# Writing-to-Think Learning Logs for Developing Critical Thinking in L2 Discussion Classes

**Megumi Uemi***

***Corresponding author's email: uemi@toyo.jp**

**Toyo University**

*Received: October 9, 2025* *Revised: April 24, 2026* *Accepted: June 7, 2026*

## Abstract

This study investigates how *writing-to-think learning logs* can support the development and articulation of critical thinking (CT) in Japanese university L2 discussion courses. *Writing-to-think* is conceptualized as using writing not only to record but to generate and clarify reasoning, and is integrated with *CT-infused* instruction, which explicitly teaches reasoning and evaluation within language learning tasks. Implemented over one academic year with CEFR A2–B1 learners (analytic sample: $n = 24$ from a larger cohort of $N = 79$), the study employed three recurring log tasks—*opinion expression (OE)*, *text summarization (TS)*, and *evaluation of viewpoints (EV)*—that prompted students to articulate reasoning in writing before and after class discussions. A sequential explanatory mixed-methods design was adopted. Quantitatively, rubric-based ratings of learning-log entries (focus, logic, and content) were analyzed across three time points (Sessions 2, 9, and 16). Results showed significant improvement in OE and EV, with larger gains observed in EV, while TS remained stable. Qualitatively, thematic analysis of representative students' entries revealed four developmental patterns: a shift from fragmented to coherent reasoning, growth in evaluative thinking, more relevant examples, and continued difficulty distinguishing reasons from examples. Overall, CT growth was observed in association with the integration of explicit reasoning demands and iterative writing cycles that slowed and structured thinking, making reasoning visible and open to reflection. Learning logs thus functioned as both scaffold and research artifact, offering a practical means to make CT teachable and traceable in L2 discussion classrooms.

**Keywords:** Critical thinking, CT-infused instruction, learning logs, L2 discussion, writing-to-think

## Introduction

Fostering critical thinking (CT) is a central goal of higher education, particularly in English as a First Language (L1) contexts where well-reasoned argumentation and effective communication are hallmarks of academic success (Facione et al., 1995). This emphasis has increasingly extended to English as a Second Language (L2) settings across Asia, where CT is viewed as essential for developing globally competent professionals (Yin et al., 2023). However, despite its recognized importance, CT remains difficult to operationalize in L2 classrooms, as models developed in L1 contexts often fail to align with learners' linguistic and sociocultural realities (Wang & Wu, 2023).

In Japan, this misalignment is particularly pronounced. Although efforts to internationalize higher education have raised expectations for students' CT performance, their actual demonstration of such skills has often fallen short. Linguistically, limited proficiency constrains the articulation of reasoning (Manalo & Sheppard, 2016). Pedagogically, instructional traditions in some contexts have been described as emphasizing memorization, offering fewer opportunities for sustained engagement with higher-order thinking (Egitim, 2022). Socioculturally, norms reported to value harmony and indirectness may discourage explicit argumentation and critique (Okada, 2017; Wang & Wu, 2023). As a result, students' reasoning may remain underdeveloped or insufficiently expressed in academic discourse. Similar patterns observed in other Asian and Southeast Asian L2 contexts suggest that these challenges may reflect broader issues in the articulation of CT in English (Namsaeng & Ambele, 2024; Tan, 2017).

This issue is especially salient in discussion-based classrooms, where students are expected to comprehend, evaluate, and synthesize ideas in real time. The cognitive and linguistic demands of spontaneous interaction can inhibit the externalization of reasoning, leading to situations where students possess ideas but struggle to articulate them clearly (Egitim, 2022; Lin et al., 2016). Thus, the issue lies not necessarily in a lack of CT ability, but in limitations on its articulation, where multiple constraints prevent learners from making their reasoning visible in academic contexts.

To address this issue, this study introduces *CT-infused learning logs* as an instructional scaffold to support the articulation and development of reasoning in L2 discussion contexts. Implemented as structured pre- and post-discussion writing tasks, the logs guide students through three recurring activities—opinion expression, text summarization, and evaluation of viewpoints—each aligned with core CT processes. The study therefore focuses not on CT or writing development in isolation, but on their interaction: how writing functions as a means of making CT visible and, consequently, developable in L2 contexts. By providing time and structure for idea organization, learning logs serve as a mediational tool that makes reasoning more explicit and transferable to discussion. Conceptually, this study positions learning logs within a *writing-to-think* framework, where writing functions not merely as a record of thought but as a means of generating and refining it. Building on the author's previous work (Uemi, 2024, 2025), the present study extends this line of inquiry through a longitudinal

mixed-methods design, enabling a more systematic examination of developmental patterns and an integrated analysis of quantitative and qualitative data. Empirically, it draws on longitudinal data from Japanese university students (CEFR A2–B1) to examine how iterative engagement with these tasks supports the articulation and development of CT over time in discussion-oriented L2 classrooms.

## Literature Review

### Conceptualizing CT as *Writing-to-Think*: The Visibility of Reasoning Through Language

Despite its global importance, CT has yet to achieve a unified definition (Liu et al., 2014). Nevertheless, it is widely understood as a higher-order mode of reasoning that is reflective, reasonable, and rational—the "three Rs" (Dummett & Hughes, 2019). These dimensions capture complementary aspects of thinking: reflection involves reconsidering assumptions, reasonableness requires justification through evidence and logic, and rationality emphasizes balanced judgment about what to believe or do. Together, they align with higher-order processes such as analysis, synthesis, and evaluation in Bloom's taxonomy (Anderson & Krathwohl, 2001; Krathwohl, 2002), as well as with definitions of CT as reasonable, reflective decision-making (Ennis, 2011) and purposeful, self-regulatory judgment (Facione, 1990).

Although CT is often described as encompassing both cognitive skills and dispositional tendencies (e.g., open-mindedness and inquisitiveness), dominant accounts emphasize its cognitive dimension (Ennis, 2011; Facione, 1990). In line with this distinction, the present study operationalizes CT as a form of performance, focusing on the observable articulation of reasoning in written discourse rather than on dispositional orientations. From this perspective, CT is conceptualized not only as a cognitive process but also as a discourse-based practice in which reasoning becomes visible through the articulation of claims, justification, and evaluation in language. Scholars have long emphasized that reasoning becomes most accessible—and most developable—through language, particularly writing (Cottrell, 2017; Wade, 1995). Ennis (2011) highlights rhetorical clarity as central to CT, while Facione and Facione (1994) operationalize CT through written performance. Similarly, Paul and Elder (2006, 2019) and Nosich (2022) argue that reasoning is inseparable from articulation, with writing offering a particularly rigorous means of testing and refining thought.

Writing, in this sense, does not merely record thinking but actively enables it. As ideas are externalized, they become available for reflection, revision, and refinement (Menary, 2007). Foundational L1 research demonstrates that composing engages recursive processes that can restructure knowledge itself (Bereiter & Scardamalia, 1987; Emig, 1977; Flower & Hayes, 1981). This function is especially salient in L2 contexts, where writing provides time and structure for learners to clarify ideas that may be difficult to access in spontaneous speech (Mehta & Al-Mahrooqi, 2015; Mineshima & Imai, 2017). L2 research similarly positions writing as a mediational tool for cognitive development, as producing text engages learners in noticing, restructuring, and justifying ideas (Cumming, 2001;

Manchón, 2011; Swain, 2006). Empirical studies further demonstrate that CT can be assessed through written performance, with rubric-based evaluations of reasoning and justification serving as indicators in applied contexts (e.g., Nejmaoui, 2019; Stapleton, 2001).

This relationship between CT and writing is best understood within the tradition of *writing-to-think* (Menary, 2007; Nosich, 2022; Paul & Elder, 2019), which views writing as a generative and transformative process rather than a neutral record of ideas (Applebee & Langer, 1987; Galbraith, 1999). For L2 learners, this function is not merely supportive but essential, enabling reasoning to take more explicit and structured form. Accordingly, improvements in textual organization and coherence may reflect increased control over the discourse through which reasoning is articulated (Manchón, 2011; Staples & Reppen, 2016). Such development often depends on explicit organizational support, particularly in contexts where discourse conventions differ across sociocultural backgrounds (Kubota & Lehner, 2004; Watanabe, 2023). Without such support, reasoning may be present but difficult to express coherently, potentially leading to an underrepresentation of students' CT performance in written output (Cumming, 2001; Hyland, 2003).

Thus, the present study defines CT as both a cognitive process—embodied in analysis, synthesis, and evaluation and guided by the reflective, reasonable, and rational "three Rs"—and a discourse practice: reasoning disciplined by evidence and made visible, teachable, and improvable through the iterative work of writing. Within this framework, CT is examined as a form of performance manifested in learners' written output, allowing for systematic analysis through rubric-based assessment of focus, logic, and content.

**CT-Infused Instructional Models in L2 Writing**

If CT is indeed a cognitive process that can be activated through writing, does simply asking students to write lead to improved CT? Research suggests otherwise: the effectiveness of writing for CT depends on instructional design. Without explicit CT-oriented guidance, extended writing practice often yields surface-level improvements in mechanics or fluency without deeper reasoning (Condon & Kelly-Riley, 2004; Seshachari, 1994). Hirose and Sasaki (2000), for instance, found that journal writing improved Japanese EFL learners' fluency but did not foster content development or organization due to its expressive, unstructured nature. Likewise, Seshachari (1994) showed in L1 contexts that unmediated journal writing resulted in limited progress in discourse-level reasoning, emphasizing the importance of instructor mediation and concept-oriented support. As Condon and Kelly-Riley (2004) put it, "Writing acts as a vehicle for critical thinking, but writing is not itself critical thinking" (p. 66). Taken together, these studies suggest that writing alone—particularly when unstructured—supports fluency and reflection but does not develop disciplined reasoning. This highlights a key distinction between writing-to-think as a pedagogical goal and the types of writing activities used to achieve it.

To address this limitation, a growing body of research has focused on CT-infused instructional models. Yin et al. (2023) identify four broad approaches—*general*, *immersion*, *infusion*, and *mixed*—

and report that infusion is the most widely adopted in L2 writing. Unlike immersion, where CT is addressed implicitly, or general approaches, where CT is taught separately, infusion integrates CT principles directly into subject-specific tasks (Ennis, 1989; Nejmaoui, 2019). Meta-analyses further show that infusion and mixed approaches—where general instruction is combined with either infusion or immersion—produce the most consistent CT gains (Abrami et al., 2015).

Empirical studies illustrate how such integration can be operationalized in classroom contexts. Lu and Xie (2019), for example, integrated Paul and Elder's (2006) CT framework into a semester-long Chinese EFL writing course using a pre–post design, reporting significant gains in reasoning and writing performance based on standardized CT measures and rubric-based assessments. Similarly, Yang et al. (2014) implemented a CT-integrated literacy course in Taiwan through an online, semester-based intervention, with outcomes evaluated through pre–post comparisons using instruments such as the California Critical Thinking Skills Test and writing rubrics. Chason et al. (2017) focused on enhancing discourse structure by extending the TSR (Topic-Support-Return) model into TBSIR (Topic-Bridge-Support-Interpretation-Return), leading to improved coherence and depth in multilingual learners' writing.

While this research demonstrates the effectiveness of CT-infused instructional design, it also reveals several limitations. Many such studies are conducted over relatively short instructional periods and assess outcomes primarily through final written products or post-intervention assessments (e.g., Lu & Xie, 2019; Yang et al., 2014), with limited attention to how reasoning develops across instructional sequences. In addition, writing is often treated as the outcome of CT rather than as a mediational process through which thinking is formed and refined (e.g., Chason et al., 2017). Consequently, less attention has been paid to how iterative, scaffolded writing can support both classroom learning and the ongoing development of reasoning over time.

The present study addresses this gap by adopting learning logs—traditionally used as reflective tools but here reconceptualized as a writing-to-think pedagogy. In this approach, writing functions not merely as a product of CT but as its medium, allowing reasoning to unfold, stabilize, and become visible across instructional cycles. Specifically, this study advances prior infusion research by treating learning logs as longitudinal, process-tracing artifacts, enabling the examination of how students' reasoning evolves across instructional cycles rather than being assessed solely through end-point outcomes.

**Learning Logs as Writing-to-Think Pedagogy for CT Development**

Learning logs (sometimes referred to as *logbooks* or *learning journals*) extend CT-infused instruction by providing a structured means of embedding reasoning into L2 learning through iterative writing. In this study, they are framed within a writing-to-think perspective (Emig, 1977; Menary, 2007), which emphasizes writing as a process that generates and refines reasoning rather than simply recording it. Importantly, learning logs must be distinguished from other forms of reflective writing. Unlike open-ended journals, which may build fluency but often focus on affective or strategic reflection

(Hammad, 2022; Hirose & Sasaki, 2000), learning logs—as defined here—are task-oriented and content-driven records that guide learners to articulate, test, and refine reasoning progressively (Khonamri et al., 2016; Uemi, 2025). In this sense, journaling is typically expressive and exploratory, whereas learning logs represent a more structured form of writing-to-think, designed to make reasoning explicit and progressively developable.

Beyond this study-specific definition, prior research on learning logs highlights their broader pedagogical value as tools for consolidating classroom experiences and monitoring learning processes (Dam, 2009; McCrindle & Christensen, 1995). In L2 contexts, such tools have been shown to promote learner autonomy and metacognitive awareness (Kaneda, 2022; Menegale, 2020), as well as deeper engagement with ideas in structured tasks (Khonamri et al., 2016), reinforcing their potential as platforms for articulating and developing reasoned thinking. Across these studies, a key strength of learning logs lies in their ability to externalize thinking in a cumulative and revisable form—an affordance that aligns directly with their use as a writing-to-think pedagogy. Successive entries preserve traces of both thought and language, enabling learners and researchers to trace developmental shifts over time. In this sense, logs function as a portfolio-like record through which students can monitor their own reasoning development.

In the present study, learning logs are infused with three CT-oriented tasks—opinion expression, text summarization, and evaluation of viewpoints—closely integrated with classroom discussions. They operate through a cyclical process of writing, discussion, and reflection: students first prepare ideas in writing, then extend and test them in peer interaction, and finally refine their entries as homework. This cycle provides all learners—including those less able to contribute spontaneously—with a "thinking space" to rehearse and articulate reasoning (Mineshima & Imai, 2017). Unlike prior implementations that emphasize written outcomes, this design explicitly integrates writing with interaction, positioning learning logs as a bridge between individual reasoning and collaborative meaning-making. The process is further supported by an expression framework guiding students to state a main idea, explain its reasons, elaborate with examples, and restate their argument. Analytically, the cumulative entries provide process-sensitive data for examining how CT becomes clearer, more justified, and more coherent over time.

Earlier implementations in Japanese university classrooms (Uemi, 2024, 2025) showed that learning logs improved students' written CT performance—particularly in clarity of argument and depth of reasoning—while also enhancing confidence in oral discussions. These findings align with prior research highlighting the role of structured, guided writing tasks in supporting the articulation and development of higher-order thinking in L2 contexts (e.g., Khonamri et al., 2016), as well as studies emphasizing the role of log-based writing in fostering reflection and learner awareness (Menegale, 2020). Building on this work, the present study positions learning logs as a means of making CT both teachable and traceable through written discourse in L2 discussion courses.

### Purpose of This Study

The present study examines learning logs as analyzable artifacts that make CT processes visible over time in an infusion-based L2 discussion course. While previous studies have demonstrated the benefits of explicit CT instruction and log-mediated scaffolding, the evidence has largely been short-term and product-focused, and the logs themselves are rarely analyzed as process-tracing artifacts. To address this gap, the study adopts a writing-to-think perspective, conducting a longitudinal analysis of learning-log writing within a CT infusion design to trace changes in speaking-focused lessons for A2–B1 learners. The aim is to investigate how structured, log-based writing supports CT across an academic year and to provide a fine-grained account of how learners' articulated reasoning develops. The study is guided by the following research questions:

**RQ1:** How does students' CT performance in writing evolve across the three CT-infused learning log tasks (opinion expression, text summarization, evaluation of viewpoints) over one academic year in a discussion-based course?

**RQ2:** How do students at different levels of CT performance (high, moderate, and low gainers) demonstrate qualitative patterns of growth and challenge in their learning logs over the instructional period?

## Methodology

### Research Design

This study employed a sequential explanatory mixed-methods design (Creswell, 2015) over one academic year. RQ1 quantified growth in CT performance via rubric-based evaluations of learning-log entries sampled at three time points across three tasks (opinion expression, text summarization, and evaluation of viewpoints). Guided by RQ1 results, RQ2 qualitatively analyzed purposively selected entries from the two tasks, showing the greatest gains to examine how students articulated and organized reasoning over time. Although data were collected as part of regular instruction, analyses proceeded sequentially, with integration at the interpretation stage through side-by-side comparison of quantitative trends and qualitative patterns. This design triangulated measurable outcomes with observable reasoning, providing a nuanced account of CT development in lower- to intermediate-level L2 learners.

## Participants

The study was conducted during the 2024 academic year in three required oral communication courses at a private university in Japan ($N = 79$), targeting CEFR A2–B1 learners. For analysis, a subset of 24 first-year students was selected based on the availability of completed paired TOEIC® Listening and Reading placement (start of semester) and progress (end of semester) scores, thereby ensuring a consistent sample with complete longitudinal data. The analytic sample did not differ substantially from the larger cohort in proficiency distribution, with TOEIC scores largely concentrated within the CEFR A2–B1 range ($M = 567$; range = 345–885). However, as inclusion was based on data completeness, potential selection bias—such as differences in attendance or task completion—cannot be fully ruled out. The use of TOEIC paired data reflects the broader research design, which initially aimed to examine relationships between learning log performance and proficiency development. In the present study, however, only learning log data are analyzed; TOEIC data will be used in a follow-up study to explore proficiency–CT relationships and generalizability. All participants provided informed consent for the use of their in-class and post-class learning logs. Participation was voluntary, and confidentiality was ensured. The study was approved by the Research Ethics Committee of the Faculty of Letters, Toyo University, and conducted in accordance with institutional ethical guidelines.

## Materials and Procedure

### *CT-Infused Learning Log Tasks as Writing-to-Think*

Three CT-infused learning log tasks were implemented as a writing-to-think pedagogy to target distinct aspects of CT performance. Embedded before and after discussions, the tasks supported the generation, clarification, and refinement of reasoning in writing, facilitating both oral contributions and subsequent reflection. The recurring tasks were: (1) opinion expression (OE), focused on articulating reasoned responses; (2) text summarization (TS), emphasizing identification of main ideas; and (3) evaluation of viewpoints (EV), encouraging appraisal of perspectives discussed in class. Together, the tasks operationalized Dummett and Hughes's (2019) "three Rs" of CT—reflective, reasonable, and rational—while engaging higher-order skills (Anderson & Krathwohl, 2001). Specifically, OE emphasized analysis and synthesis, TS combined comprehension with analysis and synthesis, and EV integrated evaluation and synthesis into judgment (see Table 1). In the learning-log template, these appeared as *Today's Question for You*, *Today's Story/Discussion*, and *Today's Reflection*, corresponding respectively to OE, TS, and EV.

**Table 1.**

*Learning-Log Tasks, Objectives, and Engagement with the "Three Rs" of CT and Higher-Order Skills*

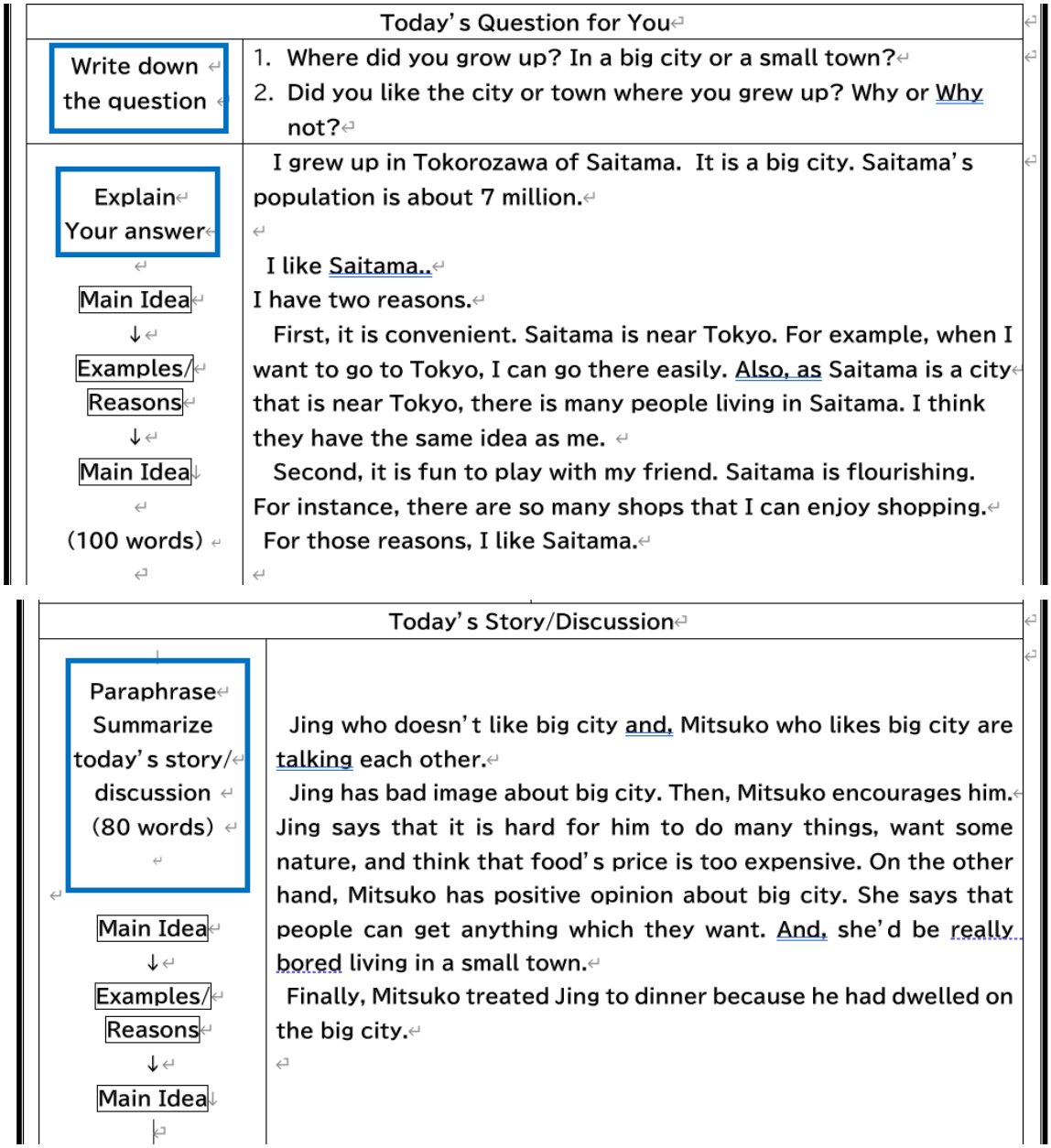
| Task | Objective | Reflective | Reasonable | Rational | Higher-Order Skills |
|---|---|---|---|---|---|
| **OE** | Articulate a relevant, well-focused response to a prompt. | Surfaces personal stance and prior knowledge. | Requires fair, well-supported claims. | Builds a coherent line of reasoning (assertion → reason → example). | • Analysis of issues<br>• Synthesis of stance, reasons, and examples |
| **TS** | Summarize the text's main ideas and arguments. | Demonstrates thorough understanding of the source. | Faithfully conveys the author's ideas without distortion. | Organizes main points and supporting details logically. | • Analysis of main ideas<br>• Synthesis in organizing arguments |
| **EV** | Reflect on a discussion and evaluate differing perspectives. | Deepens insight by revisiting views post-discussion. | Applies fair-minded criteria when judging viewpoints. | Analyzes the consistency and strength of reasons across positions. | • Evaluation of perspectives<br>• Synthesis of multiple viewpoints into reasoned judgment. |

Task selection was based on the CT-oriented discussion textbook *Impact Issues 2* (Day et al., 2019). Each unit includes (a) warm-up prompts, (b) a short text with interpretation questions, and (c) discussion prompts. These were adapted into OE, TS, and EV, respectively. This sequence—from personal stance to text comprehension to evaluation—reflects common discussion-based pedagogy, supporting both familiarity and ecological validity.

In Session 2, for example, students read a dialogue between Jing and Mitsuko debating city versus small-town life (*Impact Issues 2*, Unit 2). Jing highlights urban frustrations, while Mitsuko emphasizes convenience. The OE task elicited students' views on their hometowns, TS required a summary of the exchange, and EV prompted evaluation of both perspectives. Prompts were adapted to elicit more explicit reasoning. Occasionally, authentic materials such as TED Talks or YouTube videos replaced textbook units to expose students to more natural language. Representative prompts from Sessions 2, 9, and 16 appear in Appendix A.

Across the 2024 academic year, CT-infused instruction spanned 18 lessons (10 in spring, 8 in fall), excluding orientation and testing weeks. Each session included all three tasks—OE and EV

individually, TS individually or collaboratively, depending on text complexity. Following in-class writing, students discussed ideas, then revised and submitted digital versions via the learning management system within one week. Recommended word counts were 100 (OE), 80 (TS), and 50 (EV), with weekly totals often exceeding 230 words per student. Figure 1 shows sample entries from Student 2 in Session 2.



| Today's Question for You | |
|---|---|
| Write down the question | 1. Where did you grow up? In a big city or a small town?<br>2. Did you like the city or town where you grew up? Why or Why not? |
| Explain Your answer<br>Main Idea<br>↓<br>Examples/ Reasons<br>↓<br>Main Idea<br>(100 words) | I grew up in Tokorozawa of Saitama. It is a big city. Saitama's population is about 7 million.<br>I like Saitama..<br>I have two reasons.<br>First, it is convenient. Saitama is near Tokyo. For example, when I want to go to Tokyo, I can go there easily. Also, as Saitama is a city that is near Tokyo, there is many people living in Saitama. I think they have the same idea as me.<br>Second, it is fun to play with my friend. Saitama is flourishing. For instance, there are so many shops that I can enjoy shopping.<br>For those reasons, I like Saitama. |

| Today's Story/Discussion | |
|---|---|
| Paraphrase Summarize today's story/ discussion (80 words)<br>Main Idea<br>↓<br>Examples/ Reasons<br>↓<br>Main Idea | Jing who doesn't like big city and, Mitsuko who likes big city are talking each other.<br>Jing has bad image about big city. Then, Mitsuko encourages him. Jing says that it is hard for him to do many things, want some nature, and think that food's price is too expensive. On the other hand, Mitsuko has positive opinion about big city. She says that people can get anything which they want. And, she'd be really bored living in a small town.<br>Finally, Mitsuko treated Jing to dinner because he had dwelled on the big city. |

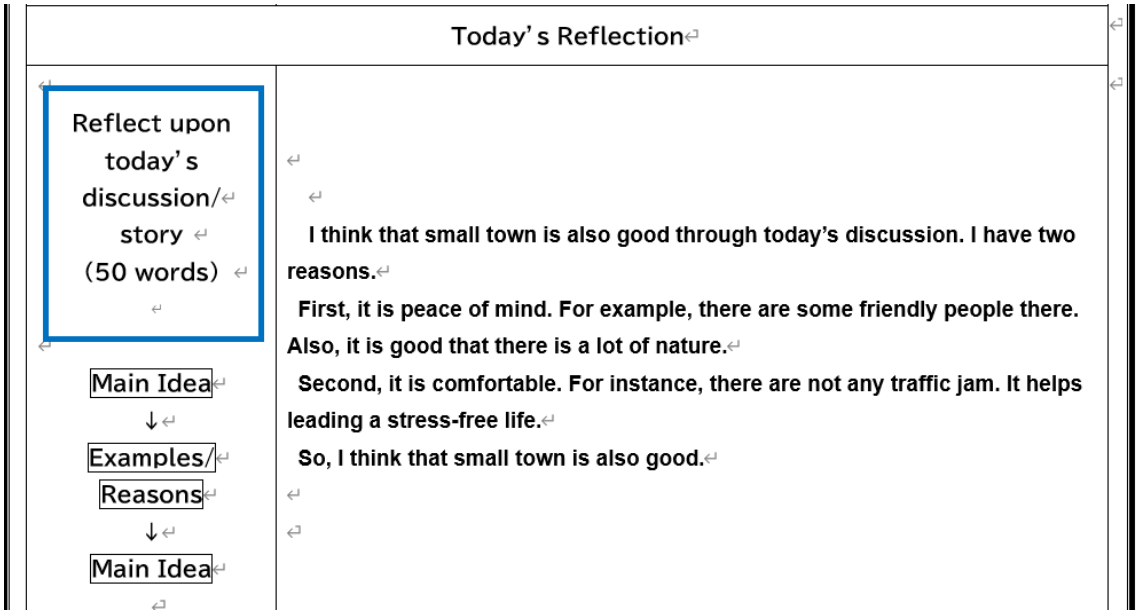



**Figure 1.** *Sample Submitted Learning-Log Entries from Student 2 in Session 2*

***CT Expression Framework***

At the outset of the course, CT was introduced as reflective, rational, and reasonable thinking (Dummett & Hughes, 2019), and CT performance as the ability to express such thinking clearly and effectively through language (Nosich, 2022; Paul & Elder, 2019). Definitions were provided in both English and Japanese to emphasize the dual aims of fostering critical thought and clear communication.

To support this, the Critical Thinking Expression Framework (CTEF) was introduced as a scaffold comprising four elements—main idea, reason, example, and restate—adapted from opinion-paragraph structure. Unlike conventional paragraph-writing instruction, which primarily emphasizes formal organization (e.g., topic sentences and supporting details), CTEF focuses on the construction of reasoning. It guides learners to differentiate key components of thought by identifying a clear claim, developing it through justification and elaboration, and refining it through restatement. In doing so, it prioritizes the articulation of reasoning over the arrangement of textual structure.

CTEF aligns with the Critical Thinking in Writing Rubric (CTWR; McLaughlin & Moore, 2012) used in this study for assessment (see the following section for details). The rubric's criteria—focus, logic, and content—correspond respectively to the clarity and relevance of the main idea, the strength of reasoning, and the depth of elaboration. Together, these criteria operationalize CT in this study by capturing how learners articulate, justify, and develop their reasoning in written form.

Because the framework targets the articulation of reasoning rather than specific content, it is adaptable across instructional contexts, allowing its core elements to be embedded in a range of tasks, including discussion prompts, reading responses, and reflective writing activities. Applied across OE, TS, and EV tasks and supported by visual aids, the framework helped A2–B1 learners reduce cognitive

load, organize ideas systematically, and communicate reasoning with greater clarity. A schematic representation is shown in Figure 2.

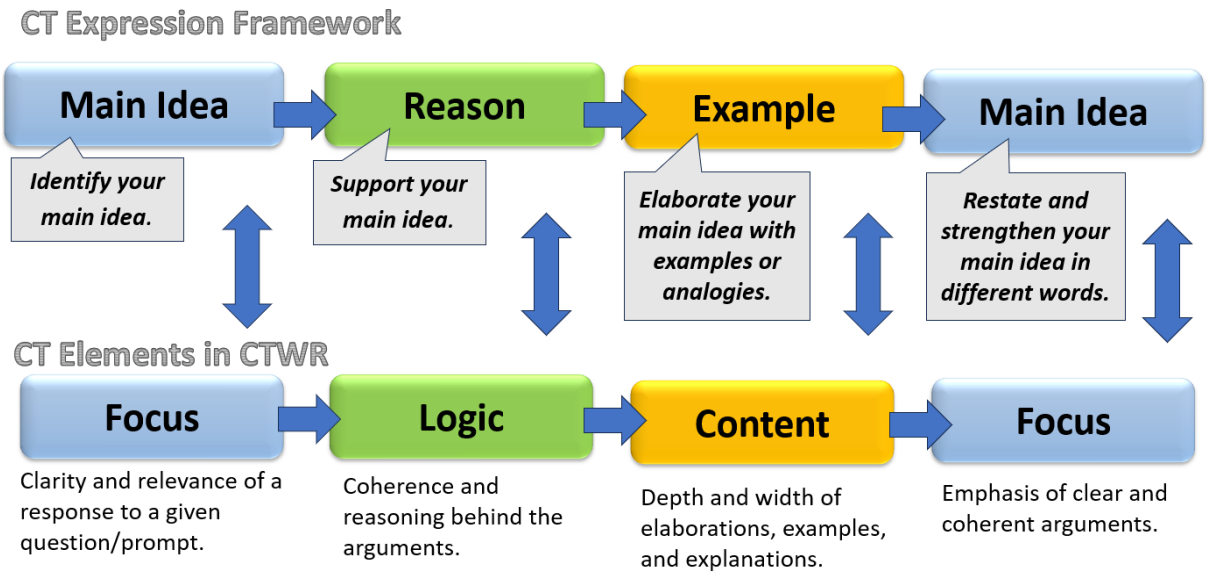



**Figure 2.** *Alignment of the CTEF with CT Elements in the CTWR*

**Data Collection and Analysis**

To address RQ1—*How does students' CT performance in writing evolve across the three learning-log tasks (OE, TS, EV) over one academic year?* —a rubric-based analysis was conducted on entries from 24 students at three time points: Session 2 (April), Session 9 (July), and Session 16 (December). An adapted version of the Critical Thinking in Writing Rubric (CTWR; McLaughlin & Moore, 2012) was employed, aligning with the study's conceptualization of CT as reasoning made visible through writing. From the original five criteria, three—focus, logic, and content—were selected as core indicators of CT, excluding style, correctness, and research as not relevant to course-based tasks. Each entry was rated on a four-point scale (1 = Inadequate to 4 = Superior; see Appendix B), and scores were averaged to yield a task-level composite representing overall CT performance. These composites demonstrated acceptable to high internal consistency ($\alpha$ = .90 for OE, .91 for TS, and .66 for EV), supporting their use as equally weighted indicators of CT.

Each student received three task scores per session (OE, TS, EV), enabling longitudinal comparison across Sessions 2, 9, and 16. Although rubric-based scores are ordinal, averaged composites were treated as interval data for parametric analysis, a common and widely accepted practice when multiple items are aggregated to represent an underlying construct (Norman, 2010). Repeated-measures ANOVA was used to examine within-subject changes across time points, given its suitability for balanced longitudinal datasets. Assumption checks were conducted prior to interpretation, and the

results were reported using multivariate Wilks' Lambda tests and linear trend analyses to examine overall task-level change and developmental trends over time.

The rating process comprised three phases. In Phase 1, two raters (including the author) completed one-hour training using benchmark entries and pilot scoring. In Phase 2, both independently rated approximately 30% of the dataset (eight students across tasks and sessions). Inter-rater reliability was high (Cronbach's $\alpha$ = .956 for OE, .947 for TS, and .935 for EV), and discrepancies were resolved through discussion. Cronbach's alpha was used to assess the internal consistency of the composite scoring structure across the three dimensions. Although intraclass correlation coefficients (ICC) are often used for inter-rater agreement, the present study focused on the reliability of the aggregated scoring structure. In Phase 3, the remaining entries were rated by the author using the finalized rubric to ensure consistency, with periodic review of scoring criteria to mitigate potential rater drift. This partial double-rating approach reflects established practice in classroom-based research.

To address RQ2—*How do students at different levels of CT performance demonstrate qualitative patterns of growth and challenge in their learning logs?* — task-level CT scores were averaged within each session to create a composite score per student. Gain scores were calculated by subtracting Session 2 composites from Session 16 composites, providing an index of CT growth. To facilitate comparison of developmental patterns, gain scores were categorized into three groups (High, Moderate, Low) using thresholds derived from the observed distribution of scores within the sample, taking into account the range and natural groupings in the data. Specifically, the thresholds were established by identifying breaks in the distribution that separated clusters of higher, mid-range, and lower gain scores. This approach enabled differentiation between students demonstrating substantial, moderate, and minimal change over time.

Based on these criteria, students were classified as High-Gain (≥1.00), Moderate-Gain (0.33–0.89), or Low-Gain (≤0.22). No students fell within the intermediate range between 0.23 and 0.29, indicating a natural break in the distribution that informed the categorization. These categories were used for analytical comparison and illustrative purposes rather than for inferential statistical analysis. The distribution of the 24 students is shown in Table 2.

**Table 2.**

*Distribution of Students by Gain Score Category*

| **Category (Gain Range)** | ***N* (students)** | **Gain Range (min–max)** |
|---|---|---|
| High-Gain (≥1.00) | 3 | 1.00–1.56 |
| Moderate-Gain (0.33–0.89) | 14 | 0.33–0.89 |
| Low-Gain (≤0.22) | 7 | -0.33–0.22 |

From these categories, two students per group ($N$ = 6) were purposively selected to illustrate representative developmental trajectories while maintaining analytical depth (Table 3), following

stratified purposeful sampling principles (Patton, 2015). This sampling approach was adopted to balance depth and variation, enabling detailed qualitative analysis while reflecting the range of developmental patterns within the cohort.

**Table 3.**

*Selected Students for Qualitative Analysis and Their Composite Scores*

| Gain Group | Student ID | Session 2 Composite | Session 16 Composite | Gain Score |
|---|---|---|---|---|
| **High-Gain** | 2 | 2.44 | 4.00 | 1.56 |
| | 11 | 1.78 | 3.00 | 1.22 |
| **Moderate-Gain** | 8 | 1.67 | 2.78 | 0.89 |
| | 5 | 3.00 | 3.78 | 0.78 |
| **Low-Gain** | 17 | 3.11 | 3.33 | 0.22 |
| | 4 | 3.44 | 3.67 | 0.22 |

Thematic analysis followed Braun and Clarke's (2006) six-phase framework. Data comprised OE and EV responses from six focal students in Sessions 2, 9, and 16, the tasks showing the clearest CT development. Coding was inductive and guided by the CT Expression Framework to interpret the structure of ideas, with themes capturing the clarity of main ideas, logical development, evaluative depth, and supporting evidence. The author and a trained rater coded collaboratively, resolving discrepancies through discussion. Inter-coder agreement was established through iterative comparison and consensus-building, with coding decisions refined through discussion to ensure consistency and transparency. CTWR dimensions (focus, logic, content) were used to contextualize themes and maintain alignment with quantitative findings.

## Results

### RQ1: Quantitative Trends in CT Performance

To address RQ1—how students' CT performance evolved across the three learning-log tasks (OE, TS, and EV) over one academic year—a repeated-measures ANOVA was conducted. Mauchly's test indicated that the assumption of sphericity was met for OE and EV ($p > .05$) but violated for TS ($p = .034$); therefore, the TS result was interpreted primarily using the multivariate test reported in Table 5. CT performance was assessed using three rubric criteria (focus, logic, and content), with the criteria averaged to yield a composite score for each task at three time points (Sessions 2, 9, and 16). Descriptive statistics are presented in Table 4. The results show consistent improvement in OE and EV across the year, whereas TS remained relatively stable.

**Table 4.**

*Descriptive Statistics for CT Performance across Tasks and Sessions*

| Task | Session 2 | Session 9 | Session 16 |
|---|---|---|---|
| | ***M* (*SD*)** | ***M* (*SD*)** | ***M* (*SD*)** |
| **OE** | 3.04 (0.47) | 3.35 (0.50) | 3.54 (0.47) |
| **TS** | 2.90 (0.72) | 3.10 (0.43) | 3.15 (0.46) |
| **EV** | 2.68 (0.37) | 3.08 (0.56) | 3.39 (0.47) |

Table 5 summarizes the repeated-measures ANOVA results. A significant effect was observed for OE (Wilks' $\Lambda$ = .578, $F(2, 22) = 8.03$, $p = .002$), with a significant linear trend ($F(1, 23) = 16.78$, $p < .001$). These findings indicate that students' ability to articulate opinions through structured reasoning improved progressively over the instructional period, with gains reflected in clearer expression and organization of reasoning in their written responses.

**Table 5.**

*Summary of Repeated Measures ANOVA Results for CT Performance across Tasks*

| Task | Wilks' $\Lambda$ | $F(df)$ | $p$ (Overall) | Linear Trend $F$ | $p$ (Linear) | Partial $\eta^2$ |
|---|---|---|---|---|---|---|
| **OE** | .578 | 8.03 (2, 22) | .002 | 16.78 | < .001 | .422 |
| **TS** | .809 | 1.36 (2, 22) | .277 | 2.81 | .107 | .110 |
| **EV** | .275 | 29.04 (2, 22) | < .001 | 55.86 | < .001 | .725 |

No significant change was observed for TS (Wilks' $\Lambda$ = .809, $F(2, 22) = 1.36$, $p = .277$). The linear trend was also nonsignificant ($F(1, 23) = 2.81$, $p = .107$), indicating that TS did not show a clear progressive increase across the three time points. These findings suggest that TS tasks were less effective at eliciting observable CT performance, as operationalized in this study, possibly because they focused on information retrieval rather than deeper analysis. In contrast, EV showed the strongest and most consistent improvement (Wilks' $\Lambda$ = .275, $F(2, 22) = 29.04$, $p < .001$), with a significant linear trend ($F(1, 23) = 55.86$, $p < .001$). This suggests that reflection tasks supported deeper engagement, enabling more explicit expression of analysis and synthesis over time.

To assess the magnitude of change, partial eta squared ($\eta^2$) was calculated for each task. Following conventional benchmarks (Cohen, 1988), values of .01, .06, and .14 indicate small, medium, and large effects, respectively. As shown in Table 5, EV demonstrated a very large effect ($\eta^2$ = .725), while OE also showed a large effect ($\eta^2$ = .422), supporting observed improvements. TS yielded a medium but nonsignificant effect ($\eta^2$ = .110), consistent with the absence of statistically significant gains. Given the relatively small sample size, these effect sizes should be interpreted with caution,

although the overall pattern suggests that EV tasks played a substantial role in supporting the development of students' articulated reasoning in writing, as reflected in gains across focus, logic, and content.

**RQ2: Qualitative Analysis of Students' Learning Log Entries**

To address RQ2, qualitative patterns of growth and challenge were examined in a purposive subsample of six students (two from each of the high-, moderate-, and low-gain groups) described in the Methods. Thematic analysis focused on opinion expression (OE) and evaluation of viewpoints (EV), which showed the clearest changes in the quality of students' articulated reasoning as an indicator of CT performance, as shown in RQ1. Four themes emerged: (1) a shift from fragmented to structured thought, (2) growth in evaluative thinking, (3) progression in the quality and depth of examples, and (4) ongoing struggles in distinguishing reasons from examples. Overall, repeated engagement with CT-infused, writing-to-think tasks appeared to support the development and clearer articulation of reasoning.

***From Fragmented to Structured Thought***

A key developmental pattern among high-gain learners was a progression from fragmented ideas to more logically sequenced arguments. Early OE responses relied on anecdotal detail without clear logical scaffolding. In Session 2, when asked to explain why she liked her hometown (see Appendix A for prompt details), Student 2 described Saitama, a prefecture adjacent to Tokyo, as "convenient" and "fun," but her support consisted mainly of personal experiences such as shopping and ease of travel. Although the response included two stated reasons, these remained loosely connected and based on sentiment rather than reasoning. By Session 9, her response to a prompt about playing a role as a world-class actor showed greater topical relevance and elaboration. She chose Cure Blossom (a heroine from the *Precure* anime series) and described the character as "shy but hardworking and ambitious." The entry included a personal connection— "When I was a child, I was moved by her"—and further contextualized this admiration through the anime's flower theme. While richer in detail, the reasoning was delivered in short, loosely connected segments, with an abrupt statement ("Like that her character is attractive") disrupting the flow and a brief digression toward her personal interest in flowers. By Session 16, however, her writing on the "secret to happiness" demonstrated a clearer argumentative structure. She began with a central claim— "the secret to being happy is the presence of friends"—followed by a cause-and-effect rationale: friends allow people to share good experiences, multiplying happiness, and to endure sadness through encouragement. The examples—"if you encounter a happy moment, you can divide it with your friends" and "if something sad, such as a broken heart, happens, your friends will encourage you"—were directly tied to her claim, reinforcing coherence and logical flow. Unlike earlier sessions, where ideas were listed without clear links, the Session 16 response illustrated a sustained assertion supported by sequential reasoning, marking a shift toward

more structured expression of critical thought. Rather than suggesting that improved organization alone constitutes CT development, this progression indicates that increased control over discourse structure enabled the student to articulate reasoning more coherently, making her CT performance more visible in writing (see Appendix C for full comparison).

***Growth in Evaluative Thinking***

Among the moderate-gain group, a key development was a shift from surface-level reflection to deeper evaluative analysis of class discussions. Early EV responses often repeated initial OE ideas rather than engaging with alternative perspectives. In Session 2, for example, Student 8 responded to a task on city versus small-town life by restating her preference: "I prefer living in a big city. That's because there [are] a lot of stores. It is convenient. For example, my town has only [one] shopping mall." This expressed a stance but offered little contextual reasoning. By Session 9, her responses became more evaluative. After reading and discussing an article about Emma Watson's choice to portray Belle in *Beauty and the Beast* rather than Cinderella, students were asked to evaluate Watson's reasoning and identify which character they resonated with. Student 8 highlighted Belle's independence and contrasted her with Cinderella, often depicted as passive and dependent on rescue:

> Belle values inner beauty, not outer appearance, and appreciates people for what is in their hearts. The process by which she comes to recognize and love the beauty of the beast's heart is inspiring to many and teaches the importance of relationships that are not limited by outward appearances…So, I resonate Belle more than Cinderella (Student 8, EV, Session 9).

Here she moved beyond personal preference to critique social values, signaling broader engagement with class themes. By Session 16, her evaluative thinking had deepened further. In response to a discussion about Jyoti—a man who left a stable career and fiancée in India to become a yoga teacher—she considered both the merits and risks of his decision:

> I think Jyoti's decision and action can't understand. His initiative is good but I don't know if it is something he should do even if it means abandoning his life and his partner. I admire his courage to leave his stable life and girlfriend to pursue his dream, but I can't relate to that because he is losing money (Student 8, EV, Session 16).

Here she weighed the ethical dimension of courage and self-determination against the practical consequences of financial instability and relationship loss. Unlike her earlier responses, which tended to affirm or reject a viewpoint outright, this reflection demonstrates a more balanced evaluation,

acknowledging both the value and the cost of Jyoti's decision. This progression suggests a developing ability to consider competing perspectives.

***Progression in the Quality and Depth of Examples***

Among the moderate-gain group, another notable improvement, particularly in the OE tasks, was the development of more meaningful and well-supported examples. Early on, students relied heavily on personal anecdotes with little abstraction. Over time, however, their examples became more detailed and evidence-based, sometimes drawing on external sources. In Session 2, Student 5 explained her preference for living in Adachi Ward, a district of Tokyo, by citing her participation in a local children's community as evidence of strong neighborhood ties. While relevant, the example was narrowly personal and lacked broader contextual support. By Session 9, when asked which role she would like to play as a famous actor, she chose Violet from Pixar's *The Incredibles*, admiring her invisibility and force-field powers. Although enthusiastic, her reasoning remained descriptive rather than analytical, echoing the reliance on preference seen in Student 2's earlier response about Cure Blossom. By Session 16, however, Student 5 provided a more sophisticated justification. In discussing the "secret to happiness," she referred to a television program explaining the positive effects of laughter and reinforced her point with the Japanese proverb *"Fortune comes in by a merry gate."* This marked a clear shift from purely subjective anecdotes to culturally grounded, evidence-based reasoning, illustrating how the writing-to-think, CT-infused process enabled students to move from preference-driven responses toward more substantiated arguments.

***Ongoing Struggle: Confusion Between Reasons and Examples***

Despite overall improvement, challenges persisted—particularly among low-gain students, who often struggled to distinguish reasons from examples, limiting the coherence of their arguments. A representative case was Student 17, whose entries across Sessions 2, 9, and 16 followed a formulaic pattern— "I agree/disagree… I have two reasons. First… Second…"—commonly taught in preparation for the EIKEN Test in Practical English Proficiency, Japan's nationwide English certification exam (Eiken Foundation of Japan, 2024). While this structure provided surface organization, it tended to encourage shallow reasoning and blur the distinction between rationale and illustration. In Session 9 for the OE task, for instance, she chose Nobita, the main character from *Doraemon*, writing: "There are two reasons why I want to play him. First of all, I want to try to battle with Gian. A thing that is important for me is to be a challenger. In the anime, I would like to struggle to become stronger, using items Doraemon lent to me." Here, fighting Gian—a strong bully character—serves as the example, while the underlying reason ("to be a challenger") follows after, making the logic appear reversed and weakening coherence. A similar issue appeared in Session 16. While she argued that "having money is the secret to happiness," she did not explain *why*; instead, she listed what money enables—travel, cram school, entertainment—treating examples as if they were reasons. This conflation limited the depth of

her reasoning and left the central claim underdeveloped. These patterns suggest that without explicit instruction in distinguishing justification (reasons) from illustration (examples), some students remain trapped in formulaic yet underdeveloped arguments, unable to progress from structured expression to critical reasoning.

## Discussion

This study examined how learning logs, designed as a writing-to-think, CT-infused pedagogy, supported Japanese university students' CT development in a discussion-oriented course. CT is conceptualized here as a discourse-based performance, operationalized through students' ability to articulate and structure reasoning in written language. Findings suggest that development in CT performance may have been supported not by writing or CT instruction alone, but by their integration within the structured learning-log format, reflecting an increased ability to articulate reasoning. By requiring short, structured entries before and after discussion, logs created iterative opportunities for students to externalize, test, and refine reasoning. This process allowed reasoning to be tracked and refined across the academic year, providing a process-sensitive account of CT development. From a writing-to-think perspective, this pattern can be interpreted as the gradual externalization and stabilization of reasoning through language (Emig, 1977; Menary, 2007), suggesting that writing supports learners in monitoring and refining their reasoning over time. This process is further shaped by the CT-infused design of tasks, in which explicit reasoning demands embedded in prompts guide learners toward the articulation of analysis, justification, and evaluation.

Quantitatively, significant gains were observed in opinion expression (OE) and evaluation of viewpoints (EV), whereas text summarization (TS) remained stable, underscoring the role of task design. OE and EV prompts explicitly embedded reasoning demands: OE required stance and justification, while EV involved evaluating others' perspectives. These aligned closely with the CT Expression Framework (CTEF), guiding students to articulate a main idea, provide reasons, support them with examples, and restate the claim. Writing appeared to slow down reasoning, making it more observable and allowing opportunities for refinement before and after discussion, which was associated with measurable gains in CT performance. These gains do not necessarily indicate improvement in underlying CT ability, but rather increased control over the expression of reasoning. Structured writing provided the time and support needed to organize ideas, making reasoning more visible—a pattern consistent with L2 writing research linking textual organization to clearer articulation of reasoning (Manchón, 2011; Staples & Reppen, 2016). In contrast, TS tasks primarily involved recall and condensation and did not operationalize reasoning processes such as justification or evaluation. As a result, students tended to paraphrase rather than demonstrate structured reasoning. This misalignment likely contributed to the absence of gains in CT performance, as the summarization functioned primarily as comprehension rather than the articulation of reasoning. However, this does not suggest that

summarization tasks are ineffective for CT development more broadly, but rather that their effectiveness depends on how reasoning demands are embedded in task design. These findings align with research showing that unguided writing supports fluency but not reasoning (Hirose & Sasaki, 2000; Condon & Kelly-Riley, 2004) and that infusion-oriented writing tasks most effectively foster CT (Abrami et al., 2015; Yin et al., 2023).

Qualitative findings revealed distinct developmental trajectories across gain groups. High-gain learners progressed from fragmented ideas to coherent, paragraph-level reasoning. Early entries were loosely connected, but the CTEF scaffold supported the integration of main ideas, reasons, and examples into cohesive arguments. The log cycle reinforced this process: writing prepared ideas, discussion tested them, and reflection consolidated them. Increased control over cohesion and structure supported clearer articulation of reasoning, suggesting that discourse-level organization enabled reasoning to become more visible in writing. This may indicate that discourse organization functions as an early marker of CT performance, as improvements in structure appear to precede or enable clearer expression of reasoning (Manchón, 2011; Staples & Reppen, 2016). This aligns with models of writing as a process that externalizes and stabilizes thought (Emig, 1977; Flower & Hayes, 1981). In line with L2 writing research that conceptualizes writing as a mediational tool for cognitive development (Cumming, 2001; Manchón, 2011; Swain, 2006), learning logs can be understood as a structured space in which reasoning is externalized, reorganized, and progressively refined.

For moderate-gain learners, evaluative reasoning and elaboration of examples emerged as key areas of growth. Initially, many entries repeated personal preferences, but over time students began to weigh alternative perspectives and provide more substantiated examples. This progression suggests a reciprocal relationship between language and reasoning: as learners gained expressive resources, they were better able to elaborate examples, while evaluative task demands supported clearer articulation of reasoning. This finding aligns with research showing that infusion-oriented writing tasks foster evaluative thinking when embedded within meaningful content (Chason et al., 2017; Nejmaoui, 2019). It also supports the claim that writing-to-think pedagogy allows learners to rehearse reasoning in manageable written segments, making evaluation and elaboration more accessible in L2 contexts (Mineshima & Imai, 2017).

For low-gain learners, however, the log cycle alone was insufficient. Many continued to conflate reasons with examples, treating illustrations as justifications. Within the CTEF, this distinction is cognitively demanding, as reasons reflect interpretation while examples illustrate it. Without mastering this distinction, arguments remained underdeveloped. Similar challenges have been reported in prior studies, suggesting that this difficulty is common among L2 learners (Lu & Xie, 2019). This pattern also resonates with instructional contexts where argumentation is less emphasized (Mulvey, 2016). Together, these findings reinforce that writing alone does not automatically lead to observable CT development (Condon & Kelly-Riley, 2004); rather, explicit instructional mediation is needed to support learners in articulating reasoning processes.

Taken together, the gain profiles suggest differential uptake of the CTEF: some learners demonstrated substantial development in organizing and articulating reasoning (high gain), others showed more selective advances (moderate gain), while a subset showed limited change, indicating the need for more explicit instructional support.

## Pedagogical Implications and Limitations

Based on these findings, several pedagogical implications can be drawn for the design and implementation of CT-oriented learning logs. First, tasks should incorporate explicit reasoning demands. Gains were strongest when tasks required stance-taking and evaluation (OE, EV), whereas recall-oriented tasks such as summarization were associated with more limited evidence of growth. Teachers should therefore design prompts that move beyond comprehension to include synthesis or evaluation.

Second, learners engaged with log-based scaffolding differently depending on their entry point into the CTEF. High-gain learners began with limited awareness of how to structure reasoning, making subsequent gains in cohesion and argument structure particularly visible. Moderate-gain learners demonstrated some initial organizational control and improved more selectively, especially in evaluative reasoning and example elaboration. Low-gain learners, while sometimes able to produce organized responses, struggled with distinguishing reasons from examples, leaving arguments underdeveloped. These patterns suggest that gains may reflect how learners adopted and extended the framework, highlighting the need for differentiated instructional support—particularly explicit modeling of the reasoning–illustration distinction.

Finally, learning logs may function as a bridge between writing and speaking. This study provides evidence of steady growth in CT performance in writing, reflected in improved reasoning and increased linguistic resources. Although spoken performance was not directly assessed, the clarification of ideas and rehearsal of language through writing-to-think tasks may support students' confidence in discussion participation. This interpretation is consistent with prior findings that log writing enhances confidence in oral interaction (Uemi, 2025) and aligns with research emphasizing the preparatory role of writing-to-think in L2 contexts (Mineshima & Imai, 2017). Accordingly, this should be understood as a pedagogical implication rather than a direct empirical finding, indicating the potential of learning logs as a scaffold for learners facing sociocultural and linguistic barriers to spontaneous oral CT.

Several limitations should be acknowledged. First, the study analyzed a relatively small subset of students ($n = 24$) with complete paired data, which may limit representativeness and generalizability. While this allowed for detailed, process-oriented analysis, caution is warranted in extending findings to broader populations. Although the analytic sample was broadly comparable to the full cohort in proficiency range, selection bias may be present due to the data completeness criterion. Accordingly, statistical findings—including effect sizes—should be interpreted as indicative rather than definitive.

Future research should extend analysis to the full cohort and examine links between CT growth and language proficiency. Second, the TS task did not yield measurable CT gains, raising questions about task design. Redesigning summarization to incorporate synthesis or critique may provide clearer insights. Third, the absence of a control group limits attribution of observed changes solely to the intervention; improvements may also reflect language development, task familiarity, or maturation over time. Fourth, although logs may have supported oral participation, spoken CT performance was not directly assessed. Future research should triangulate written, oral, and perceptual data to examine transfer effects. Finally, as a classroom-based study, the findings reflect a specific institutional context and may not generalize across L2 settings. Taken together, these limitations underscore the context-specific nature of the findings while supporting their contribution to understanding how CT performance can be made visible and supported in L2 instruction.

## Contribution and Future Directions

This study contributes to L2 CT instruction by integrating CT infusion, writing-to-think, and discussion-oriented pedagogy within a unified instructional design. It advances a discourse-based perspective of CT by conceptualizing it as the articulation of reasoning in written language within L2 classroom interaction. In doing so, it clarifies how CT can be operationalized and observed through learners' ability to articulate and structure reasoning in classroom contexts.

Methodologically, the study demonstrates how structured learning logs, combined with rubric-based evaluation, can trace the development of students' articulated reasoning over time. This process-oriented approach extends beyond traditional pre–post designs by capturing incremental changes across recurring tasks, providing a more fine-grained account of how learners develop the ability to express and refine their thinking.

Pedagogically, the study highlights the role of writing-to-think as a mediational tool in L2 discussion classes. Rather than functioning as an isolated writing activity, learning logs support learners in preparing, clarifying, and refining ideas before and after interaction, facilitating clearer articulation of reasoning. The findings further emphasize the importance of aligning task design with explicit reasoning demands, as gains were most evident in tasks requiring justification, evaluation, and synthesis. In this respect, the CTEF framework functions not as a conventional paragraph-writing model, but as a transferable scaffold for constructing and articulating reasoning across tasks and contexts.

Building on these contributions, future research may examine links to spoken CT performance, refine task designs to better elicit analytical and evaluative reasoning, and explore the applicability of the framework across proficiency levels and instructional contexts.

## Conclusion

This study demonstrates how learning logs operationalize the integration of writing-to-think and CT infusion in L2 discussion courses. Quantitative and qualitative findings indicate that reasoning-focused log tasks are associated with improvements in the articulation of CT performance, particularly in opinion expression and evaluation, while revealing persistent challenges such as distinguishing reasons from examples. The findings suggest that writing-to-think and infusion function most effectively in combination: writing makes reasoning visible, while infusion directs it toward explicit critical moves. Embedding such tasks within oral communication courses allows learning logs to function both as pedagogical scaffolds and as research tools for tracing students' reasoning over time. In this sense, learning logs represent not merely reflective activities but a practical form of CT pedagogy that supports the articulation of reasoning, making it more visible, traceable, and improvable in L2 education.

## References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85*(2), 275–314. https://doi.org/10.3102/0034654314551063

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Abridged ed.). Longman.

Applebee, A. N., & Langer, J. A. (1987). *How writing shapes thinking: A study of teaching and learning*. National Council of Teachers of English.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates.

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

Chason, L., Loyet, D., Sorenson, L., & Stoops, A. (2017). An approach for embedding critical thinking in second language paragraph writing. *TESOL Journal, 8*(3), 582–612. https://doi.org/10.1002/tesj.288

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Condon, W., & Kelly-Riley, D. (2004). Assessing and teaching what we value: The relationship between college-level writing and critical thinking abilities. *Assessing Writing, 9*(1), 56–75. https://doi.org/10.1016/j.asw.2004.01.003

Cottrell, S. (2017). *Critical thinking skills: Effective analysis, argument and reflection* (3rd ed.). Red Globe Press.

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. SAGE Publications.

Cumming, A. (2001). Learning to write in a second language: Two decades of research. *International Journal of English Studies, 1*(2), 1–23. https://doi.org/10.6018/ijes.1.2.48361

Dam, L. (2009). The use of logbooks—A tool for developing learner autonomy. In R. Pemberton, S. Toogood, & A. Barfield (Eds.), *Maintaining control: Autonomy and language learning* (pp. 125–144). Hong Kong University Press. https://doi.org/10.5790/hongkong/9789622099234.003.0008

Day, R. R., Shaules, J., & Yamanaka, J. (2019). *Impact issues 2: Student book* (3rd ed.). Pearson.

Dummett, P., & Hughes, J. (2019). *Critical thinking in ELT: A working model for the classroom*. National Geographic Learning.

Egitim, S. (2022). Do Japanese students lack critical thinking? Addressing the misconception. *Power and Education, 14*(3), 304–309. https://doi.org/10.1177/17577438221107203

Eiken Foundation of Japan. (2024). *EIKEN test in practical English proficiency*. https://www.eiken.or.jp/eiken/en/

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication, 28*(2), 122–128. https://doi.org/10.2307/356095

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18*(3), 4–10. https://doi.org/10.3102/0013189X018003004

Ennis, R. H. (2011). Critical thinking: Reflection and perspective part I. *Inquiry: Critical Thinking Across the Disciplines, 26*(1), 4–18. https://doi.org/10.5840/inquiryctnews20112613

Facione, N. C., & Facione, P. A. (1994). *Holistic critical thinking scoring rubric*. California Academic Press.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. https://eric.ed.gov/?id=ED315423

Facione, P. A., Sánchez, C. A., Facione, N. C., & Gainen, J. (1995). The disposition toward critical thinking. *The Journal of General Education, 44*(1), 1–25.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.2307/356600

Frost, K. (2017, January 17). Emma Watson on why she turned down Cinderella before accepting her *Beauty and the Beast* role. *ELLE UK*. https://www.elle.com/uk/life-and-culture/culture/news/a33458/emma-watson-turned-down-cinderella-before-belle/

Galbraith, D. (1999). Writing as a knowledge-constituting process. In G. Rijlaarsdam, E. Espéret, M. Torrance, & D. Galbraith (Eds.), *Studies in writing: Vol. 4. Knowing what to write: Conceptual processes in text production* (pp. 139–160). Amsterdam University Press.

Hammad, E. A. R. (2022). The effect of using personal experience journals on writing fluency and writing anxiety of Al-Aqsa University EFL students. *World Journal of English Language, 12*(1), 301–314. https://doi.org/10.5430/wjel.v12n1p301

Hirose, K., & Sasaki, M. (2000). Effects of teaching metaknowledge and journal writing on Japanese university students' EFL writing. *JALT Journal, 22*(1), 94–113. https://doi.org/10.37546/JALTJJ22.1-5

Hyland, K. (2003). *Second language writing*. Cambridge University Press.

Kaneda, S. (2022). Fostering learner autonomy of EFL students at a Japanese high school: Logbook research. *Language Education & Technology, 59*, 1–30. https://doi.org/10.24539/let.59.0_1

Khonamri, F., Azizi, M., & Farzanegan, M. (2016). Literature-based extensive reading accompanied by reading logs: A case for developing critical thinking skills of English literature students. *International Journal of Education, 9*(1), 58–67. https://doi.org/10.17509/ije.v9i1.3719

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Kubota, R., & Lehner, A. (2004). Toward critical contrastive rhetoric. *Journal of Second Language Writing, 13*(1), 7–27. https://doi.org/10.1016/j.jslw.2004.04.003

Lin, M., Preston, A., Kharrufa, A., & Kong, Z. (2016). Making L2 learners' reasoning skills visible: The potential of computer-supported collaborative learning environments. *Thinking Skills and Creativity, 22*, 303–322. https://doi.org/10.1016/j.tsc.2016.06.007

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). *Assessing critical thinking in higher education: Current state and directions for next-generation assessments* (ETS Research Report No. RR-14-10). ETS Research Report Series. https://doi.org/10.1002/ets2.12009

Lu, D., & Xie, Y. (2019). The effects of a critical thinking-oriented instructional pattern in a tertiary EFL argumentative writing course. *Higher Education Research & Development, 38*(5), 969–984. https://doi.org/10.1080/07294360.2019.1607830

Manalo, E., & Sheppard, C. (2016). How might language affect critical thinking performance? *Thinking Skills and Creativity, 21*, 41–49. https://doi.org/10.1016/j.tsc.2016.05.005

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.

McCrindle, A. R., & Christensen, C. A. (1995). The impact of learning journals on metacognitive and cognitive processes and learning performance. *Learning and Instruction, 5*(2), 167–185. https://doi.org/10.1016/0959-4752(95)00010-Z

McLaughlin, F., & Moore, M. (2012). Integrating critical thinking into assessment of college writing. *Teaching English in the Two-Year College, 40*(2), 145–162.

Mehta, S. R., & Al-Mahrooqi, R. (2015). Can thinking be taught? Linking critical thinking and writing in an EFL context. *RELC Journal*, *46*(1), 23–36. https://doi.org/10.1177/0033688214555356

Menary, R. (2007). Writing as thinking. *Language Sciences, 29*(5), 621–632. https://doi.org/10.1016/j.langsci.2007.01.005

Menegale, M. (2020). Using logbooks with second and foreign language learners in higher education: Learner autonomy in progress. *Philologia Hispalensis, 34*(1), 99–119. https://doi.org/10.12795/PH.2020.v34.i01.06

Mineshima, M., & Imai, R. (2017). Developing critical thinking skills in high school and university: How to make English lessons more intelligent. *Journal of the Chubu English Language Education Society, 46*, 133–140. https://doi.org/10.20713/celes.46.0_133

Mulvey, B. (2016). Writing instruction: What is being taught in Japanese high schools, why, and why it matters. *The Language Teacher, 40*(3), 3–8. https://doi.org/10.37546/JALTTLT40.3-1

Namsaeng, P., & Ambele, E. (2024). Integrating intercultural competence into the curriculum: A catalyst for critical thinking in Thai tertiary education. *THAITESOL Journal, 37*(2), 74–95. https://so05.tci-thaijo.org/index.php/thaitesoljournal/article/view/268999

Nejmaoui, N. (2019). Improving EFL learners' critical thinking skills in argumentative writing. *English Language Teaching, 12*(1), 98–109. https://doi.org/10.5539/elt.v12n1p98

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Nosich, G. (2022). *Critical writing: A guide to writing a paper using the concepts and processes of critical thinking*. Rowman & Littlefield.

Okada, R. (2017). Conflict between critical thinking and cultural values: Difficulty asking questions and expressing opinions in Japan. *Asian Education Studies, 2*(1), 91–98. https://doi.org/10.20849/aes.v2i1.122

Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). SAGE Publications.

Paul, R., & Elder, L. (2006). *The international critical thinking reading & writing test*. Foundation for Critical Thinking.

Paul, R., & Elder, L. (2019). *How to write a paragraph: The art of substantive writing*. Foundation for Critical Thinking Press.

Seshachari, N. C. (1994). Instructor-mediated journals: Raising critical thinking and discourse levels. *College Teaching, 42*(1), 7–11. https://doi.org/10.1080/87567555.1994.9926806

Staples, S., & Reppen, R. (2016). *Using corpus-based methods to investigate discourse: A practical guide*. Routledge.

Stapleton, P. (2001). Assessing critical thinking in the writing of Japanese university students: Insights about assumptions and content familiarity. *Written Communication, 18*(4), 506–548. https://doi.org/10.1177/0741088301018004004

Swain, M. (2006). Languaging, agency, and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). Continuum.

Tan, C. (2017). Teaching critical thinking: Cultural challenges and strategies in Singapore. *British Educational Research Journal, 43*(5), 988–1002. https://doi.org/10.1002/berj.3295

Uemi, M. (2024). The impact of critical thinking-focused learning logs: Enhancing Japanese students' CT performance in written communication. *The Journal of Asia TEFL, 21*(4), 797–817. https://doi.org/10.18823/asiatefl.2024.21.4.3.797

Uemi, M. (2025). Enhancing critical thinking performance in Japanese L2 learners through CT-infused learning logs. *Language Education & Technology, 62*, 207–236. https://doi.org/10.24539/let.62.0_207

Wade, C. (1995). Using writing to develop and assess critical thinking. *Teaching of Psychology, 22*(1), 24–28. https://doi.org/10.1207/s15328023top2201_8

Wang, Y., & Wu, Z. (2023). Adapting or adopting? Critical thinking education in the East Asian cultural sphere: A systematic integrative review. *Thinking Skills and Creativity, 49*, Article 101330. https://doi.org/10.1016/j.tsc.2023.101330

Watanabe, M. (2023). *Ronritekishikou no bunakatekikiban: 4 tsuno shikouhyougen style [The cultural foundations of "logical thinking": Four styles of thinking expression]*. Iwanami Shoten.

Yang, Y. T. C., Gamble, J. H., Hung, Y.-W., & Lin, T.-Y. (2014). An online adaptive learning environment for critical-thinking-infused English literacy instruction. *British Journal of Educational Technology, 45*(4), 723–747. https://doi.org/10.1111/bjet.12080

Yin, X., Saad, M. R. B. M., & Halim, H. B. A. (2023). A systematic review of critical thinking instructional pedagogies in EFL writing: What do we know from a decade of research. *Thinking Skills and Creativity, 49*, Article 101363. https://doi.org/10.1016/j.tsc.2023.101363

**Megumi Uemi** is a lecturer at the Center for International Education and Exchange, Toyo University. This research was supported by the Inoue Enryo Memorial Research Grant, Toyo University, for the project titled "Critical Thinking-Infused English Course Design at the University Level." Correspondence concerning this article should be addressed to Megumi Uemi, Center for International Education and Exchange, Toyo University, Building 8, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan.

**ORCID:** https://orcid.org/0000-0003-3568-7693

# Appendix

## Appendix A
## Learning Log Task Prompts Across Sessions

| Session (Timing) | Lesson Content Summary | OE | TS | EV |
|---|---|---|---|---|
| **2 (Beginning)** | Dialogue between Jing and Mitsuko comparing city and small-town life. | 1. Where did you grow up—city or small town?<br>2. Did you like it? Why or why not? | Describe today's discussion in your own words as if explaining it to your partner. | What's better—living in a city or a small town? Why? |
| **9 (Middle)** | Emma Watson's interview on choosing *Belle* over *Cinderella* to express empowerment. | If you were a famous actor like Emma Watson, what character would you want to play (e.g., Disney princess, superhero, anime character)? Why? Give reasons and examples. | 1. Why did Emma turn down *Cinderella*?<br>2. Why did she accept *Belle*?<br>3. What backstory did she create, and why? | Choose one:<br>• What do you think of Emma's interpretation of *Belle*?<br>• How might she describe *Cinderella*?<br>• Which do you sympathize with more, *Cinderella* or *Belle*? |
| **16 (End)** | Story of Jyoti, who leaves a successful career and fiancée to pursue a meaningful life as a yoga teacher. | What is the secret to being happy? | 1. What was Jyoti's life like before his change?<br>2. Why did he give it up?<br>3. What did he do afterward?<br>4. What is he doing now, and how does he feel? | What do you think of Jyoti's decision? Would you make the same choice? Or, if you were Callie, how would you feel about it? Why? |

*Note*. OE = Opinion Expression; TS = Text Summarization; EV = Evaluation of Viewpoints. Each task represents one component of the CT-infused learning-log cycle implemented at the beginning (Session

2), middle (Session 9), and end (Session 16) of the course. Lesson sources: Session 2 (Unit 2) and Session 16 (Unit 16) were adapted from *Impact Issues 2* (Day et al., 2019). The Session 9 material was adapted from an ELLE UK article on Emma Watson's decision to turn down the role of Cinderella before accepting the role of Belle (Frost, 2017).

**Appendix B**

**Critical Thinking Performance Rubric for Learning Log Tasks**

| Score | Focus (Main Idea & Relevance) | Logic (Reasoning & Coherence) | Content (Elaboration & Examples) |
|---|---|---|---|
| **4 (Excellent)** | Clearly and insightfully addresses the prompt; main idea is well-developed and strongly articulated. | Reasoning is logically sound, well-supported, and acknowledges alternative viewpoints when applicable. | Ideas are fully developed with relevant, detailed, and elaborated examples. |
| **3 (Proficient)** | Addresses the prompt with clarity; main idea is evident but could be more refined. | Reasoning is mostly coherent and supported, though alternative viewpoints may not be fully explored. | Ideas are developed with relevant examples, but elaboration may be somewhat limited. |
| **2 (Developing)** | Partially addresses the prompt but lacks clarity or organization; main idea is underdeveloped. | Reasoning is somewhat unclear or inconsistent, with limited support. Alternative viewpoints are not adequately considered. | Ideas are weakly developed with vague, repetitive, or minimal examples. |
| **1 (Emerging)** | Response is off-topic or lacks a clear main idea; poor organization. | Reasoning is weak, unclear, or lacks logical connections; no engagement with alternative viewpoints. | Ideas are underdeveloped or lack supporting examples; explanations are missing or unclear. |

## Appendix C

**Student 2's Responses for the Opinion Expression Task Across Sessions 2, 9, and 16**

| Session | Student Response |
|---|---|
| **2** | I grew up in Tokorozawa of Saitama. It is a big city. Saitama's population is about 7 million. I like Saitama..I have two reasons. First, it is convenient. Saitama is near Tokyo. For example, when I want to go to Tokyo, I can go there easily. Also, as Saitama is a city that is near Tokyo, there is many people living in Saitama. I think they have the same idea as me. Second, it is fun to play with my friend. Saitama is flourishing. For instance, there are so many shops that I can enjoy shopping. For those reasons, I like Saitama. |
| **9** | If I were a famous actor, I would want to play "Kyua Blossom" in Heart Catch Purikyua. I like her character of shy but hardworking and ambitious. Like that her character is attractive. When I was a child, I was moved by her. Also, the animation's theme is flowers and she likes flowers, that was same with me. In the past, I feel attraction to the language of flowers and I wanted to be a flower shop clerk. For me like that, Purikyua that the theme is flowers was very exciting. From this episode, I want to play as Kyua Blossom of my memorable animation. |
| **16** | In my opinion, the secret to being happy is the presence of friends. Because there is happiness to share something good, exciting, and sad. For example, if you encounter a happy moment, you can divide it with your friends, so there will be two people who are happy. There is nothing better than making people smile. For instance, if something sad such as a broken heart, your friends will encourage you in various way, you will notice the importance the presence of friends other than family and an ex-partner. So, I think everyone should thank the presence of friends. |

*Note.* All student responses are presented exactly as originally written, without any corrections or modifications to grammar, vocabulary, or structure, in order to preserve the authenticity of student output and accurately reflect their developmental progress.