

AI as Halfway-Exemplars: Fostering Logical Reasoning Skills in EMI English Classes

Miyuki Oikawa (Sawada)

National Kaohsiung Normal University

Abstract

This paper examines three aspects of the use of AI in EMI English education. First, AI can serve as a tool for cultivating logical reasoning. Since logic is largely context- and culture-independent, AI is effective at learning rule-based reasoning patterns and can evaluate whether sentences produced by students are logically structured. However, argumentation requires careful consideration of context in order to persuade readers with appropriate evidence, an area in which AI remains relatively weak.

Second, this study leverages this characteristic of AI to generate writing samples that students are asked to improve. In this approach, AI functions as a “halfway exemplar,” providing partially developed arguments that students must analyze, critique, and strengthen. In this way, students foster logically coherent argumentation skills.

The study compares grades from the pre-AI implementation period (September 2020–June 2022) with those from the post-AI implementation period (September 2023–June 2025); however, no significant difference is found between the two periods. That said, notable differences emerge in students’ argumentation skills and English proficiency, both of which declined markedly in the post-AI period. Nevertheless, student questionnaires indicate that AI-generated samples (all or some) helped improve their paragraph writing in both their daily work and final essays.

Finally, this paper analyzes students’ reflections on AI-generated samples in writing classes. Although no significant overall effect of AI use is observed in EMI classes, it contributes positively to the development of teaching materials and supports student learning outcomes.

1. Introduction

Currently, English education in Taiwan attempts to meet the requirements of being a bilingual country, that is, the 2030 Bilingual Project. National Development Council (NDC, 2021, cf. p. 4)¹ describes their two aims:

(1) Two aims in the 2030 Bilingual Nation Project

a. To enhance English communication skills among young people, and

1

<https://ws.ndc.gov.tw/Download.ashx?u=LzAwMS9hZG1pbmlzdHJhdG9yLzExL3JlbGZpbGUvMC8xNDUzNC9hODg1MTBkMC04YmQxLTQxZGEtYTgzZC1jOTg0NDM5Y2U3ZmMucGRm&n=QmlsaW5ndWFsIDlwMzAucGRm&icon=.pdf>, accessed 10 March 2026.

b.To bolster their global competitiveness.

Even so, (1a) is a “domestication” skill, (1b) is a “foreignization” skill (cf. translation theory). Domestication in L2 aims that a learner translates information from Chinese to English without cultural implication. In contrast, foreignization in L2 may aim to help learners convey their ideas in ways that align with the rhetorical and cultural conventions of another country. Here is a possible scenario: To sell products to British consumers, a Taiwanese should explain the benefits of the products and, if applicable, acknowledge their shortcomings while proposing possible solutions within the British context. In this process, it is important to see both sides of the item to win the global competitiveness. Essential skills for the task can be critical thinking skills. In fact, the 2030 Bilingual Project emphasizes:

(2) “Promoting English Debate Education and Cultivating Critical Thinking” (National Development Council, press release, September 18, 2019) ².

On the other hand, due to the focus on textbooks and exams in schools, teachers often cannot afford the time to incorporate debate or critical thinking activities. Therefore, it is essential for researchers to develop simple supplementary lesson plans to foster critical thinking skills in schools.

2. Critical thinking skills

Critical thinking skills are regarded as one of the essential skills, as 21st Century skills³. Ennis (1962), Facione (1990) and Hitchcock (2011) describe what kind of skills they are:

(3) Critical thinking skills (Ennis 1962; Facione 1990; Hitchcock 2011)

- a. Clarify meaning
- b. Analyze arguments
- c. Evaluate evidence
- d. Judge whether a conclusion follows
- e. Draw warranted conclusions
- f. Byproducts: logical analysis of arguments, elucidating its logical architecture, problem solving, creative component, evaluating alternatives, choosing the best of them.

Yet, the practice in education may vary across cultures and according to learners’ English proficiency. Some researchers point out that Asian students are weak in analytic and rule-based reasoning, i.e., critical thinking skills. Table 1 shows that the results of experiments in which East Asian students performed lower scores than European and American students.

² https://www.ndc.gov.tw/en/nc_8455_32705, accessed 28 Feb 2026.

³

<https://unevoc.unesco.org/home/TVETipedia+Glossary/lang=e/show=term/term=21st+century+skills>

Table 1: Ethnic oriented experiments in analytic and rule-based reasoning (logical thinking skills)

Test	Participants	The result	Discussion
Halpen Critical Thinking Assessment Lun (2010)	European students vs. Chinese students	European > Chinese	Culture, language differences, lack of training.
Formal vs. Intuitive reasoning tests (Norenzayan et al. 2002)	Asian American, European American vs. Chinese & Korean	Formal (European & American) vs. Intuitive reasoning (Chinese & Korean)	Philosophies (Early Greek vs. Chinese philosophy).
Analytic vs. Holistic cognition, foreground vs. background information (Masuda & Nisbett 2001)	American students vs. Japanese students	When the subjects watched animation, American first described the most salient object vs. Japanese described the background.	Cultural differences in socialization patterns.
Debating in L1 & L2 (Rear 2017)	Japanese	L1 (Japanese) > L2 (English)	Japanese students do have a capacity for critical thinking in their own language.

As shown in Table 1, European and American students are more likely to use rule-based reasoning whereas East Asian students routinely use intuitive reasoning system.

At the same time, as in Table 1, Rear (2017) points out that Japanese students can debate in L1, but not in L2.

The weakness of analytic and rule-based reasoning may be due to the instructional language (i.e., in English) in the experiments: East Asian students may obtain logical thinking ability in L1. In fact, Asian students are good at Math and science subjects which are counted as logical thinking skills in Inhelder and Piaget (1958).

Programme for International Student Assessment (PISA) assesses mathematical and scientific abilities among 15-year-old students. According to PISA results, since 2009, 15-year-old students in Taiwan, Japan, and Korea have outperformed those in the United States in mathematics and science⁴.

Asian students obtain critical thinking skills in L1, but they are not able to transfer the abilities to L2: It is a duty for L2 instructors to establish a teaching method to foster the skills in English.

To investigate the root of the problem, Mineshima and Chino (2013) compare English textbooks for high school in Japan, Korea, and Finland to see how critical thinking skills can be fostered by English textbooks in these countries. Although the data are not up to date, they

⁴ See Appendix 1.

analyzed types of questions which could foster critical thinking skills, as in (4), and the results are shown in Table 2.

(4) Types of questions in high school English reading textbooks (Mineshima and Chino 2013)

Type A: Content based questions (answers are obvious in the main text).

Type B: Critical thinking-based questions:

1. The answers can be inferred from the description in the textbook.
2. The answers require the judgement with reasons (*Do you think...? why?/why not?*).
3. The answers require students' attitude (disposition): How they can use what they have learned in the textbook.

Table 2: Comparisons of high school English textbooks in Mineshima & Chino (2013), translated

	Japan	Korea	Finland
	18 textbooks for English reading course which are used in Tokyo in 2011 academic year (April 2011-May 2012)	4 textbooks for English reading course, published in 2011	8 textbooks for English course, published in 2002
Type A questions	86.6%	74.8%	55.2%
Type B questions	13.4%	25.2%	44.8%

Without a statistical significance, we can see that types of questions in English textbooks in Japan and Korea are more content-oriented. As a consequence, the classroom activity may mainly focus on the content, rather than the argument.

Meanwhile, in Taiwan, the weakness of argumentation skills in English classroom has been observed in debate activities at our college. Debate can be an activity to foster argumentation skills (see Mjidai, Janssen, and Graaff 2021, etc.). I have introduced debate activities in my freshmen/sophomore English classes (around 20 or 40 students) at NKNU since 2015. Students work on debate, two sessions per semester.

However, some students overlapped the argument, and were unable to rebuttal. Some were not interested in competition, and were discouraged when they lost. Two sessions may not be sufficient to foster argumentation skills.

As we have seen, we may face a couple of problems to foster critical thinking skills through debate in Taiwan. We should find an optimal solution, considering the following facts:

(5) The current situation in English education in Taiwan

- a. Asian students are not good at analytic or rule-based reasoning.
- b. PISA data show that Taiwanese students perform well in math and science.
- c. Debate is not a practical method to improve critical thinking skills at NKNU.
- d. English textbook in Taiwan may not offer inference-based tasks.
- e. AI is good at rule-based deep learning, but not context-based arguments.

Based on the facts above, I suggest that students in Taiwan should learn logical (analytic and rule-based) reasoning using AI.

3. Logical reasoning: To foster critical thinking skills

3.1 Weakness of logical reasoning among students at NKNU

When I introduced logical reasoning tasks in my writing class, freshmen students produced the following sentences which were intuitive, not logical.

(6) Reason clauses produced by freshmen in my writing class.

Q: *Which is better, girls' /boys' school or coeducational school?*

a. *I think girls' school is better than coeducational school because I can change my clothes in the classroom.*

Q: *Which is better, digital textbooks or paper textbooks?*

b. *I think digital textbooks are better, because books are heavy.*

<Writing about their future job>:

c. *I think I can be a good teacher because I am patient.*

Students produced a variety of *because*-clauses that were not logically justified; these meet neither the necessary nor the sufficient conditions. (6a) and (6b) do not describe differences in education. In “A or B” questions, students should consider the commonality between A and B, namely education. (6c) reflects a misunderstanding of job requirements. Put differently, the necessary condition for (6a) and (6b) should concern education, while that for (6c) should concern teacher qualifications.

(6a) – (6c) suggest that students are not using analytic or logical reasoning. They need to learn the concepts of necessary and sufficient conditions, which are central notions in formal semantics.

3.2 Logical reasoning in formal semantics

Logical reasoning can be represented as [if B, then A]. The expression [A *because* B] means that B causes A, where B is the cause. If B is the only cause, then B is presupposed to be true (cf. Hooper & Thompson), and B satisfies the necessary condition. If B is irrelevant to A, then B does not cause A. Using notions from formal semantics, we can define logical reasoning as satisfying either a necessary condition, as in (7a), or irrelevance, as in (7c) (cf. Belnap 1960; Chierchia & McConnell-Ginet 1990). Moreover, as shown in (6), B may be insufficient; that is, some additional factor may be required as a necessary condition for A, as described in

(7b). The following lecture was given in a writing class in the pre-AI period (before Jan 2023).

(7) [A *because* B] is logical, when B is a necessary condition.

- a. Necessary: B satisfies the necessary condition (logical).
- b. Insufficient: the description in B is not enough. There must be something else which contributes to the event A (not logical).
- c. No contribution (irrelevant): the event in B does not contribute to A (not logical).

[A *because* B] means [B *cause* A]. When you write, make sure that your statement satisfies necessary, sufficient conditions, and it should be relevant to the main point.

Examples (7a) to (7c) are shown below; they were generated by ChatGPT (free version, see Appendix 2).

(7') Three examples of *because* clauses for (8a) to (8c)

- a. *I was absent because I was sick in bed.*
(Necessary, the *because* clause is the causer)
- b. *He became rich because he worked hard.* (Insufficient)
- c. *He passed the exam because he wore a red shirt.* (Irrelevant)

Moreover, in linguistics, three types of *because*-clauses are distinguished in terms of syntactic structure and interpretation (cf. Frey 2016 for German). In English, native speakers produce the following three types of *because*-clauses, as in (8):

(8) Three types of English *because* clauses

- a. *John came back because he loved her.* (logical)
(The *because*-clause is the causer of the main clause)
- b. *John loved her, because he came back.* (colloquial)
(The *because*-clause expresses the reason of saying)
- c. *What are you doing tonight, because there's a good movie on.* (colloquial)
(The *because*-clause provides background information for the question)

(Sweetser 1990: 77, with my explanation)

Among the three types of *because* clauses, only (8a) is considered logical reasoning. Based on the studies listed in Table 1, we can assume that students in Europe and America understand when to use (8a) and when to avoid (8b). In contrast, students in East Asia do not distinguish between (8a) and (8b). Therefore, sufficient training should be provided to help learners distinguish between the uses of *because* clauses.

However, Feng (2006) points out that these usages are mixed in junior high school textbooks in Taiwan. Feng (2026) analyzed subordinate clauses taught in junior high school English textbooks in Taiwan, as shown in Table 3.

Table 3: The time table of different types of adverbial clauses in three main textbook publishers (Feng 2026, p. 10).

Publishers Types	Kang-Hsuan (KH)	Hanlin (HL)	Nan-I (NI)
Conditional (if)	B4L6	B4U4	B4L6
Temporal (when/before/after)	B3L3	B3U2	B3L3
Causal (because)	B3L2	B2U4	B3L2
Result (so)	B3L2	B2U4	B3L2
Purpose	—	—	—

* “B” means “Book,” “L” means “Lesson,” and “U” means “Unit.”

As we can see, *because* and *so* clauses are introduced in the book 2 or 3. Feng (2026) extracted the following examples of *because* clauses in the textbooks:

(9) Causal Clauses (*because*) in junior high school textbooks (Feng 2026, p. 16)

- a. Scott and I both rested at home last week because we were sick and had a fever. (KH)
- b. I’m wearing a mask because I’m sick. (HL)
- c. Her father sold her to the old man because her family was very poor. (NI)
- d. Tanya, from Yemen, got married to a man at eleven because he gave her father money. (NI)
- e. (I didn’t see a doctor) Because people in my country don’t always see a doctor for a cold. (KH)
- f. (I need ice cream) Because it can help my sore throat. (KH)

Among those, only (9a) meets the necessary (logical) condition for a reason clause⁵. Logical reasoning should be independent from culture, or put differently, it should be true in the most cultural contexts. On the other hand, (9b) to (9f) are culturally determined.

Moreover, these textbooks explain that the order of *because* clauses are irrelevant, and *because* can be replaced with *so* (Feng 2026, p. 13).

(10) The order of *because* clauses, and *so* in junior high school textbooks (Feng 2026, p. 13, modified)

⁵ I use the examples in (9) from the textbooks as prompts to generate teaching materials in ChatGPT (free version), together with the semantic definition of *because*-clauses, as shown in (7), and their three usages, as outlined in (8). It judged that (9b) - (9f) provide background or justification rather than logical causation. Further details are provided in my lesson plan, which will be published as 'Core-competency-Based English Lesson Plan Design, Teaching Logical Reasoning through *Because* Clauses,' issued by the English Language Teaching Research Center, National Kaohsiung Normal University (supported by the Ministry of Education, Taiwan).

a. Jerry didn't come to school because he caught a cold.

=b. Because Jerry caught a cold, he didn't come to school.

=c. Jerry caught a cold, so he didn't come to school.

However, in terms of the information structure, two *because* clauses in (10) receive different interpretations. Two sentences in (10a) and (10b) are explained here:

(10a) Jerry didn't come to school **because** he caught a cold.

(10b) **Because** Jerry caught a cold, he didn't come to school.

The *because*-clause in (10a) can function as an assertion (new information), whereas the main clause, *Jerry didn't come to school*, can be interpreted as presupposed (shared, old) information (cf. Hooper and Thompson 1973). In contrast, in (10b), the *because*-clause is interpreted as presupposed (or topic, Givón 1982, Bartzokas 2024), while the main clause functions as the assertion. More precisely, (10a) and (10b) are used in different contexts and are not equivalent, contrary to what the textbooks suggest. Moreover, *so* (as in (10c)) connects two events that are only loosely related; therefore, we cannot argue that there is a strong logical connection between the two clauses.

In sum, Feng's (2026) analysis indicates that students may not learn the logical reasoning usages of *because* in these English textbooks in Taiwan.

4. AI-supported teaching materials for logical reasoning

As mentioned above, logical reasoning is not focused in English textbooks for junior high school students in Taiwan. If Taiwan aims to foster critical thinking skills, it is imperative to provide appropriate sub-materials for them. For this purpose, I utilize ChatGPT (free version, March 2026) to create lesson plans for students (junior to senior high schools, and college). Part of the activity is listed in Appendix 2⁶. However, some of the judgements offered by ChatGPT (free version) are not reliable, hence I used ChatGPT plus to evaluate inappropriate sentences. It offers an explanation and alternative expressions. At least, ChatGPT plus is more reliable in terms of to generate lesson plans for logical reasoning. Put differently, although logical reasoning is independent of context and culture, understanding across cultures changes over time. For example, as shown in Appendix 2, judgments of sentence 2 varied: *There is no Internet because there are no telephone lines*. ChatGPT Free considered it logical, whereas ChatGPT Plus did not. ChatGPT Plus described that "In the past, Internet connections often relied on telephone lines (dial-up or DSL). However, today the Internet can also be provided through fiber, cable, satellite, or mobile networks, so the statement may not always be accurate." This suggests that the common understanding of the Internet has evolved over time.

Logical reasoning is often treated as context-independent, providing the formal foundation for critical thinking and argumentation. However, effective argumentation requires adapting this reasoning to specific contexts, particularly when persuading others. In such cases, claims must be supported not only by logically valid reasoning but also by relevant data and concrete

⁶ For the complete lesson plan and the prompts used to generate the lessons, please refer to the following lesson plans. 'Core-competency-Based English Lesson Plan Design, Teaching Logical Reasoning through *Because* Clauses,' issued by the English Language Teaching Research Center, National Kaohsiung Normal University (supported by the Ministry of Education, Taiwan).

examples. In this sense, argumentation can be understood as a context-dependent application of logical reasoning, where supporting reasons should ideally be necessary and sufficient for the claim, rather than merely illustrative.

For instructional purposes, students should therefore learn to construct arguments grounded in real-world contexts. Although AI tools can assist in generating example arguments, they do not reliably provide accurate data or statistics. As a result, AI-generated responses are best used as “halfway exemplars,” which students can refine and strengthen by incorporating verified, real-world evidence.

5. AI-supported teaching materials for argumentation skills: In pre-AI period and post AI implementation period

5.1 Writing activity, “Which is better, A or B?” (pre-AI implementation period)

Logical reasoning should be independent of context. In contrast, argumentation skills require examples and data to persuade the reader. To support skill development and prevent illogical *because* clauses (see (6)), instructors should provide contexts that remain valid over time and meet the requirements in (11).

(11) Essential points to be considered to foster argumentation skills

- a. The context should be valid in Taiwan and present two sides (i.e., A and B), allowing students to judge whether each is physically accurate. This approach fosters analytic skills.
- b. Students should examine both sides and consider their advantage and disadvantage for the society, not for an individual.
- c. Students should also infer the consequences (which could happen in the future) of A and B, including one advantage and one disadvantage for each (four points in total).
- e. Between A and B, students should decide which option is more beneficial for society.
- f. Students write one paragraph, for an ease of the assessment.

To satisfy the requirements in (11), I introduced a paragraph writing activities on “which is better, A or B?” in my freshmen writing class. In the pre-AI period (before Jun 2023), the following rubric was given, as in (12):

(12) The rubric:

Focus on the topic, be informative, be sophisticated. Write more than 4 sentences.

Students should discuss the problems of his/her writing with the instructor. If she/he says ‘ok’, each student should write the paragraph until the instructor says ‘ok’.

Expected structure

Sentence 1: topic sentence (I think xxx is better).

Sentence 2: Describe a current advantage (if you want to describe disadvantages, you can).

Sentence 3: Describe a future advantage (if you want to describe disadvantages, you can).

Sentence 4: a final remark (say something to conclude your discussion)

Even so, despite these instructions, students still produced illogical reasoning, as illustrated in (6b), which is repeated below.

(6) Reason clauses produced by freshmen in my writing class.

Q: *Which is better, digital textbooks or paper textbooks?*

b. *I think digital textbooks are better, because books are heavy.*

I explained them that the reason in (6b) is not logical, and they must rewrite it. However, some students did not finish all the tasks. To improve the situation, I offered writing samples generated by ChatGPT since Feb 2023.

5.2 Writing samples: ChatGPT as hallway exemplar (post-AI implementation period)

When students write a sentence, samples may help their writing (exemplar-based ESL writing; Chong, 2019; Rexroad, 2022). Yet, when students write a topic, ‘which is better, A or B’, and we offer a good writing sample, it may limit students’ insights, since we cannot expect many variations in their answers.

For this purpose, ChatGPT can be a better option for providing argumentation writing samples. As of March 2026, ChatGPT (both free and Plus versions) do not offer data, statistics, or proper nouns. When I asked ChatGPT (free), “Which is better, digital textbooks or paper textbooks?”, it responded, “It depends on personal preference, or the environment impact⁷.” A ChatGPT sample is demonstrated in Appendix 3. For the ease of writing for students, easy steps are also provided so that students have a clear idea on their writing, as in (13):

(13) 7 steps

Step 1: Think about a specific place, like Taiwan, Green Island, etc. specific subject.

Step 2: List the current advantage of both sides. List the future advantage of both sides in the society. Which one is the most influential? Think about the order the importance (an important one should come first).

Step 3: List the current disadvantages of both sides. List the future disadvantage of both sides.

Step 4: Provide an example/data.

Step 5: Pick up one side. Then describe the most influential current advantage and the future advantage for the society. If you like, you can describe the disadvantages of the other side.

Step 6: Talk to the instructor.

Step 7: If she/he says o.k., write a paragraph (more than 4 sentences).

⁷ ChatGPT (Plus) provides more consistent answers that consider social aspects.

Using the seven steps above and the ChatGPT samples, each student wrote paragraphs on approximately six topics per semester. Here are the list of topics in the first semester (Sep 2025) and the second semester (Feb 2026):

(14) Discussion topics (1st semester, Sep 2025)⁸

1. Which is better: Paper textbooks or digital textbooks?
2. For 7th graders in Taiwan, is it better to learn English in Chinese or in English?
3. Which is better: Teamwork or solo work?
4. Is global warming good or bad?
5. For students in grades 7–12, which is better: Asking English questions to a teacher or to AI?
6. Which is better: sustainable energy or oil?

(15) Discussion topics in the 2nd semester (Feb 2026)⁹:

1. Which is better: Percentage grading or A–B–C–D letter grading?
2. Which is better: Recycling collected by city employees or by individuals who receive coins?
3. Which is better: Studying a foreign language at age 3 or at age 8 (third grade)?
4. Which is better for secondary school students: Wearing school uniforms or their own clothes?
5. Which is better for college students: Taking notes on paper or on a tablet?
6. Which is better for college students in Taiwan: Learning English-related subjects in English or bilingually (Chinese and English)?

6. The Effects of AI: On teaching

This qualitative research compared the results of students' grading, between the pre-AI and post-AI implementation periods. Each year students worked on 9-15 paragraph writing tasks per semester. Each student discussed their output one by one with the instructor. When they satisfied all requirements, they got full-mark 30 points. Most students worked on all tasks, but after the consultation with the instructor, they may have hesitated to rewrite the paragraph. Table 4 presents the completion rates for the daily performance (30 points). Meanwhile,

⁸ For the complete lesson plan and the AI samples, please refer to the following lesson plans. 'Core-competency-Based English Lesson Plan Design, AI as Halfway-Exemplars: Fostering Critical Thinking and Debating Skills in L2 Writing for the 1st semester,' issued by the English Language Teaching Research Center, National Kaohsiung Normal University (supported by the Ministry of Education, Taiwan).

⁹ For the complete lesson plan and the AI samples, please refer to the following lesson plans. 'Core-competency-Based English Lesson Plan Design, AI as Halfway-Exemplars: Fostering Critical Thinking and Debating Skills in L2 Writing for the 2nd semester,' issued by the English Language Teaching Research Center, National Kaohsiung Normal University (supported by the Ministry of Education, Taiwan).

students produced a final essay. In the first semester, they work on an argumentative essay, with a claim and its data, and in the second semester, they developed their argument with rebuttals:

(16) Instructions for Final Essay in writing class¹⁰

1st semester: Discuss an issue in your future job industry with supporting data.

2nd semester: Discuss an issue in your future job industry with supporting data and a counterargument (rebuttals).

Table 4 shows the number of students who completed daily performance tasks, and their final essay grades and the number of students who scored above 90 points, comparing the pre-AI period (Sep 2020–Jun 2022) and the post-AI period (Sep 2023–Jun 2025)¹¹. Students enrolled in my Writing 1 class also took either my English course (for freshmen) or Listening & Speaking 2 (for sophomores), where they demonstrated their speaking ability. This allowed me to compare their writing performance with their speaking abilities. However, as indicated by the five scales in Table 4, the English proficiency of students in the post-AI period (freshmen or sophomores) is lower than that of students in the pre-AI period.

Table 4: The academic performance of students with or without AI in writing classes

With shadow: Post AI implementation periods

academic year	Mean Students in Writing 1	Mean Students Scoring 30 pts in Daily Performance Sem 1	30 pts in Sem 2	Mean Number of Students Scoring >90 and Mean Final Essay Score (Sem 1)	Mean Number of Students Scoring >90 and Mean Final Essay Score (Sem 2)	Speaking Ability (5-pt Scale, Y1/Y2)
Pre-AI Sep 2020 -Jun 2022	n=20	n=17	n=9.5	n=6 M: 82.7pts	n=6 M: 81.3pts	4.5
Post AI Sep 2023- Jun 2025	n=19.5	n=10	n=2.5	n=3.5 M: 75.8pts	n=6.5 M: 82.4pts	2.5

¹⁰ Rubric for both essays:

100 pts: Meets all requirements, provides strong insight into the argument(s), and demonstrates clear and consistent reasoning.

90 pts: Meets the requirements; however, there are some inconsistencies in the arguments, or the supporting data are insufficient.

85 pts: Provides some insight, but the proposed solution (and/or counterexample) is not compatible with the problem.

80 pts: The proposed solution (and/or counterexample) is not compatible with the problem.

40–70 pts: Does not meet the assignment requirements.

¹¹ Students (Sep 2022–Jun 2023) worked without AI samples in the first semester and with them in the second.

The followings are the differences between pre-AI and post-AI implementation periods.

(17) The instructions for the daily Performance, “Which is better, A or B? in writing class in pre-AI and post-AI implementation periods:

a. Pre-AI implementation period: Sep 2020 – Jan 2023

The Rubrics in (12) covered all assignments.

b. Post AI implementation period: Feb 2023 – Jan 2026

AI halfway examples with step-by-step instructions in (13) were also introduced.

In the writing class, students are expected to meet with the instructor to improve their argumentation in both daily performance and the final essay:

(18) The interaction with the instructor in the daily performance and the final essay

a. Daily performance (30 pts, 9–15 assignments): Students must have a conference with the instructor.

b. Final essay: Students may optionally discuss the content with the instructor.

As shown in Table 4, the completion rate of daily performance is relatively higher in the pre-AI period and lower in the post-AI periods. On the other hand, the average score and the number of students who scored over 90 points on the final essay remained almost the same during the pre-AI period, whereas both increased in the second semester of the post-AI period, as described in (19):

(19) The completion rate of daily performance tasks and final essay scores

a. Pre-AI: Students worked on daily performance tasks without sample guidance. Most students completed the assignments (17 out of 20 students) in the first semester. In the second semester, the completion rate decreased (9.5 out of 20 students). The final essay scores remained nearly the same.

b. Post-AI: Students worked on daily performance tasks with AI-generated samples. About half of the students completed the assignments (10 out of 19.5 students) in the first semester. However, the completion rate further decreased in the second semester (2.5 out of 19.5 students). The final essay scores improved in the second semester.

As described in (19), the completion rate is lower in the post-AI period; therefore, it is not possible to determine whether AI samples worked for students during this period.

(20) English proficiency

a. **Pre-AI period:** Their speaking ability was stronger

b. **Post AI period:** Their speaking ability was weaker

As explained in (20b), students in the post-AI period had relatively weaker English proficiency than those in the pre-AI period. This weakness may be one reason why fewer students

completed the daily performance tasks, which required a conference with the instructor in English.

8. The Effects of AI: For students

Questionnaires on AI learning were distributed to students who had taken or take Writing 1 in their freshman year. The 10 questionnaire items related to AI use are listed in Appendix 4. Responses were received from 12 of 17 freshmen (70.6%) and 12 out of 19 sophomores (63.2%) enrolled in my writing class between Sep 2024 and Jun 2025.¹² Among respondents, 95.8% reported that AI-generated examples supported their paragraph writing in class; 25.0% found all examples helpful, while 70.8% found some helpful.

I also revised the wording to ask whether the “Which is better, A or B?” activities helped students write their final essay. In response, 87.5% of students indicated that these activities were helpful, including 33.3% who selected “helpful” and 54.2% who selected “somewhat helpful.” 4.17% of respondents indicated that the activities were not helpful, explaining that they “didn’t understand the activities.” This percentage could be higher if all students had responded.

These results suggest that, among students who responded to the questionnaire, at least 25% clearly understood the AI-supported daily activities and found them helpful in producing their final essay. For 54.2% of respondents, the daily activities were somewhat helpful and contributed to their final essay to some extent. By contrast, 4.17% of respondents appeared to have difficulty understanding the daily activities.

Although AI examples were introduced as halfway exemplars, I also asked students whether they preferred writing samples produced by the instructor or by AI. The results showed that 45.8% preferred samples written by the instructor, 41.4% preferred a mix that included some instructor-written samples, and 8.3% preferred AI-generated samples, explaining that AI encouraged them to think of better answers on their own¹³.

In addition, I included other students in other EMI classes in my survey (freshman students in an English class and junior students who had taken my writing class) regarding their use of AI in academic life (30 responses in total). More than half of the students reported using AI for multiple purposes, including writing essays (60.0%), composing sentences or paragraphs (63.3%), learning vocabulary or grammar (60.0%), translating between languages (63.3%), and checking or correcting writing (73.3%). However, students’ views on the reliability of AI were mixed: 6.7% considered AI unreliable, 6.7% considered it reliable, and the majority, 80.0%, believed that AI is reliable only in some cases.

9. Research Observation

In this qualitative study, it remains unclear whether AI improves students’ writing performance in terms of grades. Although the questionnaire responses were limited, 95.8% of students reported that AI-generated samples assisted them in paragraph writing during the writing class. 4.17% of respondents appeared to have difficulty understanding the daily

¹² The students who responded to the questionnaires were likely those who regularly submitted their daily assignments. As a result, if all students had participated, the overall responses toward AI samples might have been lower than the figures reported here.

¹³ To improve their performance, I prepared writing samples in March 2026 to be provided after students’ writing completed.

activities. The low completion rate of daily performance tasks and comprehension issues improved after introducing group activities in September 2025; however, detailed observations will be analyzed elsewhere.

Consistent with responses from other classes, while AI appears to support students' self-directed learning, the majority of participants expressed caution regarding its reliability; only 6.7% considered it reliable. Although this proportion is small, instructors should provide appropriate guidance to encourage students to double-check AI-generated responses.

10. Conclusion

This paper examined three aspects of AI use in EMI English education. First, AI functions effectively as a tool for generating teaching materials that support the development of logical reasoning skills, which are largely context-independent. However, in fostering argumentation skills, which require concrete examples and empirical data, AI demonstrates clear limitations. Despite this, AI can still serve as a useful pedagogical resource by generating incomplete writing samples that allow students to practice strengthening data-driven arguments.

The paper also highlights the importance of logical reasoning as a foundation for critical thinking and argumentation in English education in Taiwan. AI is particularly useful for teachers when preparing materials related to rule-based reasoning. Nevertheless, due to its limitations, such as insufficient use of data, statistics, and proper nouns, AI should be regarded as providing partial scaffolding rather than fully developed models for teaching argumentation.

The grading results of this study do not indicate a significant effect of AI support on students' writing development. However, AI use contributed to clearer teaching rubrics, and most students reported that AI was helpful in their writing process. These findings suggest that while AI may not directly improve measurable writing outcomes, it offers indirect pedagogical benefits in terms of instructional clarity and learner support.

References

- Bartzokas, Chrysovalantis. 2024. Procedural structures: The case of sentence-initial subordinate clause. *Journal of Pragmatics* 229:80-92.
- Belnap, Nuel D. Jr. 1960. Entailment and relevance. *The Journal of Symbolic Logic* 25: 144-146.
- Chierchia, G. & McConnell-Ginet, S. 1990. *Meaning and Grammar: An Introduction to Semantics*. Cambridge, MA: MIT Press.
- Chong, S-W. 2019. The use of exemplars in English writing classrooms: from theory to practice. *Assessment & Evaluation in Higher Education* (<https://doi.org/10.1080/02602938.2018.1535051>).
- Ennis, R. H. 1962. A concept of critical thinking: a proposed basis for research in the teaching and evaluation of critical thinking ability. *Harvard Educational Review* 32:81-111.
- Facione, Peter. A. 1990. Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. *Research findings and recommendations prepared for the Committee on Pre-College Philosophy of the American Philosophical Association*. ERIC Document # ED 315-423 (https://www.researchgate.net/publication/242279575_Critical_Thinking_A_Statement_of_Expert_Consensus_for_Purposes_of_Educational_Assessment_and_Instruction, retrived on 16/11/2020).
- Feng, Allen. 2026. Cartographic Editorial Approaches to Adverbial Clauses: Their Positions in English Textbooks for Junior High Schools in Taiwan. A term paper, English Department, NKNU.
- Frey, Werner. 2016. On some correlations between interpretative and formal properties of causal clauses. In *Linguistische Berrichte Sonderheft 21*, 153–179. Hamburg: Helmut Buske Verlag.
- Givón, Talmy. 1982. Pragmatics, with human language as the referee: Toward an empirically viable epistemology. *Journal of Pragmatics* 6(2): 81-133.
- Hitchcock, David. 2011. Critical thinking as an educational ideal. A paper presented at a conference on critical thinking at Huazhong University of Science and Technology in Wuhan, China (https://www.researchgate.net/publication/315852948_Critical_Thinking_as_an_Educational_Ideal, retrieved on 10 March 2026).
- Hooper, Joan. B., and Sandra A. Thompson. 1973. On the Applicability of Root Transformations. *Linguistic Inquiry* 4(4): 465–497.
- Inhelder, Bärbel., & Jean Piaget. 1958. *The Growth of Logical Thinking from Childhood to Adolescence*. Translated by A. Parsons & S. Milgram. New York: Basic Books, Inc., Publishers.
- Lun, Vivian Miu-Chi. 2010. Examining the influence of culture on critical thinking in higher education [Unpublished doctoral dissertation]. Victoria University of Wellington, Wellington, New Zealand.

Majidi, Abid. el., Daniel Janssen, Rick de Graaff. 2021. The effects of in-class debates on argumentation skills in second language education. *System* 101: 1-15.

MOE, *General Guideline of the 12 Year Basic Education Curriculum*. 2021. <https://www.naer.edu.tw/eng/PageSyllabus?fid=148>

MOE, *Handbook for English Education*. 2018. <https://www.naer.edu.tw/PageSyllabus?fid=197>

Masuda, Takahiko, & Richard E. Nisbett. 2001. Attending holistically versus analytically: Comparing the context sensitivity of Japanese and American. *Journal of Personality and Social Psychology* 81: 922-934.

Mineshima, Michio. & Junichiro Chino. 2013. Comparative analyses of questions and tasks in English textbooks in Japan, Korea, and Finland: How textbooks teach critical thinking. *The Chubu English Language Education Society* 42: 91-98.

Norenzayan, Ara, Edward E. Smith, Boem Jun Kim, & Richard E. Nisbett. 2002. Cultural preferences for formal versus intuitive reasoning. *Cognitive Science* 26: 653-684.

Peng, Kaiping & Richard E. Nisbett. 1999. Culture, dialectics, and reasoning about contradiction. *American Psychologist* 54(9): 741-754.

Rear, David. 2017. The language deficit: A comparison of the critical thinking skills of Asian students in first and second language contexts. *Asian-Pacific Journal of Second and Foreign Language Education* 2(13): 1-15.

Rexroad, Brent. 2022. Guiding Students to Good Writing: An Exemplar-Based Model. Ms., 玉川大学.

Sweetser, Eve. 1990. *From Etymology to Pragmatics*. Cambridge: Cambridge University Press.

Tweed, Roger. G. & Darrin R. Lehman. 2002. Learning considered within a cultural context. *American Psychologist* 57(2): 89-99.

Appendix 1

TABLE i: PISA scores in 2009, 2012, 2015, 2018, 2022

(PISA 2022 Results (Volume I) | OECD)

Country	Year	Reading	Mathematics	Science
Taiwan	2022	515	547	537
	2018	503	531	516
	2015	497	542	532
	2012	523	560	523
	2009	495	543	520
Japan	2022	516	536	547
	2018	504	527	529
	2015	516	532	538
	2012	538	536	547
	2009	520	529	539
Korea	2022	515	527	528
	2018	514	526	519
	2015	517	524	516
	2012	536	554	538
	2009	539	546	538
United States	2022	504	465	499
	2018	505	478	502
	2015	497	470	496
	2012	498	481	497
	2009	500	487	502
Finland	2022	490	484	511
	2018	520	507	522
	2015	526	511	531
	2012	524	519	545
	2009	536	541	534

Appendix 2

Learning logical reasoning by ChatGPT free version (1 March 2026):

Core Principle (Student-Friendly Version)

“A because B” is logical **only if A cannot happen without B**.
If A can still happen without B $\rightarrow$ B is **not necessary**.

You may present it visually:

If NOT B $\rightarrow$ NOT A (Necessary condition test)

Instructions for Students

Circle:

- N = Necessary
- X = Not logical

Exercise A (Simple & Clear)

1. I was absent because I was sick in bed.
2. There is no Internet because there are no telephone lines.
3. I like pizza because it is round.
4. The plant died because it had no water.
5. He passed the exam because he wore a red shirt.
6. The room is dark because the lights are off.

Teacher Key by free version by ChatGPT free

Necessary: 1, 2, 4, 6,

Not logical: 3, 5

However, #2 is strange:

2. *There is no Internet because there are no telephone lines* (i.e., context dependent). I also utilize ChatGPT Plus, to see if it can offer a reasonable answer. The following answer is generated by ChatGPT plus:

The causal relationship is partially logical but outdated or incomplete. In the past, Internet connections often relied on telephone lines (dial-up or DSL). However, today the Internet can also be provided through fiber, cable, satellite, or mobile networks, so the statement may not always be accurate.

Appendix 3

1 Which is better, a paper textbook or digital textbook?

An answer by ChatGPT free version (Feb 2023):

Both paper textbooks and digital textbooks have their own advantages and disadvantages. Paper textbooks provide a tactile and physical reading experience that many students find helpful for comprehension and retention of information. They also don't require electricity or internet access to use, which can be beneficial in certain situations.

On the other hand, digital textbooks offer several advantages such as portability, searchability, and multimedia features that can enhance learning. Digital textbooks can be accessed from various devices and can include interactive elements like videos, quizzes, and simulations that can make learning more engaging and interactive. Digital textbooks can also be updated more easily than paper textbooks, which may become outdated over time.

Ultimately, the choice between a paper textbook and a digital textbook depends on personal preference and the specific needs of the individual. Both formats can be effective for learning and it may be worth trying out both formats to see which one works best for you. Additionally, it's worth considering the environmental impact of paper textbooks and the potential cost savings of digital textbooks. END

Appendix 4

Questionnaires (10 March to 17 March 2026)

This survey investigates the effects of AI in a freshman writing class. During the course, you completed several assignments, including a midterm essay and a final essay. The purpose of this survey is to examine whether AI-supported activities helped develop your argumentation skills. Thank you for your participation.

1. Your class:

2. You worked on the topic “Which is better, A or B?” using AI-generated samples that were not well developed as arguments. Do you think the AI samples helped you or your group discussion?

Yes, all of them were helpful.

Yes, some of them were helpful.

No, they were not helpful.

I am not sure.

3. Do you think the AI samples helped you write a paragraph?

Yes, all of them were helpful.

Yes, some of them were helpful.

No, they were not helpful.

I am not sure.

8. Do you think the “Which is better, A or B?” assignments helped you write your final essay?

Yes, they were helpful.

Yes, they were somewhat helpful.

No, they were not helpful because I did not understand them.

I am not sure.

9. AI often provides undeveloped arguments due to technological limitations. Would you prefer to see a complete example written by the instructor instead of AI-generated samples?

Yes, I prefer a complete example written by the instructor.

Yes, I prefer some examples to be written by the instructor.

No, I prefer AI samples because they encourage me to think of a better answer myself.

I am not sure.

10. For which of the following purposes do you use AI when learning English? (Select all that apply.)

Writing an essay

Writing sentences or paragraphs

Practicing speaking (e.g., conversation practice)

Practicing listening

Learning vocabulary or grammar

Translating between languages

Checking or correcting writing

Other (please specify): _____

11. What do you think of the quality of the product by AI for English learning?

Very reliable.

Sometimes not so reliable.

I don't think it is reliable.

I am not sure.

End