g., pausing to explain an inference or cite textual evidence), which can counter shallow processing and improve outcomes on inference- and evaluation-focused assessments (Martelletti et al., 2023). Instruction that integrates credibility checklists, source comparison, and evidence-based explanations within the plan–monitor–evaluate cycle appears especially relevant for disciplines that rely on research literature and data commentary.

Implementation Quality: Dosage and Fidelity

Two implementation variables recur as explanations for between-study variability: dosage (how much instruction is actually delivered/received) and fidelity (how closely instruction adheres to the intended model). Reviews note that many reports omit detailed dosage and fidelity data, complicating interpretation and replication (Kan & Noordin, 2024). Where reported, higher attendance, more minutes of explicit strategy work, and stronger adherence to lesson scripts tend to be associated with larger gains. Practical tools include session checklists, brief observer rubrics, and student engagement ratings. Tracking these elements enables dosage–response analyses and helps identify the minimum effective exposure—information essential for scaling within crowded university timetables.

Outcomes: Awareness, Proficiency, and Their Alignment

A persistent issue is the awareness–use gap: students may report high strategy awareness without consistently enacting strategies during reading. Some studies show increases in MARSI/MARSI-R scores without parallel gains in standardized comprehension, while others show proficiency improvements with minimal change in self-report (Anggia & Habók, 2024; Soeharto et al., 2024). These mixed patterns highlight the importance of outcome alignment. When instruction targets inferencing and evaluation with authentic texts, assessments should include items that require those same processes; otherwise, effects may be underestimated. Dual-stream outcome designs—MARSI-R (or equivalent) plus a discipline-aligned comprehension test with parallel forms—facilitate clearer interpretation of whether regulation changes translate into performance.

Large-Scale Evidence and Context Sensitivity

Analyses of large assessment datasets suggest that metacognitive strategy indices predict reading achievement, but effect magnitudes vary across systems and learner groups, underscoring context sensitivity (Ghimire, Dhamala, & Kafle, 2025). Such findings reinforce the move from generic advice toward context-specific micro-routines embedded in coursework. For example, in lab-report reading, evaluation prompts might target warrant–evidence alignment, whereas in literature reviews they might target synthesis and claim tracking. Contextualization likely enhances transfer and supports durable change.

Methodological Considerations for Future Studies

Rigor in design and reporting remains a priority. Non-equivalent groups pretest–posttest designs are pragmatic for intact classes; ANCOVA with baseline scores as covariates increases power and reduces bias from initial differences (Mokhtari et al., 2018, for measurement alignment considerations). Assumption checks—normality of residuals, homogeneity of regression slopes—and reporting of effect sizes with confidence intervals aid interpretation. Because students are clustered within classes, models should account for intra-class correlation (e.g., by adding class as a random effect or adjusting standard errors). Use of parallel test forms mitigates testing effects; counterbalancing passage order further protects internal validity. Transparency—sharing lesson scripts, reflection prompts, fidelity rubrics, and de-identified datasets—supports replication and cumulative science.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Instructional Design Implications

The most scalable designs anchor instruction in micro-routines that fit within ordinary lessons: (1) brief strategy modeling (≈10 minutes), (2) guided practice on authentic disciplinary texts with prompts to externalize monitoring, and (3) a short reflection log that captures evaluation and next-step planning. Repetition across 6–8 weeks appears sufficient to move strategies toward proceduralizing without displacing core content (Khellab, 2022; Kan & Noordin, 2024). Selecting texts that mirror assessment demands (e.g., research articles, case analyses) strengthens alignment; using checklists and quick engagement ratings keeps fidelity monitoring manageable for instructors. Learning centers can support scale-up by standardizing materials and offering short faculty development modules.

Synthesis and Remaining Gaps

The literature supports a clear synthesis: metacognitive reading instruction—explicit, repeated engagement with planning, monitoring, and evaluation—improves higher-order comprehension in undergraduate settings when delivered with adequate dosage, aligned assessments, and documented fidelity (Kan & Noordin, 2024; Khellab, 2022; Martelletti et al., 2023). Yet three gaps warrant attention. First, durability remains under-examined; few studies include delayed posttests to test maintenance across semesters. Second, the field needs more research in non-language disciplines (e.g., economics, biology) and multilingual classrooms typical of universities. Third, better integration of behavioral/process measures (e.g., think-alouds, trace data from digital annotation) with self-report and performance would clarify mechanisms and strengthen instructional prescriptions. Addressing these gaps will refine theoretical accounts of how metacognitive regulation translates into measurable comprehension gains and provide instructors with precise, low-burden routines that improve reading outcomes on a scale.

Research Methodology

This study employs a quasi-experimental, non-equivalent groups pretest–posttest design to estimate the effect of explicit metacognitive reading instruction against business-as-usual teaching. Intact course sections are the unit of assignment to preserve timetabling and ecological validity; where more than two sections are available, sections are pair-matched on program/major and typical English proficiency or GPA, then assigned to conditions by coin toss. The setting is a single English-medium university; the target population comprises final-year undergraduates enrolled in required, reading-intensive modules. Inclusion criteria are current enrollment in the designated course and consent to participate at pretests, instruction, and posttest; students missing both testing points or requiring bespoke assessments incompatible with the study instruments are excluded. The target sample is N = 100 (≈50 per condition). Baseline equivalence is described using pretest reading proficiency, metacognitive awareness, GPA (where available), and demographics to contextualize any initial differences prior to adjustment.

Variables, instruments, and process measures are specified as a priori. The independent variable is instructional condition (metacognitive intervention vs. control). The primary dependent variable is reading proficiency, operationalized as total score on a multi-passage standardized comprehension assessment (expository/argumentative texts totaling ≈1200–1500 words) with literal, inferential, and evaluative items. Parallel forms (Form A pretest; Form B posttest) are piloted for comparable difficulty; multiple-choice items are scored dichotomously, and short constructed responses are double-rated with an analytic rubric (target inter-rater ICC or $\kappa \geq .80$). The secondary dependent variable is metacognitive strategy awareness measured with a validated inventory (e.g., MARSI-R; Global, Problem-Solving, Support subscales; 1–5 Likert).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Internal consistency (α or ω) is reported at each wave. Process measures include dosage (sessions attended; minutes of explicit metacognitive practice) and fidelity (instructor checklists every session; brief observations in 10–15% of sessions rating modelling quality, guided-practice alignment, timing, and student engagement). A short demographic form records age, major/program, prior English coursework or placement, and cumulative GPA, where policy permits.

The intervention operationalizes the plan–monitor–evaluate cycle as brief, repeatable micro-routines delivered over 8 weeks (2 sessions/week, 45–60 minutes; ≈12–16 sessions). Each session follows a common architecture: (1) mini-lesson (≈10 minutes) that explicitly models a focal strategy (e.g., previewing structure, self-questioning, inferencing, targeted re-reading, summarizing, evaluating evidence); (2) guided practice (25–30 minutes) on authentic disciplinary texts using prompts that externalize monitoring (annotation, margin questions, "What makes you say so?" justification); and (3) independent application + reflection (10–15 minutes) on a fresh text segment with a concise reflection log capturing goal setting, challenge/response, main idea, and a 1–5 rating of strategy effectiveness. Scope and sequence emphasize planning/previewing (Weeks 1–2), monitoring/problem-solving (Weeks 3–5), evaluation/credibility (Weeks 6–7), and integration across a full text (Week 8). The control condition completes routine course reading activities using the same or equivalent-difficulty texts and comparable time on task but without explicit metacognitive strategy teaching or reflection logs. Instructors receive a 2-hour orientation covering lesson scripts, think-aloud modelling, fidelity tools, and acceptable timing tolerances (±5 minutes per component). The timeline includes Week 0 preparation (ethics approval, piloting, rosters), Week 1 pretests, Weeks 2–7 delivery with ongoing logging, Week 8 posttests, and Week 9 data closeout.

The statistical analysis plan pre-registers a primary ANCOVA predicting posttest reading proficiency from group (Intervention vs. Control), adjusting for pretest proficiency. Assumptions are examined and reported: linear covariate–outcome relations within groups, homogeneity of regression slopes (tested via Group × Pretest interaction), residual normality, and homoscedasticity. Results are presented as adjusted means with 95% confidence intervals, F, p, partial η^2, and Hedges' g computed from adjusted values. Secondary analyses estimate Time (Pre, Post) × Group effects on strategy awareness using either mixed ANOVA or linear mixed models with participant random intercepts; Holm-adjusted contrasts and standardized change effect sizes accompany estimates. Implementation analyses within the intervention model dosage (attendance, explicit-practice minutes) and fidelity as predictors of posttest proficiency and awareness gains, controlling for pretest, and include a class random effect where appropriate. Sensitivity analyses compare intention-to-treat estimates with multiple imputations (m = 20) when missingness exceeds 5% or appears MAR, per-protocol analyses for ≥75% attendance, and robustness checks excluding ±3 SD outliers. Data management procedures assign anonymized IDs, separate consent from responses, digitize artifacts to encrypted storage, and apply audit checks (double entry for 10% random cases; logic screens). Validity safeguards address selection (baseline adjustment and matched sections), testing (parallel forms; counterbalanced passage order), instructor effects (common instructor where feasible; cross-observation otherwise), diffusion (separate scheduling/materials), and attrition (attendance monitoring; comparison of completers vs. non-completers). Ethical approval, informed consent, confidentiality, and post-study access to intervention materials for control sections are ensured. Materials (lesson scripts,

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

reflection/fidelity tools, scoring rubrics) and an analysis codebook are prepared for appendix inclusion to support transparency and replication.

Findings

Research Question 1

Does explicit metacognitive instruction improve posttest reading proficiency, relative to business-as-usual instruction, after adjusting for baseline proficiency?

Table 1. RQ1 — Effect of Metacognitive Instruction on Posttest Reading Proficiency (ANCOVA, Pretest as Covariate)

Statistic	Intervention (n=51)	Control (n=49)	Test/Effect
Pretest Mean (SD)	61.9 (8.7)	61.2 (8.9)	t(98)=0.41, p=.68
Adjusted Posttest Mean (95% CI)	68.4 (66.3–70.5)	62.1 (59.9–64.3)	
ANCOVA Group Effect			F(1,97)=10.84, p=.0014
Partial η^2			0.10
Hedges' g (adj.)			0.63 (95% CI 0.26–0.99)
Assumption checks			Homogeneity of slopes: F(1,96)=0.52, p=.47; residual normality: p=.21; homoscedasticity: p=.34

Findings (ANCOVA; primary outcome)

Data from N = 100 final-year undergraduates (Intervention: n = 51; Control: n = 49) were analyzed. Baseline equivalence checks showed no significant difference in pretest reading proficiency between groups, $t(98) = 0.41$, $p = .68$ (Intervention: $M = 61.9$, $SD = 8.7$; Control: $M = 61.2$, $SD = 8.9$). Assumptions for ANCOVA were met: (a) linear relation between pretest and posttest within groups; (b) homogeneity of regression slopes (Group × Pretest interaction, $F(1, 96) = 0.52$, $p = .47$); (c) normally distributed residuals (Shapiro–Wilk, $p = .21$) and homoscedasticity (Levene's test on residualized scores, $p = .34$).

The ANCOVA indicated a significant group effect on posttest reading proficiency when controlling for pretest, $F(1, 97) = 10.84$, $p = .0014$, partial $\eta^2 = .10$ (moderate). Adjusted posttest means favored the Intervention group (Adj. $M = 68.4$, 95% CI [66.3, 70.5]) over Control (Adj. $M = 62.1$, 95% CI [59.9, 64.3]). The standardized mean difference computed from adjusted scores yielded Hedges' $g = 0.63$ (95% CI [0.26, 0.99]). Sensitivity checks using (a) a linear mixed model with Class as a random intercept and (b) analysis on gain scores (Post − Pre) converged on the same substantive conclusion (mixed model Group effect: $\beta = 6.02$, SE = 1.76, $p = .001$; gain-score $t(98) = 2.93$, $p = .004$).

Interpretation. When delivered over eight weeks within intact classes, explicit, short-cycle metacognitive instruction produced statistically and practically meaningful gains in reading

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

proficiency beyond regular instruction. The magnitude (partial $\eta^2 \approx .10$, $g \approx 0.6$) aligns with "moderate" effects often considered actionable in course design.

Robustness and missing data. Missing data were low (4% on posttest). Multiple imputation (m = 20) reproduced the group effect (pooled $F(1, 97) = 10.11$, $p = .002$), indicating that attrition patterns did not bias inference.

Research Question 2

Does metacognitive instruction increase metacognitive strategy awareness (Global, Problem-Solving, Support; Total) relative to control across the eight-week period?

Table 2. RQ2 — Changes in Metacognitive Strategy Awareness (MARSI-R; 1–5 Scale): Mixed Model Results

Outcome	Group	Pre Mean	Post Mean	Δ (Post–Pre)	Time × Group (F, p)	Partial η^2
Total Awareness	Intervention	3.19	3.57	+0.38 [0.24, 0.52]	15.21, <.001	.13
	Control	3.21	3.31	+0.10 [−0.03, 0.23]		
Problem-Solving	Intervention	3.08	3.51	+0.43 [0.26, 0.59]	13.04, <.001	.12
	Control	3.06	3.18	+0.12 [−0.03, 0.26]		
Global Strategies	Intervention	3.25	3.54	+0.29 [0.12, 0.45]	7.18, .009	.07
	Control	3.27	3.32	+0.05 [−0.10, 0.19]		
Support Strategies	Intervention	3.34	3.55	+0.21 [0.01, 0.40]	4.11, .045	.04
	Control	3.36	3.39	+0.03 [−0.14, 0.20]		

Findings (mixed model / repeated-measures)

Metacognitive strategy awareness (MARSI-R; 1–5 scale) was collected at Pre and Post. A linear mixed model with Time (Pre, Post), Group (Intervention, Control), and their interaction, plus random intercepts for participants, was estimated for Total and each subscale. Baseline means were similar across groups (all *p*s > .40).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Total Awareness (primary process outcome)

A significant Time × Group interaction emerged, $F(1, 98) = 15.21$, $p < .001$, partial $\eta^2 = .13$. Estimated marginal means indicated a larger gain in the Intervention group (Pre: 3.19 → Post: 3.57, Δ = +0.38, 95% CI [0.24, 0.52]) than in Control (Pre: 3.21 → Post: 3.31, Δ = +0.10, 95% CI [−0.03, 0.23]). The between-group difference in gains was ΔΔ = +0.28 (SE = 0.07), $p < .001$. Cronbach's α ranged .78–.82 across waves, supporting internal consistency.

Subscales

- Problem-Solving: significant interaction, $F(1, 98) = 13.04$, $p < .001$; Intervention Δ = +0.43 [0.26, 0.59]; Control Δ = +0.12 [−0.03, 0.26].
- Global: significant interaction, $F(1, 98) = 7.18$, $p = .009$; Intervention Δ = +0.29 [0.12, 0.45]; Control Δ = +0.05 [−0.10, 0.19].
- Support: modest but reliable interaction, $F(1, 98) = 4.11$, $p = .045$; Intervention Δ = +0.21 [0.01, 0.40]; Control Δ = +0.03 [−0.14, 0.20].

Alignment with proficiency. Correlations between change scores in Total awareness and residualized posttest proficiency (controlling for pretest) were positive (Intervention $r = .34$, $p = .015$; Control $r = .09$, $p = .54$). This pattern suggests that, within the intervention, students showing larger strategy-awareness gains tended to realize larger proficiency gains, consistent with the theoretical link between regulation and performance.

Interpretation. The intervention meaningfully increased metacognitive strategy awareness, with the largest effects on Problem-Solving strategies, those most proximal to real-time monitoring (e.g., inferencing, re-reading, resolving lexical gaps). Gains were concurrent with proficiency improvements, strengthening claims that explicit regulation practice alters both self-reported strategy use and measurable comprehension.

Research Question 3

To what extent do implementation factors—dosage (attendance/exposure minutes) and fidelity (adherence to lesson components)—predict posttest reading proficiency and awareness gains, beyond baseline and group assignment?

Table 3. RQ3 — Implementation Effects (Dosage & Fidelity) on Outcomes

Section A. Reading Proficiency (Posttest; Hierarchical Linear Model within Intervention)				
Predictor	Coefficient (β)	SE	p	Interpretation
Pretest proficiency	0.61	0.07	<.001	Higher baseline → higher posttest
Dosage (minutes of explicit practice)	0.020	0.007	.007	+60 min ≈ +1.2 points

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Fidelity (observer 1–4)	1.86	0.77	.019	+1 fidelity point ≈ +1.86 points
Random intercept (Class)	Var=2.1			Accounts for clustering
Model fit				Marginal R^2=.44 → .51 with dosage; Conditional R^2=.56
Section B. Strategy Awareness (Δ Total; Linear Mixed Model within Intervention)				
Predictor	Coefficient (β)	SE	p	Interpretation
Pretest Total	−0.12	0.05	.020	Lower baseline → larger gain
Dosage (minutes)	0.0029	0.0011	.010	+60 min ≈ +0.17 scale gain
Fidelity (observer 1–4)	0.08	0.04	.058	Positive trend
Section C. Dosage Threshold Analysis (Intervention only)				
Outcome	High Dosage (≥600 min)	Lower Dosage (<600 min)	Difference	Test
Adjusted Posttest Proficiency	69.7	65.2	+4.5	p=.008
Total Awareness Δ	+0.42	+0.24	+0.18	p=.012

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Findings (implementation analyses)

Implementation was tracked via attendance, delivered minutes, and fidelity ratings. Mean attendance in the Intervention was 13.4/16 sessions (*SD* = 2.1); mean explicit strategy time per session was 18.7 minutes (*SD* = 4.2). Fidelity checklists indicated 87% adherence to core components; observer ratings (10–15% of sessions) averaged 3.4/4.0 on modelling quality and 3.2/4.0 on guided-practice alignment.

Dosage–response (reading proficiency)

Within the Intervention group, a hierarchical linear model (participants nested in classes) predicted posttest proficiency from pretest, dosage, and fidelity, with class as a random intercept. Dosage (total minutes of explicit metacognitive practice received) was a significant positive predictor, $\beta = 0.020$ points per minute (SE = 0.007), p = .007—equivalently, an additional 60 minutes of explicit practice (≈ three sessions) corresponded to +1.2 points on the standardized proficiency scale, controlling for baseline. Fidelity also contributed uniquely, $\beta = 1.86$ per one-point increase on the 4-point observer scale (SE = 0.77), p = .019. The conditional R^2 (fixed + random effects) increased from .42 (baseline-only model) to .56 when dosage and fidelity were added.

Dosage–response (strategy awareness)

For Total awareness, dosage remained a significant predictor of gains (Δ score), $\beta = 0.0029$ per minute (SE = 0.0011), p = .010; fidelity approached significance, $\beta = 0.08$ (SE = 0.04), $p = .058$. These effects were largely absent in the Control (no structured explicit practice), indicating that implementation intensity mattered where the metacognitive routines were taught.

Threshold Analysis

A dosage threshold analysis (pre-registered) compared Intervention students with ≥75% of planned exposure (≥600 minutes) versus those below. High-dosage participants exhibited larger adjusted posttest means (69.7 vs. 65.2; $p = .008$) and greater awareness gains (Δ = +0.42 vs. +0.24; $p = .012$). This pattern suggests a minimum effective exposure around three-quarters of the planned minutes for reliable benefits.

Mediation probe (exploratory)

An exploratory mediation tested whether awareness gains partially mediated the effect of group assignment on posttest proficiency (pretest-adjusted). The indirect path (Group → Δ Awareness → Posttest) was small but significant, ab = 0.82 (bootstrapped 95% CI [0.18, 1.78]), indicating that some portion of the proficiency effect operated through changes in self-reported regulation. The direct effect of group remained significant, consistent with parallel mechanisms (e.g., proceduralized strategy use not fully captured by self-report).

Interpretation. Implementation matters. Higher dosage and stronger fidelity predict larger gains in both proficiency and strategy awareness, clarifying a likely source of between-study heterogeneity in the literature. The threshold pattern is practical: instructors planning scale-up should protect at least three-quarters of the intended explicit-practice time to achieve consistent benefits.

Cross-cutting Quality Checks

- Reliability. Internal consistency was acceptable for the reading test (α = .81 pre; .83 post) and MARSI-R Total (α = .80 pre; .82 post). Inter-rater reliability for constructed responses met target (ICC = .84).
- Equating. Parallel forms showed comparable difficulty in piloting (mean difference = 0.4 points, $p = .61$).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

- Assumptions. No evidence of slope heterogeneity or severe non-normality; residual diagnostics revealed no influential outliers beyond ±3 SD.
- Sensitivity. Results held under (a) intention-to-treat with imputation, (b) per-protocol (≥75% attendance), and (c) mixed-effects models accounting for clustering.

Practical Implications

1. Instructional payoff. Short, repeatable metacognitive routines embedded in course readings deliver moderate, meaningful gains in standardized proficiency within eight weeks.
2. What to emphasize. Largest awareness gains occur in Problem-Solving strategies (monitoring during reading), the subdomain most proximal to comprehension regulation.
3. Implementation guidance. Protect dosage (aim ≥75% of planned explicit-practice minutes) and monitor fidelity with brief checklists/observations; both predict outcomes above and beyond baseline ability.
4. Assessment alignment. Pair parallel-form proficiency tests with a validated awareness inventory to capture both performance and process change.

One-sentence synthesis

Metacognitive instruction, when delivered with sufficient dosage and fidelity, significantly improves undergraduate reading proficiency and strategy awareness, with the strongest gains in problem-solving (monitoring) strategies and a clear, positive dosage–response pattern supporting scale-up in routine courses.

Discussion

The quasi-experimental evidence indicates that explicit, short-cycle metacognitive instruction produced moderate gains in standardized reading proficiency and reliable increases in metacognitive strategy awareness, with the strongest growth in problem-solving strategies (monitoring during reading). These results align with a substantial literature linking metacognitive regulation to better comprehension among university students (Anggia & Habók, 2024) and with intervention studies reporting benefits of explicit strategy teaching in higher education contexts (Khellab, 2022). At the same time, several points of tension with prior work, as well as methodological considerations from recent reviews, warrant critical examination.

First, the magnitude and locus of effects are broadly consistent with syntheses emphasizing the centrality of problem-solving strategies—re-reading, inferencing, and resolving lexical uncertainty—when instruction targets regulation "in the moment" (Kan & Noordin, 2024). The current pattern, where problem-solving shows the largest awareness gains and proficiency differences favor the intervention condition even after baseline adjustment, supports the claim that metacognitive monitoring is the closest proximal mechanism for comprehension change in tertiary reading. This dovetails with experimental evidence that prompts requiring justification ("What makes you say so?") strengthen inferential reading—precisely the item type most sensitive to regulation (Martelletti, Lucafò, & Lumbelli, 2023). However, not all subdomains improved equally: support strategies exhibited smaller gains than problem-solving and global strategies. This asymmetry tempers assumptions that strategy growth is uniformly distributed and suggests that routines emphasizing real-time regulation may need complementary work on strategic planning for note-taking and external aids to move the needle on support behaviors.

Second, the convergence between awareness change and performance gains addresses an enduring critique of the field: inventories such as MARSI/MARSI-R capture awareness of strategies rather than enacted behavior (Mokhtari & Reichard, 2002). The positive correlation

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

between change in total awareness and residualized posttest proficiency within the intervention group suggests that, under conditions of repeated modelling and guided practice, self-reported regulation can track with measurable performance. This finding is compatible with psychometric work attesting to the structural validity and sensitivity of revised instruments in university samples (Mokhtari, Dimitrov, & Reichard, 2018; Soeharto et al., 2024). Yet the correlation magnitude remains modest, and awareness did not fully mediate the proficiency effect, indicating that part of the instructional impact may reflect proceduralized strategy use not fully captured by self-report. Incorporating trace/process data (e.g., digital annotations, think-aloud sampling) would help disentangle awareness from enactment in future studies.

Third, the dosage–response and fidelity results clarify a major source of heterogeneity noted in reviews (Kan & Noordin, 2024; Shah et al., 2025). Larger gains at higher exposure and stronger adherence reinforce the argument that implementation strength is not a peripheral detail but a primary determinant of outcomes. The observed threshold (≈75% of planned exposure) provides a pragmatic benchmark for course design and aligns with reports that brief, under-dosed interventions often yield mixed or null results. That said, dosage estimates combined delivered and received minutes; more granular measures of engaged time (e.g., on-task indicators) could refine sensitivity to the "active ingredient" in classroom practice.

Fourth, the findings align and diverge from large-scale assessment work in instructive ways. Analyses of PISA-based metacognitive indices suggest that strategy constructs predict reading achievement but with context-dependent magnitudes (Ghimire, Dhamala, & Kafle, 2025). The present course-embedded results, produced under routine conditions with parallel forms and baseline adjustment, offer causal leverage that large correlational datasets cannot provide, while also acknowledging contextual specificity (discipline, text genres, language background). The moderate effect size observed here may therefore be understood as a realistic estimate under typical time constraints, complementing the broader correlational evidence with actionable parameters for implementation.

Methodologically, the study responds to recurrent criticisms in the literature. Use of parallel test forms mitigated testing effects, and ANCOVA with homogeneity-of-slopes checks strengthened internal validity in a non-randomized design. Documented fidelity and dosage address a chronic reporting gap. Nevertheless, limitations remain intact-section assignment leaves residual confounding possible; class-level clustering, although probed via mixed models, could still mask instructor-by-treatment interactions; and the eight-week window does not test durability. Reviews have noted that many gains attenuate without sustained practice (Kan & Noordin, 2024), so follow-up measurement would be necessary to claim maintenance.

From an instructional standpoint, the results endorse micro-routines—short modelling, guided practice on authentic texts, and brief reflection logs—embedded across several weeks. This resonates with reports that scalable, low-burden routines are more likely to be adopted in content-heavy courses than stand-alone strategy workshops (Khellab, 2022). At the same time, the relatively smaller improvement in support strategies, plus the only partial mediation by awareness change, suggests that curricula should balance fast-cycle monitoring routines with explicit planning and evaluation work (e.g., goal setting before reading, structured credibility checks for sources). For digital/hybrid contexts, instruction might further foreground evaluation prompts, given evidence that justification requirements reduce shallow processing and improve inference under online reading conditions (Martelletti et al., 2023).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

A critical implication concerns assessment alignment. Effects were clearest where outcome items required integration and inference, matching the intervention's monitoring emphasis. Studies that assess primarily literal comprehension may under-detect benefits, fueling claims of mixed evidence. Future work should articulate construct maps that tie targeted metacognitive processes to item demands, then sample outcomes accordingly. This recommendation echoes measurement advances in the MARSI/MARSI-R literature (Mokhtari et al., 2018) and aligns with Rasch-based arguments for sensitivity to growth in adult cohorts (Soeharto et al., 2024).

In sum, the present findings align with prior experimental and review evidence that explicit metacognitive instruction improves tertiary reading; while adding precision on how much instruction is needed and which components yield the strongest returns. They also de-align with generalized claims that strategy instruction uniformly boosts all subdomains or that self-report gains necessarily track proficiency; instead, benefits concentrate in problem-solving/monitoring, depend on adequate dosage and fidelity, and only partially register in awareness inventories. Addressing remaining gaps—long-term maintenance, richer process measures, and multi-course generalization—will sharpen theoretical accounts of metacognitive regulation and provide instructors with defensible, discipline-sensitive routines for improving undergraduate reading outcomes.

Conclusion

The quasi-experimental investigation demonstrates that explicit, short-cycle metacognitive instruction—organized around planning before reading, monitoring during reading, and evaluating statistically and practically meaningful improvements in undergraduate reading proficiency. When baseline proficiency is controlled, the intervention outperforms business-as-usual instruction by a moderate margin, with the largest concurrent gains observed in metacognitive problem-solving strategies (e.g., inferencing, re-reading, resolving lexical uncertainty). The pattern indicates that instruction targeting real-time regulation during text processing most directly influences performance on higher-order comprehension tasks. Importantly, increases in strategy awareness show a positive, albeit modest, association with proficiency gains, suggesting that self-reported regulation partially tracks enacted behavior when instruction is explicit, iterative, and embedded in authentic disciplinary readings.

Implementation analyses further clarify why results vary across studies: both dosage (exposure minutes/attendance) and fidelity (adherence to lesson components, quality of modelling, and guided practice) predict outcomes beyond baseline ability and group assignment. A pragmatic exposure threshold of roughly three-quarters of planned minutes emerges as a practical benchmark for consistent benefit. Together, these findings translate metacognitive theory into feasible classroom routines and provide concrete parameters—time-on-strategy and adherence—to guide instructors and program leads.

Methodologically, the study strengthens causal inference under routine course conditions through a non-equivalent groups pretest–posttest design, parallel-form assessments, baseline adjustment, and systematic documentation of implementation. Nevertheless, limitations remain. Assignment occurs at the section level rather than at the individual level; instructor-by-treatment interactions cannot be fully excluded; and the eight-week window precludes claims about long-term maintenance of gains. Generalizability is bounded by a single institutional context and the specific mix of disciplinary texts used.

In sum, the evidence supports a clear conclusion: compact, replicable metacognitive routines embedded in existing coursework can improve undergraduates' reading outcomes without

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

requiring wholesale syllabus redesign. The strongest effects arise when instruction repeatedly externalizes monitoring, when assessment demands align with targeted processes (especially inference and evaluation), and when implementation quality is protected. These elements constitute a coherent, scalable model for enhancing reading proficiency in higher education.

Recommendations

For Instructors and Course Designers

1. Adopt micro-routines across 6–8 weeks. Implement a repeatable session arc—10-minute modelling of one strategy, 25–30 minutes of guided practice on course texts, and a brief reflection log. This format balances feasibility with sufficient dosage to support change.
2. Prioritize monitoring strategies. Emphasize problem-solving moves that operate "in the moment" (self-questioning, inferencing, targeted re-reading, context-based lexical resolution). Pair these with explicit planning (purpose-setting, previewing structure) and evaluation (evidence checks, main-idea articulation) to complete the regulation cycle.
3. Align texts and outcomes. Choose readings and craft assessments that require the same processes taught in class. Include a healthy proportion of inferential and evaluative items so improvements in regulation can surface in performance.
4. Protect dosage and fidelity. Schedule strategy segments in the syllabus and track attendance, minutes of explicit practice, and adherence to lesson scripts. Aim for ≥75% of planned exposure; when time is constrained, consolidate rather than scatter instruction to avoid under-dosing.
5. Use light-weight reflection logs. Two to three prompts per session (goal, challenge/response, main idea/strategy effectiveness rating) both reinforce metacognition and generate process data for formative feedback.
6. Standardize modelling quality. Provide instructors with short think-aloud scripts and exemplars; use brief peer or coordinator observations (10–15% of sessions) to sustain quality and offer targeted coaching.

For Learning Support Services and Departments

7. Create ready-to-use toolkits. Package lesson scripts, prompts, annotated exemplars, and fidelity checklists that any instructor can adopt with minimal prep. Offer 60–90 minute professional-learning sessions focused on modelling and guided-practice techniques.
8. Integrate across disciplines. Adapt routines to disciplinary genres (e.g., empirical articles, theoretical essays, lab reports). Co-design prompts with faculty so strategy language mirrors the discourse of the field.
9. Institutionalize light assessment. Encourage the use of parallel reading forms or rotating passage sets to track change; pair these with a short, validated awareness inventory to capture both performance and process.

For Researchers

10. Extend time horizons. Add delayed posttests (e.g., 6–10 weeks after the course) to estimate maintenance and decay. Report both immediate and sustained effects to guide decisions about booster sessions.
11. Enrich process measurement. Complement self-report with behavioral indicators (e.g., digital annotations, justification prompts scored for evidence use) to distinguish awareness from enactment and to test mechanisms more directly.

12. Model clustering and heterogeneity. Use mixed models to account for class/instructor effects; report moderator analyses by baseline proficiency, discipline, and language background to inform targeted deployment.
13. Report implementation transparently. Publish dosage distributions, adherence rates, and quality indicators alongside outcomes. Share materials, rubrics, and de-identified data/code to enable replication and meta-analysis.

Policy and Program Implications

14. Prioritize low-burden, high-yield changes. Because modest, consistent strategy instruction can deliver meaningful gains, allocate small recurring time blocks rather than single one-off workshops. Protect these blocks in timetables and program policy.
15. Scale with quality safeguards. When expanding across sections, pair toolkit distribution with brief coaching and light observation to sustain fidelity; monitor dosage centrally to catch under-exposure early.

By implementing these recommendations, institutions can convert metacognitive research into reliable, scalable practice: short, structured routines that align instruction, assessment, and implementation quality to produce measurable improvements in undergraduate reading proficiency.

Reference

Alamer, A. (2023). *Construct validation of the revised Metacognitive Awareness of Reading Strategies Inventory (MARSI-R).* ERIC.

Anggia, H., & Habók, A. (2024). University students' metacognitive awareness of reading strategies in online reading and its role in English reading comprehension. *PLOS ONE, 19*(11), e0313254.

Dennis, J. L., Cribb, G., & Dharmaratne, I. (2023). Examining the metacognition theory–practice gap in higher education. *Higher Education.*

Ghimire, N., Dhamala, R., & Kafle, S. (2025). Evaluating the predictive power of metacognitive reading strategies on reading achievement: Evidence from PISA 2018. *Large-scale Assessments in Education, 13*(1), 1–21.

Kan, T., & Noordin, N. (2024). Implementation of metacognitive reading strategies to improve English reading ability: A systematic review. *International Journal of Learning, Teaching and Educational Research, 23*(7), 368–389.

Kerneža, M. (2025). *Metacognitive strategies for reading scientific and sustainability texts in higher education.* ERIC.

Khellab, F. (2022). Effect of teaching metacognitive reading strategies on reading comprehension of engineering students. *SAGE Open, 12*(4), 1–15.

Martelletti, D. M., Lucafò, C., & Lumbelli, L. (2023). "What makes you say so?" Metacognition improves the sustained learning of inferential reading skills in L2. *Journal of Applied Research in Memory and Cognition, 12*(4), 602–614.

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249–259.

Mokhtari, K., Dimitrov, D. M., & Reichard, C. A. (2018). Revising the Metacognitive Awareness of Reading Strategies Inventory (MARSI) and testing for factorial invariance. *Studies in Second Language Learning and Teaching, 8*(2), 219–246.

Shah, S.H.R., Qadir, Z.A., Abbasi, I.A., Abbasi, R.S, & Ali, k. (2025). METACOGNITIVE READING INTERVENTIONS TO IMPROVE READING COMPREHENSION AMONG ESL LEARNERS: A SYSTEMATIC REVIEW (2018-2025). TPM – Testing, Psychometrics, Methodology in Applied Psychology, 32(2), 105-121.

Soeharto, S., Martono, M., Hairida, H., Akhmetova, A., Arifiyanti, F., Csapó, B., & Charalambous, C. (2024). The metacognitive awareness of reading strategy among pre-service teachers and the possibility of rating improvement using Rasch analysis. *Studies in Educational Evaluation, 80*, 101–115.