



Castledown

 OPEN ACCESS

Technology in Language Teaching & Learning

ISSN 2652-1687

<https://www.castledown.com/journals/tltl/>

Technology in Language Teaching & Learning, 8, 103944 (2026)

<https://doi.org/10.29140/tltl.2026.103944>

Artificial Intelligence-Mediated Peer Collaboration in EFL Argumentative Writing: Exploring Learners' Inferential Reasoning Development



EKA FADILAH 

English Department, Faculty of Cultural Sciences, Universitas 17 Agustus 1945 Surabaya, Indonesia
ekafadilah@untag-sby.ac.id

Abstract

Artificial intelligence (AI) is increasingly used in writing instruction, yet its role in collaborative writing remains underexplored. While prior research has focused on individual support and automated feedback, less attention has been given to how AI can be integrated into peer interaction to support learners' reasoning. This study examines AI-mediated peer collaboration (using ChatGPT) in source-based argumentative writing in an EFL context. Using a pre-post design, 22 Indonesian EFL undergraduates completed individual writing tasks before and after three collaborative sessions in which AI use was constrained to support idea exploration. Learners' texts were analyzed using Toulmin-based measures of argument structure, with Legitimation Code Theory (LCT) used to support the interpretation of meaning movement across levels of abstraction. Peer interaction data were used to examine how AI was taken up during collaboration. Results indicate limited change in surface features but clear gains in warrants, backing, and rebuttals, indicating stronger inferential reasoning. Interactional analysis further revealed a shift from form-focused work to reasoning-oriented dialogue, with learners evaluating and selectively using AI input. These findings suggest that higher-order reasoning developed through interaction shaped by pedagogically structured, inquiry-oriented AI use, in which AI functioned as a mediational resource.

Keywords: Artificial intelligence; AI-enhanced learning; collaborative writing; argumentative writing; EFL; language teaching technology.

Copyright: © 2026 Eka Fadilah. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. **Data Availability Statement:** All relevant data are within this paper.

Introduction

Argumentative writing is a central component of academic literacy in second and foreign language (L2) education, particularly in contexts where learners are required to synthesize information from multiple sources. In EFL classrooms, source-based argumentative writing prepares students for academic study, requiring grammatical accuracy, lexical control, and the ability to construct and justify claims and engage with alternative perspectives (Bacha, 2010; Hyland, 2005; Miller & Pessoa, 2016). Despite its pedagogical importance, research consistently shows that many L2 writers struggle to articulate the inferential relationships that explain why evidence supports a claim, resulting in arguments that are more descriptive than analytically persuasive (Al Bulushi, 2022; El Majidi et al., 2020; Hartwell & Aull, 2022).

The emergence of generative AI has introduced new possibilities for writing instruction, while raising questions about its use in classroom contexts. Although AI can support drafting and revision, uncritical use risks encouraging surface-level correction and passive reliance on generated language (El-Dakhs, 2024; Guo et al., 2022; Hwang et al., 2025; Wang, 2024). The educational value of AI lies in how it is integrated and evaluated within instructional activity.

Within technology-enhanced language teaching, peer collaboration provides a productive context in which AI use can be shaped and regulated. Through dyadic interaction, learners can externalize their thinking, negotiate relationships between evidence and claims, and co-construct explanations (Storch, 2013, 2021; Storch & Aldosari, 2013). When embedded in such activity, AI can function as resource for prompting inquiry and reflection rather than a source of ready-made text. However, there remains limited empirical understanding of how peer collaboration and constrained AI mediation may work together to support higher-order reasoning in L2 source-based argumentative writing.

Argumentative writing provides a useful context for examining this issue, as it requires learners to move beyond organization toward the construction of persuasive reasoning. In particular, rhetorical functions that explain how evidence supports a claim—such as warrants, backing, and rebuttals—remain underexplored in both research and pedagogy (Toulmin, 1958/2003; Kibler & Hardigree, 2016; Qiu & Lo, 2017).

To address this gap, the present study approaches argumentation as an inquiry-oriented practice grounded in reasoning from evidence (Hillocks, 2010, 2011). From this perspective, effective argumentation depends on learners' rhetorical agency, understood as their capacity to use argumentative resources in ways that respond to audience, purpose, and alternative perspectives. The study draws primarily on Toulmin's (2003) model to examine how learners construct and justify arguments. Complementarily, Legitimation Code Theory (LCT; Maton, 2013, 2020) is used in a supporting role to interpret how meaning shifts between concrete evidence and more abstract reasoning—conceptualized as semantic agility. This dual lens allows the study to account for both the inferential structure of arguments and the regulation of meaning movement, without treating the two frameworks as equal analytic layers.

Existing studies have examined the use of generative AI in L2 writing classrooms, particularly in relation to revision, feedback, and task facilitation. For example, ChatGPT has been described as a "writing buddy" that supports noticing and self-directed revision (Strobl et al., 2024), and as a tool for consciousness-raising through guided prompts (Tran, 2024). Other work conceptualizes large language models as participants in interaction that shape how learners engage with language and knowledge (Lege et al., 2024). However, much of this work continues to focus primarily on revision and task support, with less attention to how AI-mediated interaction may contribute to the development of higher-order reasoning in writing.

The present study therefore addresses this gap by examining how rhetorical agency and semantic agility develop within structured peer–AI collaborative inquiry. It investigates how pedagogically constrained AI mediation influences learners’ inferential reasoning during collaborative writing tasks, using Toulmin’s model as the primary analytic frameworks and an LCT-informed perspective to support interpretation of how learners regulate movement between concrete and abstract meanings.

Against this backdrop, this study examines the following research questions:

RQ1. How do L2 learners enact rhetorical agency during AI-enhanced peer collaborative argumentative writing?

RQ2. How do peer–AI interactions mediate learners’ semantic agility, particularly their movement between concrete and abstract meanings?

By positioning AI (specifically ChatGPT) as a pedagogically designed mediational resource and as part of the interactional environment in which reasoning is developed, this study contributes to research on technology-enhanced language teaching and learning by showing how AI-enhanced peer collaboration can support reasoning-focused writing while maintaining learner agency in EFL contexts.

Literature Review

Argumentative Writing as Inquiry and the Development of Rhetorical Agency in L2 Contexts

A substantial body of research on L2 argumentative writing has shown that learners’ primary difficulty lies not in producing basic argument structure, but in constructing well-reasoned arguments, particularly in source-based tasks. While many L2 writers can state claims and summarize information from sources, they often struggle to articulate the inferential logic that explains why evidence supports a stance (Bacha, 2010; El-Majidi et al., 2020; Stapleton & Wu, 2015). As a result, learner texts frequently remain descriptive rather than persuasive.

Early research in L2 writing tended to emphasize global textual features such as organization, coherence, and lexical sophistication (Hyland, 2005). Although these dimensions are important, subsequent studies have shown that attention to surface-level structure alone does not explain why many learners’ arguments lack persuasive force. Increasingly, scholars argue that weaknesses in L2 argumentation are often misattributed to structural or linguistic deficiencies, when the issue lies in missing warrants, insufficient justification, and weak inferential connections between claims and evidence (Miller & Pessoa, 2016). Empirical studies grounded in Toulmin’s model similarly report that difficulties with warrants and counterarguments weaken argument quality, even when claims and evidence are present (Hartwell & Aull, 2022; Rusfandi, 2015; Stapleton & Wu, 2015).

In response to these findings, more recent scholarship has reconceptualized argumentative writing as an inquiry-driven, epistemic practice rather than a fixed textual template. Inquiry-based perspectives, such as Hillocks’ (2010, 2011) model, frame argumentation as a process in which learners pose questions, evaluate evidence, and weigh competing perspectives through interaction. This shift is particularly salient in L2 contexts, where learners are often socialized into formulaic argumentative routines that privilege textual structure over reasoning (Hirvela, 2017; Schneer, 2014). From this perspective, effective argumentative writing depends not only on familiarity with argumentative forms, but also on learners’ capacity to exercise rhetorical agency—that is, to make purposeful decisions about how claims are justified, how evidence is interpreted, and how alternative perspectives are addressed in relation to audience and purpose (Hirvela, 2017; Miller & Pessoa, 2016).

These studies suggest that the central challenge in L2 argumentative writing lies in developing learners' inferential control as part of a broader inquiry process. Approaching argumentation through Toulmin's model foregrounds rhetorical agency as functional control over how claims are justified, qualified, and contested, providing a foundation for examining how instructional designs can support learners in engaging more actively with argumentative reasoning.

Semantic Agility and Meaning Movement in Argumentative Writing

While Toulmin's model captures the functional architecture of arguments, it says less about how meaning itself is organized across levels of abstraction within a text. To complement the Toulmin-based analysis, this study draws on Legitimation Code Theory (LCT; Maton, 2013, 2020) in a supporting interpretive role.

LCT's semantics dimension distinguishes between semantic gravity (the degree to which meaning is tied to a specific context) and semantic density (the degree to which meaning is condensed or abstracted). In argumentative writing, warrants and backing typically require increased semantic density, as they involve generalization and abstraction, whereas rebuttals and qualifiers often strengthen semantic gravity by re-anchoring meaning in specific contexts. The concept of semantic agility captures writers' capacity to move flexibly and purposefully between these poles—coordinating abstraction and contextual grounding rather than sustaining one level alone (Coffin, 2004; Martin & Maton, 2017; Maton, 2013, 2020).

Empirical research operationalizing semantic agility in L2 argumentative writing remains limited, particularly in classroom-based contexts. This study uses LCT to trace how the meaning movement accompanying Toulmin-based inferential structures develops over time, rather than as a freestanding analytic system. In this way, semantic agility serves as a lens for interpreting how learners regulate abstraction in the service of argumentative reasoning.

Peer Collaboration and AI-Mediated Argument Learning

Parallel to developments in writing theory, research on technology-enhanced language learning has increasingly explored how peer collaboration supports reasoning and learning. Peer interaction provides a productive social space for externalizing thinking, negotiating meaning, and co-constructing arguments, particularly in writing tasks that require justification and evaluation (Storch, 2013, 2021). Within this collaborative space, the emergence of generative AI has introduced both new opportunities and pedagogical challenges.

A growing body of research has examined AI use in L2 writing, with much of the early work focusing on automated assessment or individual writing support. For example, Fathali and Mohajeri (2025) demonstrate that AI systems can approximate human ratings in writing assessment, while Zare et al. (2025) examine how interacting with ChatGPT in essay writing relates to learners' task motivation—highlighting that the affective and motivational dimensions of AI use are shaped by task conditions. Such studies underscore that AI's impact depends on how it is positioned within the learning activity.

Previous studies report mixed outcomes. AI's contribution to writing is shaped by how it is pedagogically aligned with learning activity and situated within pedagogical practice (Huang et al., 2022; Wang, 2025), while learners' engagement depends on how its use is guided and regulated in interaction (Wang, 2024). Furthermore, scholars increasingly emphasize that the educational impact of AI depends on how it is designed, constrained, and embedded within instructional activity (El-Dakhs, 2024; Godwin-Jones, 2022). Similarly, Creely et al. (2025) report that educators and learners hold complex,

often ambivalent attitudes toward generative AI, reinforcing the need for pedagogically grounded integration rather than unrestricted use.

Recent classroom-based research also suggests that AI use is not static; learners' interactional strategies evolve over time, underscoring the importance of pedagogical scaffolding that encourages principled rather than habitual reliance on AI (Hwang et al., 2025). However, despite growing interest in AI-enhanced writing, empirical studies examining how AI-mediated peer collaboration supports inferential reasoning in source-based argumentative writing remain limited. Particularly, little is known about how deliberately constrained AI use interacts with peer dialogue to shape learners' reasoning processes and the development of inferential relationships between claims and evidence.

Within L2 writing pedagogy, recent studies have begun to reposition AI not as a text generator but as a pedagogical resource that can support learning processes when appropriately structured. Strobl et al. (2024), for instance, show that using ChatGPT as a comparison model can prompt learners to engage in deeper revision processes, particularly when they critically evaluate differences between their own writing and AI-generated texts. Similarly, Tran (2024) demonstrates that AI can support consciousness-raising by directing learners' attention to form–meaning relationships through guided prompts. Extending this perspective, Lege et al. (2024) conceptualize large language models as interactional participants that shape how learners engage with language and knowledge during learning activities.

Taken together, these studies suggest that the pedagogical value of AI lies not in text production itself, but in how it is embedded within instructional design and interaction. However, existing work has not yet examined how such AI-mediated interaction may contribute to the development of argumentative reasoning, particularly in how learners justify, qualify, and contest claims—or how meaning is regulated across levels of abstraction in source-based writing. This gap is particularly important in source-based argumentative writing, where learners must coordinate evidence, inference, and abstraction in a principled way.

Method

Research Context and Participants

The study was conducted in an undergraduate academic writing course at a private university in Surabaya, Indonesia. Participants were 22 first-year EFL students (TOEFL ITP; approximately CEFR B1–B2). To maintain authorship, AI use was restricted to dialogic support: students were required to justify AI-informed ideas collaboratively and re-express reasoning in their own words.

Background data included TOEFL ITP scores and self-reported writing experience and familiarity with generative AI. Dyads were formed across proficiency pairings (high–high, high–low, low–low) to structure interaction, but these groupings were not used for comparison; all analyses were conducted at the individual level.

Research Design

The study employed a process-oriented classroom intervention design that combined individual outcome measures with collaborative writing activities. The design aimed to trace how rhetorical agency and semantic agility developed during peer–AI collaborative inquiry rather than to test treatment effects. Three focal dyads were selected through purposive sampling to represent contrasting proficiency pairings and were recorded across three task cycles. These dyads served as interactional contexts for process analysis, while all outcome analyses were conducted at the individual level.

Tasks and Instructional Procedure

The intervention drew on an inquiry-based approach (Hillocks, 2010, 2011), treating argumentative writing as a process of reasoning from evidence rather than completing a fixed template. Toulmin's (2003) model of argumentation served as the primary analytic framework, with semantic wave theory (Maton, 2013) used to support the interpretation of meaning movement, although these frameworks were not taught explicitly. Instead, students developed rhetorical reasoning and control of meaning through guided tasks and collaboration.

In the classroom inquiry cycles, students worked with a widely discussed public issue that required them to evaluate competing claims, assess evidence credibility, consider alternative explanations, and construct rebuttals. Instruction focused on reasoning quality, with classroom interaction emphasizing critical inquiry and evidence-based judgment.

Reasoning was supported through modelling and structured questioning. Students were guided to formulate claims, select relevant evidence, explain its significance, consider limitations, and address counterarguments. Key Toulmin elements were scaffolded through prompts such as "What does this evidence show?" and "Why is this important for your claim?", while dyadic discussion supported the externalization and negotiation of reasoning.

Instruction followed three progressive inquiry cycles. In the first cycle (coping-oriented), students worked closely with evidence under stronger scaffolding, focusing on identifying what the evidence showed. In the second cycle (mastery-oriented), support was reduced and students developed more explicit reasoning, including explaining evidence and considering limitations. In the third cycle (integrated), students coordinated evidence use, explanation, and evaluation, producing more balanced arguments that combined concrete detail with broader reasoning.

Each 100-minute session followed a fixed sequence. Students first developed arguments without AI, followed by a short AI consultation phase (15–20 minutes) in which they could ask questions about evidence, interpretation, or gaps in reasoning. This was followed by a non-AI phase in which students discussed AI responses and decided which ideas to retain, modify, or reject. AI input was treated as a resource for discussion rather than an authoritative answer, and responsibility for reasoning remained with the learners.

The AI tool was ChatGPT, accessed via a standard web interface. It was used as a short consultation tool rather than a source of text production. It was selected due to its widespread use among Indonesian university students. The system operated in its general-purpose form without customization. AI contributions functioned as prompts for reflection, and any development in reasoning depended on how learners evaluated and reformulated these inputs during peer interaction.

At the time of data collection (April 2025), ChatGPT was accessed via the public web interface using the default configuration (no custom instructions, plugins, or external tools). All interactions used short, task-focused prompts to support idea exploration rather than extended text generation.

To support semantic agility, students worked with evidence sets combining concrete cases and more abstract ethical or policy considerations. They were encouraged to move between examples and broader principles through guided prompts, supporting more controlled shifts between levels of abstraction.

In both the pre-test and post-test, students completed an individual argumentative writing task on a socio-environmental issue (e.g., whether mining activities should be permitted). Students were required

to take a clear position, use information from a provided evidence set, explain how the evidence supported their claims, and address at least one alternative perspective. The evidence set presented multiple perspectives, including economic, environmental, and social considerations, and the same task structure and level of difficulty were maintained across both tests to ensure comparability.

Data Sources and Analysis

The study drew on four main data sources: (1) dyadic interaction and ChatGPT logs, (2) individual pre- and post-writing texts, (3) background information on dyad configurations (high–high, high–low, low–low), and (4) selectively transcribed interactional excerpts. Interactional data were collected through purposive sampling, focusing on key phases of negotiation across three focal dyads. As such, these data were used as process evidence to contextualize development rather than to construct a representative corpus.

To examine written performance, texts were analyzed using Toulmin's (2003) model. Each element (claim, data, warrant, backing, qualifier, rebuttal) was coded for presence, explicitness, and functional accuracy on a five-point scale ($\kappa > .80$). On this basis, a Rhetorical Agency Index was derived from warrants, backing, qualifiers, and rebuttals, and gain scores were calculated from pre–post differences, supported by descriptive and paired-sample statistics.

Although a five-point scale was used for consistency, ratings were based on functional contribution rather than surface presence. In practice, higher scores required clear inferential links between evidence and claims, while mid-range scores reflected partial or implicit reasoning.

In parallel, semantic agility was examined through LCT-informed semantic-wave analysis. Unpacking, repacking, semantic shift, and abstraction control were rated on five-point scales and combined into a Semantic Agility Index. Here, higher scores reflected more controlled movement between concrete and abstract meanings, for example when specific cases were used to develop general claims and then qualified through contextual conditions. For interpretive clarity, semantic gravity (SG) and semantic density (SD) were treated as composite indicators, with SG gains reflecting unpacking and semantic shift and SD gains reflecting repacking and abstraction control.

To explore how these dimensions related to each other, Pearson correlations were computed between Toulmin gains and SG–SD gains. These analyses were exploratory and interpretive rather than causal, aiming to identify potential alignment between rhetorical agency and semantic agility.

Finally, dyadic interaction was analyzed for Language-Related Episodes (LREs) and Learning/Task Episodes (LTEs) to examine how peer–AI collaboration mediated development. In this stage, students were allowed to use Indonesian during interaction to support natural reasoning. All interactional data were transcribed, translated, and back translated to preserve meaning.

Operationalizing Rhetorical Agency and Semantic Agility

Rhetorical agency refers to learners' functional control over inferential reasoning in argumentation, evidenced through the explicit and purposeful articulation of warrants, backing, qualifiers, and rebuttals beyond merely stating claims and presenting data. Each element was rated on a five-point analytic scale (1 = absent/non-functional; 5 = explicit, accurate, and rhetorically effective), based on whether it fulfilled its inferential purpose rather than on formal labelling. A Rhetorical Agency Index was computed from warrants, backing, qualifiers, and rebuttals, while claims and data were examined descriptively due to their lower inferential demands.

Semantic agility was operationalized as learners' capacity to regulate movement between context-bound meanings and abstract inferential reasoning, conceptualized as purposeful oscillation consistent with semantic waves. Texts were segmented into T-units and aligned with Toulmin elements. Unpacking (abstract - contextualized explanation), repacking (concrete - generalized inference), semantic shift, and abstraction control were rated on a five-point scale and combined into a Semantic Agility Index.

Both constructs used a five-point analytic scale for scoring consistency but captured different dimensions of performance. To support LCT-informed interpretation, SG-SD was operationalized as composite proxies: SG gain derived from unpacking and semantic shift, and SD gain from repacking and abstraction control, enabling examination of alignment between meaning movement and Toulmin-based inferential development. While a five-point scale was used for analytic consistency, scoring was based on the functional role of semantic movement in supporting reasoning rather than frequency alone and is not intended to reduce the complexity of meaning-making processes across the text.

All participants provided informed consent for the use of their writing, interaction recordings, and AI chat logs for research purposes. Data were anonymized at transcription, and identifying details were removed from excerpts. Participation had no impact on course grading, and students could opt out of recording without penalty.

Results

Results are reported at the individual level for pre-post writing tasks, while interactional excerpts and ChatGPT logs illustrate how reasoning was negotiated during peer collaboration. The analyses describe patterns of development over the instructional period without attributing change to a single factor. Because all learners engaged in the same inquiry-based tasks under constrained AI use, AI is treated as part of a broader pedagogical ecology—alongside peer interaction and task progression—rather than as an isolated causal variable, with development understood as emerging from their interaction rather than from AI use alone.

RQ1: Development of rhetorical agency across Toulmin elements

RQ1 examined how L2 learners demonstrate emerging rhetorical agency in source-based argumentative writing, operationalized through the deployment of Toulmin elements (claims, data, warrants, backing, qualifiers, and rebuttals).

Descriptive development across Toulmin elements

Table 1 summarizes pre-post performance across Toulmin elements. While all components increased, the largest gains occurred in the inferential elements that require writers to justify, qualify, and contest claims. Particularly, warrants increased from $M = 1.95$ ($SD = 0.72$) to $M = 3.41$ ($SD = 0.85$), backing from $M = 2.00$ ($SD = 0.69$) to $M = 3.59$ ($SD = 0.85$), and rebuttals from $M = 2.14$ ($SD = 0.89$) to $M = 3.59$ ($SD = 1.05$). By contrast, claims and data show smaller increases, suggesting that change was concentrated in how learners explained and evaluated evidence rather than in simply stating positions or listing information.

In addition, the increase in the total Toulmin score (from 12.18 to 20.55; gain = 8.36) indicates a greater use of argumentative elements and a shift in how these elements are functionally organized within the text. Development is reflected not simply in the presence of more elements, but in how claims, evidence, and reasoning are more explicitly connected.

Table 1 *Descriptive Statistics for Toulmin Elements (N = 22)*

Element	Pre M	Pre SD	Post M	Post SD	Gain M
Claim	2.09	0.87	3.59	1.10	1.50
Data	2.18	0.85	3.32	1.09	1.14
Warrant	1.95	0.72	3.41	0.85	1.45
Backing	2.00	0.69	3.59	0.85	1.59
Qualifier	1.82	0.96	3.05	0.84	1.23
Rebuttal	2.14	0.89	3.59	1.05	1.45
Total Toulmin	12.18	2.22	20.55	2.52	8.36

To illustrate this shift more concretely, a short comparison between a lower-scoring pre-test response and a higher-scoring post-test response is presented below. The example highlights how the relationship between claim and evidence becomes more explicit, how conditions are introduced, and how alternative perspectives are incorporated into the argument.

Illustrative pre–post excerpt (annotated)

Pre-test (excerpt)

“Mining is bad for Indonesia because it damages the environment. Rivers become dirty and land is damaged, so the government should stop mining.”

[Claim] Mining is bad for Indonesia

[Data] Rivers become dirty; land is damaged

[Warrant] Implicit (environmental damage = negative)

[Backing] Not present

[Qualifier] Not present (absolute claim)

[Rebuttal] Not present

In this excerpt, reasoning remains largely assertion based. The claim is supported by general observations, but the link between evidence and conclusion is left implicit, and no conditions or alternative perspectives are introduced. Semantically, the argument stays close to a limited set of concrete statements (e.g., “rivers become dirty”, “land is damaged”) without developing these into a broader explanation of why or under what conditions mining should be stopped.

Post-test (excerpt)

“Although mining can support economic growth, it should only be permitted under strict environmental and governance conditions. In several regions, mining creates jobs and infrastructure, but cases of river contamination show that weak regulation can harm communities. This suggests that the issue is not mining itself, but how it is governed.”

[Claim + Qualifier] Mining allowed only under strict conditions

[Data] Jobs and infrastructure; environmental damage cases

[Warrant] The issue is governance, not mining itself

[Backing] Conditions of regulation and accountability

[Rebuttal] Acknowledges economic benefits before limiting the claim

In contrast, the post-test excerpt shows more controlled reasoning. The claim becomes conditional, evidence is drawn from both positive and negative perspectives, and the relationship between evidence and claim is made explicit through a governing principle (“the issue is not mining itself, but how it is governed”), which functions as a warrant. This is accompanied by a clearer movement between concrete examples (“jobs and infrastructure”, “river contamination”) and a more abstract framing of the issue (“how it is governed”), which is then used to shape a conditional policy position.

From an LCT perspective, the pre-test stays largely at strong semantic gravity (concrete consequences) with limited repacking into a governing principle, whereas the post-test explicitly repacks evidence into a higher-density warrant (‘the issue is governance’), then re-grounds it through conditions—forming a clearer semantic wave. This shift reflects a move from implicit, linear justification to a more explicit and balanced argument, where claims are qualified and supported through clearly articulated reasoning. Similar patterns are seen across other students’ responses, where categorical claims (e.g., “mining should be stopped”) are reformulated into more conditional and evidence-based positions.

Pre-post contrasts

Paired-sample comparisons indicated that gains across Toulmin elements were statistically reliable ($p < .001$), with large within-participant effect sizes (Cohen’s d ranging from 1.82 to 2.93). These results point to substantial within-participant change over time; however, they do not isolate the relative contribution of specific instructional components, including peer interaction, task design, or AI consultation.

RQ2: Development of semantic agility in argumentative writing

RQ2 examined how learners’ semantic agility developed during peer–AI collaborative writing, focusing on their ability to move between concrete source-based meanings and abstract inferential reasoning.

Descriptive development across semantic agility indicators

As shown in Table 2, learners demonstrate increases across all four semantic agility indicators. Gains in repacking indicate improved ability to generalize from concrete evidence to abstract reasoning, while gains in unpacking suggest greater capacity to ground abstract claims in contextualized explanation.

Importantly, development was not characterized by a unidirectional shift toward abstraction. Instead, gains across both unpacking and repacking indicate more regulated movement between semantic levels. In LCT terms, this reflects the emergence of more discernible semantic waves, rather than relatively “flat” patterns of meaning.

Table 2 *Descriptive Statistics for Semantic Agility Indicators (N = 22)*

Indicator	Pre M	Pre SD	Post M	Post SD	Gain M
Repacking	1.91	0.92	3.36	1.05	1.45
Unpacking	1.73	0.83	3.27	1.12	1.55
Semantic shift	1.86	0.71	3.36	0.90	1.50
Abstraction control	1.91	0.87	3.41	0.96	1.50
Total semantic agility	7.41	1.84	13.41	2.06	6.00

Semantic Gravity and Semantic Density Proxies

To summarize this movement, SG-SD was modelled as composite proxies derived from the four indicators. Across the post-test texts, both proxies increased, consistent with improved ability to generalize from concrete evidence while also returning to contextual constraints when qualifying claims.

The parallel increases in SG- and SD-based indices are consistent with gains in both unpacking and repacking, indicating more coordinated movement between context-bound and abstract meanings rather than a shift toward abstraction alone. Post-test trajectories show higher values on both SG- and SD-based indices compared with pre-test trajectories.

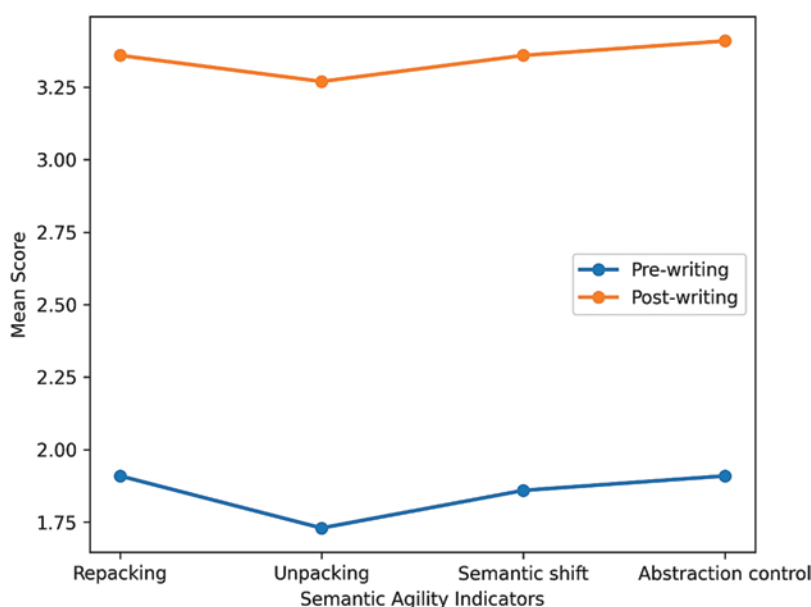


Figure 1 *Semantic Wave Trajectory Across Pre- and Post-Writing Tasks.*

Note. The figure represents LCT-informed semantic gravity (SG) and semantic density (SD) proxies derived from repacking, unpacking, semantic shift, and abstraction control. Post-writing scores indicate more coordinated movement between concrete and abstract meanings.

Illustrative excerpt of semantic movement

The same post-test excerpt further illustrates how semantic movement becomes more controlled. In the pre-test, meaning remains relatively flat. The argument stays at a general evaluative level (“mining is bad”) with limited elaboration and no clear movement between concrete situations and abstract reasoning.

By contrast, the post-test shows a more developed semantic pattern. The text moves from concrete cases (jobs, infrastructure, river contamination) to a more abstract principle (“the issue is governance rather than mining itself”) and then implies policy-level conditions under which mining may or may not be permitted. This pattern reflects more controlled movement between levels of meaning, with writers moving from specific cases to broader generalizations and returning to contextual conditions to qualify their claims.

Linking rhetorical agency and semantic agility

Exploratory correlations were conducted to examine associations between gains in Toulmin elements and gains in semantic agility indices. Table 3 reports correlations between gain scores.

Table 3 *Correlations Between Semantic Gravity-Density and Toulmin Elements*

Toulmin Element (gain)	SG Gain	SD Gain
Claim	-.06	-.06
Data	.11	-.26
Warrant	.26	-.06
Backing	.20	-.13
Qualifier	-.12	-.19
Rebuttal	-.34	.28
Toulmin total	-.02	-.16

Overall correlations were modest. However, selective alignment was observed for elements with higher inferential demands. Gains in warrants show a positive association with SG gain ($r = .26$), and gains in rebuttals indicate a positive association with SD gain ($r = .28$). Claims and data exhibit negligible associations with both indices. These correlations are reported descriptively and are not interpreted as causal relationships.

Interactional Episodes: LREs and LTEs as Process Evidence

Across dyads, ChatGPT was not consistently treated as an authoritative source of argument content. Instead, its responses entered the dialogue as proposals that learners evaluated, reformulated, or rejected, making the locus of reasoning more visible in peer interaction. This interactional visibility is important, as it provides a process-level explanation for the gains in rhetorical agency reported earlier, particularly in learners' ability to articulate warrants, backing, and rebuttals through dialogue.

Variation was also evident across dyad types. Although all groups moved toward more reasoning-oriented engagement across cycles, they did so with different interactional dynamics and levels of regulation. L-L pairs show collaborative orientation but with limited linguistic and epistemic confidence. In T1, they relied heavily on AI for reassurance and often accepted suggestions uncritically. Under the structured constraints of T2 and T3, however, ChatGPT increasingly functioned as a thinking prompt rather than a solution provider. Their LTEs increasingly involved mutual questioning and shared construction of reasoning, although development remained slower and more tentative.

“AI says this is strong evidence... but why? Maybe because it's from the court? Yes, so we should say the court is more valid than opinions.” (L-L, T3)

These dyads exhibit clearer articulation of warrants and backing, although this often emerged after AI prompted initial exploration. Their development reflects what Storch identified as mutuality with developing equality, where reasoning competence grows through joint effort rather than expertise.

H-L pairs commonly reflected an expert–novice pattern, where the higher-proficiency learner often assumed regulatory control over reasoning and decisions, while the lower-proficiency partner contributed reactively. In T1, the high-proficiency member typically directed AI interaction (“Let's ask AI this”), while the low-proficiency member agreed or sought clarification. By T3, however, many H-L dyads displayed productive pedagogic collaboration, with the stronger learner explaining warrants and backing to the weaker peer:

- H: “The reason the court decision works as backing is because it’s an authoritative source.”
 L: “So we say...it is stronger than Roy Suryo’s opinion?”
 H: “Yes, exactly. And we can use AI to help phrase that, not to think for us.”

AI supported conceptual scaffolding, but *human scaffolding remained primary*. These dynamics align with Storch’s claim that expert–novice pairings can be highly generative when expertise is used to mediate thinking rather than dominate.

H–H dyads most frequently displayed collaborative orientation with high equality, engaging in rapid reasoning negotiation, co-regulated decision making, and more complex rebuttal development. AI was rarely accepted at face value; instead, it was evaluated, contested, and strategically appropriated:

- “AI’s rebuttal is okay, but it sounds too neutral. We need to show that scepticism exists while arguing that it doesn’t outweigh verified legal evidence.” (H–H, T3)

These dyads demonstrate the clearest semantic regulation, moving fluidly between concrete case details and abstract evaluative principles, and returning to context to refine stance. This movement between concrete cases and abstract evaluation also reflects the development of semantic agility, as learners shift between context-bound meanings and more generalized reasoning in constructing their arguments. In Storch’s terms, they exemplified high mutuality and high equality, but here extended to high epistemic agency supported by AI as a dialogic, not authorial, partner.

Overall, these interactional patterns suggest that reasoning development was not only reflected in written products, but also co-constructed through dialogue, with AI contributing as a mediational resource rather than a source of answers.

Table 4 *Patterns and Roles of Dyadic Interactions*

Dyad Type	Storch Pattern	AI Role	Agency Pattern	Reasoning Outcome
L–L	Collaborative but tentative	Cognitive opener/reassurance	Shared but fragile	Warrants emerging; slower regulation
H–L	Expert–Novice	Scaffolding amplifier	Guided co-agency	Strengthening justification; mediated learning
H–H	Collaborative (high equality–mutuality)	Dialogic thinking partner	Strong co-agency	Clear warrant–backing–rebuttal + semantic control

Discussion

This study examined how Indonesian EFL learners developed rhetorical agency and semantic agility in source-based argumentative writing through peer–AI collaborative inquiry. In this study, the AI system refers to ChatGPT; broader references situate the findings within AI-supported language learning. The analysis centered on Toulmin’s (2003) model of argumentation, with LCT-informed semantic analysis (Maton, 2013, 2020) used in a supporting role to trace meaning movement across levels of abstraction. This combination allowed the study to examine both the inferential architecture of learners’ arguments and the regulation of meaning that accompanies it. The findings point to development primarily at the level of inferential reasoning and meaning regulation, rather than surface textual features. The findings are discussed in relation to prior research on L2 argumentative writing, meaning-making, and AI-mediated collaboration, with attention to their pedagogical intersections.

AI did not simply support reasoning from the outside; it functioned as a mediational resource within the interaction through which reasoning developed, with its contribution shaped by how it was guided and socially mediated (Wang, 2024) and by how it was integrated and evaluated within the instructional context (Wang, 2025).

Developing Rhetorical Agency Through Inquiry-Oriented Collaboration

Results from RQ1 indicate substantial gains in learners' rhetorical agency, particularly in Toulmin elements associated with inferential reasoning. As shown in Table 1, warrants, backing, and rebuttals demonstrate the largest mean gains (1.45–1.59), whereas claims and data showed comparatively smaller increases. This pattern suggests that learners' development was not driven by improved ability to state positions or select evidence per se, but by enhanced control over how evidence was justified, qualified, and contested.

A notable pattern in these results is the concentration of gains in inferential elements rather than in claims or data. This pattern speaks to long-standing concerns in L2 writing research that learners often produce structurally complete arguments that nevertheless lack persuasive force due to weak inferential explanation (El-Majidi et al., 2020; Hartwell & Aull, 2022; Stapleton & Wu, 2015). The observed gains in rebuttal writing are particularly noteworthy, as counterargument has been repeatedly identified as one of the most persistent challenges for L2 writers (Al Bulushi, 2022; Nussbaum & Schraw, 2007). Rather than remaining underdeveloped, rebuttals showed gains comparable to warrants and backing, indicating that learners became more willing and able to engage with alternative perspectives. This pattern is also reflected in the illustrative pre–post comparison presented earlier, where a categorical claim (“mining is bad”) is reformulated into a conditional, principle-based position grounded in governance and regulation.

What matters here is that these developments appear to be pedagogically responsive, rather than constrained by fixed proficiency limits. Interactional evidence suggests that inquiry-oriented collaboration enabled learners to treat Toulmin elements not as obligatory textual slots but as rhetorical resources for interrogating claims and refining positions. This supports Toulmin's (2003) original conception of argumentation as an inferential practice and extends prior L2 research by demonstrating that rhetorical agency—defined here as functional control over inferential decision-making—can be cultivated through principled instructional design. Peer collaboration played a central role by externalizing reasoning, allowing inferential choices to be negotiated, challenged, and revised (Storch, 2021). In this sense, rhetorical agency emerged as a relational and developmental capacity (van Lier, 2008), rather than as an individual trait tied solely to linguistic proficiency.

Semantic Agility and Regulation of Meaning in Argumentative Writing

Findings from RQ2 further indicate that learners developed greater semantic agility over the instructional period. As shown in Table 2, gains were observed across all four indicators—unpacking, repacking, semantic shift, and abstraction control—with the total semantic agility score increasing by six points from pre-test to post-test. Importantly, these gains were distributed across both unpacking and repacking processes, suggesting improved regulation of movement between concrete evidence and abstract reasoning rather than a unidirectional shift toward abstraction.

These patterns are consistent with LCT's account of effective meaning-making, in which the ability to form clear semantic waves—moving between contextual grounding and principled abstraction—is a hallmark of cumulative knowledge-building (Maton, 2013). The semantic gravity–density proxy trajectories (Figure 1) show parallel increases in both dimensions, indicating that learners became

more adept at grounding abstraction in evidence while also generalizing from specific cases. A similar movement is explicated in the illustrative excerpt, where concrete examples (*e.g., environmental damage, economic benefits*) are reorganized into a more abstract framing of governance and then applied to conditional policy reasoning.

From a pedagogical perspective, these findings speak directly to a persistent dilemma in L2 writing instruction: learners often oscillate between summarizing sources without interpretation and producing decontextualized generalizations (Hirvela, 2017; Qiu & Lo, 2017). The emergence of clearer semantic wave patterns suggests that inquiry-oriented collaboration helped learners manage this tension by making abstraction an explicit object of reasoning. Abstraction was not treated as a stylistic expectation; instead, it became something learners had to justify and work through in interaction.

Moreover, exploratory correlations between Toulmin gains and semantic indices (Table 3) suggest selective alignment between inferential rhetorical functions and meaning movement. Gains in warrants aligned with increases in semantic gravity, while gains in rebuttals aligned with semantic density. Although modest, these associations indicate that semantic agility is most relevant to micro-rhetorical functions that require justification and qualification rather than to claims or data alone. This supports earlier work arguing that abstraction in academic writing develops through rhetorical practice and discourse organization, not lexical sophistication in isolation (Coffin, 2004; Martin, 2013).

Peer–AI Collaboration as Mediated Reasoning Practice

The interactional findings further clarify how rhetorical agency and semantic agility were mediated through peer–AI collaboration. Contrary to concerns that generative AI inevitably encourages passivity or textual dependence, the results show that AI can function as a mediational resource that prompts and sustains reasoning when its use is pedagogically constrained and dialogically framed. Across dyad types, interactional excerpts revealed a shift from form-focused Language-Related Episodes toward reasoning-oriented Learning/Task Episodes, particularly in later task cycles.

These findings suggest that AI-mediated peer collaboration can support the development of inferential reasoning, not as a direct effect of AI use, but as part of a mediational resource shaped by task design, peer engagement, and constrained AI mediation. As summarized in Table 4, dyad configurations shaped how AI mediation unfolded. High–high dyads most consistently engaged in co-regulated reasoning, critically evaluating and selectively appropriating AI suggestions. High–low dyads displayed productive asymmetry, with the more proficient learner mediating inferential reasoning for the partner while maintaining shared decision-making. Low–low dyads initially relied on AI as a source of reassurance but increasingly used it as a prompt for questioning and justification under tighter instructional constraints. These patterns align with Storch’s (2013, 2021) interactional framework and reinforce sociocultural claims that tools support learning only when agency remains with learners (Lantolf & Thorne, 2006).

Importantly, AI did not replace reasoning; rather, it helped create conditions in which reasoning became necessary. The observed development in both rhetorical agency and semantic regulation suggests that AI-mediated collaboration supported the co-development of inferential control and meaning movement. These findings challenge deficit-oriented views of L2 writers and simplistic narratives of AI harm, highlighting that pedagogy—rather than technology itself—determines whether AI constrains or deepens learners’ engagement with argumentative reasoning.

This pattern resonates with recent work in AI-supported language learning. Strobl et al. (2024) show that ChatGPT can function as a “writing buddy” that supports revision by prompting learners to compare, notice, and refine their own texts. Similarly, Tran (2024) demonstrates that AI can facilitate consciousness-raising when learners are guided to engage with prompts rather than passively accept outputs.

The present findings extend these perspectives by suggesting that AI can also support the development of inferential reasoning when embedded within collaborative inquiry rather than individual text production. At the same time, the results align with Lege et al. (2024), who conceptualize large language models as interactional participants that shape how learners engage with language and knowledge. In the present study, however, AI participation did not determine the direction of reasoning; instead, its contributions were taken up, negotiated, or rejected within peer interaction, making the process of reasoning more visible and accountable.

Across dyad types, interactional excerpts revealed a shift from form-focused Language-Related Episodes toward reasoning-oriented Learning/Task Episodes, particularly in later task cycles. These patterns reflect a broader principle: that the role of AI in learning is not fixed but emerges through how it is positioned within instructional design and taken up by learners in interaction.

Contributions to Technology-Enhanced Language Teaching and Learning

This study contributes to research on technology-enhanced language teaching and learning by demonstrating how AI-mediated collaboration can be designed to support higher-order reasoning in L2 argumentative writing without diminishing learner agency. While prior studies have shown that ChatGPT can function as a revision-oriented “writing buddy” (Strobl et al., 2024) and as a tool for consciousness-raising (Tran, 2024), the present findings extend this work by showing how AI can be integrated into inquiry-oriented collaborative tasks that foreground reasoning rather than text production. Rather than treating argumentative writing as primarily a linguistic or structural skill, the findings provide empirical support for conceptualizing it as a site of rhetorical and epistemic development, where learners learn to justify claims, regulate abstraction, and engage with counterarguments through inquiry-oriented activity (Manchón, 2011).

Methodologically, the study illustrates the analytic value of using Toulmin’s model of argumentation as a primary lens complemented by LCT-informed semantic analysis in a supporting interpretive role, to capture both inferential structure and meaning movement in learner writing. This approach responds to recent calls for more fine-grained analyses of argument quality that move beyond surface-level indicators such as organization or lexical sophistication (Qiu & Lo, 2017). By linking Toulmin elements to semantic gravity and density proxies, the study demonstrates how rhetorical reasoning and meaning regulation can be examined as interrelated dimensions of writing development.

Pedagogically, the findings highlight the importance of instructional design in shaping how AI functions in the writing classroom. Consistent with recent work that conceptualizes large language models as interactional participants in learning (Lege et al., 2024), the results show that peer collaboration combined with deliberately constrained AI mediation can promote reasoning-oriented engagement, shifting interaction from form-focused repair toward justification, evaluation, and conceptual negotiation. In this study, AI did not operate as a source of answers, but as a mediational resource whose contributions were taken up, negotiated, or rejected within peer interaction. This supports a view of AI as part of a broader interactional ecology in which learning emerges through dialogue rather than through text generation alone.

The study also contributes a contextual perspective by focusing on an underrepresented EFL setting. In contexts such as Indonesia, prior research has often examined debate activities or genre-based instruction in isolation (Sari & Wahyuni, 2018). By triangulating rhetorical, semantic, and interactional analyses within a classroom-based intervention, the present study extends writing research by showing how technology-enhanced pedagogy can support complex argumentative practices in resource-constrained EFL contexts.

Limitations and Directions for Future Research

Several limitations should be acknowledged. First, the study was conducted with a single intact class over a relatively short instructional period, which limits the generalizability of the findings. Future research could employ longitudinal designs, involve learners across a wider range of proficiency levels, or examine diverse EFL and ESL contexts to explore the durability and transferability of rhetorical and semantic development.

Second, although rich interactional data were collected, the analysis of Language-Related Episodes and Learning/Task Episodes was selective, prioritizing explanatory depth over comprehensive quantification. Subsequent studies may adopt more systematic coding schemes or mixed-method designs to examine interactional dynamics at scale. Third, the study focused on rhetorical agency and semantic agility, leaving other dimensions of writing development—such as lexical complexity, syntactic control, or fluency—outside its analytic scope.

Fourth, the study integrated multiple theoretical perspectives—Toulmin’s model, LCT, and sociocultural mediation—which, while complementary, contribute to analytic density. Future research may benefit from sharpening the analytic focus around a single primary framework while retaining supplementary perspectives in a more streamlined manner.

Finally, while AI use was deliberately constrained to support reasoning, different AI designs and degrees of learner autonomy were not systematically compared. In addition, the study did not include a comparison condition without AI mediation, limiting the ability to isolate the specific contribution of AI relative to peer interaction and task design. Future research could investigate how varying forms of AI scaffolding influence learners’ reasoning processes, agency, and ethical engagement with AI across instructional contexts.

Conclusion

This study demonstrates that the challenges learners face in source-based argumentative writing extend beyond surface-level language production to include difficulties in articulating inferential relationships and regulating movement between concrete evidence and abstract reasoning. Across quantitative and interactional analyses, the findings show that these capacities can develop when argumentative writing is approached as an inquiry-oriented practice supported by peer collaboration and constrained AI mediation.

Despite the modest sample size ($N = 22$), the findings indicate that higher-order reasoning can develop in classroom settings where ChatGPT functions as a participant in dialogic reasoning processes rather than as a text-producing authority. In this study, AI formed part of the mediational environment within peer interaction by introducing prompts that learners evaluated, reformulated, or rejected during peer discussion. Rather than attributing development to AI alone, the results suggest that reasoning was shaped through the interaction between task design, peer collaboration, and constrained AI mediation. In this sense, AI contributed to learning by supporting processes of questioning, justification, and evaluation within guided and socially mediated interaction (Wang, 2024).

For technology-enhanced language teaching and learning, the key implication is that AI’s value lies not in generating text, but in how it is used to support reasoning and engagement with evidence. Learning environments that foreground rhetorical agency and semantic control can help learners engage more critically with both sources and AI. As AI becomes more integrated into writing instruction, the focus should remain on pedagogical design that strengthens rather than substitutes learners’ argumentative reasoning.

Use of Generative AI

This manuscript made use of QuillBot to assist in paraphrasing and improving the clarity of text. No AI tools were used for the generation of research hypotheses or the writing of conclusions. The final manuscript was authored and reviewed by the listed authors without additional AI input.

References

- Al Bulushi, Z. (2022). Source choice and use in counterargument writing. *TESOL Journal*, 13(2), e646. <https://doi.org/10.1002/tesj.646>
- Bacha, N. N. (2010). Teaching the academic argument in a university EFL environment. *Journal of English for Academic Purposes*, 9(3), 229–241. <https://doi.org/10.1016/j.jeap.2010.05.001>
- Coffin, C. (2004). Learning to write history: The role of causality. *Written Communication*, 21(3), 261–289. <https://doi.org/10.1177/0741088304265476>
- Creely, E., Barnes, M., Tour, E., Henderson, M., Waterhouse, P., Pena, M. A., & Patel, S. V. (2025). Exploring attitudes to generative AI in education for English as an additional language (EAL) adult learners. *ReCALL*, 37(2), 174–190. <https://doi.org/10.1017/S0958344024000314>
- Dornburg, A., & Davin, K. J. (2024). ChatGPT in foreign language lesson plan creation: Trends, variability, and historical biases. *ReCALL*, 1–16. <https://doi.org/10.1017/S0958344024000272>
- El-Dakhs, D. A. S. (2024). Artificial intelligence in EFL writing: Threats to learner agency and opportunities for pedagogy. *System*, 120, 103214. <https://doi.org/10.1016/j.system.2023.103214>
- El Majidi, A., De Graaff, R., & Janssen, D. (2020). Debate as L2 pedagogy: The effects of debating on writing development in secondary education. *The Modern Language Journal*, 104(4), 804–821. <https://doi.org/10.1111/modl.12673>
- Fathali, S., & Mohajeri, F. (2025). Artificial intelligence in international English language testing system writing assessments: A comparative study of human ratings and DeepAI. *Technology in Language Teaching & Learning*, 7(4), 103131. <https://doi.org/10.29140/tl.v7n4.103131>
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, 26(2), 5–24. <https://doi.org/10.10125/73474>
- Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, 54, 100666. <https://doi.org/10.1016/j.asw.2022.100666>
- Hartwell, P., & Aull, L. (2022). Constructs of argumentative writing in assessment tools. *Assessing Writing*, 54, 100675. <https://doi.org/10.1016/j.asw.2022.100675>
- Hillocks, G. (2010). *Teaching argument writing, grades 6–12: Supporting claims with relevant evidence and clear reasoning*. Heinemann.
- Hillocks, G. (2011). *The critical role of inquiry in teaching argument*. National Council of Teachers of English.
- Hirvela, A. (2017). Argumentation and second language writing: Are we missing the boat? *Journal of Second Language Writing*, 36, 69–74. <https://doi.org/10.1016/j.jslw.2017.05.002>
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38, 237–257. <https://doi.org/10.1111/jcal.12610>
- Hwang, H., Chang, X., & Sun, J. (2025). Generative AI is useful for second language writing, but when, why, and for how long do learners use it? *Journal of Second Language Writing*, 69, 101230. <https://doi.org/10.1016/j.jslw.2025.101230>
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Kibler, A. K., & Hardigree, C. (2016). Using evidence in L2 argumentative writing: A longitudinal case study across high school and university. *Language Learning*, 66(1), 75–109. <https://doi.org/10.1111/lang.12198>

- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- Lege, R., Bonner, E., & Aikawa, T. (2024). Pecha, a language practice peer: Guiding language learning interactions through large language models. *Teaching and Learning in Technology and Language*, 6(3), 1716. <https://doi.org/10.29140/tltl.v6n3.1716>
- Manchón, R. M. (2011). Writing to learn the language: Issues in theory and research. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 61–82). John Benjamins.
- Martin, J. R. (2013). Embedded literacy: Knowledge as meaning. *Linguistics and Education*, 24(1), 23–37. <https://doi.org/10.1016/j.linged.2012.11.006>
- Martin, J. R., & Maton, K. (2017). *Climbing the hierarchy of abstraction*. Routledge.
- Maton, K. (2013). Making semantic waves: A key to cumulative knowledge-building. *Linguistics and Education*, 24(1), 8–22. <https://doi.org/10.1016/j.linged.2012.11.005>
- Maton, K. (2020). Semantic waves: Context, complexity and academic discourse. In J. R. Martin, K. Maton, & Y. J. Doran (Eds.), *Accessing academic discourse: Systemic functional linguistics and Legitimation Code Theory* (pp. 59–85). Routledge. <https://doi.org/10.4324/9780429280726-3>
- Miller, R. T., & Pessoa, S. (2016). Where's your thesis statement and what happened to your topic sentences? Identifying organizational challenges in undergraduate student argumentative writing. *TESOL Journal*, 7(4), 847–872. <https://doi.org/10.1002/tesj.260>
- Nussbaum, E. M., & Schraw, G. (2007). Promoting argument–counterargument integration in students' writing. *The Journal of Experimental Education*, 76(1), 59–92. <https://doi.org/10.3200/JEXE.76.1.59-92>
- Qiu, X., & Lo, Y. Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, 21(6), 681–698. <https://doi.org/10.1177/1362168816684368>
- Rusfandi. (2015). Argument–counterargument structure in Indonesian EFL writing. *TEFLIN Journal*, 26(2), 181–199. <https://doi.org/10.15639/teflinjournal.v26i2/181-199>
- Sari, M. W., & Wahyuni, S. (2018). Indonesian EFL students' difficulties with academic writing. *Indonesian Journal of Applied Linguistics*, 8(1), 126–137. <https://doi.org/10.17509/ijal.v8i1.11520>
- Schneer, D. (2014). Rethinking the argumentative essay. *TESOL Journal*, 5(4), 619–653. <https://doi.org/10.1002/tesj.123>
- Stapleton, P., & Wu, Y. A. (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes*, 17, 12–23. <https://doi.org/10.1016/j.jeap.2014.11.006>
- Storch, N. (2021). Collaborative writing: Promoting languaging among language learners. In M. García Mayo (Ed.), *Working Collaboratively in Second/Foreign Language Learning* (pp. 13–34). De Gruyter Mouton. <https://doi.org/10.1515/9781501511318-002>
- Storch, N. (2019). Collaborative Writing. *Language Teaching*, 52 (1), 40–59. <https://doi.org/10.1017/S0261444818000320>
- Storch, N. (2013). *Collaborative Writing in L2 Classrooms*. Multilingual Matters.
- Storch, N., & Aldosari, A. (2013). Pairing learners in pair work activity. *Language Teaching Research*, 17(1), 31–48. <https://doi.org/10.1177/1362168812457530>
- Strobl, C., Menke-Bazhutkina, I., Abel, N., & Michel, M. (2024). Adopting ChatGPT as a writing buddy in the advanced L2 writing class. *Teaching and Learning in Technology and Language*, 6(1), 1168. <https://doi.org/10.29140/tltl.v6n1.1168>
- Toulmin, S. (2003). *The uses of argument* (Updated ed.). Cambridge University Press.
- Tran, T. (2024). Exploring the potential of ChatGPT in facilitating consciousness-raising tasks: A promising solution for alleviating the burden on educators. *Teaching and Learning in Technology and Language*, 6(1), 1104. <https://doi.org/10.29140/tltl.v6n1.1104>

- van Lier, L. (2008) Agency in the classroom. In J. P. Lantolf & M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 163–186). Equinox.
- Wang, Y. (2025). Research methods and AI. In Y. Wang, A. Alm, & G. Dizon (Eds.), *Insights into AI and language teaching and learning* (pp. 115–132). Castledown Publishers. <https://doi.org/10.29140/9781763711600>
- Wang, Y. (2024). Cognitive and sociocultural dynamics of self-regulated use of machine translation and generative AI tools in academic EFL writing. *System*, 126, 103505. <https://doi.org/10.1016/j.system.2024.103505>
- Zare, J., Al-Issa, A., & Ranjbaran Madiseh, F. (2025). Interacting with ChatGPT in essay writing: A study of L2 learners' task motivation. *ReCALL*, 1–18. <https://doi.org/10.1017/S0958344025000035>