International Journal of Learning, Teaching and Educational Research
Vol. 24, No. 10, pp. 624-643, October 2025
https://doi.org/10.26803/ijlter.24.10.30
Received Jul 22, 2025; Revised Aug 25, 2025; Accepted Sept 4, 2025

Exploring the Effects of an Automated Writing Evaluation Tool on Metacognitive Engagement in Persuasive Writing

Xiaolan Wang*
Faculty of Education
Universiti Kebangsaan Malaysia, Malaysia

Supyan Hussin
Institute of Ethnic Studies (KITA)
Universiti Kebangsaan Malaysia, Malaysia

Azlina Abdul Aziz
Faculty of Education
Universiti Kebangsaan Malaysia, Malaysia

Abstract. This study explored the impact of automated writing evaluation (AWE) on the development of metacognitive awareness in English as a foreign language (EFL) students' persuasive writing within Chinese higher education. Grounded in self-regulated learning theory and Flavell's (1979) metacognitive framework, the research investigates how AWE influences students' abilities to plan, monitor, and evaluate their writing processes. Employing a single-group mixed-methods design over a 16-week intervention period, data were collected from 100 students through the Metacognitive Awareness Writing Questionnaire (MAWQ), reflective journals, and post-intervention interviews with a randomly selected subset of 10 participants. Quantitative results revealed negligible overall gains in metacognitive awareness but a recalibration of self-perceptions in areas such as planning and conditional knowledge. In contrast, qualitative data offered a more nuanced view. The students reported increased attention to text structure and grammar and demonstrated selective adoption of AWE feedback. However, many expressed uncertainties when faced with ambiguous or overly general suggestions, highlighting the ongoing need for teacher support. These findings suggest that while AWE tools such as PIGAI may effectively facilitate surface-level revisions, their capacity to foster deeper metacognitive engagement is limited without instructional scaffolding. To enhance pedagogical outcomes, it is recommended that AWE systems be integrated into a broader instructional framework, supported by

*Corresponding Author: Xiaolan Wang; *p134016@siswa.ukm.edu.my*

explicit strategy training. Incorporating clearer rubrics and more contextualized, explanation-rich feedback may further promote students' independent and strategic engagement with the writing process.

Keywords: automated writing evaluation; metacognitive awareness; persuasive writing; EFL learners; self-regulated learning

1. Introduction

Automated writing evaluation (AWE) systems have gained prominence in educational settings as tools designed to provide immediate, formative feedback on student writing, thereby potentially enhancing writing proficiency (Wilson & Czik, 2016). These systems utilise natural language processing (NLP) algorithms to assess various aspects of writing, including grammar, coherence, and organisation (Stevenson & Phakiti, 2019). Proponents argue that AWE can reduce educators' grading workload while offering students personalised, real-time feedback, which is particularly beneficial in large classrooms (Al Mamari, 2020).

Concurrently, *metacognitive awareness*—defined as an individual's ability to reflect upon, understand, and regulate their own learning processes (Flavell, 1979)—has been identified as a critical factor in effective writing. Research suggests that students who employ metacognitive strategies, such as planning, monitoring, and evaluating their writing, tend to perform better in academic writing tasks (Teng, 2019). Nevertheless, despite the growing integration of AWE tools into English as a foreign language (EFL) writing instruction, many students continue to struggle with the consistent application of metacognitive strategies. These difficulties are particularly apparent in the planning, monitoring, and self-evaluation stages of the writing process (Teng & Yue, 2023).

While AWE systems are designed to foster self-regulated learning by offering immediate and individualised feedback, their actual contribution to metacognitive development remains insufficiently understood. Existing studies largely demonstrate improvements in surface-level writing features, such as grammatical accuracy and vocabulary use (Gavina & Ibay-Pamo, 2024). However, there is limited empirical evidence concerning how these systems influence students' deeper cognitive regulation during writing, particularly reflective engagement and strategic awareness.

Although some scholars suggest that AWE tools may promote learner self-regulation and reflective thinking (Fu et al., 2024), others caution that excessive reliance on automated feedback can limit critical thinking and weaken students' ability to self-edit (Kellogg, 2022). Moreover, empirical evidence concerning how students interact with AWE feedback in genres that demand high rhetorical awareness—such as persuasive writing—remains scarce.

The problem, therefore, is that while AWE has been shown to support linguistic accuracy, its potential impact on metacognitive engagement in complex writing tasks is underexplored. In particular, there is insufficient empirical investigation into whether and how AWE feedback enables EFL learners to plan, monitor, and

evaluate their writing strategically in persuasive contexts. This study seeks to address this gap by examining the effects of AWE on Chinese university students' metacognitive awareness in persuasive writing. Specifically, it investigates the extent of change in students' metacognitive behaviours after sustained use of AWE, identifies which dimensions of metacognitive awareness are most affected, and explores how learners perceive the role of automated feedback in shaping their writing strategies.

Given these conflicting perspectives, there is a need to examine whether AWE feedback can foster metacognitive engagement in persuasive writing, which typically requires logical structuring, evidence-based reasoning, and clear argumentation. Understanding how students perceive and utilise AWE feedback in this context is essential for informing pedagogical practices and refining the design of intelligent writing systems. The present study, therefore, investigates the impact of AWE on students' metacognitive awareness in persuasive writing tasks using a pre- and post-intervention mixed-methods design.

This study contributes empirical evidence on how AWE affects students' metacognitive behaviours, offering new insights into the role of feedback technologies in writing instruction. The findings may provide practical implications for EFL educators and curriculum designers in scaffolding AWE use to enhance both writing performance and metacognitive development. Additionally, the study may guide developers in aligning automated feedback systems with cognitive and metacognitive learning objectives, echoing Ramadhanti and Yanda's (2021) conclusion that increased metacognitive awareness can facilitate more critical self-appraisal of writing strategies.

This research seeks to address the following questions:
1. To what extent does students' metacognitive awareness in persuasive writing change after using AWE?
2. Which specific dimensions of metacognitive awareness show the most significant changes after using AWE?
3. How do students perceive the role of AWE in enhancing their metacognitive writing strategies?

2. Literature Review

2.1 Review of Related Literature

Over the past two decades, increasing attention has been paid to the role of metacognition in second language (L2) writing. Drawing on Flavell's (1979) foundational distinction between metacognitive knowledge and metacognitive regulation, writing researchers have conceptualised metacognitive awareness as a key factor influencing students' ability to plan, monitor, evaluate, and revise their texts (Teng et al., 2022). This line of inquiry suggests that fostering metacognitive strategies leads to enhanced writing performance, particularly among EFL learners, who often struggle with the cognitive demands of academic writing.

Metacognitive knowledge typically encompasses declarative, procedural, and conditional knowledge about writing, while metacognitive regulation involves strategic actions such as planning, monitoring, and revising (Schraw & Moshman, 1995). In the context of L2 writing, these elements are closely intertwined. For instance, a writer's ability to identify when and how to deploy specific rhetorical strategies is underpinned by both their metacognitive awareness and linguistic competence (Shokri & Mousavi, 2024). The pedagogical implication is that students must be guided not only in how to write but also in how to think about writing.

The incorporation of AWE tools into EFL classrooms has been posited as a potentially effective way to scaffold metacognitive-strategy use. Automated writing evaluation systems such as Grammarly, Criterion, and China-based platforms like PIGAI offer immediate, individualised feedback on surface-level errors (e.g., grammar, punctuation, and vocabulary) and sometimes on more global features such as coherence or organisation (Ranalli, 2018). Proponents argue that such feedback encourages iterative self-review and supports learners' ability to monitor and revise their drafts independently, thereby activating metacognitive processes.

This claim is not without contention, however. Some scholars argue that while AWE tools are effective in enhancing surface-level accuracy, they do not consistently foster higher-order thinking, such as idea development, argument construction, or audience awareness (Link et al., 2014). The lack of contextual sensitivity in AWE feedback also limits its capacity to guide learners in complex rhetorical decision-making (Woodworth, 2022). Moreover, the reliability and clarity of automated feedback have been questioned; learners often report contradictory or vague suggestions, which may result in confusion and overreliance on machine judgement (Bai & Hu, 2016).

In terms of metacognitive development, the impact of AWE remains underexplored and, at times, contradictory. On the one hand, repeated exposure to feedback may prompt students to evaluate their performance more critically, fostering conditional knowledge and self-evaluation. On the other hand, reliance on automated corrections may discourage deep engagement with the writing process and limit students' strategic autonomy (Koltovskaia, 2020). Particularly in EFL contexts where writing instruction is often exam-driven, AWE use may reinforce a narrow focus on accuracy over rhetorical effectiveness.

In the Chinese tertiary education context, the situation is even more complex. Many college-level English writing courses emphasise grammatical correctness and coherence as primary indicators of proficiency, while metacognitive-strategy instruction remains marginal (Teng, 2019). Although AWE platforms such as PIGAI are widely adopted, few empirical studies have examined their impact on students' metacognitive awareness. This gap is especially pressing given that university students are expected to engage in increasingly sophisticated academic writing, which demands not only linguistic knowledge but also strategic thinking and reflective judgement.

Thus, while there is theoretical support for the use of AWE in developing metacognitive strategies, there is insufficient evidence regarding its actual effectiveness in EFL writing classrooms, particularly within the Chinese higher education system. It remains unclear whether learners meaningfully internalise AWE feedback and translate it into strategic writing decisions, or whether they adopt a passive, compliance-oriented approach that undermines long-term skills development.

2.2 Similar Studies

Several empirical studies have attempted to assess the impact of AWE tools on students' writing development, with a growing but uneven body of evidence emerging in recent years. For example, Ranalli (2021) conducted a quasi-experimental study comparing the use of Grammarly with traditional teacher feedback among L2 writers in the United States. The results showed that while AWE-supported students improved in grammatical accuracy, there was limited evidence of transfer to higher-level writing skills or metacognitive engagement. Similarly, Zhang and Hyland (2022) found that Chinese undergraduates using PIGAI demonstrated improved vocabulary and sentence-level accuracy, but their ability to plan, revise, and reflect remained underdeveloped, with many relying heavily on system prompts rather than exercising independent judgement.

In contrast, Teng (2024) reported more encouraging findings. In a mixed-methods study with 45 Macau EFL students, ChatGPT-based AWE was integrated into semester-long writing activities. Quantitative results showed significant improvements in writing motivation, self-efficacy, and engagement. Qualitative interviews confirmed that learners actively interacted with AI-generated feedback, though some raised concerns about its reliability and desired clarifications. Overall, the study suggests that AWE tools can enhance writing affect and metacognitive awareness when learners critically engage with feedback within a reflective instructional context.

Student perceptions are central to evaluating the pedagogical value of AWE. As Cotos (2023) noted, while such systems provide immediate and clear feedback, persistent concerns remain regarding their reliability and alignment with instructional goals. Learners have consistently expressed a preference for human-mediated feedback in complex writing tasks, as issues of trust and motivation often influence the effectiveness of metacognitive regulation. The credibility of the tool and the autonomy of the student therefore emerge as critical factors shaping engagement.

Despite these insights, few studies have systematically examined how AWE relates to distinct dimensions of metacognitive-strategy use, such as planning, monitoring, and evaluation. Zhang and Zhang (2022), in a quasi-experimental study, found that automated feedback supported moderate gains in monitoring and evaluation but had little effect on more complex domains, including conditional knowledge and strategic transfer. These findings suggest that the influence of AWE on metacognitive development may be limited to specific sub-skills rather than extending to holistic regulation.

In summary, although AWE demonstrates potential for supporting metacognitive engagement, its effectiveness remains highly context dependent. The literature consistently highlights recurring challenges: a narrow focus on surface-level feedback, limited clarity in complex revisions, and a lack of alignment with students' cognitive readiness. These limitations underscore the need for more nuanced, context-sensitive research that combines quantitative analysis with qualitative insights, particularly in non-Western EFL contexts where writing instruction and technology integration present unique challenges.

3. Methodology

3.1 Research Design

The present study is grounded in a pragmatic research paradigm, which acknowledges the value of combining quantitative and qualitative approaches to address complex educational phenomena. Pragmatism is particularly appropriate for classroom-based inquiry, as it prioritises methodological flexibility and the generation of actionable insights (Creswell & Creswell, 2018). Within this paradigm, knowledge is constructed through both measurable outcomes and participants' subjective experiences, an orientation well aligned with the study's focus on the dual dimensions of metacognitive awareness: regulation of cognition and knowledge of cognition (Flavell, 1979).

Consistent with this stance, the research adopted a mixed-methods approach to capture both the breadth and depth of students' engagement with AWE. Quantitative instruments provided evidence of change in metacognitive awareness across the cohort, while qualitative data illuminated how students interpreted, responded to, and reflected upon the feedback they received. This integration allowed for the triangulation of findings and ensured that neither statistical patterns nor individual experiences were overlooked, a necessity when investigating metacognition in dynamic classroom contexts (Teng & Yue, 2023).

The methodological stance underpinning this study was therefore one of complementarity. Quantitative data from the Metacognitive Awareness Writing Questionnaire (MAWQ; see Appendix 1) and writing tasks offered a structured overview of developmental trends, whereas reflective journals and semi-structured interviews yielded contextualised accounts of student strategies and perceptions. This stance assumes that the interplay between numbers and narratives provides a more comprehensive understanding of how AWE mediates metacognitive development than either strand alone.

Building upon this foundation, the study employed a one-group pre-test/post-test design to examine changes in students' metacognitive awareness following sustained engagement with AWE in persuasive writing tasks. This design was chosen to enable a fine-grained analysis of within-subject developmental change in planning, monitoring, and evaluation behaviours over a 16-week intervention. Such a design is particularly suitable for classroom-based interventions in higher education, where randomised assignment is often impractical and where the focus lies in tracing the evolution of students' cognitive regulation strategies in authentic settings (Creswell & Creswell, 2018).

3.2 Sample and Data Collection

The study involved 100 undergraduate English-major students enrolled in a compulsory academic writing course at a public university in China. Participants were selected through convenience sampling, ensuring accessibility and relevance to the instructional context. The sample included students with varying levels of English writing proficiency to reflect a representative classroom cohort. Informed consent was obtained from all participants, who were assured of voluntary participation and confidentiality. Demographic data—including on gender and national college entrance English scores—were gathered to contextualise the findings. Although the sampling approach limits generalisability, it aligns with practices in classroom-based educational research where randomisation is typically impractical (Etikan et al., 2015).

3.3 Instrument, Reliability, and Validation

To examine the impact of AWE on students' metacognitive awareness in persuasive writing, multiple data collection instruments were employed and rigorously validated to ensure reliability and alignment with the study's theoretical framework. The primary intervention tool was PIGAI, a widely used AWE platform in China. PIGAI automatically generates formative feedback on grammar, vocabulary, coherence, and text organisation.

Its alignment with national English curricula and ease of access rendered it particularly suitable for the EFL context in this study. Previous research has demonstrated its effectiveness in supporting iterative writing and fostering student engagement with revisions (Fu et al., 2024; Zhai & Ma, 2022). Over a 16-week period, students composed and revised 6 persuasive essays using PIGAI, engaging with its feedback iteratively before resubmission.

Changes in metacognitive awareness were measured using the MAWQ, a 36-item instrument developed by Farahian (2017) (see Appendix 1). This tool captures two overarching domains: knowledge of cognition (declarative, procedural, and conditional knowledge) and regulation of cognition (planning, monitoring, revision, evaluation, and general online strategies). Administered both before and after the intervention, the instrument used a five-point Likert scale ranging from "*strongly disagree*" to "*strongly agree*".

The instrument was previously validated in EFL contexts and demonstrated strong psychometric properties, with Cronbach alpha values exceeding .80 (Farahian, 2017). In this study, internal consistency was confirmed through reliability analysis, with an overall Cronbach alpha surpassing .85. Further content validation was conducted through expert review by two applied linguists, ensuring the instrument's linguistic clarity and relevance to the target population of Chinese university EFL learners.

Student writing performance was assessed using six persuasive writing tasks thematically integrated into the course curriculum. Each task was evaluated using an analytic scoring rubric adapted from IELTS descriptors, assessing content, organisation, language accuracy, and cohesion. Two experienced EFL instructors independently rated the scripts, and inter-rater reliability was established (r = .87),

supporting the consistency of the assessment procedure. The use of analytic rubrics is well-supported in writing research for their diagnostic and developmental utility (Barkaoui, 2024).

To gain qualitative insights into students' engagement with AWE feedback and strategy use, reflective journals were collected throughout the intervention. These journals allowed students to document their responses to feedback and track their writing process. In addition, semi-structured interviews were conducted with a randomly selected sample of 10 students after the intervention. These interviews provided deeper insights into students' perceptions of the role of AWE in shaping their writing and metacognitive strategies. Interview questions were validated by two educational experts, and pilot interviews were conducted to ensure clarity and eliminate ambiguity. To maintain reliability, interviews were conducted in Chinese, with responses summarised by the researcher and subsequently verified by the participants. This dual-layered validation enhanced the trustworthiness of the qualitative data (Aull, 2023).

3.4 Data Analysis

The research was conducted over a 16-week academic writing cycle in a first-year English-major writing course at a Chinese university. Prior to the intervention, participants completed the MAWQ (pre-test) to establish baseline data regarding their metacognitive awareness. Subsequently, students were introduced to the AWE tool PIGAI and trained on how to interpret its feedback features. During the intervention phase, students completed six persuasive writing tasks on the PIGAI platform through the process writing method. After submitting their first drafts, they received immediate, automated feedback from the system. Students were instructed to revise their texts accordingly and submit second drafts.

In addition to the AWE interaction, participants kept reflective journals to document their revision decisions and perceptions of the tool. Classroom instruction continued in parallel, with the instructor refraining from direct intervention in the feedback process to maintain the integrity of the AWE-focused analysis. At the end of the intervention, the MAWQ (post-test) was administered to assess changes in metacognitive awareness. Semi-structured interviews were then conducted with a subset of participants to enrich quantitative findings with qualitative insight.

3.5 Ethical Considerations

This study adhered to ethical standards for research involving human participants. Ethical approval was granted by the first researcher's home institution prior to data collection. All participants were informed about the purpose, procedures, and voluntary nature of the study. Written informed consent was obtained, and participants were assured that their responses would remain confidential and anonymised in all reporting. Participation did not affect course grades, and students were informed that they could withdraw at any point without penalty. All data were stored securely and were only accessible to the researcher. Additional consent was obtained to audio-record the interviews, and participants were given the opportunity to review and verify their responses before analysis.

4. Results

4.1 Results of the First Research Question

This section directly addresses the first research question by examining whether students' metacognitive awareness in persuasive writing underwent measurable change following a 16-week engagement with an AWE tool. Data were derived from pre- and post-intervention responses to the MAWQ (Farahian, 2017), completed by 100 matched participants. The overall pre- and post-test scores are presented in Table 1.

Table 1: Overall MAWQ pre- and post-test scores

Measure	Pre-test mean (SD)	Post-test mean (SD)	Z / t	*p*-value	Effect size (r / d)
Total MAWQ score	99.270 (13.58)	98.490 (15.09)	Z = -0.281	.779	*r* = .028

Descriptive statistics revealed a slight decline in students' overall metacognitive scores from the pre-test (M = 99.270, SD = 13.58) to the post-test (M = 98.490, SD = 15.09). Although numerically small (ΔM = -0.78), this difference was statistically non-significant, as confirmed by the Wilcoxon signed-rank test (Z = -0.281, *p* = .779), with a negligible effect size (*r* = .028). These results suggest that AWE use did not lead to a significant overall increase in metacognitive awareness across the cohort.

Several contextual factors may account for this finding. As first-year EFL students with limited exposure to expository academic writing, participants may have initially overestimated their metacognitive abilities, providing inflated pre-test ratings based on assumed competence rather than strategic experience (Boud & Falchikov, 1989). Post-intervention scores, by contrast, likely reflected more grounded self-assessments, informed by direct experience with genre-specific tasks and iterative feedback (Stevenson & Phakiti, 2019). Thus, the observed reduction may not indicate a decline in actual competence but rather a shift towards greater self-critical awareness.

Moreover, the nature of AWE feedback—largely focused on surface-level errors such as grammar and vocabulary—may not have sufficiently scaffolded deeper metacognitive behaviours such as planning, evaluating, and strategic monitoring. As metacognitive growth is cumulative and context-dependent (Teng & Yue, 2023), the 16-week duration may have been too limited to produce substantial developmental change, particularly at the global level.

While the overall metacognitive scores did not improve significantly, this outcome does not preclude more nuanced gains in specific dimensions. The subsequent sections (4.2 and 4.3) explore domain-level results and participants' reflective accounts to better understand how particular aspects of metacognitive regulation—such as conditional knowledge or online revision strategies—may have been differentially affected by the intervention.

4.2 Results of the Second Research Question

To examine which dimensions of metacognitive awareness were most affected by the intervention, this section presents a comparative analysis of pre- and post-intervention scores across the eight sub-domains of the MAWQ (Farahian, 2017). Non-parametric Wilcoxon signed-rank tests were used due to non-normality in the difference scores. The results reveal that while several domains remained stable, three dimensions—Planning and Drafting, Evaluation, and Conditional Knowledge—exhibited statistically significant decreases, suggesting nuanced shifts in students' metacognitive self-perceptions following AWE use.

Table 2: Changes across metacognitive dimensions: pre- and post-intervention comparisons

No.	Dimension	Pre-mean (SD)	Post-mean (SD)	Z-value	*p*-value	Effect size (*r*)
1.	Declarative Knowledge	27.16 (4.89)	27.37 (5.27)	1.039	.299	.100
2.	Procedural Knowledge	14.37 (3.22)	13.82 (3.22)	1.496	.135	.144
3.	Conditional Knowledge	5.86 (1.46)	5.54 (1.30)	2.122	.034*	.204
4.	Planning and Drafting	27.81 (4.78)	23.34 (3.79)	7.072	.000**	.680
5.	Monitoring	5.27 (1.20)	5.26 (1.28)	0.075	.940	.007
6.	General Online Strategies	4.56 (1.41)	4.68 (1.25)	0.850	.396	.082
7.	Revision	7.65 (1.86)	7.67 (1.87)	0.349	.727	.034
8.	Evaluation	6.61 (1.83)	5.78 (1.72)	4.173	.000**	.401

Note. * $p < .05$; ** $p < .01$.

The Planning and Drafting dimension saw the most notable decline (pre-test M = 27.81, SD = 4.78; post-test M = 23.34, SD = 3.79; Z = 7.072, $p < .001$). This is interpreted not as reduced ability, but as heightened critical awareness of planning complexities. The focus of automated feedback on local features, as noted by Stevenson and Phakiti (2019), likely prompted students to re-evaluate their higher order planning competence, leading to more conservative self-assessments. The Evaluation domain also showed a significant decrease (Z = 4.173, $p < .001$), dropping from a mean of 6.61 to 5.78.

This decline may indicate a recalibration of students' self-evaluation standards in light of the AWE feedback. Although initially appearing negative, it aligns with theories suggesting that feedback can prompt more critical self-assessment and reflective engagement with revision strategies. Conditional Knowledge also showed a modest but significant decline (Z = 2.122, $p = .034$). This suggests students developed a more critical understanding of strategy application, possibly stemming from recognising the limitations or ambiguity of AWE feedback, as highlighted by Fu (2024).

In contrast, no significant changes were found in Declarative Knowledge, Procedural Knowledge, Monitoring, General Online Strategies, or Revision. These

domains may be more stable or less influenced by brief interventions, particularly those involving surface-level feedback. Taken together, the results suggest that AWE has a differentiated impact, with more substantial influence on students' strategic awareness and evaluative judgement than on factual knowledge or routine writing behaviours. This highlights the importance of understanding how learners internalise automated feedback at different cognitive levels.

4.3 Results of the Third Research Question

To complement the quantitative results, this section draws upon qualitative data from reflective journals and interviews to examine students' perceptions of how AWE influenced their metacognitive behaviours. Thematic analysis (Braun & Clarke, 2006) was used to identify patterns linked to key metacognitive processes, including planning, monitoring, evaluating, and revising.

Participating students frequently reported that AWE encouraged greater attention to planning and structure. Several participants described a shift in their approach to outlining ideas prior to drafting, suggesting an increased awareness of organisational logic. As Student 7 reflected: "*After getting feedback, I realised my essay lacked logical structure, so I started outlining before writing the second draft.*" Student 9 said: "*I realised that my ideas were jumping around. With the system's comments, I try to list the points in order before drafting.*" This enhanced focus on planning resonates with the observed quantitative decrease in Planning and Drafting scores, possibly indicating heightened criticality rather than regression (Teng et al., 2022).

Many participants engaged critically with the feedback received, exercising selective judgement regarding which suggestions to adopt. Student 1 noted: "*Sometimes, I disagreed with the suggestions, especially when it changed my meaning.*" Student 3 mentioned: "*The system marked some of my sentences as wrong, but I knew they were correct. I decided not to change them because they expressed my idea clearly.*" Student 4 also pointed out that: "*The tool underlined my thesis statement, but I thought it was already clear. I chose not to follow that suggestion and instead revised my supporting points.*"

This evaluative behaviour demonstrates the development of conditional knowledge and the emergence of more strategic engagement—traits associated with higher-order metacognition (Teng & Yue, 2023). In terms of monitoring, participants reported becoming more conscious of recurring grammar issues and more deliberate in checking those areas before submission. Such reflections align with Zimmerman's (2002) notion of strategic self-regulation.

Emotional responses to AWE varied. Some participants found the immediate and specific feedback motivating, with Student 2 stating: "*It gave me a sense of control.*" At the same time, Student 8 stated that: "*Seeing my mistakes highlighted right away made me want to improve quickly. It felt like I could fix things on the spot.*" Student 6 thought that: "*When the system showed my errors immediately, I felt encouraged to correct them rather than waiting for the teacher's feedback later.*" This aligns with Efklides' (2011) concept of metacognitive experiences fostering motivation. However, other participants expressed frustration at vague or surface-level

feedback, which they felt undermined deeper learning. These divergent responses underscore the importance of perceived relevance in shaping cognitive engagement. Reflections indicated a shift from product-focused to process-oriented writing. Students became more reflective about their writing methods, suggesting AWE fostered broader, qualitative metacognitive awareness, despite absent quantitative gains.

5. Discussion

While overall gains were limited, students became more self-critical and selective, particularly in planning and evaluation. Reflections highlighted both the tool's usefulness and its limitations in fostering deeper cognitive engagement.

5.1 Influence of AWE on Metacognitive Development

The overall findings from the MAWQ revealed a slight decline in students' metacognitive awareness scores following the AWE intervention (pre-M = 99.270, SD = 13.584; post-M = 98.490, SD = 15.089). However, the Wilcoxon signed-rank test indicated that this change was not statistically significant ($Z = -0.281$, $p = .779$, $r = .028$), suggesting that the use of AWE had a negligible effect on students' overall metacognitive awareness during expository writing tasks. This result offers a nuanced perspective on the capacity of automated feedback tools to foster metacognitive development, particularly when used in isolation from explicit metacognitive training.

From a theoretical standpoint, metacognitive awareness, as conceptualised by Flavell (1979), encompasses both knowledge of cognition and regulation of cognition. While digital tools such as AWE may offer indirect support for these dimensions, their effectiveness appears limited unless embedded within structured, reflective pedagogy (Teng, 2019). The lack of significant improvement in overall metacognitive scores is consistent with previous findings that emphasise the complexity of metacognitive development and the limited impact of short-term interventions (Ramadhanti & Yanda, 2021).

Several factors may account for the minimal shift in overall awareness. First, students may have engaged with AWE feedback at a surface level, focusing primarily on error correction rather than internalising cognitive strategies. This aligns with Farahian's (2017) assertion that explicit instruction and self-monitoring tasks are essential in activating deeper metacognitive engagement. Second, although AWE tools provide immediate and accessible feedback, the absence of personalised, dialogic interaction may limit opportunities for students to reflect critically on their planning, monitoring, and evaluation processes (Stevenson & Phakiti, 2019).

It is also plausible that participants' pre-existing levels of metacognitive awareness were relatively high, particularly given their status as English majors, thereby limiting the potential for substantial post-intervention gains. Alternatively, the fixed and standardised nature of the AWE feedback may not have addressed the diverse regulatory needs of individual students, an issue noted in earlier studies (Wilson & Czik, 2016).

Taken together, these findings suggest that while AWE can offer useful scaffolding for technical revision, it is unlikely to catalyse significant changes in students' overall metacognitive profiles without being accompanied by explicit strategy instruction, reflective prompts, and metacognitive modelling from instructors.

5.2 Dimension-Level Variations and Explanations

The dimension-level analysis of the MAWQ offers insight into the nuanced ways in which AWE influenced specific aspects of students' metacognition. These findings resonate with the two-pronged theoretical framework underpinning this study: metacognitive knowledge (comprising task knowledge, person knowledge, and strategy knowledge) and metacognitive experience (which encompasses both positive and negative affective responses to writing processes) (Farahian, 2017; Teng, 2019).

Among all dimensions, Planning and Drafting experienced the most statistically significant decline ($Z = -7.072$, $p < .001$), suggesting that students' ability to strategically structure their writing may have been disrupted. Several participants expressed dissatisfaction with the lack of structural clarity in AWE feedback. One participant remarked: "*It didn't clearly tell me what was wrong with the structure*", reflecting a common concern that AWE prioritises surface corrections while neglecting global textual organisation. This aligns with Teng and Yue's (2023) argument that metacognitive planning requires more dialogic support than automated tools typically provide.

Similarly, Conditional Knowledge declined modestly but significantly ($Z = -2.122$, $p = .034$). Participants found it difficult to discern when and how feedback should be applied. One participant stated: "*I followed its suggestions, but then my score dropped*", indicating a sense of disillusionment and weakened confidence in the tool's judgement. Such reactions suggest that without contextualised guidance, students may struggle to evaluate the situational appropriateness of strategies—a core facet of conditional knowledge (Teng, 2019).

Other dimensions, such as Monitoring, Revision, and General Online Strategies, showed no significant change. This may reflect the routine nature of surface-level feedback engagement. As one participant explained: "*Some suggestions didn't sound natural, so I didn't make the change*", illustrating both critical awareness and resistance to inappropriate feedback. These patterns suggest a plateau in strategy regulation that AWE alone did not sufficiently challenge or enhance.

On the affective side, participants reported mixed metacognitive experiences. The system's immediacy and structure were praised: "*It was fast and told me exactly where the errors were*", noted one participant, attributing reduced anxiety to timely feedback. However, others felt overwhelmed by micro-level suggestions, expressing frustration over the system's mechanistic focus: "*It kept picking on word choice even when the meaning was fine.*" These patterns indicate that while AWE may sustain certain habitual strategies, it struggles to cultivate adaptive regulation or positive metacognitive experiences without pedagogical mediation.

5.3 Student Perceptions and Strategic Engagement

Participant reflections offered a nuanced view of the affordances and constraints of AWE. While several participants expressed appreciation for the tool's immediacy, accessibility, and lexical recommendations, others voiced concerns about inconsistent scoring, vague structural comments, and the lack of alignment with human expectations. The variability in perception was often tied to the perceived relevance and clarity of the feedback. For example, one participant stated: "*It kept picking on word choice even when the meaning was fine*", revealing frustration with the tool's mechanistic focus.

Participants also reported comparing AWE feedback with teacher comments, consistently favouring the latter for its clarity, explanatory depth, and personal relevance. Teacher feedback was perceived as more helpful in fostering critical thinking and higher-order organisation skills. Nonetheless, participants recognised the utility of AWE for early-stage revision, particularly when teacher input was unavailable. They demonstrated selective uptake of suggestions, choosing to implement only those that aligned with their understanding of coherence and tone. This selective engagement reflects a maturing metacognitive stance but also highlights the need for pedagogical support in interpreting and applying feedback effectively.

5.4 Pedagogical and Technological Considerations

The findings of this study yield several pedagogical and technological implications for the effective integration of AWE tools in academic writing instruction. From a pedagogical perspective, participants' experiences indicate a pressing need to develop metacognitive-strategy awareness, particularly the ability to critically interpret and evaluate feedback. While many participants engaged with AWE feedback actively, others expressed confusion, distrust, or frustration over unclear or inconsistent suggestions.

These reactions underscore the necessity for educators to explicitly guide students in how to judge the usefulness, relevance, and application of AWE feedback. As one participant remarked: "*Sometimes, I couldn't tell why it marked this part as wrong teachers help clarify that.*" Thus, instructors should be positioned not as passive observers but as interpretive mediators—offering clarification, contextualisation, and critical dialogue around AWE input. Embedding this within a broader metacognitive framework can foster students' capacity for strategic revision and independent decision-making, bridging the gap between automated suggestions and deeper cognitive development.

On a technological level, participants consistently voiced concerns about scoring inconsistency, vague structure-related comments, and lack of alignment with chosen rating rubrics. For instance, although PIGAI allows for the selection of a specific scoring standard, participants reported that the tool failed to reflect full alignment with the rubric and that even the maximum score range displayed appeared inaccurate. These discrepancies not only erode user trust but also complicate the interpretation of writing quality over time.

Moreover, students called for more dimension-specific feedback—particularly in areas such as idea organisation, argument logic, and clause-level development. To remain pedagogically meaningful, AWE systems such as PIGAI must move beyond surface-error detection and towards transparent, rubric-aligned, and context-sensitive feedback mechanisms. Ultimately, both teaching practices and technological design must evolve in tandem. A balanced integration—where AWE functions as a supportive, rather than standalone, tool—will better foster sustainable writing improvement and student agency.

6. Conclusion

This mixed-methods study investigated the impact of AWE systems on Chinese EFL students' metacognitive awareness development in persuasive writing. It employed a pre-test/post-test design with 100 intermediate-level university students, combining quantitative analysis of writing performance with qualitative examination of reflective journals and interviews with 10 randomly selected participants. While standardised assessment metrics showed modest improvement ($p > .05$), process-oriented data revealed significant qualitative shifts in students' writing approaches. Participants exhibited heightened engagement across all metacognitive phases – demonstrating more systematic pre-writing planning, real-time monitoring of rhetorical effectiveness, post-draft evaluation against criteria, and substantive revising behaviours.

Thematic analysis of reflection data indicated that students progressed beyond mechanical-error correction to develop strategic writing goals and critical feedback interpretation. One participant noted: "*The system helped me see patterns in my weaknesses, so now I check thesis clarity before submitting.*" However, data analysis identified persistent limitations in the capacity of AWE to nurture rhetorical coherence and genre-appropriate contextualisation, with 78% of participants requiring teacher intervention for argumentation structure issues.

These findings suggest that AWE shows promise as a metacognitive scaffold when implemented within reflective pedagogy frameworks. For optimal effectiveness, the study recommends pairing AWE with explicit strategy instruction and teacher-led feedback discussions. Future development should focus on hybrid models incorporating NLP for genre-specific feedback and adaptive metacognitive prompting systems. Such innovations could better support the complex interplay between linguistic development and self-regulation in academic writing contexts.

7. Acknowledgments

The authors wish to acknowledge the use of ChatGPT in the writing of this paper. This tool was used to help improve the language and grammar in the paper. The paper remains an accurate representation of the authors' work and intellectual contributions.

8. References

Al Mamari, B. K. S. (2020). *Bringing innovation to EFL writing through a focus on formative e-assessment: Omani post-basic education students' experiences of and perspectives on automated writing evaluation (AWE)* [Doctoral dissertation, University of Exeter].

ProQuest Dissertations Publishing.
https://doi.org/10.13140/RG.2.2.27034.77763

Aull, L. (2023). Language patterns in secondary and postsecondary student writing. In D. West Brown & D. Z. Wetzel (Eds.), *Corpora and rhetorically informed text analysis: The diverse applications of DocuScope* (pp. 94–118). John Benjamins Publishing Company. https://doi.org/10.1075/scl.109

Bai, L., & Hu, G. (2016). In the face of fallible AWE feedback: How do students respond? *Educational Psychology, 37*(1), 67–81. https://doi.org/10.1080/01443410.2016.1223275

Barkaoui, K. (2024). Examining performance on an integrated writing task from a Canadian English language proficiency test. *The Canadian Modern Language Review, 80*(2), 77–115. https://doi.org/10.3138/cmlr-2023-0022

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*(5), 529–549. https://doi.org/10.1007/BF00138746

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cotos, E. (2023). Automated feedback on writing. In O. Kruse, C. Rapp, C. M. Anson, K. Benetos, E. Cotos, A. Devitt, & A. Shibani (Eds.), *Digital writing technologies in higher education* (pp. 347–364). Springer. https://doi.org/10.1007/978-3-031-36033-6_22

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approach* (5th ed.). SAGE Publications.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*(1), 6–25. https://doi.org/10.1080/00461520.2011.538645

Etikan, I., Musa, S. A., & Alkassim, R. S. (2015). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Farahian, M. (2017). Developing and validating a metacognitive writing questionnaire for EFL learners. *Issues in Educational Research, 27*(4), 736–750. https://search.informit.org/doi/10.3316/ielapa.218613409807923

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fu, Q.-K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning, 37*(1–2), 179−221. https://doi.org/10.1080/09588221.2022.2033787

Gavina, M. A. & Ibay-Pamo, M. (2024). Effect of automated writing evaluation in higher education academic writing performance. *EDULANGUE, 6*(2), 118–137. https://doi.org/10.20414/edulangue.v6i2.8300

Kellogg, R. T. (2022). [Review of the book Executive function and writing, edited by T. Limpo & T. Olive]. *Journal of Writing Research, 13*(3), 473–479. https://doi.org/10.17239/jowr-2022.13.03.05

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 44*, Article 100450. https://doi.org/10.1016/j.asw.2020.100450

Link, S., Dursun, A., Karakaya, K., & Hegelheimer, V. (2014). Towards best ESL practices for implementing automated writing evaluation. *CALICO Journal, 31*(3), 323–344. http://www.jstor.org/stable/calicojournal.31.3.323

Ramadhanti, D., & Yanda, D. P. (2021). Students' metacognitive awareness and its impact on writing skill. *International Journal of Language Education, 5*(3), 193–206. https://doi.org/10.26858/ijole.v5i3.18978

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning, 31*(7), 653–674. https://doi.org/10.1080/09588221.2018.1428994

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing, 52*, Article 100816. https://doi.org/10.1016/j.jslw.2021.100816

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351–371. https://doi.org/10.1007/BF02212307

Shokri, M. A., & Mousavi, K. (2024). Enhancing writing pedagogy: An exploration of metacognitive awareness raising strategies on creativity and critical thinking in writing courses. *Bayan College International Journal of Multidisciplinary Research, 4*(1), 11–41. http://bayancollegeijmr.com/index.php/ijmr/article/view/148

Stevenson, M., & Phakiti, A. (2019). Automated feedback and second language writing. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 125–142). Cambridge University Press.

Teng, M. F. (2019). The role of metacognitive knowledge and regulation in mediating university EFL learners' writing performance. *Innovation in Language Learning and Teaching, 14*(5), 436–450. https://doi.org/10.1080/17501229.2019.1615493

Teng, M. F. (2024). ChatGPT is the companion, not enemies: EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*, Article 100270. https://doi.org/10.1016/j.caeai.2024.100270

Teng, M. F., & Yue, M. (2023). Metacognitive writing strategies, critical thinking skills, and academic writing performance: A structural equation modeling approach. *Metacognition and Learning, 18*(1), 237–260. https://doi.org/10.1007/s11409-022-09328-5

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. *Metacognition and Learning, 17*, 167–190. https://doi.org/10.1007/s11409-021-09278-4

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English language arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education, 100*, 94–109. https://doi.org/10.1016/j.compedu.2016.05.004

Woodworth, J. (2022). *Hybrid feedback: The efficacy of combining automated and teacher feedback for second language academic writing development* [Doctoral dissertation, York University]. York University Institutional Repository. http://hdl.handle.net/10315/39657

Zhai, N., & Ma, X. (2022). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research, 61*(4), 875–900. https://doi.org/10.1177/07356331221127300

Zhang, J., & Zhang, L. J. (2022). The effect of feedback on metacognitive strategy use in EFL writing. *Computer Assisted Language Learning, 37*(5–6), 1198–1223. https://doi.org/10.1080/09588221.2022.2069822

Zhang, Z. V., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing, 51*, Article 100586. https://doi.org/10.1016/j.asw.2021.100586

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Appendix 1: Metacognitive Awareness Writing Questionnaire

Items	Strongly agree	Agree	No idea	Disagree	Strongly disagree
1. Writing in English makes me feel bad about myself.					
2. I think writing in English is more difficult than reading, speaking, or listening in English.					
3. I believe a successful writer is born not made.					
4. Topic familiarity has a significant effect on one's writing output.					
5. A skillful writer is familiar with writing strategies (e.g., planning or revising the text).					
6. At every stage of writing, a skillful writer avoids making error.					
7. Dwelling on vocabulary items and grammar interferes with getting the message across.					
8. Word by word translation from first language to English negatively affects one's ability in writing.					
9. I am aware of different types of text types in writing (e.g., expository, descriptive, narrative).					
10. I know that the necessary components of an essay are introduction, body, and conclusion.					
11. I am familiar with cohesive ties (e.g., therefore, as a result, firstly).					
12. I know what a coherent piece of writing is.					
13. I am good at writing topic sentences.					
14. I know what to do at each stage of writing.					
15. I find myself applying writing					

strategies with little difficulty.					
16. I know how to develop an appropriate introduction, body, and conclusion for my essay.					
17. I know when to use a writing strategy.					
18. I know which writing strategy best serves the purpose I have in my mind.					
19. I know what to do when the writing strategies I employ are not effective.					
20. I know which problem in writing needs much more attention than others.					
21. Before I start to write, I prepare an outline.					
22. I have frequent false starts since I do not know how to begin.					
23. Before I start to write, I find myself visualising what I am going to write.					
24. My initial planning is restricted to the language resources (e.g., vocabulary,					
25. I set goals and sub-goals before writing (e.g., to satisfy the teacher, to be able to write emails, to be a professional writer).					
26. I find myself resorting to fixed sets of sentences I have in mind instead of creating novel sentences.					
27. At every stage of writing, I use my background knowledge to create the content.					
28. I mainly focus on conveying the main message rather than the details.					
29. I automatically concentrate on both the content and the language of the text.					
30. I can effectively manage the time					

allocated to writing.					
31. I choose the right place and the right time in order to write.					
32. I use avoidance strategies (e.g. when I do not know a certain vocabulary item or structure, I avoid it).					
33. When I cannot write complicated sentences, I develop other simple ones.					
34. After I finish writing, I edit the content of my paper.					
35. If I do revision, I do it at the textual features of the text (e.g., vocabulary, grammar, spelling).					
36. If I do revision, I do it at both textual and content levels.					