

*Article*

## **Enhancing Critical Thinking Skills with ChatGPT-Powered Activities in Chinese Language Classrooms**

**Xinliang Jiang\***

**Jing Li**

**Chiu-Hung Chen**

University of Toronto Mississauga, Canada

Received: 19 September, 2023/Accepted: 7 January, 2024/Published: 20 February, 2024

### **Abstract**

ChatGPT, an Artificial Intelligence (AI) chatbot, emulates human-like conversations and is built upon OpenAI's GPT language model introduced in 2018. In the context of language classroom activities, our observations indicate that language learners employ a spectrum of strategies during interactions with ChatGPT, reflecting Bloom's taxonomy. These strategies encompass lower-order thinking (e.g., remembering, understanding, applying) and higher-order thinking (e.g., analyzing, evaluating, creating). Utilizing Bloom's revised taxonomy (cited in [Anderson & Krathwohl, 2001](#)) as a framework for defining critical thinking (CT) skills, this paper seeks to investigate the impact of ChatGPT-powered activities on CT skill application within three Chinese language classrooms: introductory and intermediate courses for second language (L2) learners and an academic writing course for native Chinese speakers. It also investigates whether the nature of tasks contributes to learners' application of CT. The following two research questions frame the current study: 1) Which CT skills are utilized in ChatGPT-empowered learning activities for both L2 learners of Chinese and native speakers of Chinese? 2) How do task types in ChatGPT-empowered learning activities relate to the application of CT skills among L2 learners of Chinese and native speakers of Chinese? This research adheres to an action research framework, marked by a cyclic process involving planning, execution, observation, and reflection ([Kemmis and McTaggart, 2005](#), as cited in [Burns, 2011](#)). Qualitative data have been gathered to assess the application of CT skills among Chinese learners. The findings revealed that six levels CT skills including remembering, understanding, applying, analyzing, evaluating and creating, were consistently applied across all three language courses, irrespective of their levels of language proficiency. Second, task types in ChatGPT-empowered learning activities and the application of students' CT abilities are intricately linked. The types of tasks play a crucial role in shaping the utilization and cultivation of different levels of CT skills. Consequently, these factors significantly impact the attainment of learning objectives.

### **Keywords**

Critical thinking (CT), ChatGPT, Chinese as a foreign language, computer-assisted language learning

---

\*Corresponding author. Email: [xinliang.jiang@utoronto.ca](mailto:xinliang.jiang@utoronto.ca)

## 1 Introduction

Critical thinking (CT) is often regarded as an important skill for the 21st century, prominently featured in educational curricula worldwide. A study conducted by Voogt and Roblin (2012), which scrutinized international frameworks pertaining to 21st-century competences, highlighted the consensus on the significance of CT across these frameworks. In the realm of L2 learning, CT holds significance as language learners consistently participate in communicative tasks necessitating applications, analyses, evaluations, and more. Its integration into the classroom has been substantiated by studies demonstrating its capacity to enhance the learning experiences of L2 learners (Zhou, 2022; Yuan et al., 2022).

Incorporating CT skills into the curriculum can be met with challenges like time constraints and the availability of teaching staff. However, the emergence of generative AI, especially with ChatGPT gaining prominence in the fall of 2022, holds promise for potentially mitigating these obstacles. ChatGPT is a language model created by OpenAI, a United States-based startup, with a focus on natural language understanding and generation. Its capability for intelligent conversation is achieved through a two-step process: pre-training on extensive dialogue datasets and fine-tuning in real chat scenarios. This model excels at understanding user inputs and generating coherent responses, effectively simulating human-like conversations.

Within language classroom activities supported by ChatGPT, our observations reveal that language learners employ a diverse range of strategies when interacting with ChatGPT, aligning with the principles of Bloom's taxonomy. These strategies encompass both lower-order thinking skills (such as remembering, understanding, and applying) and higher-order thinking skills (including analyzing, evaluating, and creating).

Against this backdrop, this paper reports the impact of ChatGPT-powered activities on CT skill application within three Chinese language classrooms: introductory and intermediate courses for L2 learners and an academic writing course for native Chinese speakers. It also investigates whether the nature of tasks contributes to learners' application of CT. In the following sections, we first present the background of this study. It is followed by the research methodology, results, discussions and conclusion.

## 2 Background

This section begins with a discussion on the definition of CT, and its importance and application in language classrooms. It is followed by an introduction of ChatGPT, a language chatbot investigated by the current study, and its relationship to the application of CT skills in Chinese language classrooms.

### 2.1 CT and its application in L2 classroom

What is critical thinking? According to the National Council for Excellence: "Critical thinking is the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action." At its core, CT denotes the capacity to effectively solve problems and analyze information, leveraging high-order cognitive skills (Halpern 1993). These cognitive skills, which constitute the foundational framework for the conceptualization of CT, are notably delineated within Bloom's taxonomy and were subsequently refined and expanded in Bloom's revised taxonomy (cited in Anderson & Krathwohl, 2001) which the present study is based on.

Bloom's revised taxonomy (cited in Anderson & Krathwohl, 2001) redefines cognitive domain as the convergence point between the cognitive process dimension and the knowledge dimension. The knowledge, from tangible to conceptual, include four types: factual, conceptual, procedural and

metacognitive knowledge. The cognitive process dimension is employed to describe six cognitive processes that encompass a spectrum from lower order thinking skills to higher order thinking skills, representing the various ways in which individuals engage with knowledge: remembering, understanding, applying, analyzing, evaluating and creating. There are subcategories associating with each type of knowledge and cognitive process. Table 1 below describes the intersections of the knowledge and the six cognitive processes (Fisher, 2007):

Table 1

*The Cognitive Process Dimension*

| The Knowledge Dimension  | The Cognitive Process Dimension |            |            |               |          |           |
|--------------------------|---------------------------------|------------|------------|---------------|----------|-----------|
|                          | Remember                        | Understand | Apply      | Analyze       | Evaluate | Create    |
| Factual Knowledge        | List                            | Summarize  | Classify   | Order         | Rank     | Combine   |
| Conceptual Knowledge     | Describe                        | Interpret  | Experiment | Explain       | Assess   | Plan      |
| Procedural Knowledge     | Tabulate                        | Predict    | Calculate  | Differentiate | Conclude | Compose   |
| Meta-Cognitive Knowledge | Appropriate Use                 | Execute    | Construct  | Achieve       | Action   | Actualize |

These intersections are typically employed for crafting learning objective statements, delineating the desired cognitive processes and the knowledge students should gain. For example, in an intermediate level Chinese as a Foreign Language (CFL, henceforth) class, a learning objective statement might read as follows: Students will demonstrate the ability to analyze (cognitive process) the concept of group-oriented consciousness in Chinese culture (conceptual knowledge).

CT skills have been integrated into L2 learning classrooms because of its importance. Zhou (2022) argues that it enhances the overall learning experience for students, rendering the process of acquiring a foreign language more meaningful and engaging. It is also closely associated with students' academic achievements. Since CT is a skill that requires cultivation, strategies of its application have been discussed in the literature. Yuan et al. (2022) discuss several instructional strategies of CT instructions in the EFL (English as a foreign language, henceforth) context. These strategies which emerged from their system review of 25 empirical studies on K-12 and college teachers' perceptions of and engagement with CT include:

- Triggering: triggering students' motivation for addressing critical issues.
- Exposing: Exposing students to educatively "messy" information, experiences, and perspectives.
- Guiding: Guiding students to form and defend justifiable stances.
- Extending: Extending students to think beyond contextualized experiences (p.7).

To assess the application of CT using strategies such as above, Zhou (2022) posits that CT fundamentally involves mental processes, and as such, it can only be discerned by observing its direct manifestations. This includes assessing a student's willingness to engage in CT and their instance, when a student can assess validity of a claim with supporting evidence, it showcases his pability to demonstrate mastery of this skill as shown in Bloom's taxonomy discussed earlier. For iroiciency in CT skills.

However, the impact of CT instructional strategies may be contingent on various factors. The study by Yuan et al (2022), which was discussed earlier, identified the following factors that could facilitate or hinder the impacts of the instructions: student-, teacher- and context-related factors. On one hand,

students' attitudes, expectations, self-efficacy, and the pre-existing cognitive framework exhibited strong correlations with the implementation and outcomes of EFL teachers' CT instruction. On the other hand, the level of education, years of teaching experience, attitude, understanding of CT, and choice of instructional strategies all played substantial roles in influencing their approach to CT instruction. Finally, institutional conditions, such as time constraints, mandatory curricula, and rigid textbooks, and cultural traditions, such as "exam-focused and spoon-fed education" had a significant impact on the way EFL teachers approached CT instruction (p.43).

In the era of AI, can contextual challenges like time constraints that impede the cultivation of CT skills be mitigated through the support of chatbots such as ChatGPT? We will discuss such possibilities in the following section.

## **2.2 Leverage AI for CT in L2 classroom**

The history of chatbots in the context of language learning dates back to the 1960s and 1970s when researchers began exploring the potential of using computers to aid in language acquisition (Hubbard, 2022). With the advancement of AI technology, chatbots underwent significant improvements, incorporating natural language processing and machine learning techniques to simulate human-like conversations. These modern chatbots are better equipped to understand and respond to natural language inputs accurately, and they possess more sophisticated conversational capabilities. These technological strides have paved the way for the integration of chatbots into language learning (Fryer et al., 2019; Kuhail et al., 2023).

ChatGPT, the AI chatbot used in the current study, is a language model developed by the United States-based startup OpenAI, designed for natural language understanding and generation. It achieves the functionality of intelligent conversation by undergoing pre-training on a large dataset of dialogues, followed by fine-tuning in actual chat tasks. The ChatGPT model can comprehend user inputs and produce a coherent response, thereby simulating conversations between humans.

According to Cai (2023), ChatGPT offers a variety of features that can support L2 acquisition in accordance with L2 acquisition theories. First, as mentioned above, ChatGPT possesses an extensive knowledge base and retrieves pertinent information to understand user inputs and generate a cohesive reply. With proper prompts, ChatGPT can provide comprehensible input tailored to the learner's proficiency level. Second, through interacting with ChatGPT, L2 learners can negotiate meaning, receive clarifications and notice discrepancies between their L2 and L1. Third, ChatGPT can offer diverse forms of feedback to learners in response to their prompts, including explicit and corrective feedback. Fourth, ChatGPT can provide various types of input enhancement, such as frequent use of target vocabulary or structures in the input, thereby heightening L2 learners' awareness of language forms and thus aiding L2 acquisition. Finally, ChatGPT offers an optimal environment for L2 learners to produce a substantial volume of output through interactive exchanges and feedback.

Throughout the interaction process with ChatGPT, as demonstrated above, L2 learners would utilize a range of strategies to comprehend, negotiate meanings, and refine their output based on the feedback to ensure effective communication. These strategies are reflected in Bloom's taxonomy, where L2 learners may utilize lower-order thinking skills, such as remembering, understanding and applying, as well as higher-order thinking skills like analyzing, evaluating and creating. As pointed out by Yuan et al. (2022), there has been a strong call to integrate CT into foreign language curricula, given the close connection between language, culture, and thought. However, we have predominantly observed this integration in the EFL context, with less emphasis on the CFL context, and the context in which native speakers of Chinese who take an academic writing course. Furthermore, given the advancements in AI technology, there is potential to address the factors that hinder the integration of CT into language classrooms as discuss above. In the context of these considerations, the current study intends to answer the following questions:

1. Which CT skills are utilized in ChatGPT-empowered learning activities for both L2 learners of Chinese and native speakers of Chinese?
2. How do task types in ChatGPT-empowered learning activities relate to the application of CT skills among L2 learners of Chinese and native speakers of Chinese?

This study follows an action research model, characterized by a cyclical process encompassing planning, implementation, observation, and reflection (Kemmis and McTaggart, 2005, as cited in Burns, 2011).

### 3 Research Methods

This section will provide a detailed overview of the research methodology, including participant background information, research material and activity design, as well as data collection and analysis.

#### 3.1 Participants

All participants in this study are L2 and native Chinese learners enrolled in relevant Chinese language courses at a public university in Canada. They come from Beginner Chinese class, Intermediate-High Chinese class, and Chinese Academic Writing class. Among them, students in the first two categories are L2 Chinese learners, including both Chinese heritage and non-heritage students, while the Chinese Academic Writing class is designed for native Chinese speakers.

##### 3.1.1 Second-language learners

Students in the Beginner Chinese class have a Chinese proficiency level approximately equivalent to the Novice-Mid level defined by ACTFL (American Council on the Teaching of Foreign Languages). They can discuss self-related topics using simple vocabulary and sentence structures, understand basic spoken and written texts with the help of contextual cues and repetition. Students in the Intermediate-High Chinese class have a proficiency level approximately at the Intermediate-High level. Compared to the Beginner class, these students are more adept at using diverse vocabulary and sentence patterns, as well as comprehending more complex texts and abstract topics.

**Novice Level.** A total of 21 students participated in the Beginner Chinese class in this study, with six of Chinese ethnic backgrounds, and 15 of non-Chinese ethnicities, primarily come from Canada, South Korea, Thailand, the Philippines, Mexico, and Colombia. The course they were taking covers various functional Chinese topics, including discussing daily activities and geographical locations. Tasks encompass making plans for daily activities, asking about time and age, inviting others, asking for directions, and making phone calls, among other things. The course duration is 12 weeks, with classes held twice a week of two hours each. For this study, 16 participants were recruited from the Novice-level Chinese class. However, data from one participant were deemed invalid, resulting in a total of 15 valid participants. Among these participants, there are 10 male and five female students, representing various majors from first to fourth year in their academic studies.

**Intermediate-High Level.** The Intermediate-High Chinese class has seven students, six of whom have Chinese heritage backgrounds, while two are from Thailand and South Korea. They used the textbook “Expressive Chinese” covering topics like daily life, professions, music, and architecture. The course spans 12 weeks, meeting once a week for three hours. On the day of data collection, four students initially participated in the activity; however, only two students, both female with Chinese heritage and majoring in Economics, became the final valid participants.

### 3.1.2 Native speakers of Chinese

Ten native Chinese speakers enrolled in the *Chinese for Academic Purposes II* course participated in the study. Compared with students who complete their high school education in China, these students demonstrated a relatively lower proficiency in the Chinese language, marked by notable disparities in their language competencies. Despite their shared Chinese heritage, their proficiency in the language had generally plateaued or regressed following their relocation to Canada during primary or secondary school. After their arrival in Canada, most of them discontinued formal Chinese language instruction, resulting in their Chinese language proficiency remaining largely static at the level it had reached when they departed China.

Evident in their written work were recurrent issues, including imprecise vocabulary, punctuation inaccuracies, and challenges related to sentence structures, which involved foundational grammatical and expressive hurdles. Given these distinctive participant attributes, the provision of prompt, tailored writing assessments, and feedback became particularly critical in this course, aimed at fostering their CT abilities.

## 3.2 ChatGPT-powered activity design

Relevant teaching activities were designed to address the research questions mentioned above in 2.2 and understand whether ChatGPT can assist Chinese L2 learners and native speakers in promoting and enhancing CT, as well as which CT skills may be employed. The primary instructional activities include:

- A. Novice-level Chinese Class: Students use ChatGPT to resolve questions arising during quizzes, enhancing their self-correction and self-guided Chinese language learning abilities. It also helps in fostering the application of CT.
- B. Intermediate-high Chinese Class: ChatGPT is employed to provide necessary information on various topics. Students extract valuable information to expand their understanding on the topics, eventually forming clear and well-reasoned written or oral expressions.
- C. Chinese Native Speakers' Academic Writing Class: ChatGPT is used to evaluate introductions, word choices, and writing styles in students' papers. Students then improve their drafts based on a comparison, reflection, and evaluation of two versions, ultimately revising their introductions.

### 3.2.1 Second language learners

While L2 learners are in the process of acquiring the target language and continually developing their language skills, learners at different stages have varying levels of language acquisition and competence. Especially at the beginner stage of learning Chinese, learners may be influenced by various factors such as pronunciation, vocabulary, grammar, and culture. Therefore, the design of research activities for novice-level and intermediate-high Chinese language learners is different. The activities in L2 research are closely aligned with students' existing knowledge and abilities, as well as the practical needs of language teaching.

Students at the novice level are typically limited to discussing straightforward topics related to their personal experiences, utilizing basic vocabulary and grammatical structures. They frequently struggle to seek assistance when encountering difficulties. ChatGPT, as a powerful language model based on real text data, can engage in interactive conversations with users. Consequently, within the novice-level class, the primary emphasis is placed on analyzing and resolving questions that arise during quizzes. It involves cultivating students' self-correction skills, enhancing their understanding of errors, and expediting the

correction process by actively engaging with ChatGPT through posing inquiries. Furthermore, in the course of comprehending and analyzing the responses furnished by ChatGPT, students are anticipated to exercise and foster their CT abilities. This activity is designed to enable students to achieve the following learning objectives:

- Understanding and applying knowledge: Students are adept at posing questions to ChatGPT, exhibiting memory of previously learned materials, and comprehending the new information provided by ChatGPT to solve problems in new situations.
- Cultivating analytical and reflective skills: Students are expected to employ their previously acquired knowledge to critically compare and analyze ChatGPT's information, reflecting on its answers by conducting analyses, evaluations, and judgments, ultimately resulting in the resolution of doubts and identification of quiz question errors.

In contrast, intermediate-high level Chinese class students demonstrate proficiency in utilizing diverse vocabulary and complex sentence structures. They can grasp intricate texts and abstract topics to some extent. While most students can articulate their viewpoints effectively, they may encounter difficulties due to limited background knowledge. Hence, the activity designed for this level Chinese class primarily concentrates on nurturing students' ability to elaborate on topics, concurrently fostering CT skills through active engagement with ChatGPT, particularly at the higher-order level, aligning with the following learning objectives:

- Comprehension and Application of Knowledge: In the absence of background information, students are able to center discussions around the topic, ask relevant questions, and engage in reading, memorizing, and comprehending the information provided by ChatGPT. They demonstrate the ability to apply their previously acquired knowledge.
- Cultivation of Analytical, Evaluative, and Creative Abilities: Students are able to establish connections between the information supplied by ChatGPT and the topic at hand. They are able to conduct analyses and evaluations, distilling valuable information for incorporation into existing paragraphs as effective support material for topic development, ultimately resulting in the creation of new paragraphs.

Novice Level. The specific research activity conducted in the Novice-level Chinese class using ChatGPT involve the following steps after a classroom quiz. Students were first required to self-assess and correct any uncertainties they had regarding the quiz content. For questions they could not resolve independently, they were encouraged to seek assistance from ChatGPT. Subsequently, students were tasked with comparing and analyzing ChatGPT's answers, reflecting on their validity, and finally providing feedback on whether ChatGPT had aided in resolving their doubts and sharing their user experiences. The detailed steps of this research activity and the template for student feedback are outlined below:

1. After the quiz, students are instructed to self-assess.
2. They review the quiz questions and identify those for which they are unsure and could easily find the answers. They list the question numbers for these queries.
3. For these questions, students perform self-analysis, identifying the key issues, such as grammar, character writing, pronunciation, punctuation, or expression, and make necessary corrections. They then remove the question numbers from the list.
4. If they are unable to self-correct certain questions, students can seek assistance from ChatGPT by individually addressing the source of their doubts. They can ask ChatGPT questions, such as: "Is the following Chinese sentence correct in grammar, characters, punctuation, or expression?"

5. Next, students are required to compare and analyze their own answers with those provided by ChatGPT and contemplate the appropriateness of ChatGPT's answers. For example, do they agree with ChatGPT's response? Where were their issues (grammar, typing, pinyin, pronunciation, punctuation, or expression)? Have they made corrections as indicated by ChatGPT? Have they resolved their doubts about the quiz question?
6. Students repeat steps four and five until all questions are resolved.
7. Finally, they conduct a comprehensive assessment and provide feedback. Students share their experiences of using ChatGPT, answer prompting questions (see Appendix 1) about how ChatGPT has assisted them in learning Chinese and in which aspects it has been helpful. They also discuss whether they intend to use ChatGPT for Chinese language learning in the future and provide reasons for their decisions.

These activities are closely aligned with the practical needs of beginner-level students, allowing them to independently address questions that arise during their Chinese language learning or quizzes/tests/exams using ChatGPT. The guidance provided in the activity steps aims to facilitate the application and training of students' CT skills.

Intermediate-High level. The specific research activity involving the use of ChatGPT in the intermediate-high Chinese class was conducted during a classroom topic discussion. Students were tasked with selecting a topic, analyzing and contemplating its various aspects, and then, crafting a short essay of approximately 100 words or forming a spoken expression of around one minute in duration based on the chosen topic. Students were required to employ the three-step outlining strategy skills taught to them for analyzing the topic and expanding it into paragraphs. Additionally, they were expected to extract valuable information from the resources provided by ChatGPT for use in their oral or written expressions. The ultimate goal is to create a well-structured and well-reasoned written or oral expression on the chosen topic. The specific steps for implementing ChatGPT in teaching are as follows:

1. Students are required to reflect on a profession that is not easily replaceable, present viewpoint, and brainstorm to provide reasonable and reliable evidence to support their stance. If lacking background information, students can seek guidance from ChatGPT.
2. After contemplating the answers, they should select a profession they agree with and wish to express, proceed to the next step of thinking, and use the outlining strategy to list reasons that support their viewpoint.
3. If they still lack background information, they can request ChatGPT to provide a reason. For example, "Please tell me one reason why teachers are not easily replaced by AI."
4. They should critically evaluate the provided reason to determine if it can be an effective argument.
5. After writing a paragraph, they need to check it for grammar or structural errors and make necessary revisions.
6. They can ask ChatGPT to help recheck the paragraph for errors. They should compare their paragraph with the questions raised by ChatGPT, reflect on their correctness, and decide whether further changes are needed. If unsure, they can consult their teacher.
7. Finally, they should formulate a written paragraph or oral discussion expression that satisfies their expectations.

The activity is closely aligned with the teaching content of Lesson Six, "Future Directions," which focuses on the topic of professions. Specifically, the question, "Which profession is not easily replaceable?" directly relates to the current discussion about professions that ChatGPT might easily replace.



### 3.2.2 Native speakers of Chinese

As previously mentioned, the participants in the *Chinese for Academic Purposes II* course exhibited generally low proficiency in Chinese, with varying levels of competence, compared with students who complete their high school education in China. This emphasized the importance of providing timely and personalized writing assessments and feedback, as it was through these assessments and feedback that participants could reflect on their strengths and weaknesses in writing, fostering CT and ultimately improving their practical writing skills. This was the primary motivation for the researchers to incorporate ChatGPT into the teaching activities.

ChatGPT's attributes, including its capacity to provide timely feedback and its role as a non-authoritative entity (akin to a peer rather than an authoritative figure), rendered it an ideal companion for offering writing advice to participants and encouraging their self-reflection. The ChatGPT-Powered Activity in the classroom of native Chinese speakers was meticulously designed to align with the curriculum of the fifth week in the *Chinese for Academic Purposes II* course. This specific week centered on the nuances of tone in Chinese academic papers and involved practical exercises pertaining to crafting paper introductions. In this activity, ChatGPT assumed the role of an editor, furnishing participants with alternative renditions of introductory paragraphs that shared similar content but varied in expression. Participants employed these provided versions as points of reference to undertake self-assessment and evaluation of their own writing, subsequently proceeding to revise their introductions. The anticipated learning objectives for this activity were two-fold:

1. Understanding, memorizing, and applying knowledge: By contrasting the revised introductions generated by ChatGPT with their initial drafts, participants will identify and explain the differences between the two versions. This process will facilitate the recall, comprehension, and application of the criteria and structural characteristics governing Chinese academic writing, which are essential for the revision of their introductory drafts.
2. Cultivating CT skills and enhancing writing proficiency: As participants refer to the revised introductions offered by ChatGPT, they will examine and reflect on their own initial drafts. This analytical process encompasses analyzing, evaluating, and decision-making, empowering them to discern areas of improvement, amend their drafts, elevate their writing skills, and concurrently develop CT abilities.

The process of the ChatGPT-Powered Activity is straightforward and is facilitated through a structured worksheet (see Appendix 2). The steps involved are as follows:

1. Participants submit their initial introduction draft to ChatGPT for revision. Through multiple interactions with ChatGPT, they work towards achieving the most refined introduction version provided by ChatGPT.
2. Participants carefully observe and list all modifications made by ChatGPT to the original text, including vocabulary, sentence structure, punctuation, paragraph structure, academic writing tone, and other comparative aspects.
3. In alignment with the Chinese academic writing standards supplied by the researchers, participants assess whether ChatGPT's revisions adhere to and exemplify these established criteria.
4. Participants evaluate the strengths and weaknesses of ChatGPT's revisions compared to their own initial drafts and explain the reasons for these assessments.
5. Participants engage in reflective analysis to identify and list the shortcomings present in their own introduction writing. They also explore potential strategies for improvement.

6. Drawing upon the comparative assessments and evaluations of the introduction writing details outlined above, participants proceed to revise their paper's introduction, aiming to produce an enhanced version.

### 3.3 Measures and data collection

The data for this study were collected in the classroom under the guidance of the researchers. Due to limited classroom time, data collection for Chinese as a foreign language learners took approximately half an hour.

First, before collecting data, the researchers explained the purpose of the activity to the participants and informed them that the data provided would be used for research purposes, seeking their consent.

Second, the researchers explained the specific implementation steps of using ChatGPT in teaching, as shown in the previous section, with the hope that participants who agreed to take part in the research could provide effective data.

Third, participants answered questions based on the provided Word document, providing data. During data collection, if participants had any questions, the researchers provided timely answers. Participants who completed all data entry were allowed to submit immediately, and those who did not finish within the specified time were asked to submit later in the evening.

In the end, data materials were received from 16 participants in the novice-level Chinese class. However, one submission did not meet the requested criteria and was therefore excluded, resulting in a total of 15 valid submissions. For the intermediate-high Chinese class, some students did not provide the content as required, making their data unusable for this study. As a result, data from the intermediate-high Chinese class was largely excluded from the study. Ultimately, two valid data submissions were received from the intermediate-high Chinese class.

In the ChatGPT-Powered Activity for native Chinese speakers, spanning a one-and-a-half-hour classroom session, ten participants submitted their worksheet: *Chinese Academic Writing Tone Practice-ChatGPT's Introductory Paragraph Revisions, My Reflections, and Rewrites*.

These submissions constituted the primary dataset for this section of the study. Subsequently, participants, equipped with insights garnered during the exercise, proceeded with their regular coursework and ultimately submitted their final papers. From each participant's final paper, the researchers extracted the introductory section, juxtaposing it beneath the original introductory draft from the activity worksheet. This comparative analysis facilitated the identification of potential changes and improvements in the participants' writing skills resulting from the activity.

### 3.4 Data scoring and coding

To safeguard participant privacy, pseudonyms are employed for all individuals who provided valid data. For instance, N1 denotes the initial participant from the Novice-level Chinese class, while N15 signifies the final participant. Likewise, IH designates the Intermediate-High level Chinese class, reflecting an intermediate-high level, with IH1 as the code for the initial participant with valid data. In the case of native Chinese participants within the Chinese academic writing class, A is utilized, signifying academic native, and A1 corresponds to the first participant with valid data. A similar pattern is applied to assign codes to other valid participants.

Furthermore, the researchers adopted a mixed-methods approach to analyze the collected valid data. Two researchers assessed the CT demonstrated by participants in the ChatGPT-Powered Activity on a scale ranging from CT-L1 (Remember) to CT-L6 (Create), aligning with Bloom's taxonomy of educational objectives (cited in [Anderson & Krathwohl, 2001](#)). In the realm of qualitative analysis, each

researcher independently scrutinized the gathered valid data, pinpointing instances where participants applied CT during their interactions with ChatGPT. They employed distinct colors to distinguish and categorize responses from CT-L1 (Remember) to CT-L6 (Create). To ensure the precision and dependability of data analysis, upon completing their individual assessments, both researchers conducted a comprehensive cross-validation check of their work, and reached a final agreement, resolving minor differences after discussions. In the context of quantitative analysis, the researchers compiled data regarding the frequency and distribution of varying levels of CT exhibited by participants in their utilization of ChatGPT.

## 4. Results

### 4.1 Research question 1

Which CT skills are utilized in ChatGPT-empowered learning activities for both L2 learners of Chinese and native speakers of Chinese?

To address this question, the researchers conducted a qualitative analysis of all valid data texts below, based on Bloom's revised taxonomy (cited in [Anderson & Krathwohl, 2001](#)), rating them according to CT levels. Quantitative statistics were also employed to assess the frequency of use for each thinking level, the overall frequency, and the percentage of usage. In ChatGPT-supported learning activities, both novice-level and intermediate-high-level L2 learners, as well as native academic class participants, utilized CT skills. However, there were differences in the frequency of use among these thinking levels. Novice-level students primarily employed thinking skills at levels from CT-L1 (Remember) to CT-L5 (Evaluate), whereas intermediate-high-level students and native academic class participants utilized thinking skills across all levels. In terms of thinking level frequency, novice-level students concentrated more on the use of higher-order thinking skills at CT-L4 (Analyze) and CT-L5 (Evaluate), while intermediate-high-level students, although they used a considerable amount of intermediate-high-level thinking skills like CT-L4 (Analyze) and CT-L5 (Evaluate), focused more on lower-order thinking skills, such as CT-L1 (Remember) and CT-L3 (Apply). In contrast to novice-level students, intermediate-high-level students also employed the highest-level L6 (Create) thinking skills. Native academic writing class participants concentrated more on the use of higher-order CT-L5 (Evaluate) and lower-order CT-L1 (Remember) thinking skills. See tables 2, 3 and 4 below.

First, let's take a closer look at the frequency statistics for the novice-level Chinese class regarding the use of thinking skills in ChatGPT activities. As shown in Table 2 below, participants in the novice-level Chinese class employed the first five levels of Bloom's thinking levels, from the lowest level, CT-L1 (Remember) while using ChatGPT as an auxiliary tool for learning Chinese, including remembering, understanding, applying, comparing and analyzing, as well as evaluating and assessing information. Looking at the numerical values for each level, we can see that the 15 participants engaged in higher-order thinking activities. In CT-L4 (Analyze), which involves comparing and analyzing, they had a total of 58 thinking activities. Even more notably, in CT-L5 (Evaluate), which involves evaluation, they had a remarkable 126 thinking activities. In contrast, in the lower-order thinking levels (from CT-L1 Remember to CT-L3 Apply), there were only 20 occurrences for each level. On average, participants thought about 1.3 times per level per person for CT-L1 (Remember) to CT-L3 (Apply), 3.9 times for CT-L4 (Analyze), and as high as 8.4 times for CT-L5 (Evaluate). In total, the 15 participants engaged in thinking 224 times while using ChatGPT for Chinese-assisted learning, with an average of 16.3 times per person. CT-L4 (Analyze) thinking accounted for 23.8% of the total thinking occurrences, while CT-L5 (Evaluate) comprised a high percentage of 51.6%. In contrast, lower-order thinking levels (from CT-L1 Remember to CT-L3 Apply) only accounted for 8.2% of the total occurrences. Participants' thinking

was predominantly concentrated in CT-L4 (Analyze) and CT-L5 (Evaluate); however, the highest level of thinking, CT-L6 (Create), did not occur among novice-level class participants.

Table 2

*Frequency of Use of CT at Various Levels for Novice-Level Participants*

| CT Participants               | L1 Remember | L2 Understand | L3 Apply | L4 Analyze | L5 Evaluate | L6 Create | Total |
|-------------------------------|-------------|---------------|----------|------------|-------------|-----------|-------|
| N1                            | 1           | 2             | 1        | 2          | 11          |           | 17    |
| N2                            |             |               | 2        |            | 6           |           | 8     |
| N3                            |             |               | 2        |            | 7           |           | 9     |
| N4                            |             | 1             |          |            | 5           |           | 6     |
| N5                            | 1           | 3             | 1        | 1          | 7           |           | 13    |
| N6                            | 2           | 3             | 2        | 5          | 11          |           | 23    |
| N7                            | 1           | 3             |          | 10         | 24          |           | 38    |
| N8                            | 2           | 2             | 2        | 4          | 4           |           | 14    |
| N9                            |             | 4             |          | 6          | 11          |           | 21    |
| N10                           | 1           |               | 1        | 5          | 8           |           | 15    |
| N11                           | 2           |               | 2        | 6          | 7           |           | 17    |
| N12                           | 6           |               | 3        | 9          | 6           |           | 24    |
| N13                           | 1           | 1             | 2        | 4          | 8           |           | 16    |
| N14                           | 2           |               |          | 3          | 5           |           | 10    |
| N15                           | 1           | 1             | 2        | 3          | 6           |           | 13    |
| Total of L-frequency          | 20          | 20            | 20       | 58         | 126         | 0         | 244   |
| Average-frequency/person      | 1.3         | 1.3           | 1.3      | 3.9        | 8.4         | 0         | 16.3  |
| Percentage of total frequency | 8.2%        | 8.2%          | 8.2%     | 23.8%      | 51.6%       | 0.0%      | 100%  |

Note: N represents Novice Level; L represents CT levels, with L1 signifying CT Level 1 Remember. This pattern continues for L2 through L6.

Table 3 presents the frequency of cognitive activities performed by intermediate-high-level Chinese class participants using ChatGPT. The data reveals that participants in the intermediate-high-level Chinese class engaged in cognitive activities spanning across almost all cognitive levels when using ChatGPT for expanding topics into paragraphs during discussions. Unlike novice-level participants, intermediate-level participants exhibited a higher frequency of cognitive activities at the lower cognitive levels, specifically CT-L1 (Remember) and CT-L3 (Apply), which accounted for 29.2% and 25% of the total cognitive activities, respectively. Following this were the higher cognitive levels, CT-L5 (Evaluate) and CT-L4 (Analyze), which constituted 16.7% and 12.5% of the total activities, respectively. Among the two valid datasets, one participant utilized the highest cognitive level, CT-L6 (Create), once, accounting for 4.2% of the total activities. The cognitive activities of intermediate-high-level participants primarily centered on the lower levels CT-L1 (Remember) and CT-L3 (Apply), followed by the higher levels, CT-L4 (Analyze) and CT-L5 (Evaluate).

Table 3

*Frequency of Use of CT at Various Levels for Intermediate-high-Level Participants*

| CT Participants               | L1 Remember | L2 Understand | L3 Apply | L4 Analyze | L5 Evaluate | L6 Create | Total |
|-------------------------------|-------------|---------------|----------|------------|-------------|-----------|-------|
| IH1                           | 6           | 1             | 1        | 2          | 1           | 1         | 11    |
| IH2                           | 1           | 2             | 5        | 1          | 3           |           | 12    |
| Total of L-frequency          | 7           | 2             | 6        | 3          | 4           | 1         | 24    |
| Average frequency /person     | 3.5         | 1.5           | 3        | 1.5        | 2           | 0.5       | 12    |
| Percentage of total frequency | 29.2%       | 12.5%         | 25.0%    | 12.5%      | 16.7%       | 4.2%      | 100%  |

Note: IH represents Intermediate-High Level; L represents CT levels, with L1 signifying CT Level 1 Remember. This pattern continues for L2 through L6.

Likewise, the researchers conducted a quantitative analysis to summarize the frequency and percentage of CT levels utilized during the ChatGPT-Powered Activity in the *Chinese for Academic Purposes II* course (see Table 4). The findings revealed that participants engaged in varying degrees of training across six distinct CT levels: remember, understand, apply, analyze, evaluate, and create.

Table 4

*Frequency of Use of CT at Various Levels for Academic Native Level Participants*

| CT Participants               | L1 Remember  | L2 Understand | L3 Apply     | L4 Analyze | L5 Evaluate | L6 Create | Total |
|-------------------------------|--------------|---------------|--------------|------------|-------------|-----------|-------|
| A1                            | 1            | 0             | 2            | 1          | 4           | 0         | 8     |
| A2                            | 4            | 0             | 1            | 4          | 9           | 0         | 18    |
| A3                            | 2            | 3             | 2            | 1          | 5           | 0         | 13    |
| A4                            | 4            | 0             | 1            | 3          | 8           | 3         | 19    |
| A5                            | 8            | 3             | 4            | 6          | 17          | 0         | 38    |
| A6                            | 5            | 0             | 1            | 2          | 6           | 3         | 17    |
| A7                            | 5            | 1             | 6            | 7          | 16          | 0         | 35    |
| A8                            | 3            | 1             | 0            | 4          | 11          | 0         | 19    |
| A9                            | 1            | 1             | 0            | 1          | 11          | 0         | 14    |
| A10                           | 4            | 1             | 0            | 0          | 6           | 2         | 13    |
| Total of L-frequency          | 37           | 10            | 17           | 29         | 93          | 8         | 194   |
| Average-frequency/person      | 3.7          | 1             | 1.7          | 2.9        | 9.3         | 0.8       | 19.4  |
| Percentage of total frequency | L1-L3: 6.4   |               | L4-L5: 12.2  |            |             |           |       |
|                               | 19.1%        | 5.2%          | 8.8%         | 14.9%      | 47.9%       | 4.1%      |       |
|                               | L1-L3: 33.1% |               | L4-L5: 62.8% |            |             |           |       |

Note: A represents Academic Native Level; L represents CT levels, with L1 signifying CT Level 1 Remember. This pattern continues for L2 through L6.

Overall, during the 1.5-hour ChatGPT-Powered Activity, each participant demonstrated an average engagement in CT approximately 19.4 times. This signifies a notably intensive level of CT practicing within the scope of this activity. Examining the frequency and distribution of participants' utilization of different CT levels, the lower-intermediate CT levels (L1-Remember, L2-Understand, and L3-Apply) were observed to be less frequently employed. On average, participants utilized these CT levels 3.7, 1, and 1 time, respectively, comprising 19.1%, 5.2%, and 8.8% of the total instances, collectively contributing to 33.1%. In contrast, intermediate-high CT levels (L4-Analyze and L5-Evaluate) emerged as the predominant modes of thinking in this activity. Participants engaged in these CT levels approximately 2.9 and 9.3 times on average, respectively, constituting 14.9% and 47.9% of the total instances, respectively, cumulatively amounting to 62.8%.

In summary, the data and analysis presented above highlight that the participants engaged in intensive practice of CT across mostly all levels in the ChatGPT-powered activities conducted in various Chinese language courses. Furthermore, the frequency of use of CT at each level displayed considerable variation among different courses, showcasing the diversity of these instructional contexts.

## **4.2 Research question 2**

How do task types in ChatGPT-empowered learning activities relate to the application of CT skills among L2 learners of Chinese and native speakers of Chinese?

In order to address this question, the researchers conducted a statistical analysis of the frequency and distribution of CT levels used in different task types during the ChatGPT activities. Based on the qualitative analysis of data and statistical results regarding the distribution of levels, the researchers concluded that there is a close relationship between the design of ChatGPT activities and the application of students' CT abilities. Task types, including the number of prompting questions in the L2 context, can, to some extent, influence the usage and training of CT skill levels, ultimately affecting the achievement of learning objectives.

First, let's examine the L2 part. To demonstrate the relationship between CT levels and task types more clearly, the researchers categorized the prompts from the instructional documents used in the L2 research activities. Through analysis and summarization, it was evident that the prompts in the activities provided some guidance for participants, aligning with Bloom's taxonomy of thinking skill levels. For instance, when participants encountered problems or had doubts about quiz questions, the instructional prompts guided them to consult ChatGPT. When participants posed questions like 'If you can't or are not sure, you can ask ChatGPT,' the term 'ask' in the sentence directed participants to ask ChatGPT questions, process the information provided by ChatGPT, including memorizing and extracting information, and comprehend new information. Therefore, the term 'ask' carried a certain level of directive instruction. Consequently, questions with directive characteristics like 'ask' were categorized under the category of lower-level thinking, CT-L1 (Remember).

Similarly, in the prompt 'Comparing your answers with ChatGPT's answers, do you agree with ChatGPT?' terms like 'compare' and 'agree' also had directive qualities. When participants encountered the term 'compare,' they were prompted to compare and analyze the existing information with that provided by ChatGPT, falling within the category of higher-order thinking activity, CT-L4 (Analyze). When participants encountered 'agree,' they were guided to assess and judge whether they agreed with ChatGPT's viewpoint and express their own opinions. Here, participants not only employed existing knowledge for comparative analysis but also processed acquired information extensively, involving evaluation and judgment, which falls under the higher-order thinking category, CT- L5 (Evaluate).

The researchers used this method to classify thinking levels for prompts in the L2 activities. The results shown in Table 5 below (See the detailed one with instructional questions in Appendix 1.) indicate

that the prompts in the instructional documents for the novice-level class align with thinking levels of CT-L1 (Remember) to CT-L5 (Evaluate). However, for the intermediate-high level class, the prompts in the instructional documents cover all levels except for CT-L3 (Apply). Tasks related to lower-order thinking in the novice-level class have three prompts, while the intermediate-high level class has as many as eight. Conversely, regarding prompts related to higher-order thinking, the results are reversed. The novice-level class has 12 prompts, including eight for L5 evaluation thinking activities, while the intermediate-high level class has only four. Another distinction is that the intermediate-high level class has a prompt word for the highest-level thinking activity, CT-L6 (Create), a feature absent in the novice-level class.

Table 5

*Number of Assigned Task Questions into CT Levels*

| L2 Groups<br>CT Levels | No. of instructed questions<br>in L2 Novice Level | No. of instructed questions<br>in L2 Intermediate High Level |
|------------------------|---|--|
| Task-L1 Remember       | 3   | 7  |
| Task-L2 Understand     |   | 1  |
| Task-L3 Apply          | 1   | 0  |
| Task-L4 Analyse        | 4   | 2  |
| Task-L5 Evaluate       | 8   | 2  |
| Task-L6 Create         | 0   | 1  |

Note: L represents CT levels, with L1 signifying CT Level 1 Remember. This pattern continues for L2 through L6.

Next, the researchers examined the relationship between the use of CT at various thinking levels and the types and quantity of prompting questions in different task types in the L2 activities. The statistical results for novice-level and intermediate-high-level classes reveal a certain relationship between the usage of CT skills and the task types and quantity of prompting questions in these activities (as shown in Tables 6 and 7).

Now, let's delve into the specific details of the novice-level class regarding the distribution of task instructional prompts among thinking activity levels and the number of prompts. The primary tasks for the novice-level class were to enable learners to use ChatGPT to self-correct questions they found problematic in the quiz, develop skills for autonomous learning, and cultivate CT abilities during interactions with ChatGPT.

The primary task instructional questions for the novice-level class can be categorized into four types, each corresponding to certain levels of thinking skills. The first type Task-L1 constitutes lower-order thinking activities, corresponding to thinking levels CT-L1 (Remember) and CT-L2 (Understand). Participants were required to pose questions, read, memorize, and understand the information provided by ChatGPT. There were three questions of this type in the instructional documents, and the percentage of thinking activities derived from these questions accounted for 1.5% of the total thinking activities.

The second type of task (Task-L3) can be grouped under thinking level CT-L3 (Apply), where participants were expected to apply existing knowledge and skills to solve new situations. For example, participants used the information acquired from ChatGPT to address new situational questions. Only one question in the activity design corresponded to this type. The thinking activities generated by participants based on this question accounted for a mere 0.5% of the total thinking activities.

The third and fourth types of tasks correspond to thinking activity levels CT-L4 (Analyse) and CT-L5 (Evaluate). These tasks (Task-L4 and Task L5) required participants to analyze ChatGPT's

information in conjunction with their existing knowledge, make comparative analyses, and make decisions through information assessment. For instance, they were asked whether they agreed with ChatGPT's viewpoint. These two types of questions played a significant role in the novice-level class activities, with four questions corresponding to CT-L4 (Analyze) thinking activities and a substantial eight questions corresponding to CT-L5 (Evaluate) thinking activities. Under the guidance of these questions, participants produced a rich array of thinking activities, constituting a significant portion. Notably, thinking activities arising from the third type of task (Task-L4-Analyze) were not limited to CT-L4 (Analyze) thinking activities (36 instances, accounting for 17.8% of the total thinking frequency) but also encompassed numerous thinking activities in the higher level of CT-L5 (Evaluate).

The fourth type of task (Task-L5-Evaluate) led to a peak in thinking activities, with participants engaging in thinking activities across all these five CT levels. The thinking activities related to this thinking level (CT-L5 Evaluate) reached a remarkable 76 instances, accounting for 37.6% of the total thinking frequency. As depicted in Table 6, it can be seen that the more instructional questions related to CT levels a task type has, the higher the corresponding thinking frequency might be. For instance, the task type of Task-L5-Evaluate with eight questions achieved the highest corresponding thinking frequency at 37.6% of the total frequency. Following closely, the task type of Task-L4-Analyze with four questions also achieved the second-highest corresponding thinking frequency at 17.8% of the total frequency. Consequently, we can infer a strong relationship between the task design of the novice-level ChatGPT activities and the application of participants' CT abilities.

Table 6

*Task Types Relating to Different Level CT Skills for Novice L2 Chinese learners*

| CT Task Type        | No. of Questions | L1 Remember | L2 Understand | L3 Apply | L4 Analyze | L5 Evaluate | Total |
|---------------------|------------------|-------------|---------------|----------|------------|-------------|-------|
| Task-L1-Remember    | 3                | 1           | 2             |          |            | 1           | 4     |
| Percentage of total |                  | 0.5%        | 1.0%          |          |            |             |       |
| Task-L3- Apply      | 1                |             |               | 1        |            |             | 1     |
| Percentage of total |                  |             |               | 0.5%     |            |             |       |
| Task-L4-Analyse     | 4                |             | 8             | 3        | 36         | 28          |       |
| Percentage of total |                  |             |               |          | 17.8%      | 13.9%       |       |
| Task-L5- Evaluate   | 8                | 8           | 3             | 13       | 22         | 76          | 122   |
| Percentage of total |                  |             |               |          | 10.9%      | 37.6%       | 202   |

Note: L represents CT levels, with L1 signifying CT Level 1 Remember. This pattern continues for L2 through L6.

Now, let's examine the specific situation of the distribution of task instructional questions among thinking activity levels and the number of questions in the intermediate-high-level class. The primary tasks for the intermediate-high-level class were to acquire the ability to select the necessary information with ChatGPT's assistance and develop the capability to expand topics into paragraphs.

The task instructional questions for the intermediate-high-level class corresponded to all thinking levels except for CT-L3 (Apply). Table 7 also reveals the same pattern as observed in the novice-level activities. Types of task instructional questions with greater quantity guided participants to engage in more corresponding thinking activities. Specifically, there are seven instructional questions corresponding to CT-L1 (Remember) in the type of Task-L1-Remember. This type had the highest number of task questions and consequently generated the highest frequency of memory-related thinking,



accounting for 33.3% of the total frequency. Following closely were Task-L4-Analyze) and Task-L5-Evaluate task types, each with two instructional questions, resulting in relatively higher frequencies of thinking activities at their respective levels, both accounting for 12.5% of the total frequency.

Table 7

*Task Types Relating to Different Level CT Skills for Intermediate-high L2 Chinese learners*

| CT Task Type        | No. of Questions | L1 Remember | L2 Understand | L3 Apply | L4 Analyze | L5 Evaluate | L6 Create | Total |
|---------------------|------------------|-------------|---------------|----------|------------|-------------|-----------|-------|
| Task-L1-Remember    | 7                | 8           |               | 4        |            |             |           | 12    |
| Percentage of total |                  | 33.3%       |               | 16.7%    |            |             |           |       |
| Task-L2 Understand  | 1                |             | 2             |          |            |             |           | 2     |
| Percentage of total |                  |             | 8.3%          |          |            |             |           |       |
| Task-L4-Analyse     | 2                |             |               | 2        | 3          | 1           |           | 6     |
| Percentage of total |                  |             |               | 8.3%     | 12.5%      | 4.2%        |           |       |
| Task-L5 Evaluate    | 2                |             |               |          |            | 3           |           | 3     |
| Percentage of total |                  |             |               |          |            | 12.5%       |           |       |
| Task-L6 Create      | 1                |             |               |          |            |             | 1         | 1     |
| Percentage of total |                  |             |               |          |            |             | 4.2%      | 24    |

Note: L represents CT levels, with L1 signifying CT Level 1 Remember. This pattern continues for L2 through L6.

In summary, it can be observed that the task types and quantity of task-related instructional questions in the novice and intermediate-high-level activity instructions are directly proportional to the utilization of thinking activities at their corresponding levels. The more task-related instructional questions there are, the more thinking activities at the corresponding levels tend to be employed. This further addresses the research question, highlighting that in L2 activities, the types of tasks and the quantity of task-related instructional questions utilized with ChatGPT can affect the utilization of thinking level skills.

Next, let's examine the situation in the activities involving native Chinese-speaking students. Within the ChatGPT-empowered learning activities conducted in the *Chinese for Academic Purposes II* course, a notable relationship emerges between task types and the application of CT skills. It should be noted that the quantity of prompting questions was consistent across CT levels in the different task types; thus, the relationship between the quantity of prompting questions and the application of CT skills is not explored in the native Chinese context.

As indicated in the Appendix 2, Chinese Academic Writing Tone Practice- ChatGPT's Introductory Paragraph Revisions, My Reflections, and Rewrites worksheet, the primary objective of this activity was to engage participants in the comparative evaluation and selection of ChatGPT-revised versions of their

initial draft essays. This evaluation encompassed various aspects, including vocabulary usage, sentence structures, punctuation, paragraph structure, and academic writing tone.

These tasks can be categorized into two distinct types. The first type, relatively straightforward in nature, required participants to locate and comprehend their previously acquired knowledge. These tasks are designed as Task 1, focusing on "...Changes in (a specific comparative aspect)," and Task 2, centered on "Rewrite Effectiveness." In contrast, the second type of task was more intricate, demanding participants to evaluate and make decisions about specific elements of their writing within the context of real writing scenarios. This complexity was primarily concentrated within Task 3, the "Your Evaluation" section of the worksheet. (See in Appendix 2)

Upon scrutinizing the collected data (see in Table 8), it becomes apparent that when participants completed the first, more fundamental type of task, they predominantly applied low to intermediate-order CT, namely CT-L1 (Remember), CT-L2 (Understand), and CT-L3 (Apply). The frequency of use for these CT levels was 30, 10, and 15 times, respectively, constituting 30.6%, 10.2%, and 15.3%. The cumulative total of these three CT levels was 56.1%.

Conversely, when tackling the second, more complex type of task, participants demonstrated a substantial inclination toward high-order CT, specifically CT-L4 (Analyze) and CT-L5 (Evaluate). These CT levels were invoked 18 and 61 times, respectively, comprising 20.5% and 69.3% of the instances, resulting in a combined total of 89.8%.

Table 8

*Task Types Relate to Different Level CT Skills for Academic Native Level Participants*

| Task                                | L1<br>Remember | L2<br>Understand | L3<br>Apply | L4<br>Analyze | L5<br>Evaluate | Total |
|-------------------------------------|----------------|------------------|-------------|---------------|----------------|-------|
| Task 1: ... 改变                      | 20             | 6                | 3           | 2             | 1              | 32    |
| Task 2: 改写效果                        | 10             | 4                | 12          | 9             | 31             | 66    |
| Total of Task 1+ Task 2             | 30             | 10               | 15          | 11            | 32             | 98    |
| Percentage                          | 30.6%          | 10.2%            | 15.3%       | 11.2%         | 32.7%          |       |
| Total Percentage of Task 1 + Task 2 | 56.1%          |                  |             | 43.9%         |                |       |
| Task 3: 你的评价                        | 7              | 0                | 2           | 18            | 61             | 88    |
| Percentage of Task 3                | 7.9%           | 0                | 2.3%        | 20.5%         | 69.3%          |       |
|                                     | 10.2%          |                  |             | 89.8%         |                |       |

Note: L represents CT levels, with L1 signifying CT Level 1 Remember. This pattern continues for L2 through L6.

In the following sections, we will provide a detailed exploration of various task types, elucidating and scrutinizing the mechanisms governing the diverse levels of CT demonstrated by participants throughout the task completion process.

Firstly, when participants undertake the first type tasks, they engage in tasks of lower CT levels, specifically CT-L1 (Remember), CT-L2 (Understand), and CT-L3 (Apply). These tasks involve identifying differences between two introductory versions, spanning from broader divisions such as paragraphs and sections down to finer elements like individual characters, punctuation marks, or sentence structures. Subsequently, when describing these differences, participants need to remember (CT-L1 Remember) and apply (CT-L3 Apply) their knowledge of Chinese grammar, classify (CT-L3 Apply) the nature of these changes, and subsequently explain (CT-L4 Analyze) whether these modifications align with established Chinese academic writing standards.

For instance, when completing tasks related to “sentence structure changes” and “rewrite effectiveness,” five participants identified alterations made by ChatGPT in sentence structure, described and defined (CT-L1 Remember) this modification as changing from a question in the original text to a declarative sentence in ChatGPT’s revision. Then, participants recalled (CT-L1 Remember) two key points of knowledge: on the one hand, that question sentences can arouse readers’ interest in academic writing, on the other hand, that academic writing should maintain a formal tone, favoring declarative sentences. As participants realized that different knowledge aspects led to varying judgment criteria, they classified (CT-L3 Apply) the transformation of declarative sentences into questions as aligning with the first point but conflicting with the second. Ultimately, participants predicted (CT-L2 Understand) the implications of this alteration on their introductory writing. Evidently, during the completion of tasks in the first type, CT levels CT-L1 (Remember), CT-L2 (Understand), and CT-L3 (Apply) were the prevailing modes of cognitive engagement.

Following the completion of the first type tasks, participants progressed to the second type of tasks, specifically Task 3. They were tasked with evaluating the differences identified in the introductions of the two versions and providing their own evaluation. In this context, participants primarily invoked two higher-order CTs, namely CT-L4 (Analyze) and CT-L5 (Evaluate). After determining whether ChatGPT’s alterations conformed to general Chinese academic writing standards, participants further contextualized these changes within their own papers, analyzing (CT-L4) and evaluating (CT-L5) the impact of these alterations on the effectiveness of their introductory writing. Additionally, they were expected to explain (CT-L4 Analyze) their reasoning.

Taking the earlier example of sentence structure changes, conventional standards dictate that declarative sentences better align with the required objectivity and formality in academic papers compared to question sentences. However, considering the first key point that question sentences can stimulate reader interest, aiding in the introductory function, participants provided diverse assessments when evaluating ChatGPT’s transformation of question sentences into declarative ones. Three participants deemed ChatGPT’s conversion to declarative sentences as ineffective since, as one participant expressed, “*I intentionally retained the question sentences to underscore my awareness of the issues. Nevertheless, ChatGPT altered all my questions into declarative sentences, leaving me quite speechless.*” Conversely, two participants believed that declarative sentences were more suitable, as they “*made the paper more formal*” (A1) and “*enhanced the objectivity and academic tone of the text*” (A3). Interestingly, the two participants who commended ChatGPT’s use of declarative sentences in their evaluations did not subsequently change their question sentences to declarative ones in their final paper. The researchers conjectured that this discrepancy arose from the fact that their evaluations were grounded in general standards, while their decision to retain question sentences in the final paper was influenced by the contextual nuances of the introductions, demonstrating a flexible application of these criteria to optimize their writing.

In conclusion, when examining the different task types and CT levels used in ChatGPT-Powered Activities across all courses, there is a strong relationship between the two. Moreover, through statistical data and task process analysis, we can gain a clear understanding of the mechanisms and application of different CT levels in various tasks. In the L2 contexts, we observe that the number of prompts within a task also plays a crucial role. Specifically, tasks with a higher quantity of guiding questions in L2 course activities tend to elicit more CT activities at the same proficiency level.

## 5 Discussion

In this discussion section, we will delve deeper into three key issues: CT independent of language proficiency, the relationship between task types and CT application, and the distinctive attributes of ChatGPT as a pedagogical tool in supporting the application of CT skills.

### 5.1 CT independent of language proficiency

According to the principles governing language acquisition and cognitive development, it is commonly presumed that more advanced language learners would engage in a greater number of CT activities or employ higher-order cognitive processes compared to the lower-grade counterparts. Nevertheless, does empirical evidence support this presumption? This study's results suggest otherwise, as they reveal that learners' language proficiency does not necessarily correspond to the application of CT skills.

Leveraging the diversity of participants across three distinct Chinese language courses in the ChatGPT-Powered Activities, this research encompasses individuals spanning a spectrum of Chinese language proficiency, including novice level learners, intermediate-high level learners and academic native level speakers. By summarizing the average frequency and proportion of CT usage among participants at varying Chinese language proficiency levels in Table 9, the researchers seek to uncover their potential relationships between language proficiency and the application of CT skills.

As observed in Table 9, with respect to the aggregate utilization of lower-intermediate CT levels (L1 Remember, L2 Understand, and L3 Apply), beginners demonstrated the lowest frequency of usage at 24.6%, while intermediate-high level learners exhibited the highest frequency at 65.2%, and native speakers fell in between at 32.7%. Conversely, when considering the total deployment of intermediate-high CT levels (CT-L4 Analyze, CT-L5 Evaluate, and CT-L6 Create), beginners with rudimentary Chinese language proficiency displayed the highest frequency of usage at 75.4%, while intermediate-high level learners showed the lowest frequency at 34.7%, and native speakers once again occupied an intermediate position at 67.3%. It seems that beginners, despite their limited Chinese proficiency, predominantly employed high-order CT levels. Intermediate learners, conversely, predominantly engaged in lower-order CTs. Academic native learners, regardless of their heightened language proficiency, exhibited a relatively balanced utilization of CT across all levels. These atypical and non-corresponding data sets imply that there may not be inherent link between participants' Chinese language proficiency and the levels and frequencies of CT usage. In essence, the employment and enhancement of CT do not fluctuate in accordance with participants' Chinese language proficiency. Even at the beginner Chinese proficiency level, various CT levels, including academic native ones, should be integrated into the practicing. Conversely, participants' high language proficiency does not necessarily translate to the use of advanced CT levels.

Table 9

*Frequency of Use of CT at Various Levels for All Participants*

| Language proficiency | L2 learners<br>Novice Level (15 participants) |            |       | Intermediate-High Level (2 participants) |            |       | Academic Native Level (10 participants) |            |       |
|----------------------|---|------------|-------|--|------------|-------|---|------------|-------|
|                      | Average                                       | Percentage |       | Average                                  | Percentage |       | Average                                 | Percentage |       |
| CT                   |   |            |       |  |            |       |   |            |       |
| L1-Remember          | 1.3   | 8.2%       |       | 3.5                                      | 29.2%      |       | 3.7                                     | 19.1%      |       |
| L2-Understand        | 1.3   | 8.2%       | 24.6% | 1.5                                      | 12.5%      | 66.7% | 1                                       | 5.2%       | 33.1% |
| L3-Apply             | 1.3   | 8.2%       |       | 3  | 25%        |       | 1.7                                     | 8.8%       |       |
| L4-Analyze           | 3.9   | 23.8%      |       | 1.5                                      | 12.5%      |       | 2.9                                     | 14.9%      |       |
| L5-Evaluate          | 8.4   | 51.6%      | 75.4% | 2  | 16.7%      | 33.2% | 9.3                                     | 47.9%      | 66.9% |
| L6-Create            | 0   | 0          |       | 0.5                                      | 4.2%       |       | 0.8                                     | 4.1%       |       |

Note: L represents CT levels, with L1 signifying CT Level 1 Remember. This pattern continues for L2 through L6.

Novice Chinese learners, despite being in the initial phases of language acquisition and proficiency development, can effectively engage in CT activities when interacting with ChatGPT. Whether using English or their native language as a medium, they can acquire and comprehend new information, apply existing knowledge to analyze and compare data, assess accuracy, make informed judgments, and successfully address their concerns to accomplish various tasks. In this process, they engage in various levels of CT, particularly at higher levels. This engagement holds significant value. Moreover, the CT activities conducted using other languages can also lay the groundwork for future thinking in Chinese.

## 5.2 The relationship between task types and CT application

From the frequency of CT usage in novice and intermediate-high Chinese classes, it is evident that novice participants used higher-level thinking skills such as CT-L4 (Analyze) and CT-L5 (Evaluate) more frequently compared to intermediate-high participants. Conversely, when it comes to lower-level thinking skills like CT-L1 (Remember) and CT-L2 (Understand), intermediate-high students used them more frequently than novice students.

The main reason for this significant difference can be attributed to the analysis in Section 4.2, which highlights that task types, including the number of task-related prompting questions, seem to affect the utilization of various CT abilities. Even intermediate-high participants, with higher language proficiency, may focus their thinking on lower-level CT skills when provided with numerous lower-level prompting questions. Similarly, novice Chinese learners can engage in higher-level CT activities when assigned many higher-level task-related prompting questions. This insight underscores the importance for educators, when designing activities for students, to carefully select task types and provide meaningful prompting questions, while also considering the intended level of CT training and improvement for students. Of course, the quantity of prompting questions can also play a role when aiming to cultivate specific levels of CT, aligning with Yuan et al.'s (2022) discussion that the choice of teaching strategies is a crucial factor influencing CT instruction from a teacher's perspective.

In the ChatGPT-powered activities conducted with native Chinese speakers, the researcher made the similar observation as in the L2 Chinese classes: the frequency of utilizing various levels of CT abilities seems to be affected by task types.

In the results section, the researchers has already elucidated how task types affect the mechanisms of CT. To elaborate further, the types and frequencies of CT application are influenced by several factors, including task steps, difficulty levels, and types. The core tasks in the ChatGPT activities for the *Chinese for Academic Purposes II* course revolved around two rounds of "comparisons" between different introductory versions. The first step involved a relatively simple comparison, wherein participants were required to locate differences between the two texts and provide brief explanations. This necessitated the use of CT levels from CT-L1 (Remember) to CT-L3 (Apply), encompassing skills like finding, labeling, listing, identifying, describing, recognizing, simple comparisons and classifications. In contrast, the second step entailed a more intricate and challenging comparison. Participants needed to not only compare but also evaluate the adaptability and quality of the two versions within the specific context of academic writing. They had to make optimal choices under conflicting principles. This demanded the utilization of advanced CT levels, including ranking, assessing, choosing, comparing, criticizing, concluding, selecting, supporting, and other high-level abilities. It becomes evident that the characteristics of task types affect the direction and emphasis of participants' CT application.

Through an analysis of the relationship between task types and the application of CT in the ChatGPT activities across the three courses, the researchers posits that task types influence the type and frequency of CT practicing for participants. Various aspects of task content, such as the choice of wording, exemplified by distinct instructional terms like "ask," "identify," "list," "consult," "compare," and "choose" in L2 course activity prompts, guide participants towards different areas of work and thought. Additionally, the number of prompts within a task also plays a role; tasks featuring a higher quantity of guiding questions in L2 course activities tend to elicit more CT activities at the same proficiency level.

Moreover, task design factors such as the structure of task steps, their levels of difficulty, and their specific types exert their influence. For instance, in the context of the *Chinese for Academic Purposes II* course, Task 1 (identifying changes made by ChatGPT to the original introduction) and Task 2 (analyzing the effectiveness of ChatGPT's rewriting) focus primarily on training students in CT-L1 (Remember), CT-L2 (Understand), and CT-L3 (Apply), while Task 3 (assessing modifications made by ChatGPT) emphasizes the aspects of CT-L5 (Evaluate). Variations in the interpretation of 'comparison,' encompassing both sequencing and content, lead to distinct emphases in CT application.

### 5.3 ChatGPT as a pedagogical tool in supporting the application of CT skills

As a pedagogical tool, ChatGPT provides a range of features that cater to the development of CT (Cai, 2023). One foundational aspect lies in its unique peer-like identity, which not only significantly fuels participants' enthusiasm and confidence in CT but also poses challenges to their CT abilities.

On one hand, ChatGPT's non-authoritative nature fosters an environment conducive to the application of CT skills, offering participants a friendly atmosphere and strong motivation. When interacting with ChatGPT, participants do not exhibit the hesitancy and reservations often associated with seeking guidance from teachers. On the other hand, as a recently developed language AI, ChatGPT enables participants to engage with it on an equal footing, promoting a relaxed exchange. Participants approach the information provided by ChatGPT with a sense of skepticism, evaluating its authenticity from a critical standpoint. Unlike the unquestioning acceptance frequently observed in response to teacher feedback, participants are more inclined to critically assess the information offered by ChatGPT. This inclination is particularly evident among advanced-level participants. For example, within the context of the *Chinese for Academic Purposes II* course activities, certain participant comments illustrate this approach:

- Agreement: "The vocabulary is more precise, making the article more persuasive." (A9)
- Partial agreement: "The expansion is very successful, helping me focus my introduction on the specific aspects I want to explore. However, it removed a sentence that included the crucial concept of 'documentary aesthetics,' making the original introduction less complete." (A4)
- Strong disagreement: "Not favorable, intelligent but not useful. From a scientific perspective, the article has not changed. Furthermore, in terms of precision, whether it's the precision of word choice or sentence structure revision, it becomes verbose and does not align with my original intent." (A2)

The above comments reflect participants' peer-based assessment of ChatGPT's revised versions, characterized by objectivity, robust reasoning, and adherence to CT traits. Statistically, among 45 ChatGPT revisions, 63.3% garnered agreement from participants, while 32.2% were met with disagreement. This demonstrates that participants did not unquestioningly embrace ChatGPT's modifications but critically assessed and made informed decisions. This underscores the role of task design in ChatGPT activities, guiding the direction of CT skill application and ensuring the attainment of course objectives.

Conversely, ChatGPT's peer-like identity and the relatively immature nature of its Chinese language proficiency pose challenges for participants striving for profound CT. Specifically, participants, particularly those with lower Chinese language proficiency levels, may need to delve more profoundly and pose more insightful questions to ChatGPT. This is because they lack the ability to discern the accuracy of ChatGPT's responses. Inadequate inquiry may result in the risk of being misinformed or obtaining incomplete information.

Consequently, they needed the ability to ask ChatGPT follow-up questions to obtain desired answers. Without a conscious effort to engage in further analysis and continuous questioning, participants risked being misled by incorrect information provided by ChatGPT, thus increasing the chances of cognitive errors. Incomplete information from ChatGPT could prevent them from discovering the true solution to a problem. See an example from N4 in the novice-level Chinese class.

Problem to be solved: Question 14 (N4)

**Your question on ChatGPT:** Is this the correct way to say “Please say that again?”. 你说一次

**ChatGPT’s answer:** The sentence “你说一次” can be translated to “Please say it once” or “Say it once, please”, but it is not the most natural or polite way to ask someone to repeat what they said. A more polite and natural way to say “Please say that again” in Chinese is “请再说一遍” (qǐng zài shuō yī biàn). This sentence uses the polite word “请” (please) to make the request more polite and respectful, and “再说一遍” (say it again) to indicate the request to repeat what was said.

What do you think? Do you agree with it? Where is your problem (grammar, typo, pronunciation, punctuation or expression)? Did you revise as the ChatGPT pointed out?

**Your idea:** This uses characters that were not in the list of characters I could use so it did not help.

From the responses of the N4, it is clear that ChatGPT provided valuable information. However, due to the participant’s familiarity with the Chinese word for “次” (measure word for "times") but not “遍” (measure word for "iterations"), he abandoned the opportunity to obtain the final answer to resolve his doubts. If the participant had exercised further cognitive awareness and posed follow-up questions to ChatGPT, such as, "Can you say '请再说一次' (Can you please say it again)?" or "Can I say '请你再说一次' (Can you please say it again)?" he would likely have received additional explanations from ChatGPT. This not only would have helped N4 obtain the desired answer and resolve his doubts in Question 14 but also allowed him to learn new vocabulary and grammar, specifically the usage of the word “遍.” This would also have further enhanced his Chinese language learning. If the participant has a mindset for further reflection and inquiry, he should not judge ChatGPT as unhelpful just because he did not get an answer for one word he was not familiar with.

From this analysis, it is evident that as language educators and facilitators introducing AI technology to assist language learning, we should not only guide students on how to use AI technology for assistance but also nurture their ability to think critically and ask further questions when faced with uncertain information provided by ChatGPT. This approach can prevent them from overlooking precise answers that could have been obtained.

In summary, when introduced as a pedagogical tool, ChatGPT can effectively support the application of CT skills if used correctly. Due to ChatGPT’s limitations in Chinese expression, teachers should emphasize and cultivate participants’ reflective awareness when introducing ChatGPT activities in the classroom. Participants should not blindly trust the tool but rather focusing on enhancing their questioning skills. This approach allows us to maximize its benefits and minimize its limitations, thereby enabling ChatGPT to better assist in Chinese language learning.

## 6 Conclusions

In summary, this study delved into the impact of ChatGPT-powered activities on CT skill application across a spectrum of Chinese language courses, addressing two primary research questions. It explored the application of CT skills across varying language proficiency levels and examined the relationship between task types and the enhancement of CT abilities.

Our findings unequivocally suggest that CT skills were actively harnessed by participants, regardless of their language proficiency levels, showcasing the potential of ChatGPT-powered activities to nurture CT, both among L2 learners and native speakers of Chinese. Furthermore, task types affect CT application.

In terms of pedagogical implications, this research underscores the efficacy of ChatGPT-powered activities in cultivating CT within Chinese language classrooms. It accentuates the pivotal role of task design in shaping and honing CT skills, emphasizing that language proficiency levels do not dictate one’s

CT capacity. Educators and curriculum designers should consider incorporating diverse tasks to foster CT skills among language learners. As AI-driven tools continue to evolve, this study opens avenues for further exploration of AI's contribution to cognitive development within educational settings.

However, the current study has certain limitations, particularly due to the relatively small sample size for data collection. We hope that future research can address this by increasing the volume of effective data. Additionally, we anticipate future studies to delve deeper into the role of ChatGPT in fostering the development of critical thinking skills. This includes exploring internal development within groups and variations across different groups. We also hope that future research will examine the relationship between the use of CT at various levels and the quantity of prompting questions in different task types within the native Chinese context.

## Appendix 1

|                       | L2 Novice Level<br>Questions   | No. | L2 Intermediate High Level<br>Questions   | No. |
|-----------------------|--|-----|---|-----|
| Task-L1<br>Remember   | You can ask the ChatGPT<br>What are your questions?<br>What does ChatGPT reply?  | 3   | You can ask the ChatGPT like ...<br>You can ask the ChatGPT for one<br>reason ...<br>You can ask it in either Chinese or<br>English.<br>It's best to ask the ChatGPT to<br>answer ...<br>Please describe how you used the<br>ChatGPT.<br>What questions did you ask?<br>What information did you get? | 7   |
| Task-L2<br>Understand |  |     | Please try to think of the reasons on<br>your own.  | 1   |
| Task-L3<br>Apply      | Will you use ChatGPT and why?  | 1   |   |     |
| Task-L4<br>Analyse    | Comparing your answers with<br>ChatGPT's answers.<br>What are the differences between<br>your answers and ChatGPT's?<br>What do you think?<br>Where is your problem?   | 4   | What do you think of it?<br>Whether do you think it can help?   | 2   |
| Task-L5<br>Evaluate   | Do you agree with the ChatGPT?<br>If you think ChatGPT is right,...?<br>Where exactly did you go wrong?<br>Do you agree with it? Why?<br>Where is your problem?<br>Did you revise as the ChatGPT<br>pointed out?<br>...and explain whether ChatGPT can<br>help you?<br>Will you use ChatGPT and why? | 8   | Whether you have got your own<br>essay ...<br>Your comments on using it.  | 2   |
| Task-L6<br>Create     |  | 0   | You can form your own essay ...   | 1   |



## Appendix 2

### Chinese Academic Writing Tone Practice

#### - ChatGPT's Introductory Paragraph Revisions, My Reflections, and Rewrites

|  | Your Academic Writing | ChatGPT Revised Academic Writing |
|--|-----------------------|----------------------------------|
| Introduction   |                       |                                  |
| Changes in Vocabulary  |                       |                                  |
| Rewrite effectiveness  |                       |                                  |
| Your Evaluation  |                       |                                  |
| Changes in Sentence Structure  |                       |                                  |
| Rewrite effectiveness  |                       |                                  |
| Your Evaluation  |                       |                                  |
| Changes in Punctuation   |                       |                                  |
| Rewrite effectiveness  |                       |                                  |
| Your Evaluation  |                       |                                  |
| Changes in Paragraph Organization  |                       |                                  |
| Rewrite effectiveness  |                       |                                  |
| Your Evaluation  |                       |                                  |
| Changes in writing tone  |                       |                                  |
| Rewrite effectiveness  |                       |                                  |
| Your Evaluation  |                       |                                  |
| Other Changes you found  |                       |                                  |
| Rewrite effectiveness  |                       |                                  |
| Your Evaluation  |                       |                                  |
| Comparative Criteria for Rewriting: Evaluation Standards for Chinese Academic Writing                            |                       |                                  |
| · Scientificness   |                       |                                  |
| · Precision  |                       |                                  |
| · Logicality   |                       |                                  |
| · Standardization  |                       |                                  |
| · Specialization   |                       |                                  |
| · Objectivity  |                       |                                  |
| · Formality  |                       |                                  |
| · Creativity   |                       |                                  |
| Your Overall Evaluation (Did you find ChatGPT's revisions successful? Ratio of success to failure?)              |                       |                                  |
| Your Summary of Experience (In your academic paper writing, where do you need improvement? How can you improve?) |                       |                                  |

## References

- Anderson, L. W. & Krathwohl, D. R. (2001). *A Taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Burns, A. (2011). Action research in the field of second language teaching and learning. In E. Hinkel (Ed.). *Handbook of Research in Second Language Teaching and Learning* (pp. 237-253). Routledge.
- Cai, W. (2023). ChatGPT 环境下的汉语学习与教学 [Chinese learning and teaching in the ChatGPT environment]. *语言教学与研究 [Language Teaching and Research]*, 222(04), 13-23.222(04), 13-23.
- Fisher, D. (2007). *Instructional design: The taxonomy table*. Oregon State University. Retrieved from: <http://oregonstate.edu/instruct/coursedev/models/id/taxonomy/#table>
- Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, 93: 279–289. <https://doi.org/10.1016/j.chb.2018.12.023S>
- Halpern, D. F. (1993). Assessing the effectiveness of critical-thinking instruction. *The Journal of General Education*, 42(4), 238–254.
- Hubbard, P. (2022). Bridging the gap between theory and practice: technology and teacher Education. In by N. Ziegler & M. González-Lloret (Eds.). *The Routledge Handbook of Second Language Acquisition and Technology* (pp. 21-35). Routledge. <https://doi.org/10.4324/9781351117586>
- Kemmis, S., & McTaggart, R. (2005). Participatory action research: Communicative action and the public sphere. In N. Denzin & Y. Lincoln (Eds.). *Handbook of Qualitative Research* (pp. 559–604). Sage.
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1): 973–1018. <https://doi.org/10.1007/s10639-022-11177-3>
- Luo, H, & Gui, M. (2019). Developing an effective Chinese-American telecollaborative Language learning program: An action research study. *Computer Assisted Language Learning* 34 (5-6): 609-636. <https://doi.org/10.1080/09588221.2019.1633355>
- National Council for Excellence in Critical Thinking. “*Defining Critical Thinking*”. Retrieved from: <https://www.criticalthinking.org/pages/defining-critical-thinking/766>
- Voogt, J. & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, 44 (3): 299-321. <https://doi.org/10.1080/00220272.2012.668938>
- Yuan, R., Liao, W., Wang, Z., Kong, J., & Zhang, Y. (2022). How do English-as-a-foreign-language (EFL) teachers perceive and engage with critical thinking: A systematic review from 2010 to 2020. *Thinking Skills and Creativity*, 43: 1-15. <https://doi.org/10.1016/j.tsc.2022.101002>
- Zhou, H. (2022). Developing critical thinking skills in Russian language studies: Online learning tools in Chinese universities. *Foreign Language Annals*, 55: 98–115. <https://doi.org/10.1111/flan.12583>

**Xinliang Jiang** is Sessional Lecturer in the Department of Language Studies at the University of Toronto Mississauga. His main research interests include Chinese language acquisition and education, second language acquisition, phonetics and phonology, and Chinese dialects.

**Jing Li** is Sessional Lecturer in the Department of Language Studies at the University of Toronto Mississauga. Her research interests focus on Chinese literature, diaspora culture, and Chinese for academic purposes.

**Chiu-Hung Chen** is Associate Professor in the Department of Language Studies at University of Toronto Mississauga. Her primary research interests include Chinese as a foreign language: pedagogy, assessment and literacy development, second language pedagogy and computer-assisted language learning.

# 通过 ChatGPT 支持的中文课堂活动提高批判性思维技能的研究

蒋新亮 \*

李晶

陈秋宏

多伦多大学密西沙加分校, 加拿大

## 摘要

ChatGPT 是一款人工智能 (AI) 聊天机器人, 它使用基于 2018 年 OpenAI 推出的 GPT 架构的语言模型, 并可模拟人类对话。在语言课堂活动中, 研究者发现语言学习者在与 ChatGPT 互动时采用了多种策略。这些策略涵盖了布鲁姆认知目标分类的不同层次, 既包括较低层次的思维 (如记忆、理解、应用), 也包括较高层次的思维 (如分析、评价、创造)。本文采用布鲁姆修订后的认知目标分类 (见于 Anderson 和 Krathwohl, 2001 年) 作为批判性思维 (CT) 类型的框架, 旨在调查在三门汉语课程中, ChatGPT 支持的学习活动对学生批判性思维的应用的影响。这三门课程包括面向第二语言 (L2) 学习者的入门级汉语课程和中级汉语课程, 以及面向中文母语者的学术写作汉语课程。本文同时还研究了任务性质对学习者的批判性思维运用的影响。研究包括以下两个主要问题: 1) 在 ChatGPT 支持的学习活动中, 中文二语学习者和中文母语者使用了哪些批判性思维技能? 2) 在 ChatGPT 支持的学习活动中, 任务类型与中文二语学习者和中文母语者的批判性思维技能的运用具有怎样的联系? 这项研究采用了行动研究框架, 其特点是一个涉及规划、执行、观察和反思的循环过程 (Kemmis 和 McTaggart, 2005 年, 引自 Burns, 2011 年)。本研究通过对定性数据的分析, 评估了中文学习者批判性思维技能的运用情况。研究结果显示, 在三门语言课程中, 学习者对记忆、理解、应用、分析、评估和创造这六种类型的批判性思维都有一定的应用, 且运用情况不受语言熟练程度水平的影响。其次, 在 ChatGPT 支持的学习活动中, 任务类型与学生批判性思维的运用情况密切相关, 任务类型对不同层次批判性思维技能的培养和运用起到了关键作用。总之, ChatGPT 支持的学习活动及不同的任务类型对学习者的学习目标的实现有着显著的影响。

## 关键词

批判性思维 (CT), ChatGPT, 对外汉语, 计算机辅助语言学

蒋新亮是多伦多大学密西沙加分校语言研究系的兼职讲师。他的主要研究兴趣包括: 汉语习得与教学、二语习得、语音学和音系学, 以及中国方言。

李晶是多伦多大学密西沙加分校语言研究系的兼职讲师。她的研究兴趣集中在中国文学、侨民文化和学术汉语。

陈秋宏是多伦多大学密西沙加分校语言研究系副教授。她的主要研究兴趣包括对外汉语: 教学法、评估和读写能力发展, 二语教学法和计算机辅助语言学习。