p-ISSN 2252-7427, e-ISSN 2541-3635
https://journal.uniku.ac.id/index.php/IEFLJ/index

p-ISSN 2252-7427 e-ISSN 2541-3635
Indonesian EFL Journal
AISEE
The Association of Indonesia Scholars of English Education

VALIDATING A METACOGNITIVE ACADEMIC WRITING STRATEGIES SCALE FOR THE AI ERA: A META-SYNTHETIC ANALYSIS

Fahrus Zaman Fadhly
Universitas Kuningan, Kuningan, Indonesia

Corresponding author: fahrus.zaman.fadhly@uniku.ac.id

APA Citation: Fadhly, F. Z. (2026). Validating a metacognitive academic writing strategies scale for the AI era: A meta-synthetic analysis. *Indonesian EFL Journal, 12*(3), 1125-1136. https://doi.org/10.25134/ieflj.v12i3.232

Received: 11-03-2026 Accepted: 21-06-2026 Published: 30-09-2026

Abstract: Metacognitive academic writing strategies help writers plan, monitor, evaluate, and regulate the production of scholarly texts. Generative artificial intelligence (AI) changes the conditions under which these strategies operate: writers can delegate language-level tasks, consult AI for feedback, compare alternative formulations, and receive rapid explanations, but they must also evaluate inaccurate outputs, preserve authorial agency, and disclose assistance appropriately. This article develops an evidence-informed framework for validating a Metacognitive Academic Writing Strategies Scale for the AI era (MAWSS-AI). It synthesizes the supplied 2022–2025 bibliography on metacognitive writing, academic writing pedagogy, AI-assisted writing, AI literacy, and research integrity. Because the available source is a reference list rather than a complete corpus of full-text studies, the synthesis is conceptual and bibliography-led; it does not claim a completed systematic review, effect-size meta-analysis, or empirical psychometric validation. The proposed framework retains the core regulatory cycle of planning, monitoring, evaluating, and revising, while adding AI-task orchestration, critical verification, human–AI decision-making, and transparency/ethical accountability. A provisional 30-item pool is organized into six dimensions and accompanied by a construct map, synthesis matrix, content-validation rubric, and staged validation protocol. The article recommends expert review, cognitive interviews, pilot testing, exploratory and confirmatory factor analyses, reliability estimates, convergent and discriminant validity, criterion-related evidence, measurement invariance, and response-process checks. The central proposition is that AI-era metacognitive competence is not measured by frequency of AI use, but by the quality of writers' strategic decisions before, during, and after interacting with AI. The scale should therefore assess intentionality, epistemic vigilance, ownership, and responsible regulation rather than technological enthusiasm alone.
Keywords: *metacognition; academic writing; scale validation; generative artificial intelligence; AI literacy; self-regulated learning; academic integrity*

INTRODUCTION

Academic writing is not simply the conversion of ideas into grammatical sentences. It is a recursive activity in which writers define purposes, identify audiences, organize evidence, monitor the developing argument, respond to feedback, and evaluate whether the final text meets disciplinary expectations. Metacognitive strategies matter because writers need to understand not only what they are doing but why they are doing it, whether it is working, and what adjustment is required. Research on metacognitive academic writing strategies provides a direct foundation for treating these processes as measurable constructs rather than as vague descriptions of "good writing" (Teng et al., 2022). Reviews of academic writing have likewise emphasized the complexity of planning, drafting, revising, and learning the conventions of scholarly communication (Chauhan, 2022; Barasa, 2024).

Generative AI introduces a consequential change to this regulatory environment. Writers may use a large language model to brainstorm, clarify a concept, restructure a paragraph, explain feedback, suggest search terms, or edit language. These affordances may reduce some barriers and make feedback more available, but they do not guarantee that the writer understands the content or

can judge the reliability of the output. Studies and discussions of AI-assisted writing describe both opportunities and limitations, including productivity, language support, critical-thinking concerns, misinformation, authorship, and integrity (Dergaa et al., 2023; Lingard, 2023; Nguyen et al., 2024; Kim et al., 2025). The resulting challenge is metacognitive: a writer must decide when to use AI, what to ask, how to assess the response, whether to accept or reject it, and how to remain accountable for the resulting text.

Metacognition is commonly understood as knowledge about cognition and regulation of cognition. In academic writing, metacognitive knowledge includes awareness of task demands, audience, genre, available strategies, and personal strengths or limitations. Regulation includes planning goals and procedures, monitoring comprehension and argument development, evaluating progress against criteria, and adapting strategies when a problem emerges. These processes are recursive rather than strictly linear: evaluation may lead to a new plan, and monitoring may trigger a return to sources or an earlier draft. Teng et al. (2022) provide a direct precedent for validating a metacognitive academic writing strategies measure and examining its relationship with writing performance in a foreign-language context. This work makes it important to preserve continuity with established writing-strategy constructs instead of creating an entirely technology-centered measure.

Writing development is also social and feedback-mediated. Dialogue, peer response, collaborative assessment, and feedback can help writers externalize judgments that would otherwise remain implicit (Schillings et al., 2023; Taufiqulloh et al., 2024; Fadhly et al., 2025). A scale should consequently capture how writers interpret and use feedback, not merely whether they receive it. A writer may obtain multiple comments but demonstrate little metacognition if the comments are accepted without evaluation. Conversely, a writer may reject a suggestion after comparing it with the assignment, source evidence, disciplinary conventions, and intended meaning.

AI-assisted writing can be understood as a sequence of human decisions around task framing, prompt construction, output inspection, selective adoption, verification, revision, and disclosure. Human–AI collaboration is not one uniform behavior; the writer may alternate between independent composing, AI consultation, and human review (Nguyen et al., 2024). Prompting itself can be strategic when the writer defines a bounded purpose, supplies relevant context without exposing sensitive information, and asks for a response that supports learning rather than replaces judgment (Giray, 2023). A reflexive workflow also requires writers to examine how AI shaped their thinking and the text, not just to record that a tool was used (Johnson & Paulus, 2024).

The literature identifies recurring tensions. AI may support fluency, motivation, and language development, while also producing plausible but inaccurate claims, flattening voice, encouraging overreliance, or obscuring authorship (Song & Song, 2023; Elsayed, 2024; Hidayatullah et al., 2025). Transparency and responsible use are therefore not external administrative add-ons: they are part of the writer's regulatory competence (Tang et al., 2024; Balalle & Pannilage, 2025). Critical AI literacy includes the ability to understand a tool's limits, judge the relevance of its outputs, and decide whether a particular use is acceptable for the task and institutional setting (Cardon et al., 2023; Hossain et al., 2025; Wang & Wang, 2025).

A useful instrument must distinguish at least four things that are often conflated: (a) general metacognitive writing regulation; (b) technical familiarity with AI tools; (c) attitudes toward AI; and (d) ethical and epistemic judgment when AI is used. High confidence in AI does not imply accurate verification. Frequent use does not imply effective planning. A writer may be highly metacognitive and deliberately choose not to use AI. The proposed scale treats AI-era competence as regulation across a writing task, including the decision to use, limit, or avoid AI. It also treats human responsibility as an observable strategy—for example, checking a claim against a source—rather than a self-description such as "I am an ethical writer."

Existing measures of metacognitive writing strategies cannot automatically be assumed to cover this new context. A scale may adequately capture conventional planning and revision while failing to represent AI-specific regulation, such as checking claims generated by a model, recognizing when a task should remain unaided, protecting confidential material, or disclosing substantive assistance. Conversely, simply adding items about frequency of AI use would confuse tool adoption with strategic competence. The need is for a construct-valid extension that retains the theoretical core of metacognition while capturing the changed decisions writers face in AI-supported environments.

This article addresses four questions: (1) Which established metacognitive academic writing processes should remain central to a scale for the AI era? (2) Which additional dimensions are justified by the 2022–2025 literature on AI-assisted academic writing? (3) How can these dimensions be translated into a provisional item pool and content-validity process? and (4) What empirical evidence is required before the instrument can be described as validated? The article's contribution is a proposed measurement framework and validation roadmap, not a claim that the proposed items have already passed psychometric testing.

METHOD

The source base was the bibliography supplied with the request. Entries published in 2022, 2023, 2024, or 2025 were retained for the conceptual synthesis; entries dated 2021, 2018, or 2026 were excluded from the eligible reference set, and repeated bibliographic entries were deduplicated where titles were clearly identical. The supplied list contains work on metacognitive writing strategy validation, academic writing pedagogy and assessment, AI-assisted writing processes, AI literacy, ethical use, and academic integrity. The date restriction is applied to the reference set, not to a claim that every eligible publication was retrieved or fully analyzed.

A key limitation must be explicit: the supplied material contains bibliographic records, not the full texts, abstracts, search histories, or study-level data needed for a reproducible systematic review. Consequently, the present procedure is a bibliography-led conceptual meta-synthesis. It organizes the themes signaled by the supplied records and uses them to generate a construct map and provisional item pool. It does not report a PRISMA flow, claim exhaustive database coverage, calculate pooled effects, or attribute specific empirical findings to papers unless such claims are supported by the bibliographic information and established topic of the work. The proposed framework should be followed by full-text retrieval and formal screening before journal submission as a systematic review.

The conceptual synthesis proceeded through four interpretive stages. First, the literature was grouped by its apparent focus: metacognitive writing strategies; writing pedagogy and feedback; AI-assisted writing processes; AI literacy and human–AI collaboration; and ethical, transparency, and integrity concerns. Second, these groups were mapped onto a regulatory sequence: anticipate and plan; enact and monitor; evaluate evidence and outputs; revise and transfer learning; and account for decisions. Third, candidate dimensions were checked against two inclusion principles: each dimension had to represent a strategy or decision process rather than a general attitude, and each AI-related dimension had to connect to a recognizable writing task. Fourth, the resulting dimensions were translated into a provisional pool of self-report items, with wording designed to refer to behavior during a recent academic writing task.

No numerical synthesis was conducted. No participants were recruited, no expert ratings were collected, and no factor analysis was performed. Accordingly, all item counts, dimensional structures, and validation thresholds presented below are design proposals rather than observed results. This distinction protects the manuscript from presenting an instrument as validated before the necessary evidence exists. The overall procedure, from source filtering through the four interpretive stages to the provisional outputs, is summarized in Figure 1.

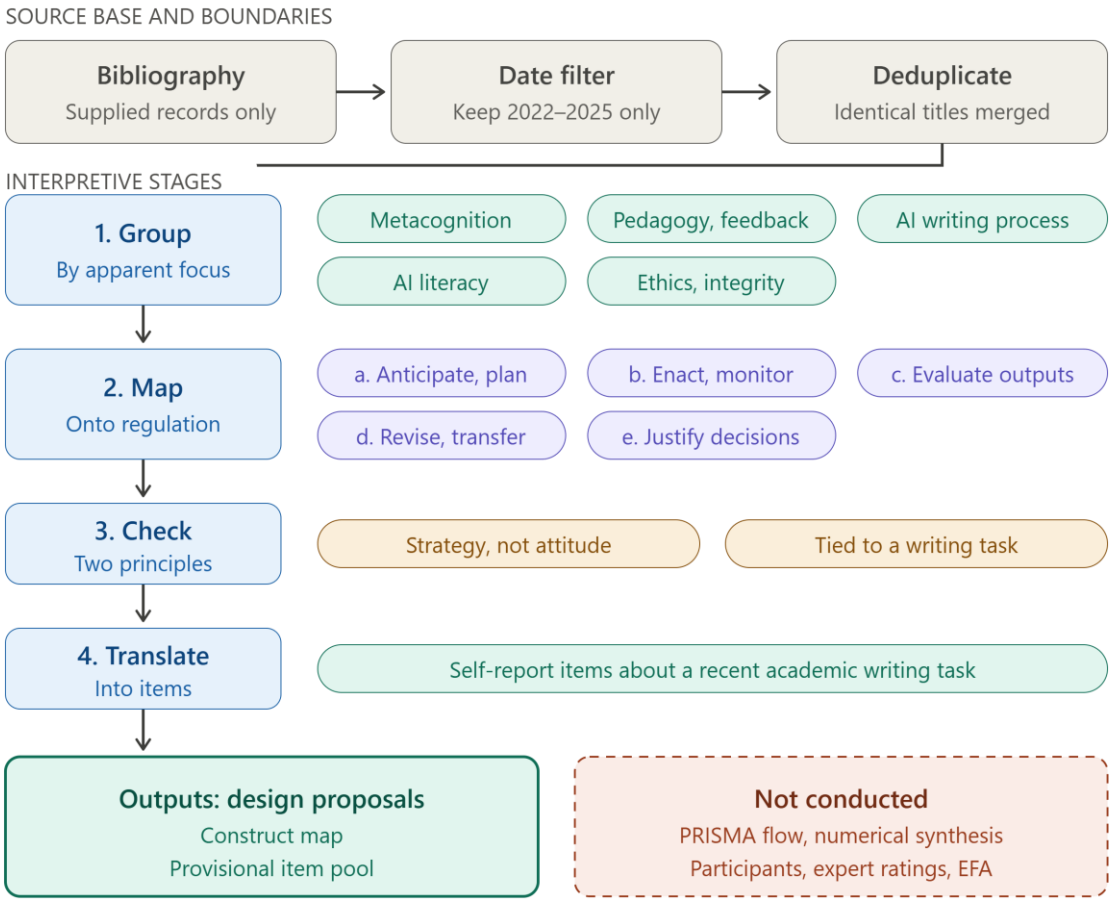


Figure 1. *Bibliography-led conceptual meta-synthesis procedure. Outputs are design proposals; no empirical validation was conducted*

RESULTS AND DISCUSSION

Proposed construct framework

Synthesis themes

The bibliography supports a coherent conceptual direction: the established regulatory cycle remains necessary, but AI changes the objects of monitoring and evaluation. Writers must monitor not only their own comprehension and argument but also the contribution, uncertainty, and limitations of AI-generated suggestions. They must evaluate not only whether a paragraph reads smoothly but whether its claims are source-grounded, relevant, and appropriately attributed. Finally, they must regulate the boundary between assistance and delegation. These themes lead to six proposed dimensions.

Table 1. *Bibliography-led synthesis matrix*

Theme cluster	Core concern	Metacognitive interpretation	Scale implication
Conventional writing regulation	Goals, planning, monitoring, revision	Control of the writing process	Retain planning and monitoring items
Feedback and dialogue	Interpreting comments and revising	Evaluate feedback rather than obey it	Measure reasoned acceptance/rejection
Human–AI collaboration	Different patterns of tool use	Allocate tasks intentionally	Measure task–tool fit and bounded prompting
AI literacy and criticality	Reliability, limitations, hallucinations	Epistemic vigilance	Measure claim/source verification
Integrity and transparency	Ownership, disclosure, confidentiality	Accountability for decisions and text	Measure policy awareness and transparent reporting

p-ISSN 2252-7427, e-ISSN 2541-3635
https://journal.uniku.ac.id/index.php/IEFLJ/index

Learning and transfer	Skill development beyond one assignment	Reflect and adapt strategy	Measure post-task reflection and independent capability

The table is an interpretive map, not a frequency count of coded full-text studies. Its purpose is to make the reasoning from literature clusters to proposed measurement dimensions visible and auditable.

The six proposed dimensions

Dimension 1, strategic planning, concerns goal setting, task analysis, audience awareness, and selecting a writing plan before drafting or consulting AI. AI-era planning includes deciding whether AI is appropriate for the task and identifying what the writer must do independently. Dimension 2, metacognitive monitoring, concerns checking understanding, argument coherence, evidence use, and progress during drafting. In AI-assisted work it also includes monitoring whether an AI suggestion changes the intended claim or obscures the writer's meaning.

Dimension 3, evaluative judgment and revision, concerns comparing drafts with task requirements, disciplinary standards, feedback, and evidence. It includes reasoned acceptance, adaptation, or rejection of suggestions. Dimension 4, AI-task orchestration, concerns choosing a bounded role for AI, framing useful prompts, and managing transitions between independent writing and AI assistance. This is not a test of prompt-engineering sophistication; it is a measure of intentional task management.

Dimension 5, critical verification and epistemic vigilance, concerns checking AI-generated claims, references, quotations, and explanations against credible sources and the writer's own knowledge. Dimension 6, agency, transparency, and ethical accountability, concerns maintaining intellectual ownership, protecting confidential information, following applicable policies, and disclosing AI assistance when required. The dimensions are related but should not be presumed to form six statistically distinct factors until data support that interpretation.

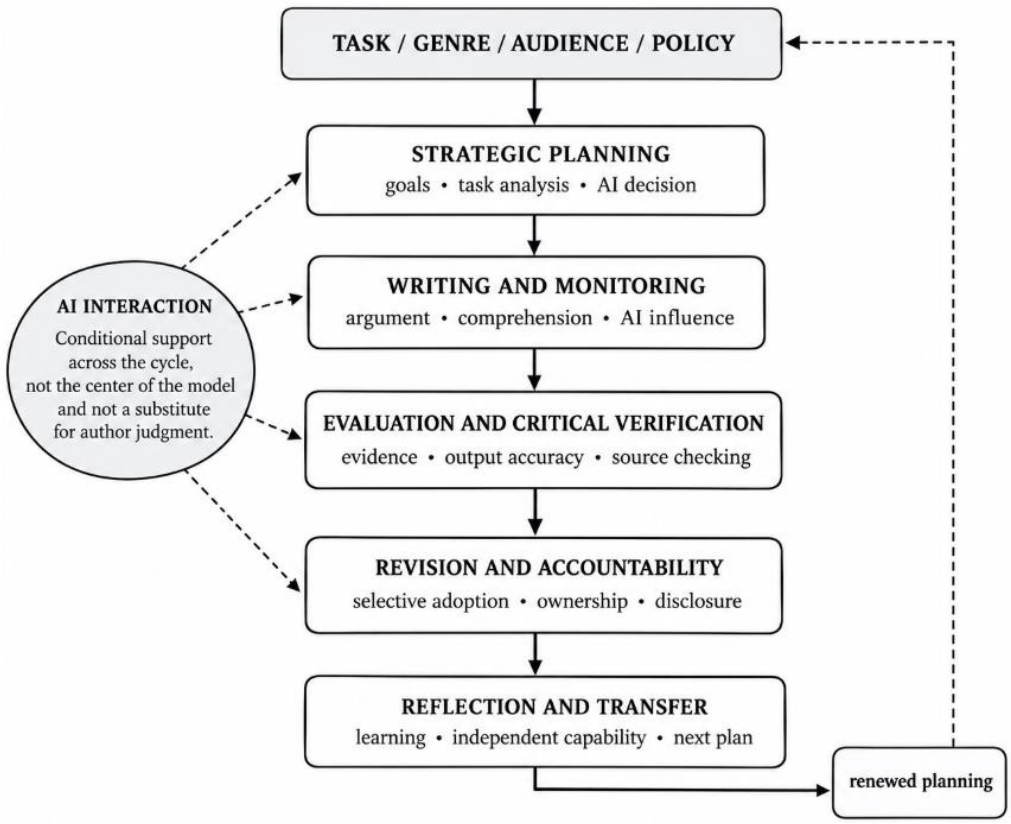


Figure 2. *Proposed conceptual model: metacognitive regulation in AI-Assisted academic writing*

Table 1. *Construct dimensions and behavioral indicators of the metacognitive academic writing strategies scale for the AIera, with potential contamination threats*

Dimension	Operational definition	Illustrative behavior	Potential contamination to avoid
1. Strategic planning	Setting goals and selecting procedures in relation to the writing task	Identifies purpose, audience, criteria, and whether AI is suitable	General confidence or frequency of AI use
2. Metacognitive monitoring	Tracking comprehension, argument development, and process quality	Checks whether each section supports the central claim	Self-rated writing ability alone
3. Evaluative judgment and revision	Judging feedback and revising on the basis of evidence and criteria	Accepts, adapts, or rejects suggestions with reasons	Automatic compliance with feedback
4. AI-task orchestration	Allocating bounded roles to AI and self across task stages	Requests critique of a draft but retains final decisions	Technical prompt fluency alone
5. Critical verification	Testing claims and outputs against credible evidence	Verifies generated references and factual claims in source records	Trust in polished language or model confidence
6. Agency, transparency, and ethics	Maintaining responsibility and following disclosure/privacy expectations	Records substantive AI assistance when policy requires it	Social desirability statements with no behavioral anchor

Provisional metacognitive academic writing strategies scale for the AI era

The proposed instrument is called the Metacognitive Academic Writing Strategies Scale for the AI Era (MAWSS-AI). It is intended for university students and postgraduate writers completing academic assignments, research papers, theses, or manuscripts. Respondents should be instructed to think about one clearly defined recent writing task, preferably completed within the previous four weeks. A task-specific instruction reduces the risk that respondents answer according to general beliefs about themselves rather than recallable behavior.

Suggested response scale: 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Almost always. A "not applicable" option may be considered for items that depend on AI use, but it should not automatically be scored as the lowest response because non-use may reflect a deliberate and appropriate choice. For an initial study, items can be administered as a 30-item pool and reduced only after content review and empirical analysis.

Table 3. *Provisional item pool (30 items)*

Code	Dimension	Draft item
P1	Planning	Before I begin writing, I identify the purpose and intended audience of the text.
P2	Planning	I translate the assignment or research goal into specific writing objectives.
P3	Planning	I identify the evidence and sources I will need before drafting major claims.
P4	Planning	I decide which parts of the task, if any, would benefit from AI assistance.
P5	Planning	I set boundaries for what I will do myself and what I may ask an AI tool to support.
M1	Monitoring	While drafting, I check whether each section contributes to my central argument.
M2	Monitoring	I notice when I no longer understand a claim or explanation in my draft.
M3	Monitoring	I check whether AI suggestions preserve the meaning I intended to express.
M4	Monitoring	I pause and change my approach when my writing plan is not working.
M5	Monitoring	I keep track of which ideas or wording came from sources, AI suggestions, or my own reasoning.
E1	Evaluation/revision	I compare my draft with the task criteria before considering it complete.
E2	Evaluation/revision	I evaluate feedback before deciding whether to use it.

E3	Evaluation/revision	I revise AI-generated or AI-edited wording when it does not fit my argument or voice.
E4	Evaluation/revision	I can explain why I accepted, modified, or rejected an important suggestion.
E5	Evaluation/revision	After revising, I check whether the changes improve meaning and evidence, not just fluency.
A1	AI orchestration	When I use AI, I state a specific purpose for the assistance I request.
A2	AI orchestration	I provide only the context needed for the task rather than submitting unnecessary sensitive information.
A3	AI orchestration	I change my request when the response does not address the writing problem I intended to solve.
A4	AI orchestration	I use AI feedback as one input among others rather than as the final authority.
A5	AI orchestration	I choose to work without AI when using it would conflict with the task or learning goal.
V1	Verification	I verify factual claims suggested by AI against credible sources.
V2	Verification	I check AI-suggested references in the original publication or a reliable bibliographic record.
V3	Verification	I distinguish a plausible-sounding explanation from one supported by evidence.
V4	Verification	I check quotations, statistics, and specific details instead of assuming they are accurate.
V5	Verification	I revise or remove AI-assisted content when I cannot verify its accuracy.
T1	Agency/ethics	I take responsibility for the claims and conclusions in the final text, including AI-assisted parts.
T2	Agency/ethics	I follow the relevant course, university, publisher, or journal rules for AI use.
T3	Agency/ethics	I disclose AI assistance when the applicable policy or assignment requires it.
T4	Agency/ethics	I retain enough records of AI assistance to explain how it affected my work when necessary.
T5	Agency/ethics	After using AI, I reflect on what I learned and what I still need to be able to do independently.

Scoring and interpretation

For exploratory use, calculate mean scores for each proposed dimension only after checking item direction, missingness, and dimensional structure. A total score should not be used until evidence supports a higher-order or sufficiently coherent general factor. Higher scores would indicate more frequent self-reported use of the listed strategies, not objectively demonstrated writing quality, ethical character, or AI literacy. Scores should be interpreted alongside writing samples, process logs, interviews, or task performance where possible. AI-dependent items require careful handling: respondents who do not use AI can still demonstrate strong metacognitive planning and may choose non-use intentionally.

No cut scores are recommended at this stage. Categorizing students as "low," "moderate," or "high" metacognitive writers without norming or criterion evidence would be premature. The instrument should initially be treated as a research measure for investigating patterns, not as a high-stakes screening tool or a basis for misconduct judgments.

Content validation and psychometric validation plan

Stage 1: Expert review and content validity

A panel of approximately 6–10 reviewers is proposed, with expertise spanning second-language writing, educational measurement, metacognition/self-regulated learning, AI literacy, and research integrity. Reviewers should rate each item for relevance, clarity, representativeness, and age/task appropriateness using a four-point scale. They should also flag double-barreled wording, unnecessary technical vocabulary, overlap between dimensions, and policy-dependent assumptions. Item-level content validity indices (I-CVI) and scale-level averages (S-CVI/Ave) may be reported, but values must be calculated from actual expert responses rather than asserted in advance. A modified kappa can help account for chance agreement. Numerical cutoffs should be selected and justified from the chosen content-validity methodology, not treated as universal laws.

Experts should also examine whether the six dimensions adequately cover the construct without overrepresenting AI use. In particular, the pool must not imply that using AI is always desirable or that disclosure rules are identical across assignments, institutions, disciplines, and journals. Items should refer to applicable policies rather than impose a universal disclosure requirement for every minor spelling correction. Ethical judgment includes knowing when to consult the relevant policy.

Stage 2: Cognitive interviews and response-process evidence

Conduct think-aloud interviews with approximately 8–15 students representing relevant proficiency levels, disciplines, and AI-use experiences. Ask participants to paraphrase each item, recall a specific writing episode, explain how they chose a response, and identify any uncertainty. Cognitive interviews can reveal whether "verify," "AI assistance," "central argument," and "disclose" are understood consistently. Revise items when respondents interpret them as attitude statements, technology-use frequency, or moral self-assessments rather than concrete strategies. Translate and back-translate if a local-language version is needed, followed by reconciliation and cognitive testing rather than literal translation alone.

Stage 3: *Pilot testing and item analysis*

A pilot sample of around 100–200 respondents may be useful for assessing completion time, missing responses, response distributions, floor/ceiling effects, item-total relationships, and preliminary reliability. These numbers are planning suggestions, not a guarantee of adequate power. Sample size should be based on model complexity, communalities, expected factor loadings, estimator, and intended subgroup analyses. Examine whether respondents can distinguish "not applicable" from "never," and whether AI-specific items create a separate factor simply because they are conditional on tool use.

Stage 4: Exploratory and confirmatory factor analysis

Use an exploratory factor analysis (EFA) to investigate the dimensional structure, with a factor-retention approach that considers parallel analysis, theoretical interpretability, and item quality. For ordinal Likert items, polychoric correlations and an estimator appropriate for categorical data may be suitable. An oblique rotation is reasonable because metacognitive dimensions are expected to correlate. The EFA sample should not be reused uncritically as the sole confirmatory sample. A separate sample should test the proposed six-factor model and plausible alternatives, including a correlated-factor model and, only if theoretically defensible, a higher-order model.

Report factor loadings, cross-loadings, factor correlations, residuals, and fit indices with confidence intervals or uncertainty where possible. Commonly reported indices include CFI, TLI, RMSEA, and SRMR, but fit should not be reduced to mechanical threshold chasing. Model decisions should combine statistical evidence, content coverage, interpretability, and parsimony. Items should not be deleted only to improve fit if their removal creates a serious content gap. If the structure differs by AI-use status or language background, investigate the source of that difference before treating the same score as comparable.

Stage 5: Reliability and validity evidence

Internal consistency should be estimated for each supported subscale using omega and, where appropriate, alpha. Test–retest reliability can assess score stability over a short interval when no major writing intervention is expected. Reliability is a property of scores in a particular population and use, not a permanent feature of an instrument. Convergent evidence may be examined through associations with established metacognitive writing strategy measures, self-regulated learning, reflective writing behaviors, or independent ratings of process quality. Discriminant evidence should test whether the scale is distinguishable from general AI enthusiasm, confidence, frequency of use, and social desirability. Criterion-related evidence may include writing quality assessed by blinded raters, source-verification accuracy, quality of revision decisions, or process-trace indicators.

Predictive analyses should be preregistered where possible and should control for relevant factors such as prior writing proficiency, language background, discipline, genre, task stakes, and access to AI. Correlation with writing performance would not alone prove that the scale causes better writing. Likewise, an association between ethical accountability scores and self-reported disclosure does not establish compliance with every institutional policy.

p-ISSN 2252-7427, e-ISSN 2541-3635
https://journal.uniku.ac.id/index.php/IEFLJ/index

Stage 6: Measurement invariance and fairness

If the scale is used to compare groups, test configural, metric, and scalar invariance across relevant groups, potentially including undergraduate and postgraduate writers, L1 and L2 English writers, disciplines, language versions, and AI-use profiles. Differential item functioning may be particularly important for AI-dependent items because access, institutional policy, and prior experience vary. Invariance should not be assumed because the wording appears identical. Where invariance is not supported, report the limitation and avoid unsupported mean comparisons. The instrument should also be checked for accessibility and for bias against writers who intentionally limit AI use.

Table 4. *Validation evidence matrix*

Evidence source	Procedure	Evidence to report	Decision enabled
Content	Expert ratings and qualitative comments	I-CVI, S-CVI/Ave, reviewer rationale	Retain, revise, or remove items
Response process	Think-aloud/cognitive interviews	Misinterpretations, recall strategies, burden	Improve item clarity and instructions
Internal structure	EFA and CFA in suitable samples	Loadings, factor correlations, fit, residuals	Support or revise dimensional model
Reliability	Omega/alpha; test–retest where feasible	Estimates with confidence intervals	Judge score precision for intended use
Convergent/ discriminant	Correlations with related and distinct constructs	Expected and unexpected association pattern	Assess construct boundaries
Criterion-related	Writing samples, verification tasks, revision traces	Associations with external performance	Test practical relevance
Invariance/fairness	Multi-group models and DIF analyses	Invariance evidence and item bias	Bound group comparisons and improve equity

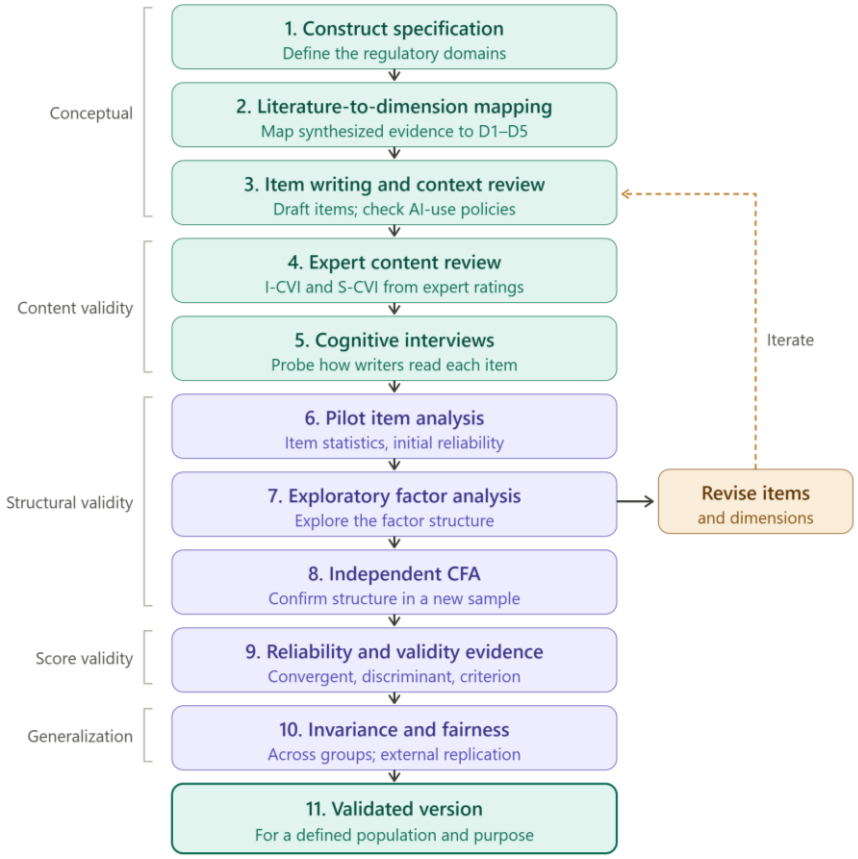


Figure 3. *Proposed validation pathway*

The proposed MAWSS-AI framework reframes metacognition for an environment in which writing processes may be distributed across a human writer, digital resources, peers, and AI systems.

The key theoretical continuity is regulation: writers still need to plan, monitor, evaluate, revise, and reflect. The key change is that the object of regulation expands. Writers monitor the impact of AI suggestions on meaning, evaluate generated claims against sources, decide which tasks to delegate, and account for the contribution of tools to the final text. This approach is consistent with research that treats AI writing as human–AI collaboration rather than a single undifferentiated activity (Nguyen et al., 2024) and with work emphasizing critical AI literacy and human agency (Cardon et al., 2023; Wang & Wang, 2025; Han, 2025).

A practical implication is that scale development should not begin by asking whether students use AI "often." Frequency is an inadequate proxy for strategic quality. A student who uses AI at every stage without verifying output may score high on use frequency but low on critical regulation. A student who chooses not to use AI for a sensitive or tightly constrained task may be making a sophisticated metacognitive decision. The proposed items therefore emphasize purpose, boundaries, verification, and accountability. This design also reduces the risk that the instrument will become obsolete as particular platforms or interface conventions change.

The framework gives a central place to verification because fluent language can create an illusion of reliability. Writers need to verify citations, claims, quotations, and statistics rather than treating polished output as evidence. This aligns with concerns about misinformation and ethical risks in AI-supported academic work (Elsayed, 2024; Mohamed et al., 2024; Tang et al., 2024). Verification should be measured behaviorally: "I check references in the original publication" is more informative than "I know AI can make mistakes." The latter captures awareness; the former captures a regulatory strategy.

The agency and transparency dimension likewise requires nuance. Policies vary by institution, journal, discipline, and task, and disclosure expectations can change. A good item does not impose one universal rule; it asks whether writers consult and follow the applicable rule and disclose when required. This is compatible with broader debates about academic integrity, authorship, and responsible use (Balalle & Pannilage, 2025; da Veiga, 2025; Gulumbe et al., 2025). Importantly, self-report scales cannot certify ethical conduct. Scores should never be used alone to accuse a student of misconduct or infer that a writer has violated a policy.

The proposed framework also has pedagogical value. Teachers can use the dimensions as a framework for reflective prompts: What was your goal? Why did you consult AI at this point? Which suggestions did you reject? How did you verify claims? What did you learn that you can now do independently? Such questions turn AI use from a hidden shortcut into an inspectable learning process. Feedback and collaborative assessment research suggests that writing improves through dialogue and evaluation, not simply through the receipt of comments (Schillings et al., 2023; Taufiqulloh et al., 2024). AI feedback can be integrated into that ecology, but it should not displace teacher, peer, source-based, or self-evaluation.

For multilingual writers, AI may provide valuable language support, but a scale must distinguish language polishing from intellectual delegation. An item such as "I revise AI-assisted wording when it changes my intended meaning" allows language support without equating polished English with strong reasoning. This matters because academic writing is shaped by genre, disciplinary norms, language resources, and unequal access to feedback. Scale users should examine whether items are interpreted differently across language backgrounds and whether the instrument unfairly rewards access to paid tools or prior prompt-engineering experience.

Limitations

This manuscript has several limitations. First, it is based on the supplied bibliography, not a complete set of retrieved full texts. The source list includes duplicates, variable citation formatting, and works whose titles suggest relevance but whose detailed methods and findings cannot be confirmed from the record alone. Second, the synthesis was not conducted through a preregistered systematic search and screening protocol; no database-level search counts or exclusion logs are available. Third, the six-dimensional structure and 30 items are provisional. No experts have rated the items, no students have completed cognitive interviews, and no quantitative data have been collected. Fourth, the item pool relies on self-report and is vulnerable to recall error, social desirability, and differences in how respondents interpret AI policies. Finally, AI systems and institutional rules evolve rapidly, so the scale will require periodic review without compromising construct stability.

Future research

The next phase should retrieve full texts for the eligible studies and conduct a transparent systematic or meta-synthetic review with documented search strings, databases, inclusion criteria, study characteristics, and a traceable coding framework. Researchers should distinguish empirical studies from conceptual essays, editorials, reviews, and policy analyses. Each included source should be coded for population, writing context, AI function, metacognitive process, reported outcome, and ethical dimension. Two or more reviewers should independently code a subset and resolve disagreements through documented discussion. The resulting full-text synthesis could refine or remove dimensions before expert validation.

A subsequent mixed-method instrument-development study should combine expert review, cognitive interviews, pilot testing, and independent factor-analytic samples. To avoid common-method bias, self-report scores should be compared with task-based indicators such as whether a participant verifies AI-generated citations, identifies fabricated claims, provides reasons for revision decisions, or maintains a transparent process log. Longitudinal studies could examine whether explicit metacognitive instruction changes strategy use and writing quality over time. Intervention studies could compare conventional writing instruction with instruction that explicitly teaches AI-task orchestration, verification, and reflective accountability.

Cross-cultural adaptation is especially important. The scale should be tested in different educational systems, language backgrounds, and disciplinary writing contexts. A culturally responsive version should retain the construct while adapting policy examples and terminology. Researchers should also examine whether AI access, disability accommodations, disciplinary conventions, and institutional restrictions affect item responses. Finally, score interpretations should be linked to clearly stated purposes: formative reflection, research, or program evaluation. High-stakes decisions require substantially stronger evidence than exploratory classroom use.

CONCLUSION

This article proposes a metacognitive academic writing strategies scale for the AI era that preserves core writing regulation while extending measurement to AI-task orchestration, critical verification, and ethical accountability. Its central claim is that responsible AI-assisted writing depends less on the amount of tool use than on the quality of the writer's decisions. The proposed six dimensions and 30-item pool offer a transparent starting point for scale development, not a completed validation. Before the MAWSS-AI can be described as validated, researchers must establish content coverage, response-process clarity, internal structure, reliability, relations with relevant constructs and performance criteria, and fairness across populations. The ultimate goal is to measure writers' capacity to remain purposeful, critical, reflective, and accountable when AI becomes part of the academic writing environment.

AI DISCLOSURE STATEMENT

Generative AI tools (ChatGPT-4) were used for language editing, manuscript formatting, content structuring, and consistency checking.

REFERENCES

Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. *Social Sciences & Humanities Open*, *11*, 101299.

Barasa, D. (2024). Demystifying the discourse: Techniques to effective academic writing. *Journal of Research and Academic Writing, 1*(1), 13-21.

Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly, 86*(3), 257-295.

Chauhan, P. (2022). Fundamentals of academic writing: A literature review. *Journal of NELTA, 27*(1-2), 161-180.

da Veiga, A. (2025). Ethical guidelines for the use of generative artificial intelligence and artificial intelligence-assisted tools in scholarly publishing: a thematic analysis. *Science Editing, 12*(1), 28-34.

Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. B. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport, 40*(2), 615-622.

Elsayed, H. (2024). The impact of hallucinated information in large language models on student learning outcomes: A critical examination of misinformation risks in AI-assisted education. *Northern Reviews on Algorithmic Research, Theoretical Computation, and Complexity, 9*(8), 11-23.

Fadhly, F. Z. (2022). Enhancing the academic writing of EFL learners: An analysis of effective strategies

through meta-synthesis. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics), 6*(2), 397–410.

Fadhly, F. Z., Solihat, D., & Wihadi, M. (2025). Unraveling the power of written feedback: A meta-synthesis on student academic writing. *Elsya: Journal of English Language Studies, 7*(3), 331–346.

Gulumbe, B. H., Audu, S. M., & Hashim, A. M. (2025). Balancing AI and academic integrity: What are the positions of academic publishers and universities?. *AI & Society, 40*(3), 1775-1784.

Han, Y. (2025). Beyond the algorithm: Reconciling generative AI and human agency in academic writing education. *International Journal of Learning and Teaching, 11*(1), 39-42.

Hidayatullah, M. H., Suryati, N., Cahyono, B. Y., & Mawaddah, N. (2025). A systematic literature review of artificial intelligence in academic writing: Challenges and opportunities. *Journal of Research on English and Language Learning (J-REaLL), 6*(2), 249-266.

Hossain, Z., Çelik, Ö., & Hınız, G. (2025). Exploring EFL students' AI literacy in academic writing: Insights into familiarity, knowledge and ethical perceptions. *Journal of Theoretical Educational Sciences, 18*(1), 157-181.

Johnson, C. W., & Paulus, T. (2024). Generating a reflexive AI-assisted workflow for academic writing. *The Qualitative Report, 29*(10), 2772-2792.

Lingard, L. (2023). Writing with ChatGPT: An illustration of its capacity, limitations & implications for academic writers. Perspectives on medical education, 12(1), 261-270.

Mohamed, Y. A., Mohamed, A. H. H., Khanan, A., Bashir, M., Adiel, M. A., & Elsadig, M. A. (2024). Navigating the ethical terrain of AI-generated text tools: A review. *IEEE Access, 12*, 197061-197120.

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in higher education, 49*(5), 847-864.

Schillings, M., Roebertsen, H., Savelberg, H., & Dolmans, D. (2023). A review of educational dialogue strategies to improve academic writing skills. *Active Learning in Higher Education, 24*(2), 95-108.

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in psychology, 14*, 1260843.

Tang, A., Li, K. K., Kwok, K. O., Cao, L., Luong, S., & Tam, W. (2024). The importance of transparency: Declaring the use of generative artificial intelligence (AI) in academic writing. *Journal of Nursing Scholarship, 56*(2), 314-318.

Taufiqulloh, T., Fadhly, F. Z., & Rosdiana, I. (2024). Effects of collaborative assessment on undergraduate students' writing performance. *Studies in English Language and Education, 11*(3), 1574–1595

Taufiqulloh, T., Fadhly, F. Z., Rosdiana, I., Ferrer, C. N., Ratsamemonthon, C., Nindya, M. A., & Irawan, N. (2025). Comprehensive review of writing assessments in EFL contexts: A meta-synthetic study. *International Journal of Language Testing, 15*(1), 193–213.

Teng, M. F., Qin, C., & Wang, C. (2022). Validation of metacognitive academic writing strategies and the predictive effects on academic writing performance in a foreign language context: Validation of metacognitive academic writing strategies and the predictive effects on academic writing performance in a foreign language context. *Metacognition and learning, 17*(1), 167-190.

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing, 67*, 101187.