Journal of Technology and Chinese Language Teaching

Volume 16 Number 1, June 2025
二〇二五年六月　第十六卷第一期

ISSN: 1949-260X
http://www.tclt.us/journal

科技与中文教学

Journal of Technology and Chinese Language Teaching

A peer-reviewed online publication with in-print supplement
ISSN: 1949-260X http://www.tclt.us/journal

Volume 16 Number 1, June 2025

Managing editor for this issue: Jun Da

Articles

科技与中文教学

Journal of Technology and Chinese Language Teaching

A peer-reviewed online publication with in-print supplement
ISSN: 1949-260X http://www.tclt.us/journal

Sponsor

Department of World Languages, Literatures, and Cultures, Middle Tennessee State University

Editorial board

Editorial staff

Contacts

URL: http://www.tclt.us/journal
Email: editor@tclt.us

An Explorative Case Study of the Integration of a Fully Immersive VR Game Into the Chinese Classroom (融入沉浸式虛擬實境遊戲於中文課程之探索性個案研究)

Cheng, Hsiu-Jen
(鄭琇仁)
National Kaohsiung Normal University
(國立高雄師範大學)
t3645@mail.nknu.edu.tw

Zhan, Hong
(戰紅)
Embry-Riddle Aeronautical University
(安伯瑞德航空大學)
zhan121@erau.edu

Abstract: This case study aims to integrate a story-driven virtual reality (VR) game into a Chinese lesson to investigate the dynamic interactions between teacher facilitation, student engagement, and language learning within a VR environment. Based on the theory of affordance, three VR learning tasks were designed to engage the student with a virtual environment to learn Chinese. Data was analyzed using descriptive analysis and chi-square tests. The findings reveal that while technical assistance played a crucial role in the initial VR game learning stages, questioning strategies and story-driven facilitation techniques proved to be more effective in fostering meaningful language learning. The study highlights the importance of balancing technical guidance with immersive, content-driven questioning to maximize student engagement and language acquisition in VR-based instruction.

摘要: 此個案研究旨在將一款以故事導向的虛擬實境遊戲(VR)融入中文課程並探究此學習環境下教師輔助、學生參與、語言學習間的互動關係。依據賦能理論，研究者設計了三項沉浸式虛擬實境的學習任務，以促進學生在虛擬環境的中文學習。資料分析採用敘述性統計以及卡方檢定，顯示在 VR 遊戲學習的初期階段，教師的技術協助扮演重要角色，而教師的提問策略以及故事內容相關的教學指引更能有效促使有意義的語言學習。研究結果強調，在 VR 融入中文教學中，需在教師技術指導與內容導向的提問策略之間取得平衡，使學習者參與以及語言習得的效果最大化。

Keywords: Virtual reality, story driven game, Chinese learning, affordance theory, case study

關鍵詞: 虛擬實境，故事導向遊戲，中文學習，賦能理論, 個案研究

1. Introduction

Virtual reality (VR) is often referred to as a three-dimensional digital environment in which users can immerse themselves within and interact with the environment. VR has been increasingly applied in foreign language education in recent years. Through different VR apps and games, language learners can practice vocabulary, speaking, reading, and writing in immersive and interactive environments (Alfadil, 2020; Huang et al., 2021; Tai et al., 2020). Fully immersive VR games with wearable devices can create more immersive learning experiences through simulations (Jensen & Konradsen, 2018; Alfadil, 2020). Through interacting with objects offered in the VR game environment, learners can perceive possibilities for doing something with the target language (Aronin, 2014; Kordt, 2018). This practice is well supported by affordance theory which values human-computer and human-environment interactions. In this theory, affordance is not just a physical property but also a relational concept—it depends on both the object's characteristics and the individual's abilities to interact with it.

In the VR game-based learning process, a teacher's facilitation can encourage learners, guide them in identifying affordance, and help them take actions for learning. With the support of well-designed VR learning procedures, learners can engage in the learning process more effectively, cognitively, and behaviorally (Kordt, 2018; Jong, 2023).

Although the number of studies on digital game-based learning of second languages has increased (Dixon et al., 2022), many studies mainly focus on learning English as a second or foreign language (ESL or EFL). Learning Mandarin Chinese as a foreign language (CFL) through fully immersive VR games still needs more investigation. This research aims to provide a picture of how a Chinese teacher facilitated a single student's understanding and completion of an authentic story-driven game, and how the student engaged and performed in the fully immersive VR game-based learning process.

2. Literature Review

2.1 Theoretical framework – Affordance Theory in VR game-based learning design

Affordance theory was introduced by American psychologist James J. Gibson in his article "The Theory of Affordance" (Gibson, 1986). The original focus of affordance theory values animal-environment interaction, that is, the understanding of how animals perceive and interact with their surroundings based on opportunities or affordance that the environment provides. Gibson's theory suggests that the environment offers various affordances or action possibilities that are directly perceivable by the observer, without the need for complex cognitive processing. These affordances guide behaviors and facilitate interactions, shaping the way individuals navigate and engage with their environment.

Educators have extended his classic concept of animal-environment relations to human-computer and human-environment interactions (Wang et al., 2018; Lee et al., 2018). With this developed definition, affordance theory expands its observation to what physical

or virtual environment offers or affords, and what actions humans can take when interacting with the objects provided in the environment.

Affordance theory has been applied in many disciplines as a theoretical and instrumental guideline for learning design to increase users' interactions and engagement with learning environments. Instructional designers often apply affordance theory in VR game-based learning to create affordance for possibilities of action and interaction. To increase human-computer or human-digital environment interaction, instructional designers must consider cognitive, physical, sensory, and functional affordance in design (Hartson, 2003). These affordances are design features that help users to know something, do something, sense something, and accomplish the task (functional affordance). Hartson (2003) further explains the four affordance types using the example of clicking a button labeled "SORT". Once a user sees the SORT button, its text label tells the user what will happen if he clicks on it (cognitive affordance). The button is large, so the user can click on it accurately (physical affordance). In addition, the label font size is large enough for the user to read it easily (sensory affordance). Lastly, the user can get the information sorted (functional) by clicking on the SORT button.

Affordance theory has also been adopted to support multilingual education (Aronin, 2014; Kordt, 2018). Affordance theory explains how the objects perceived by individual learners and their interaction with the learning environment create many unique opportunities for language and cultural learning. Sun (2016) also pointed out that a language teacher should understand the affordance of the teaching environment as well as learners' perceptions toward it to ensure language learners efficiently interact with the language environment and to provide more opportunities for language input and output. The affordance theory guides the design for interaction, usability, and learning experience, but when using VR games in learning, how teachers facilitate students in the VR-based game to impact learning is worth discussing.

2.2 VR games in teaching and learning of foreign languages

VR has gained increasing attention in foreign language education because of its immersive and interactive features, which are essential factors contributing to the success of foreign language learning (Peixoto et al., 2021; Parmaxi, 2020). Language learners can apply their language skills when experiencing real-life scenes or tasks in the simulated environment, interacting with objects afforded in the environment, or communicating with other people as they do in real life.

Based on the level of immersion, VR is generally classified into three types: non-immersive, semi-immersive, and fully immersive (Villena-Taranilla et al., 2022; Peixioto et al., 2021). Non-immersive VR is the lowest type of immersion, and users can only view and interact with a virtual world through the screen of a digital device. Semi-immersive VR gives users the feeling of being slightly or partially immersed in experiencing a virtual environment. Fully immersive VR allows users to experience and interact in the virtual world within the environment itself as if it were real and authentic.

Each type of VR has valuable and unique merits for the development of foreign languages, but recent reviews have reported that fully immersed VR is more efficient for learning (Alfadil, 2020; Villena-Taranilla et al., 2022; Parmaxi, 2020). Through head-mounted displays, headphones, and gloves, users can experience the virtual world and perform tasks as if they were real using the target language. However, current research findings are mainly based on learning English as a foreign language and integrating fully immersive VR into uncommonly taught foreign languages, like Mandarin Chinese, requires further exploration. In addition, current studies mainly observe the learning of specific skills (for example, speaking and vocabulary, Parmaxi, 2020), user perceptions, satisfaction, and motivation (Peixoto et al., 2021); while other essential elements of learning interactions and teacher facilitation in the VR experience have remained under-investigated.

2.3 Facilitation in VR game-based learning

The affordances mentioned above in VR game-based learning environment design may be perceived differently based on a person's attributes, interests, attention, and level of context awareness. Researchers argue that teachers must first perceive and recognize the affordances and then develop pedagogical strategies to enhance learning opportunities to help students complete game tasks (Sun, 2016). Teachers' mediation becomes more important when experiencing VR games that are not specially designed for subject-based teaching and learning (Kordt, 2018; Poole et al., 2022).

However, the existing research investigating teacher facilitation in VR game-based learning is still scarce (Jong et al., 2017; Kordt, 2018; Poole et al., 2022). Among a few identified studies relevant to teachers' facilitation in game-based learning, Jong et al. (2017) proposed a teacher-supported pedagogical framework that specified three sequential phases to provide facilitation and scaffolding before, during, and after experiencing VR games. In Phase 1, the teacher scaffolds students with initial abstract knowledge related to the subject matter covered in the game. In Phase 2, the teacher engages students in gaming and reflection upon their learning in the virtual world. In Phase 3, the teacher debriefs students during and after the gaming process, either providing encouragement or discussing gaming strategies and game learning experiences.

In a more current study conducted by Poole et al. (2022) on the affordance and effectiveness of a digital game in the dual Chinese language immersion classroom, seven types of supports (which were named pedagogical affordance in the study) were identified during digital gameplay: quest management (direct learners towards to the next quest), combat strategy discussion (discuss strategies related to winning a battle), meaningful communication (discuss current status/object in the game), encouragement (encourage a player to explore the game or read a text on their own), technology support (show a player how to play the game or help with a technical problem related to the game), confirmations (confirm a question or belief about the game), and linguistic support (help learners read a text or provide a translation to a word). The learning procedures and types of support can guide teachers through a VR game-based learning process, but how a teacher can facilitate student learning of Chinese in a non-combat-, narrative-based, and decision-making VR game in the classroom needs further exploration.

2.4 Engagement and performance in VR game-based learning

Teachers' facilitation, as discussed above, helps students in the learning process, but student engagement during the process directly leads to their learning performance. The importance of engagement in learning can never be over-emphasized. As Perry and Booth (2021) stated, no learning occurs without student engagement. Systematic reviews of the literature on gamified learning in higher education revealed that gamification and game-based learning have a positive effect on student engagement in many disciplines (Subhash & Cudney, 2018), and digital, narrative, game-based learning has positive effects on engagement, motivation, and learning (Breien & Wasson, 2020).

Learning engagement often refers to the ability to cognitively, affectively, and behaviorally be involved in the learning process (Dubovi, 2022). Following these three dimensions of engagement, Xin (2022) found that cognitive engagement and emotional engagement in VR game learning environments contributed to better learning results. Dubovi (2022) found that engagement, particularly during the active VR procedural learning phase, is associated with positive emotions and increased mental effort (cognition). These research findings are reported mainly from English or science education studies. It is still unclear in the literature how teachers guide learners in completing VR game-based tasks, how learners of the Chinese language engage in the learning process, and how their engagement connects to their learning performance.

3. Purpose and Research Questions

The purpose of the study was to investigate how to integrate a story-driven VR game into a Chinese language class. Through the observation of a teacher's facilitation of a single student completing VR game tasks, this study particularly aimed to answer the following two questions.

1. How could the teacher facilitate the completion of an authentic story-driven VR game?
2. What is the relationship between the teacher's facilitation and the student's performance?

4. Methodology

4.1 Design of the study

A single case study is defined by Stake (1995) as the study of the particularity and complexity of a single case (an individual, group, organization, or event), coming to understand its activity within important circumstances. With a pre-selected case or subject, a case study in the natural context can capture the uniqueness and specific characteristics of the case through in-depth observations and analysis. Accordingly, this study adopted a single-case design to explore uniqueness regarding how to integrate a story-driven game into an intermediate-level Chinese class. This class was an independent learning course,

not associated with any formal Chinese language curriculum. This design enables teachers and researchers interested in the use of VR games in language instruction to understand the structure of a VR game lesson, how a VR game can be carried out during class, and what type of insight can be gained from VR game-based learning.

4.2 Participants

The exploratory single case investigated in this study was an intermediate Chinese language student, Mike (pseudonym), enrolled in an intermediate Chinese language course during a 4-week summer study abroad program in Taiwan. Mike, a 19-year-old American college freshman, had completed 12 credits of Chinese courses prior to the study abroad program. Mike's oral language proficiency was rated intermediate low based on the Oral Proficiency Interview (OPI)[1] guidelines. During the study abroad program, Mike attended 4 hours of Chinese instruction daily and was the sole student in the class.

Jean (pseudonym), a professor with 15 years of experience teaching Chinese as a foreign language (CFL) and training CFL teachers, served as Mike's VR game-learning instructor. Jean is very experienced in using technology in teaching Mandarin Chinese. During the program, Jean offered three 1-hour sessions of VR game-based class to Mike outside of his regular class time.

4.3 The VR-game–*The Price of Freedom*

The Price of Freedom, a single-player room-scale VR game, was selected for this research. Different from other VR games which feature combat, this game requires users to follow narrative information to make a decision. This game was released by Construct Studio in 2016 and is based on a real story that occurred when the United States was deeply involved in the Cold War with Russia. The player takes the role of CIA Agent Zero, and the mission is to find and assassinate Benjamin Miller to prevent the classified chemical weapons research valuable to nuclear war he has stolen from falling into enemy hands.

The selection of this game was predicated on two principal considerations. Firstly, from a language acquisition standpoint, the material is intended for native speakers, and this study seeks to utilize authentic virtual reality (VR) learning resources for learners of Chinese as a Second Language (CSL). The linguistic complexity of the game was not consistently adapted for specific CSL learners. According to the vocabulary outlined for the learning stages in the study's 4.4.1 session, it is more appropriately suited for students at an intermediate level of Chinese proficiency. Secondly, from a digital game design perspective, the game is supported in both English and Mandarin Chinese in terms of the texts and audio voices. The game is well designed with many authentic documents, historical graphics, theatrical narration, and interactive objects in the game scenes. The game storyline proceeds through entering three rooms in sequence after completing tasks in each room.

[1] https://www.actfl.org/assessments/postsecondary-assessments/opi

The first room is the target Benjamin's office, where Agent Zero killed Benjamin and receives a command to look for a stolen documents that Benjamin hided. The second room is a chamber of secrets full of newspapers posted on the wall and documents stored in drawers. Agent Zero follows the built-in audio commentary, reads the documents and news, finds secret codes with the flashlight, and decodes clues to locate the security box where the important information is located, and the user will learn the truth behind Benjamin's motives. The third room is the ward of a mental hospital to which the players return. There the players will find an unexpected twist. Figure 1 illustrates the game space.

Figure 1 Sample game spaces for *The Price of Freedom*

4.4 Procedures

4.4.1 Instructional stages for each lesson

Based on the development of the VR game storyline and changes in the scenes in three rooms, three one-hour lessons were designed for in this case study for three consecutive days. Each of the three lessons proceeded through four learning stages: (1) vocabulary learning, (2) image-based guidance, (3) virtual reality gaming interactions, and (4) story-telling.

The following section outlines the design of the learning tasks for each stage, with a particular emphasis on Stage 3, where the student was fully immersed in the VR game, which is also the focus of the study.

Stage One was vocabulary learning. Vocabulary was selected based on the story, resulting in an unequal number of vocabulary words across the three lessons (see Table 1 on next page). In this stage, the instructor first presented new words related to the game story to the student. As the pretest, the student was asked to tell the meaning of the words. Then, the instructor explained the new words to help the student understand the vocabulary by using flash cards / word cards to enhance the student's comprehension and retention.

Through explicit vocabulary learning, the student was prepared to interact with the VR game.

Table 1 Descriptions of three tasks

Tasks	Learning Objectives	Task Complicity	Amount of Vocab
Task 1	1. Collecting the personal background information of the victim	Name, passport	23
Task 2	1. Examining the documents in drawer 2. Examining the documents on walls (medical records, CIA documents, news, special symbols) 3. Finding the location of the security box	Recap details Medical records, CIA documents, news, special symbols document format	34
Task 3	1. Finding the password 2. Decoding the symbol 3. Finding out the truth	Numbers. Interpretation of data/graph	18

Stage Two provided image-based guidance for the VR game story. Using the screenshot images captured from the VR game, the teacher helped the student describe the story of the VR game. The learning tasks in this stage served two purposes: to provide a learning context for repeated use of the vocabulary learned in Stage One, and to preview the VR game that the student would experience in the next stage.

Stage Three was set aside for VR game interactions, which lasted for approximately 10-13 minutes in each session. Wearing an HTC Vive mounted headset, the student played the game following the instructor's audial guidance. The student walked into one room in each lesson: lesson one was in Benjamin's office (Room 1), lesson two was in the secret room (Room 2), and lesson three was in the ward of the mental hospital (Room 3). In each room, the student player searched for the document, looked for clues through signs and documents, and asked / answered questions related to the missions.

Stage Four required the student to take off the headset and meet with the instructor. Using visual aids comprised of screenshot images, the student retold the VR game story that they just experienced. The participant's oral report was considered to be the learning assessment and their vocabulary use served as the post-test of vocabulary learning.

4.4.2 The design of VR game-based learning activity based on Affordance Theory

The three tasks designed for the VR game environment were based on the Hartson's four types of affordances: function affordance to accomplish tasks, physical affordance to trigger physical reactions, cognitive affordance to comprehend and acquire information, and sensory affordance to help learners feel something (Hartson, 2003). During each lesson,

the student needs to complete one task. Specific learning tasks and each type of affordance in the VR game environment are described in the following section.

Task 1 During Lesson 1

The student explored Room 1 (Benjamin's office) to collect Benjamin's personal background information and kill Benjamin. The following affordances embedded in Room 1 of the VR game were identified to help the student complete Lesson 1.

Functional affordance: A clickable button in the elevator, an audio commentary, a key for opening the door, a triggerable gun, documents and a passport could be moved. The instructor provided audial guidance. These functional affordances induce the learner's physical behavior.

Physical affordance: A clickable button with lights in the elevator was available for the learner to click and documents were movable. VR hands were able to move items such as documents. The small room limited the area for exploration, which was designed to help focus the learner's attention on specific tasks or interactions within the room. It also allowed for a detailed and dense environment where everything can be explored.

Cognitive affordance: Information was provided by documents written in Chinese as well as by audio commentary in Chinese, which helped learners to search for information or listen to commentary. The instructor's audial guidance provided extra assistance in navigating tasks and the environment, ensuring that the learner stayed on track. This game design aims to help players truly understand Benjamin.

Sensory affordance: Sound effects were triggered for when doors opened, elevators ran, guns were used, and documents were turned so that the player felt immersed and realistic. When the player heard familiar sounds, such as a door opening or an elevator bell, it helped create a sense of presence in the virtual world. A suspenseful soundtrack kept the player on edge and heightened their emotional engagement, making them more invested in the experience. By using headsets, the player was surrounded by the VR environment both visually and audibly. The combination of visual elements and detailed sound effects ensured that the player was fully immersed in the story, making the experience more engaging and convincing.

Task 2 During Lesson 2

The student pressed a button on a statue and entered a secret room to continue finding information. The student discovered secret codes appearing on many documents when they scanned the documents with a special flashlight. The following affordances designed in Room 2 of the VR game were identified to help the student complete Task 2.

Functional affordance: A clickable button on a statue, audio commentary, movable drawers from a cabinet, a movable flashlight, audio sounds from a player, movable documents, invisible codes on documents and news, a board with secret symbols, and a rotating lock were available. These effects allowed the learner to navigate the VR

environment. The instructor's audial guidance provided additional support through verbal instructions, helping the player navigate tasks and understand the game's objectives.

Physical affordance: VR hands to move anything such as documents, a board with secret symbols, a flashlight, and the ability to do anything in the VR space that the participant could do in real life (such as rotating a lock, turning on the flashlight, and turning on an audio player) were possible.

Cognitive affordance: Documents written in Chinese as well as audio commentary in Chinese helped the learner to search for information or listen to commentary. A board with secret symbols helped them to find a password. The instructor provided audial guidance.

Sensory affordance: Sound effects for turning documents were triggered so that the student could feel immersed, and suspicious background music was played so that they could feel tensions. With the assistance of a headset device, the learner could be fully immersed in the VR story visually and verbally.

Task 3 During Lesson 3

The student decoded symbols to open a security box hiding the valuable documents that Benjamin stole from the CIA. The student read the documents, discovered Benjamin's motives, and finally burned the documents, returning to the ward of a mental hospital. Finally, the student identified the killer. The following affordances designed in Room 3 of the VR game were identified to help the student complete Task 3.

Functional affordance: Movable drawers from a cabinet, movable documents, invisible codes on documents and news, a rotating lock, audio commentary, a burning fire, and a talking VR character were provided. The instructor provided audial guidance.

Physical affordance: VR hands allowed the learner to move anything such as documents and to do anything in the VR world they could do in real life, such as burning a document.

Cognitive affordance: Documents written in Chinese as well as audio commentary in Chinese helped the learner to search for information or listen to commentary. The instructor provided audial guidance.

Sensory affordance: Sound effects for burning documents allowed the learner to feel immersed. Suspicious background music let the learner feel tension. Using headset devices, the learner could be fully immersed in the VR story visually and verbally.

The three tasks described above were designed with different learning objectives, levels of task complicity, and amounts of vocabulary. More details are listed in Table 1 above.

4.5 Data collection

To answer the two research questions, data were collected from multiple sources, including the teacher's audial guidance during the student VR learning activities, transcriptions of videotaped teacher-student interaction when the teacher provided audial guidance, time logs and counts of appropriateness of the student's oral responses, and the student's reaction to the VR game environment (see Table 2). The unit of analysis is semantic unit in phrases or sentence.

Table 2 Data resources and research questions

Research Questions	Data Sources	Data Type
Question 1: Teacher's facilitation	Teacher's audial guidance	Text, categorical
Question 2: Relationship between teacher's facilitation and student's performance	Transcriptions of videotaped teacher's facilitation and student's performance	Categorical

4.6 Data coding

In this study, two researchers collaboratively coded the same data set concurrently—a process known as pair-coding (Paul et al., 2021)—to strengthen the reliability and consistency of the findings, minimize individual bias, and ensure a more rigorous and well-rounded interpretation of the data. This collaborative approach not only aligns with best practices in qualitative research but also reinforces the methodological soundness of the study (Gao et al.. 2024). To answer the research questions, data were divided into two sets: *the teacher's facilitation* and the *student's performance*.

4.6.1 Teacher's facilitation

Data Set 1 is the teacher's facilitation, which is mainly investigated for Research Question 1. A total of 184 pieces of facilitation were identified (70 from Task 1, 70 from Task 2, and 44 from Task 3) to help researchers to analyze the instructor's instructional facilitations. The observations related to the teacher's facilitation were further categorized as oral response and questioning.

The teacher's oral responses were statements that the teacher used to interact with the student to *offer technical assistance, provide hints, confirm the student's answers*, and *give instructions.* The teacher's questioning were the questions the teacher asked when offering *technical assistance, confirming student's answers, encouraging student's further explorations*, and *helping the student to delve into more details for story-related content*.

4.6.1.1 Coding of variables

Examples, definitions and coding are provided below and listed in Table 3.

Table 3 Examples, definitions, and coding of instructors' responses

Facilitations	Variables	Values		Examples
		1	0	
Teacher's oral response		Provided	Not provided	
	(TtechAS)- Technical assistance	14	170	就是你戴着的时候,可以按这里, 它就会动(Task 1) (When you are wearing it, you can press here. It will move.)
	(Thints)- Hints	27	157	找找看密碼 (Task 3) (Please look for the password.)
	(Tconf)- Confirm	42	142	對, 入眠。好，那你的手電筒 (Task 2)。 (Yes, fall asleep. Good. Then use your flashlight.)
	(Tins)- Instructions	19	165	慢一點 (Task 3) Slowly please.
Questions				
	(TQtechAS)- Technical assistance	29	155	你看得到了吗？（TA is testing the function of the remote control hands and gave the hands to S） (Task 1) Can you see it?
	(TQconf)- Confirm	9	175	你在找甚麼? (Task 2) What are you looking for?
	(TQexp)- Explorations	29	155	那另外一张是什么？(Task 1) What is the other piece of paper?
	(TQcontent)- Content	71	113	你觉得密码是什么 (Task 3) What do you think the password is?

Teacher's oral responses

1. Technical assistance (TtechAS): Refers to the teacher's oral responses for technical assistance. A code of 1 is assigned when the response addresses technical assistance, while a code of 0 is assigned when no such response is provided.

2. Hints (Thints): Refers to the teacher's oral responses that provide hints. A code of 1 is assigned when the response provides a hint, while a code of 0 is assigned when no hint is provided.

3. Confirm (Tconfirm): Refers to the teacher's oral responses for confirmation. A code of 1 is assigned when the response includes confirmation, while a code of 0 is used when no confirmation is provided.

4. Instructions (Tins): Refers to the teacher's oral responses for providing instructions. A code of 1 is assigned when the response includes instructions, while a code of 0 is used when no instruction is provided.

Teacher's questions

1. Technical assistance: Refers to the teacher's question related to technical assistance (TqtechAS). A code of 1 is assigned when the teacher's question involves technical assistance, while a code of 0 is used when no such question is provided.

2. Confirm (Tqconf): Refers to the teacher's questions for confirmation. A code of 1 is assigned when the teacher's question seeks confirmation, while a code of 0 is applied when no such question is provided.

3. Exploration (Tqexp): Refers to the teacher's questions that encourage students to explore further. A code of 1 is used when the teacher's question prompts exploration, while a code of 0 is applied when no such question is provided.

4. Content (Tqcontent): Refers to the teacher's questions related to VR content. A code of 1 is assigned when the question pertains to the VR content, while a code of 0 is used when no such question is asked.

4.6.2 Student's performance

Data Set 2 includes the student's reaction to the teacher's responses, which was investigated for Research Question 2. Based on the video recording of the three VR-based learning tasks, 187 instances of the student's reactions to the teacher's responses were identified (71 from task 1, 71 from Task 2, and 45 from Task 3). Seven variables were extracted from these reactions: *Appropriateness, Promptness, Verbal Reaction, Responsive Action, Active Action, Reading VR Prompt*, and *Listening VR Prompt*. These variables provide insights into how the student engaged with and responded to teacher feedback in a VR learning environment. They are defined below and presented in Table 4 (see next page).

1. Appropriateness (ApproP): Refers to whether the student correctly answers the question. Responses were coded as 1 for appropriate responses and 0 for inappropriate answers, based on errors in sentence structure or word use.
2. Promptness (Promp): Refers to the length of a pause between the end of the teacher's question / responses and the beginning of the student's response. Measured in seconds.
3. Verbal Reaction (VerbR): Refers to whether the student responds verbally to the teacher's questions. Coded as 1 if the student provided a verbal response and 0 if no verbal response is required.

4. Non-verbal Action (NonVerb): Refers to non-verbal actions taken by the student in response to direct or indirect teacher questions or instructions. Coded as **1** when a non-verbal action is observed and **0** when no such action occurs.
5. Active Action: Refers to actions initiated by the student. Coded as 1 if the student takes an active initiative and 0 if no active action is necessary.
6. Reading VR Prompt (RVRP): Refers to whether the student reads VR prompts to obtain information. Coded as 1 if VR reading is involved and 0 if no VR reading occurs.
7. Listening VR Prompt (LVRP): Refers to whether the student listens to VR prompts or instructor's verbal facilitations to obtain information. Coded as 1 if VR listening is involved and 0 if no VR listening occurs.

Table 4 Coding of the response variable and six explanatory variables

Variables	Values	
	1	0
Response Variable		
(*APProP*) – Appropriateness (*Y*)	145	39
Explanatory Variables:		
(Promp) – Promptness (X_1)	Scale	
(*VerbR*) – Verbal Reaction (X_2)	145	39
(Non-Verb) –Nonverbal Reactions (X_3)	142	42
(*ActA*) – Active Action (X_4)	156	28
(RVRP) – Reading VR Prompt (X_5)	148	36
(LVRP) – Listening VR Prompt (X_6)	149	35

Note. Promptness was measured by seconds.

5. Data Analysis and Results

This study integrated a story-driven VR game into a Chinese lessen to investigate instructor's facilitations, a student's performances, and the student-teacher interactions. This section describes the results of the data analysis.

5.1 Data analysis and results for Question 1

Research Question 1. How could the teacher facilitate the completion of an authentic story-driven VR game?

This session delineates the findings pertaining to research question 1. A descriptive analysis was employed to address this research question. The teacher's facilitation strategies across the three tasks predominantly involved oral responses and inquiries. The data were analyzed by calculating the frequencies and percentages of various strategies. The results of this analysis are presented below.

1. The types of instructor facilitation: Oral responses and questions

Among the instructor's audial guidance, oral responses and questions emerged as the primary facilitation strategies. The frequencies of oral responses and questions were

54% and 52%, respectively, indicating that both types of guidance occurred in more than half of the instructor's interactions with the student (see Figure 2).

During the VR gaming activities, where the student wore VR goggles and lacked visual contact with the instructor, questioning and oral guidance appeared to be more effective in eliciting feedback. This finding suggests that the instructor played a **proactive role** in facilitating and guiding the learning process, actively engaging the student through a combination of responses and questions.

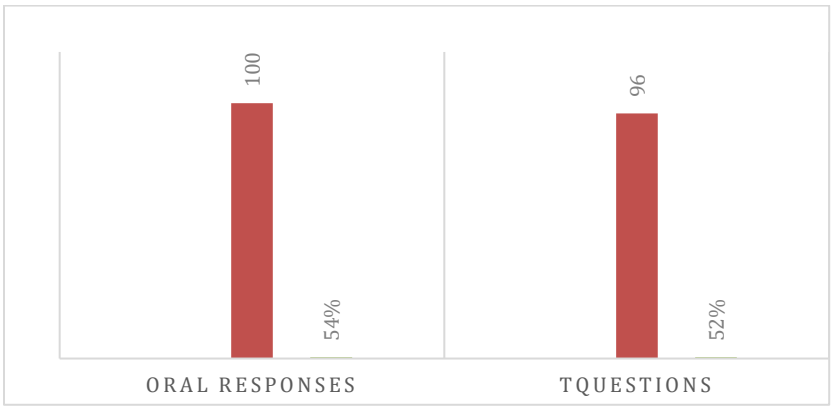


Figure 2 Frequency of instructor's audial guidance

2. The uses of oral responses and questions among the three tasks

The line chart below (Figure 3 on next page) illustrates a decreasing trend in the instructor's use of oral responses for technical assistance, hints, confirmation, and instruction across tasks. The frequency dropped from 45% in Task 1 to 35% in Task 2 and further to 20% in Task 3. In contrast, the frequency of questions did not follow a consistent trend across tasks, with occurrences at 34% in Task 1, increasing to 42% in Task 2, and then declining to 24% in Task 3.

This variability suggests that the instructor's questioning strategy was influenced by evolving instructional guidance and learning objectives rather than a fixed pattern. Notably, questions were used more frequently than oral responses in Tasks 2 and 3, indicating the important role of questioning in guiding the student through the VR tasks. As the student became more familiar with the VR game environment, the instructor's approach shifted toward questioning rather than direct oral responses, reflecting a more natural, authentic, and conversational interaction within the VR story game.

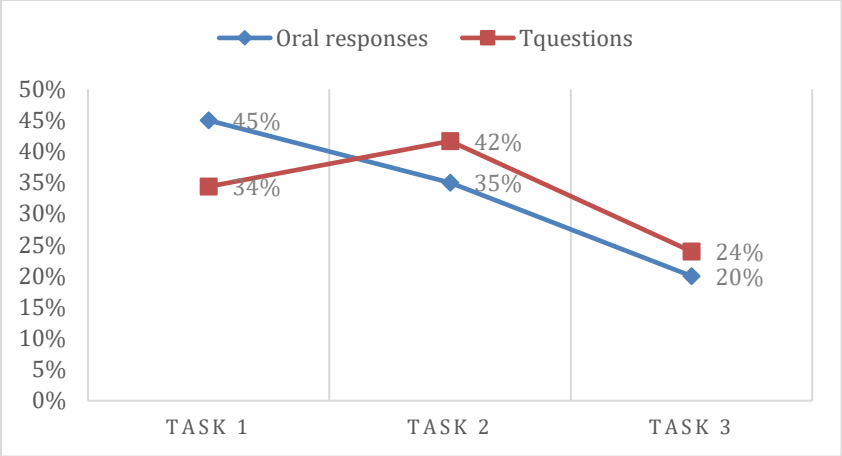


Figure 3 The variable frequency of the instructor's responses among three tasks

3. The uses of oral response for technical assistance, hints, confirmation, and instruction

When oral responses were used, instructors aimed to offer technical assistance, provide hints, confirm the answers from the students, and give instructions. The line chart in Figure 4 below illustrates these four facilitation strategies in oral response. In these instances, it was not always necessary for the student to respond verbally to the instructor's guidance. Instead, the student typically processed the information and reacted non-verbally, demonstrating comprehension through actions rather than spoken responses. This finding suggests that oral responses served as a direct instructional tool, facilitating the student's engagement without requiring continuous verbal interaction.

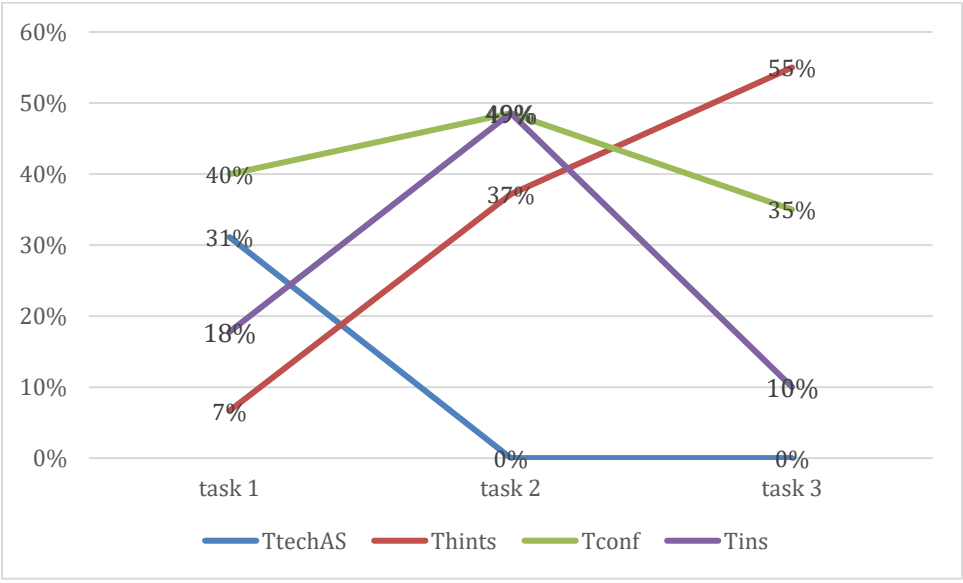


Figure 4 Instructor's frequency of oral response used for technical assistance, hints, confirmation, and instruction

The four facilitation strategies used across different tasks were further analyzed to understand their contributions to different learning activities. The findings of the three tasks are provided below.

Task 1. Technical assistance (31%) was only provided in Task 1. At the beginning of the VR games, the instructor ensured that the student understood how to operate the VR devices, making technical assistance a crucial strategy in Task 1. However, as the VR techniques were not challenging for young learners, further technical guidance was not necessary in Tasks 2 and 3. Confirmation strategy was a predominant (40%) strategy, allowing the instructor to interact with and support the student. Instructions strategy (18%) was used less frequently than confirmation but still played a role in guiding the student through the task. Hints (7%) were rarely provided in Task 1, which focused on collecting personal background information about the victim. This task is relatively less challenging, so the instructor offered minimal hints to encourage the student to explore and uncover information independently.

Task 2. Strategies of hints and confirmations were used most frequently in Task 2. Among the four strategies, hints increased significantly from 7% in Task 1 to 37% in Task 2, and confirmations remained the most frequently applied strategy (49%). In Task 2, the student needed to (a) examine the documents in a drawer, (b) examine the documents on walls (medical records, CIA documents, news articles, special symbols), and (c) locate the security box. These specific tasks made Task 2 the most challenging because the student had to understand and analyze multiple details within the storyline and comprehend and process 34 Chinese words to successfully accomplish the tasks. Given these complexities, the instructor relied more heavily on hints and confirmations than on other strategies to facilitate learning and task completion.

Task 3. Hints (55%) were the most frequently used strategy, while confirmation (35%) and instructions (10%) were applied less frequently. The use of instructions decreased significantly compared to previous tasks, dropping to 10%. In Task 3, the student needed to (a) find the password, (b) decode the symbol, and (c) find out the truth. Since these tasks involved working with codes, numbers, and documents to unlock the security box and piece together the storyline, the instructor relied heavily on hints (55%) to guide the student toward discovering the truth in the VR story. As a result, hints and confirmations were prioritized over direct instructions, allowing the student to engage more actively in problem-solving and critical thinking within the VR environment.

4. The use of instructor questioning for technical assistance, confirmations, explorations, and content

Similar to the oral responses, technical assistance questions were used only in Task 1. Across all tasks, the majority of questions focused on story content, with content-related questions increasing steadily across three tasks (Task 1 at 48%, Task 2 at 83%, Task 3 at 96%). The use of exploration and confirmation questions among three tasks did not show a consistent pattern across tasks. Similar to the instructor's oral responses, the use of exploration and confirmation questions may have been influenced by the learning objectives of each task. In Task 2, which required more complex messages and information,

the instructor applied exploration questions (38%) more frequently than confirmation questions (3%), suggesting a greater emphasis on encouraging deeper engagement and discovery.

The instructor's questioning approach was greatly influenced by the nature and complexity of each task. A detailed analysis of the four questioning strategies used in each task is provided below (and see Figure 5).

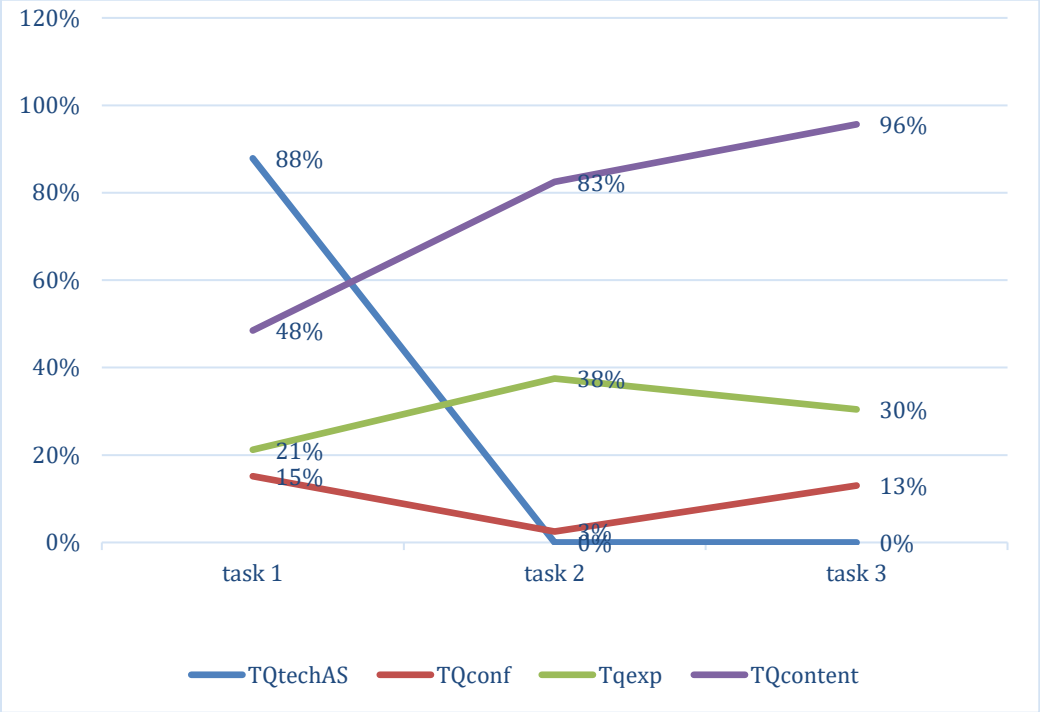


Figure 5 Instructor's frequency of questioning used for technical assistance, confirmations, explorations, and content

Task 1. In Task 1, many questions focused on technical issues (88%), followed by VR content related (48%), exploration (21%), and confirmation (15%). Since Task 1 was the student's first VR task, the instructor asked more technical questions to help him familiarize himself with the VR environment. Additionally, story-content questions were frequently used to aid in navigation and guide the student through the task. Like the instructor's oral responses, questions related to technical issues were the most frequently used in Task 1, reinforcing the need for initial support as the student adapted to the VR learning environment.

Task 2. In Task 2, the instructor primarily focused on content-related questions (83%), followed by exploration (38%) and confirmation (3%). No technical questions were asked, as the learner did not encounter any technical challenges. The presence of various functional VR objects (e.g., movable drawers, flashlights, documents, and game sounds) and the high challenge level of this session across all three tasks led the instructor to reply more content-related questions than exploration and confirmation questions to assist the learner in navigating and completing the tasks effectively.

Task 3. In Task 3, the instructor primarily used content-related questions (96%), followed by exploration (30%) and confirmation (13%). The tasks required decoding symbols to open a security box and solve mysteries, making guiding questions essential for effective exploration of the VR environment. Consequently, content-related questions played a crucial role in helping the learner navigate the challenges and complete the tasks successfully. Given the nature of the activity, all the instructor's questions were directly related to the VR game itself, reinforcing the importance of content-driven questioning in facilitating problem-solving and engagement within the VR experience.

5.2 Data analysis and results for Question 2

Research Question 2. What is the relationship between the teacher's facilitation and the student's performance?

This session presents the findings and Chi-square tests pertaining to research question 2. To examine the relationship between teacher facilitation and student performance, 2 x 2 Chi-Square ($\chi 2$) tests were conducted. These tests determined whether a statistically significant association existed between one variable for teacher facilitation (oral responses and questions) and one variable for the student performance. A total of 48 Chi-Square ($\chi 2$) tests were performed, with eight variables identified in teacher facilitation and six in student performance. Table 5 and Table 6 report significant variables. In both tables,

- rows in Column A = Each variable for student performance, categorized as 0= absent, 1= present, and
- variable in Column B = Each variable for teacher's oral responses or teacher's question variable, categorized as 0= absent, 1= present.

Table 5 Teacher oral responses

Variable A Student Performance		Variable B Teacher's Oral Responses		Chi-Square Results df=1, n=184		
		TtechAS				
		0a n=170	1b n=14	X2	P	Phi(ϕ)
*NonVerb	0c	35 (21%)	7(50%)	6.352	0.012	-0.186
	1d	135 (79%)	7(50%)			
*ActA	0	19 (11%)	9(64%)	28.278	<0.001	-0.392
	1	151 (89%)	5(36%)			
*RVRP	0	22(13%)	14(100%)	62.29	<0.001	-0.582
	1	148(87%)	0 (0%)			
		Thints		Chi-Square Results df=1, N=184		
		0 n=157	1 n=27			
*VerbR	0	28(18%)	11(41%)	7.237	0.007	-0.198
	1	129(82%)	16 (59%)			
*ActA	0	28(18%)	0(0%)	5.68	0.017	0.176
	1	129 (82%)	27(100%)			
		Tconf		Chi-Square Results		

		0 n=142	1 n=42	df=1, n=184		
*RVRP	0	34(24%)	2(5%)	7.578	0.006	0.203
	1	108(76%)	40(95%)			
		Tins		Chi-Square Results df=1, N=184		
		0 n=165	1 n=19	X2	P	Phi(ϕ)
*VerbR	0	309(18%)	9(47%)	8.689	0.003	-0.217
	1	135(82%)	10(53%)			

Notes:

(*) significant student performance

a. responses are not provided; b. responses are provided; c. student's performance is found; d. student's performance is not found.

After analyzing all 48 Chi-Square ($\chi 2$), 16 significant correlations were identified, including both positive and negative associations between teacher facilitation and student performance. The results are further discussed from two key perspectives: the teacher's oral responses and the teacher's questions.

Teacher's Oral Responses

Table 5 reports the significance variables on teacher's oral responses. A total of seven correlations were identified in Table 5, including five significant negative correlations and two significant positive correlations.

***Negative correlations*.** Five negative correlations indicate that certain teacher facilitation strategies may hinder specific aspects of student performance.

1. Technical support (TtechAS) is negatively correlated with the student's non-verbal reactions (NonVerb) ($\chi 2(1, N=184) = 6.352$, $p=0.012$, $\varphi = -0.186$, small effect), active actions (ActA) ($\chi 2(1, N=184) = 28.278$, $p<0.001$, $\varphi = -0.392$, medium effect), and reading VR prompts (RVRP) ($\chi 2(1, N=184) = 62.29$, $p<0.001$, $\varphi = -0.582$, strong effect). These results indicated that when the teacher provided technical support, the student was less likely to engage in non-verbal interactions, active participation, and reading VR prompts. This finding implies that direct technical guidance might limit student autonomy in the VR environment.
2. Hints (Thints) and verbal reactions (VerbR) ($\chi 2(1, N=184) = 7.237$, $p=0.007$, $\varphi = -0.198$) are negatively correlated. The more hints the instructor provided, the less likely the student was to respond verbally, possibly indicating that excessive guidance reduced opportunities for students to articulate their own thoughts.
3. Instructions (Tins) also show a modest negative correlation with verbal reactions (VerbR) ($\chi 2(1, N=184) = 7.237$, $p=0.007$, $\varphi = -0.198$). When the instructor provided explicit instructions, the student was less likely to respond verbally, suggesting that structured guidance may limit spontaneous verbal engagement in the VR learning environment.

Positive correlations. Two positive correlations suggest that certain teacher strategies enhanced student engagement in VR activities:

1. Teacher's hints (Thints) and student's active actions (ActA) ($\chi 2(1, N=184) = 5.68$, p=0.007, effect size Phi(φ) =0.176) are positively correlated with modest effects. When the instructor provided hints, the student was more likely to take active actions, suggesting that subtle guidance encouraged exploratory learning and participation.
2. Teacher's instructions (Tins) and student's reading of VR prompts (RVRP) ($\chi 2(1, N=184) = 7.578$, p=0.006, effect size Phi(φ) =0.176) with modest effect. When the instructor provided instructions, the student was more likely to read VR prompts, indicating that structured guidance helped direct students' attention to relevant in-game information.

The significant findings suggest that teacher-provided technical support may hinder students' non-verbal interactions, active engagement, and VR reading activities, possibly by reducing their autonomy and problem-solving opportunities. However, hints and instructions can positively influence students' engagement by fostering active participation and VR-related reading behaviors. These insights highlight the importance of balancing guidance with opportunities for independent learning in VR-based instruction, ensuring that facilitation strategies enhance rather than limit student interaction and engagement.

Teacher's Questions

Table 6 (next page) shows significant variables on teacher's questions. As can be seen in Table 6, A total of nine significant correlations were identified, including three negative correlations and six positive correlations.

Negative correlations. Technical questions (TQtechAS) are negatively associated with nonverbal reactions (NonVerb) ($\chi 2(1, N=155) = 9.460$, $p = 0.002$, $\varphi = -0.227$, small effect), active actions (ActA) ($\chi 2(1, N=155) = 9.904$, $p = 0.002$, $\varphi = -0.232$, small effect), and reading VR prompts (RVRP) ($\chi 2(1, N=155) = 4.868$, $p = 0.027$, $\varphi = -0.163$). These three negative correlations suggest that technical questions (TQtechAS) may hinder certain aspects of student engagement. When the instructor asked technical questions, the student was less likely to engage in non-verbal reactions, active participation, and VR reading prompts. This finding suggests that technical questions primarily serve a problem-solving function rather than promoting immersive engagement.

Positive correlations. Three types of questions created six positive correlations, suggesting that certain questioning strategies enhanced student interaction and engagement.

1. Technical questions (TQtechAS) are positively related to students' verbal responses (VerbR) and the appropriateness of their answers (Approp) (both $\chi 2(1, N=155) = 6.492$, $p = 0.011$, $\varphi = 0.188$, small effect). While technical questions reduced non-verbal engagement, they positively influenced verbal responses and the appropriateness of student answers. This finding indicates that the student

understood these questions and responded correctly, showing that technical questioning reinforced comprehension and accuracy in verbal responses.

2. Exploratory questions (TQexp) are positively linked to the student's VR reading prompts (PVRP) ($\chi 2(1, N=155) = 5.682$, $p = 0.017$, $\varphi = 0.176$, small effect). When the instructor used exploratory questions, the student was more likely to read VR prompts, suggesting that exploratory questioning fostered deeper engagement and active exploration within the VR environment.
3. VR content-related questions (TQcontent) are positively associated with verbal reactions (VerbR), active actions (ActA), and VR reading prompts (PVRP), with the strongest effect on reading prompts (PVRP) ($\varphi = 0.363$, medium effect). Content-related questions were strongly associated with verbal engagement, active student participation, and VR reading prompts. This finding highlights the effectiveness of content-related questions in guiding student attention, facilitating exploration, and creating a more immersive learning experience.

Table 6 Teacher's questions

Variable A Student Performance		Variable B Teacher's Questions		Chi-Square Results df=1, N=184		
		TQtechAS				
		0 n=155	1 n=29	X2	P	Phi(ϕ)
*VerbR	0	38	1	6.492	0.011	0.188
	1	117	28			
*NonVerb	0	29	13	9.460	0.002	-0.227
	1	126	16			
*ActA	0	18	10	9.904	0.002	-0.232
	1	137	19			
*Approp	0	38	1	6.492	0.011	0.188
	1	117	28			
*RVRP	0	26	10	4.868	0.027	-0.163
	1	129	19			
		TQexp		Chi-Square Results df=1, N=184		
		0 n=155	1 n=29	X2	P	Phi(ϕ)
*RVRP	0	35	1	5.682	0.017	0.176
	1	120	28			
		TQcontent		Chi-Square Results df=1, N=184		
		0 n=1137	1 n=71	X2	P	Phi(ϕ)
*VerbR	0	34	5	13.865	<0.001	0.275
	1	79	66			
*ActA	0	22	6	4.103	0.043	0.149
	1	91	65			
*RVRP	0	35	1	24.218	<.001	0.363
	1	78	70			

Notes: (*) significant student performance

These findings indicate that there is a complex interplay between question types and student engagement in VR learning environments. Technical questions (TQtechAS) demonstrate a negative impact on the student's nonverbal reactions, active actions, and reading of VR prompts, suggesting that questions that are associated with technical issues aiming to resolve the technical problems the student encountered. However, these technical questions positively increase the frequency of the student's verbal responses and answer appropriateness, indicating that the student understands these questions and also properly responds to the instructor. Additionally, exploratory questions (TQexp) enhance student interaction with VR reading prompts, fostering deeper engagement. Content-related questions (TQcontent) show the strongest positive correlations with verbal reactions, active actions, and reading prompts, particularly highlighting their effectiveness in guiding student attention and facilitating immersive experiences.

6. Summary of findings

Throughout the story-driven VR game used for this case study, the instructor progressively relied more on questioning rather than simply responding orally to the student's reactions, especially in the later VR sessions.

Among the three tasks:

1. Technical support responses were most frequently used during the first Chinese learning session, as the student needed guidance on operating the VR system.
2. In later sessions, the instructor relied more on hints, confirmations, story-content questions, and exploration questions to facilitate learning and engagement.

Key findings on question types and responses:

1. Technical questions and responses did not promote non-verbal interactions, active participation, or VR reading activities. However, they were positively associated with the student's verbal responses and the appropriateness of their answers, suggesting that technical facilitation enhanced comprehension but reduced autonomous exploration.
2. Story-content and exploration questions showed significant positive correlations with verbal reactions, active participation, and reading prompts, highlighting their effectiveness in guiding students through the VR environment.
3. Student verbal engagement was linked to answer appropriateness—the more verbal actions a student took, the more accurate their responses tended to be.

These findings suggest that questioning strategies, rather than direct technical support, play a crucial role in fostering active learning and engagement in VR-based Chinese language instruction. Figure 6 (see next page) provides a visual map of key findings of this research.

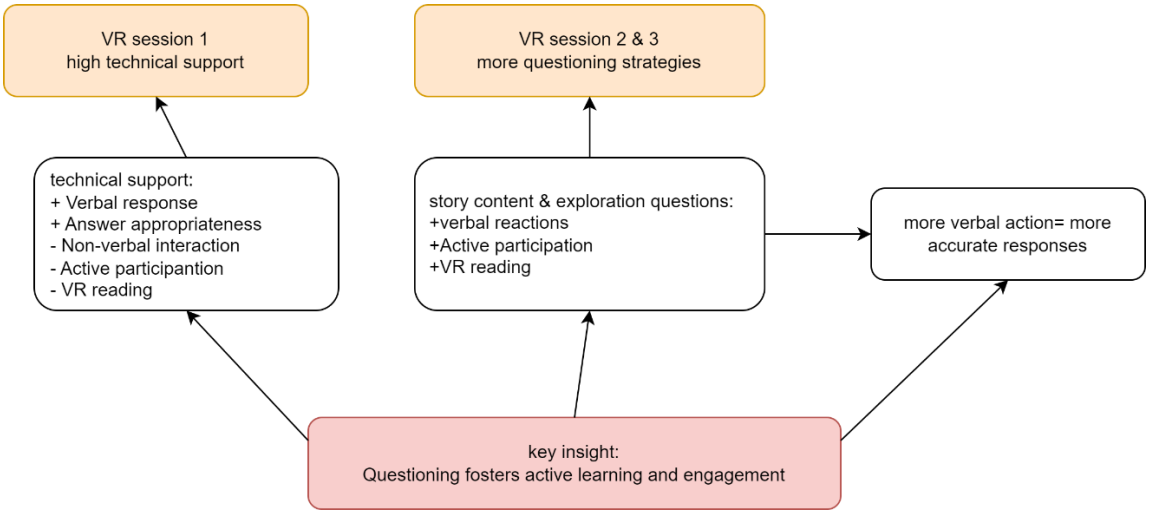


Figure 6 A Summary of key findings

7. Discussion and conclusion

This study explored a Chinese lesson that integrated a story-driven VR game with three learning tasks. It aimed to investigate how a teacher facilitated a Chinese learner in VR-based learning, and how the teacher's facilitation influenced the student's performance.

This section first presents the findings of current study, followed by a discussion of its implications, limitations, and suggestions for future research.

7.1 Findings of the interactions between teacher and student in a story-driven VR game

7.1.1 Teacher's facilitation strategies during the VR learning game

For language learners, an effective instructional approach involves creating a rich learning environment that enables students to explore and engage with language elements freely. Research suggests that teacher guidance plays a crucial role in capturing students' attention and fostering motivation in the learning process (Kordt, 2018). In this case study of a story-driven VR learning environment in the VR game, *The Price of Freedom*, the student was immersed in a virtual space where the teacher's facilitation and guidance served as a key learning affordance.

A teacher's facilitation aims to scaffold learning, which "leads to new affordances—after a period of practice—that emerges even without teacher or peer support" (Kordt, 2018, p. 141). Additionally, scaffolding helps develop the learner's ability to self-scaffold. Due to the absence of physical eye contact in a fully immersive virtual game, the student in this study not only searched independently for information, documents, and objects within the VR environment, but also relied on social interactions with the instructor to navigate challenges throughout the VR experience. Such VR environment also empower instructors to design more interactive language learning environment. Echoing

the findings of Martin et al. (2020) and Çakır (2024), the instructor's support served as a key learning affordance, helping to sustain students' motivation and interest within the learning environment. This aligns with Çakır's literature review, which highlights the importance of instructor involvement. In turn, this affordance encouraged students to engage more actively with virtual objects, messages, and information within the virtual rooms.

From the conclusion of Çakır's (2024) literature review, the teacher's role shifts towards facilitator of immersive learning experiences, which naturally influence the interactions with students. In this study, the instructor used two primary facilitation strategies: oral responses and questions in the nature of authentic communications in the VR environment. The oral responses strategy was used to provide *technical assistance, hints, confirmation*, and *instruction*, and the questions strategy was used to provide *technical assistance, confirmations, explorations*, and *content guidance*. These facilitation strategies provided evidence that the teacher created new learning affordances which enabled the learner to accomplish the three VR tasks. Among the three VR tasks, the instructor relied heavily on technical assistance to ensure the student could fully operate the VR system in Task 1. As a result, oral responses and questions related to technical issues were applied only in Task 1.

Tasks 2 and 3 required deeper cognitive engagement due to their complexity, so the instructor increased the use of hints, confirmations, and exploratory questions. Unlike traditional second / foreign language learning materials, this study integrated authentic VR content rather than pre-designed, language-specific instructional material. The three learning tasks were structured based on the natural story flow and scene settings. Due to the commercial nature of the VR game, the complexity of tasks was not intentionally controlled for language learners. Task 2 was the greatest challenge, requiring the most instructor support.

7.1.2 Relationship between teacher facilitation and student performance

From an affordance theory perspective, learning environments provide numerous affordances that student perceive to guide their behaviors (Cheng et al., 2017; Wang et al., 2018). Similarly, the instructor's perceptions of learning affordances and facilitation also directly affects the student's learning achievement (Martin et al., 2020). In the context of VR language learning environments, this study focuses on how teacher facilitation—through oral responses and questions—interacts with student performance, with the support of virtual affordances.

Chi-square tests revealed the following key findings:

1. The teacher's technical oral responses and questions significantly hindered the student's nonverbal interactions, active actions, and VR reading activities. However, when the teacher' oral responses were used to provide hints and instructions, they positively fostered student's VR actions and VR reading activities.

2. Technical questions positively correlated with the students' verbal responses and the accuracy of their answers. Content-related questions significantly enhanced students' verbal responses, active participation, and engagement with VR reading prompts, highlighting their effectiveness in immersive learning.

7.2 Implications

When applying digital devices into Chinese language learning lessons, the instructor may initially focus on technical support to ensure that the devices are functioning properly. However, excessive focus on technical troubleshooting could shift the attention from the main lesson content. While these technical responses and questions may not necessarily foster nonverbal interactions, active learning actions, or immersive VR reading activities, they can promote verbal communication by encouraging students to speak and respond in the target language, which in this study is Mandarin Chinese. To ensure that learners focus on language practice rather than technical issues, it is recommended that technical troubleshooting be conducted in the students' native language rather than in Chinese. This approach minimizes lesson disruptions and maximizes the time spent on meaningful language learning rather than on resolving technical difficulties.

Besides technical support facilitations, the instructor may incorporate hints, confirmations, story-content questions, and exploration questions to enhance VR learning sessions. Story-content questions focus on the narrative elements of the VR experience, helping students engage with the storyline. Exploration questions encourage students to discover new advantages within the virtual environment, fostering active learning and problem-solving skills. The combined use of story-content questions and exploration questions encourages verbal responses from students, helping them to practice using the target language (e.g., Chinese) more actively. These types of questions also foster active engagement and reading virtual prompts, which means that students interact more dynamically with the content and practice language skills in a context that is linked to the VR story.

The immersive nature of the VR environment is particularly effective for language learning, as it creates a more engaging and contextually rich setting where students can practice both verbal and nonverbal communication in the target language. The instructor's role in using strategic questioning techniques is critical for maximizing the learning potential of VR sessions.

The instructor in this case study that used this story-driven VR game tended to ask more questions rather than simply providing oral responses to the student's reactions across the three VR tasks. When the student responded more frequently to the instructor's oral guidance (likely in response to more instructor questions), the responses tended to become more appropriate. This finding implied that encouraging students to practice language as much as possible during VR learning experiences creates more opportunities for negotiation and problem-solving interactions. In the immersive single-player virtual learning environment, students without eye contact support with the instructor become more focused on the affordances of the virtual learning environment and on the instructor's oral guidance, and they can thus obtain more opportunities to produce language output.

Such interactions help students adjust their language use, leading to more appropriate responses. This contributed several perspectives on the use of VR in language learning, as highlighted by Parmaxi (2023) and Çakır (2024). These include the need for further research on the design of real-life tasks—such as the story-driven VR game featured in the study—to promote greater language production, the social dynamics of collaborative learning (particularly teacher-student interactions), and valuable insights into the pedagogical dimensions of using VR games in language education, such as teachers' responses and questioning techniques.

7.3 Limitations and future studies

There are several limitations to this study. First, as a single case study, it collects one Chinese learner's VR learning experience, which did not include a sufficient sample to generate broadly generalizable results. Future research should include more empirical studies with larger sample sizes, incorporating different learning themes and diverse student learning backgrounds. Second, this study focused on a single-player VR game, which did not include data associated with group interactions or language practice with other learners. For further research, a multiple-player game for language learning is suggested to observe the impact of learners' interactions on language performance. Finally, this study integrated a story-driven virtual game into a Chinese lesson facilitated by an instructor, rather than employing a self-directed learning mode. Since language learners progress at varying paces, implementing a virtual game in a classroom setting may not accommodate all learners equally. For future investigations, a self-paced game design could be explored to assess the effectiveness of different learning models.

References

Aronin, L. (2014). The concept of affordances in applied linguistics and multilingualism, In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism* (pp. 157- 174). Springer International Publishing.

Alfadil, M. (2020). Effectiveness of virtual reality game in foreign language vocabulary acquisition, *Computers & Education, 153,* 1-13. https://doi.org/10.1016/j.compedu.2020.103893

Breien, F., & Wasson, B. (2020). Narrative categorization in digital game-based learning: Engagement, motivation & learning. *British Journal of Educational Technology, 52*, 91-111.

Çakır, N. (2024). Virtual reality games in foreign language education: A systematic review. *Language Education and Technology, 4*(1), 13-37.

Cheng, A., Yang, L., & Andersen, E. (2017). Teaching language and culture with a virtual reality game. In G. Mark & S. Fussell (Ed.), *Proceedings of the 2017 conference on human factors in computing systems* (pp. 541-549). Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3025453.3025857

Drouhard, M., Aragon, C. R., Zheng, X., Kocielnik, R., Cen, K., Pena-Araya, V., Suh, J., & Chen, N.-C. (2017). *Aeonium: Visual analytics to support collaborative qualitative coding*. 220–229. https://doi.org/10.1109/pacificvis.2017.8031598

Dubovi, I. (2022). Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology, *Computers & Education, 183*, 1-16

Dixon, D. H., Dixon, T., & Jordan, E. (2022). Second language (L2) gains through digital game-based language learning (DGBLL): A meta-analysis. *Language Learning & Technology, 26*(1), 1–25.

Gao, J., Li, T. J.-J., Zhang, T., Perrault, S. T., Zhang, Z., Guo, Y., & Lim, G. (2024). CollabCoder: A Lower-barrier, Rigorous Workflow for Inductive Collaborative Qualitative Analysis with Large Language Models. In F. F. Mueller & P. Kyburz (Eds.), *Proceedings of 2024 Conference on Human Factors in Computing Systems* (pp. 1-29). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642002

Gibson, J. J. (1986). The Theory of Affordances. In James Gibson, *The ecological approach to visual perception (*pp. 127-137). Lawrence Erlbaum Associates, Inc.

Hartson, H. (2003). Cognitive, physical, sensory and functional affordances in interaction design. *Behaviour & Information Technology 22*(5), 315-338.

Huang, X., Zhou, D., Cheng, G., & Xie, H. (2021). A systematic review of AR and VR enhanced language learning. *Sustainability*, *13*(9), 1-28.

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies, 23*(4), 1515-1529. https://doi.org/10.1007/s10639-017- 9676-0

Jong, M. S. (2023). Pedagogical adoption of SVVR in formal education: Design-based research on the development of teacher-facilitated tactics for supporting immersive and interactive virtual inquiry fieldwork-based learning. *Computers & Education*, *207*, 104921.

Jong, M. S. Y., Dong A., & Luk, E. (2017). Design-based research on teacher facilitation practices for serious gaming in formal schooling. *Research and Practice in Technology Enhance Learning*, *12*(1), 1-16.

Kordt, B. (2018). Affordance theory and multiple language learning and teaching. *International Journal of Multilingualism, 15*(2), 135–148. https:// doi. org/ 10. 1080/ 14790 718. 2016. 12230 81

Lee, J., Bang, J., Suh, H. (2018). Identifying affordance features in virtual reality: How do virtual reality games reinforce user experience? In Schmorrow, D. & Fidopiastis, C. (Eds.), *Augmented cognition: Intelligent technologies.* AC 2018. Lecture Notes in Computer Science, 10915. Springer.

Martin, F., Wang, C., & Sadaf, A. (2020). Facilitation matters: Instructor perception of helpfulness of facilitation strategies in online courses. *Online Learning, 24*(1), 28-49. https://doi.org/10.24059/olj.v24i1.1980

Parmaxi, A. (2020). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments, 31*, 1-13.

Parmaxi, A. (2023). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments, 31*(1), 172-184.

Paul, J. W., Rodrigues, R. B., & Cicek, J. S. (2021). Pair-coding as a method to support intercoder agreement in qualitative research*. 2021 IEEE frontiers in education*

conference (FIE), Lincoln, NE, USA. https://doi.org/10.1109/FIE49875.2021.9637248

Peixoto, B., Pinto, R., Melo, M., Cabral, L., & Bessa, M. (2021). Immersive virtual reality for foreign language education: A PRISMA systematic review. *IEEE Access*. 9, 48952 – 48962. https://doi.org/10.1109/ACCESS.2021.3068858

Perry, E., & Booth, J. (2021). The practices of professional development facilitators, *Professional Development in Education*, *50*(1), 144-156.

Poole, F., Clarke-Midura, J., & Ji, S. (2022). Exploring the affordances and effectiveness of a digital game in the Chinese dual language immersion classroom, *Journal of Technology and Chinese Language Teaching*, *13*(1), 46-73.

Stake, R. E. (1995). *The art of case study research*. Sage Publications.

Subhash, S. & Cudney, E. (2018). Gamified learning in higher education: A systematic review of the literature. *Computers in Human Behavior*, *87*, 192-206.

Sun, Y. (2016). Second language acquisition feature exploration based on affordance theory. In Q. Dong & M. Guo (Eds.), *the 2015 3rd International Conference on Education, Management, Arts, Economics and Social Science* (pp. 889-892). Atlantis Press.

Tai, T-Y., Chen, H. J., & Todd, G. (2020). The impact of a virtual reality app on adolescent EFL learners' vocabulary learning, *Computer Assisted Language Learning*, *35*(4), 892-917. https://doi.org/10.1080/09588221.2020.1752735

Villena-Taranilla, R., Tirado-Olivares, S., Cózar, R., & González-Calero, J. (2022). Effects of virtual reality on learning outcomes in K-6 education: A meta-analysis. *Educational Research Review*, *35,*100434

Wang, H., Wang, J., & Tang, Q. (2018). A review of application of affordance theory in information systems. *Journal of Service Science and Management*, *11*, 56-70. https://doi.org/10.4236/jssm.2018.111006

Xin, Y. (2022). Influence of learning engagement on learning effect under a Virtual Reality (VR) environment. *International Journal of Emerging Technologies in Learnin*g, *17*(05), 226–237

Diffusion of Chatbots as Conversational Partners in Chinese Language Teaching: Insights From a Case Study
(聊天机器人作为虚拟语伴在中文教学中的传播——基于教师视角的个案研究)

Lin, Jiajia
(林佳佳)
University of Auckland
(奥克兰大学)
jiajia.lin@auckland.ac.nz

Abstract: The emergence of advanced and versatile chatbots, represented by ChatGPT, has reignited discussions on their use in language education. However, the extent to which using chatbots as virtual language partners can benefit language learning is still controversial. To provide more insights into this question, this study investigates chatbot diffusion among three Chinese language teachers through a bottom-up perspective. Grounded in the Diffusion of Innovation (DOI) theory, the findings revealed the dynamic impact of communicative channels and innovation attributes on teachers' chatbot adoption, which enriched the accountability of DOI in this context. The findings also revealed the teachers' attitude shift from reluctant to willing to use chatbots, which highlighted using the curriculum-informed approach to tailor chatbot use to specific teaching demands. Overall, the study demonstrates the sustainable potential of chatbots as language partners, as this approach innovates the conventional practice of Communicative Language Teaching by facilitating socialization-oriented practice.

摘要：以 ChatGPT 为代表的新型多功能聊天机器人的出现，再次引发了在语言教育中应用聊天机器人的热烈讨论。然而，将聊天机器人作为虚拟语言伙伴这一策略在多大程度上能够真正促进语言教学仍存在争议。为进一步探讨这一问题，本研究从自下而上的视角出发，考察了聊天机器人的使用在三位汉语教师中的传播过程。基于创新扩散理论（Diffusion of Innovation, DOI）视角，本研究论证了传播渠道与创新特征对教师使用聊天机器人的动态影响，因而拓展了该理论在这一研究语境下的解释力。本研究同时揭示了三位教师从犹豫到积极使用聊天机器人的态度转变，突出了以课程为导向，将聊天机器的使用与具体教学需求相结合的重要性。总体而言，本研究论证了聊天机器人作为语言伙伴在语言教学中的可持续发展潜力，这一使用促进了以社交为导向的语言实践练习，创新了传统的计算机辅助交际式语言教学模式。

Keywords: Chatbot; Diffusion of Innovation; teachers' perspectives; Communicative Language Teaching; Computer-assisted language learning

摘要：聊天机器人、创新扩散理论、教师研究、交际式语言教学、计算机辅助语言教学

1. Introduction

Chatbots are software interfaces or computer-based dialogue systems that simulate human-like conversations (Chen et al., 2020; Rodríguez Cardona et al., 2019). The term "chatbot" is a "neologism" (Rodríguez Cardona et al., 2019, p. 2) derived from "chat" and "bot", highlighting their capability to facilitate real-time conversational interaction between humans and software robots.

The approach of using chatbots as language partners has a long history, dating back to the 1980s (Bibauw et al., 2019). This approach aligns with the "broad assumption" (Bibauw et al., 2019, p. 829) of Communicative Language Teaching (CLT), which prioritizes learners' proficiency in communication and abilities to manage negotiation and communication strategies (Savignon, 1991; Swain & Suzuki, 2008). However, its effectiveness remained limited for several decades, as reflected in the relatively low number of related publications—fewer than 20 per year over the subsequent 30 years (Bibauw et al., 2019). This limited diffusion can primarily be attributed to the technological constraints of early chatbots (Coniam, 2008; Qian et al.,2023). They relied on pattern-matching rules to interpret and respond to user inputs based on predefined responses, which often led to decontextualized responses. Thus, they were regarded as "idiots" (Gallacher et al., 2018, p. 70) rather than partners by language learners. Recent advances in artificial intelligence chatbots have improved their ability to generate contextually appropriate and relevant responses. Despite these improvements, concerns remain. Studies have shown that these chatbots can still provide inaccurate information (Lo et al., 2024) and often produce wordy and repetitive language (Kohnke et al., 2023), which may not serve as ideal input for language learners. Consequently, they are not yet considered capable of teaching independently (Van Horn, 2024). In this regard, technological advancement alone is insufficient to render chatbots competent language partners, and the question of their effectiveness in language learning remains open.

Language teachers' perspectives are important but underexplored in answering this question. Teachers are considered the "lynchpin" (Arnold & Ducate, 2015, p. 1) who play an essential role in the success of technology-involved learning events. Their attitudes toward technology decide whether it can be used in classes and thus influence students' attitudes and use of it. However, teachers' perspectives based on hands-on experience with chatbots were not fully explored. According to the exploration of human-chatbot collaboration in 24 studies, Ji et al. (2023) found that teachers either played a minimal role or were entirely absent. In several studies focusing on teachers' perspectives, most did not involve teachers' hands-on use of chatbots, as noted by Timpe-Laughlin et al. (2022). This

led to a neglect of the dynamics between teacher perspectives and actions, as teachers' perspectives might change while using them (Van Den Branden, 2009). Therefore, this study aims to fill this gap by revealing teachers' perspectives based on their hands-on experiences with chatbots in the contextualized teaching process.

2. Literature review

2.1 Diffusion of Innovation Theory

Everett Rogers' Diffusion of Innovations (DOI) theory (2003) is an empirical framework that explains a universal process of how an innovation is adopted in a population from a social and dynamic perspective (Rodríguez Cardona et al., 2019; Frei-Landau et al., 2022; Grgurović, 2014; Kaminski, 2011). It explains "the process in which an innovation is communicated through certain channels over time among the members of a social system" (Rogers, 2003, p. 5). It includes three basic constructs: adopter categories, innovation attributes, and the innovation-decision process, as shown in Figure 1.

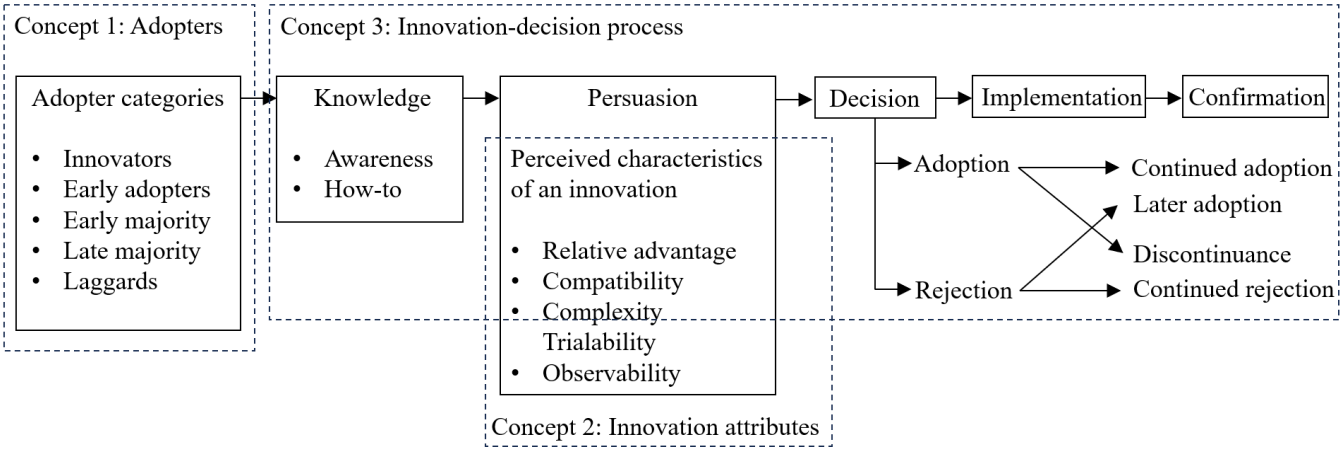


Figure 1 The framework of DOI (Derived from *Diffusion of Innovations, Fifth Edition* by Everett M. Rogers. Copyright (c) 2003 by The Free Press. Reprinted with permission of the Free Press: A Division of Simon & Schuster.)

Adopters are the subjects during the diffusion process. Rogers (2003) divided a population into five categories according to their propensity to adopt a specific innovation: innovators, early adopters, early majorities, late majorities, and laggards. These groups are distinguished based on their willingness to take risks to try innovation and the desire to be trendsetters in an industry (Kaminski, 2011). The further along in this sequence, the more cautious and sensitive these groups are about the boundaries of the social system, peer impact, and costs (Kaminski, 2011; Sahin, 2006). Five attributes describe the object—innovation. Relative advantage is defined as the extent to which an innovation is perceived as superior to the idea it replaces, which frequently co-occurs with disadvantages. Compatibility refers to the extent to which an innovation is seen as aligning with the values, prior experiences, and demands of potential adopters. Complexity denotes the intricate nature of comprehending and integrating an innovation. It frequently co-occurs with

simplicity and ease of use (Jwaifell & Gasaymeh, 2013). The fourth attribute, trialability, quantifies the extent to which an adopter can experiment with the innovation. Observability refers to the extent to which the results or benefits of adopting an innovation are visible or easily noticeable to others. It encompasses the idea that innovations that yield tangible and visible outcomes are more likely to be adopted.

The innovation-decision process in DOI consists of five distinct stages, demonstrating the comprehensive process that individuals or organizations go through from the awareness of innovation to full integration (Bax, 2003; Chambers & Bax, 2006). The knowledge stage focuses on individuals becoming aware of the innovation and learning how it works. Two key types of knowledge are emphasized here: awareness knowledge and how-to knowledge. The former involves basic awareness of the existence of innovations, while the latter relates to understanding the practical steps required to use them effectively. The persuasion stage is when potential adopters form opinions about the innovation. In the decision stage, individuals decide whether to adopt or reject the innovation. Decisions are influenced by various factors, such as personal desires, peer pressure, social situations, and researchers' intervention in empirical settings (Frei-Landau et al., 2022; Grgurović, 2014). The implementation stage is when the innovation is used. Users begin to examine its outcomes and adapt it to their specific needs. Feedback during this phase is crucial, affecting whether adopters will continue to use the innovation. If the outcomes are favorable, the innovation is likely to be more deeply integrated into daily routines. In the confirmation stage, the innovation becomes a part of the users' routine practices. Promoting innovation to others is an important indicator of this stage (Grgurović, 2014).

The implementation of DOI in computer-assisted language learning (CALL) conceptualizes the process of the diffusion of technological innovations by offering a structured approach to understanding how they are adopted and evaluated (Grgurović, 2014; Markee, 1992). Accordingly, it informs language teaching by providing a cohesive set of "guiding principles for the development and implementation of language teaching innovations" (Markee, 1992, p. 229), ultimately facilitating effective organizational reforms (Stoller, 1994). The five attributes can be used to understand adopters' perspectives, which play a mediating role in influencing adopters' attitudes (Ayanwale & Ndlovu, 2024; Fatemi Jahromi & Salimi, 2013). For example, Jwaifell and Gasaymeh (2013) used the five attributes to explain the participants' use of interactive whiteboards and the factors that might affect their decision to use them. The results revealed that relative advantages were the strongest predictor of the teachers' adoption. Moreover, it could be found that these attributes carried different weights in varying contexts. While some studies found that relative advantage had the strongest influence on adopters' attitudes (Rodríguez Cardona et al., 2019; Jwaifell & Gasaymeh, 2013), some others demonstrated that observability and trialability were the most closely related attributes to adopters' attitudes (Martins et al., 2004). In this regard, investigating the different weights these attributes carry contributes to identifying specific and practical issues in unique contexts, ultimately leading to strategies to better address the specific needs or concerns of adopters in different educational settings. It should be noted that DOI is a sociological theory that regards the diffusion process as a form of communication, which is influenced by social factors, such as communication channels and change agents. The channels of this communication are

categorized as mass media and interpersonal interaction. Interpersonal interaction is usually more powerful to create or change strong attitudes held by an individual (Sahin, 2006). Change agents also play a pivotal role in facilitating the shift in users' attitudes. For instance, in Grgurović's (2014) study, the author assumed the role of a change agent by helping teachers recognize their need for a learning management system and persuading them to integrate it into their teaching. Similarly, El Shaban and Egbert (2018) identified professional development facilitators as key change agents, who equipped teachers with the necessary knowledge to use the "English Center" and fostered positive attitudes toward its implementation. It can be seen that in these examples, different individuals took on the role of change agents, such as the author in Grgurović's (2014) and the professional development facilitators in El Shaban and Egbert's (2018). This highlights the importance of social intervention in shaping teachers' attitudes toward technology use.

2.2 CLT and computer-facilitated CLT

CLT has inspired the practice of language teaching process, from the teaching aims and methods to the assessment of learning outcomes. The theoretical foundation of CLT is the concept of communicative competence, introduced by Hymes (1972). This concept emphasizes that learning a language involves more than just knowledge of grammar and structures; it requires understanding how to use language appropriately in various social contexts. CLT aims to develop learners' communication proficiency in real-life situations. To achieve this goal, this approach emphasizes learning through communication rather than rote memorization or isolated grammar exercises. Consequently, the assessment of this approach emphasizes a process-oriented perspective (Teh, 2021), focusing on learners' progress and performance in communicative tasks rather than traditional tests.

Developing the practice of CLT is one of the focal points of related studies. Nunan (1991) listed five features of the communicative approach, which include "a focus on interactive communication, usage of authentic materials, the availability of chance for learners to work on, the inclusion of own experiences to aid in learning, and the linkage between classroom learning and real-world application" (Teh, 2021, p. 66). In classroom settings, activities such as role-plays and discussions are frequently adopted (Littlewood, 2013). However, these conventional approaches often fail to address the issue that students might feel "embarrassed and lazy to speak" (Teh, 2021, p. 65) in language classes. Computer-facilitated CLT is a supplement to conventional methods. Huang and Liu (2004) defined the connection between CLT and CALL as "computer simulation" (Huang & Liu, 2004, p. 1) and further categorized it into instruction-oriented and fun-oriented. However, they also have limitations: the former follows predefined instructions, limiting learners' initiated interaction and control during the conversations, while the latter usually involves virtual tasks, such as establishing a town in the virtual world, which may distract from real-life communication settings (Huang & Liu, 2004).

Computer simulation facilitated by chatbots may have the potential to overcome the limitations of conventional approaches. Interaction with chatbots has been evidenced to be meaningful for language learners. For example, Yin and Satar (2020) demonstrated that interaction between chatbots and language learners included the patterns of negotiation of meaning, which provide opportunities for second language development (Yin & Satar,

2020). Fryer et al. (2019) found that learners' interest in conversing with chatbots rebounded after the temporary novelty effect as they found in Fryer et al. (2017), which contributed to sustainable learning motivation. Besides, chatbots have the potential to engage learners in communications that incorporate the five features of CLT practice listed by Nunan (1991). Conversations with chatbots have been proven to occur in an anxiety-free environment (Hapsari & Wu, 2022; Hsu et al., 2021), which helps learners to focus on communication. Given that chatbots provide individualized responses, learners must create their own responses rather than repeating predefined ones. This requires them to use their language knowledge flexibly. Meanwhile, it has also been found that users might establish emotional connections (Yuan et al., 2024) and friendships (Skjuve et al., 2021) with chatbots, which may influence their social interaction and relationships. Although the broader social impact of this is uncertain, it highlights that users do have self-disclosure when communicating with chatbots. This self-disclosure is meaningful for language learning, as it requires learners to use the language to discuss their real-life experiences, as emphasized by Nunan (1991).

Given chatbots' potential in CLT practice and limited insights into chatbot-facilitated CLT, the following research questions will be addressed in this study:

1. How does the approach of using chatbots as language partners currently diffuse among language teachers?

2. How does chatbot use align with the principles of CLT?

3. Methodology

3.1 Participants

This study recruited three language teachers who teach Chinese as a second language, ranging from 25 to 35 years of age. All participants taught in the same beginning-level Chinese course at a university in New Zealand while the interviews were conducted. This is a blended learning program that integrates both online and in-person instruction. Teacher A is a Malaysian and Chinese heritage speaker with two years of experience teaching Chinese, while Teachers B and C are both Chinese native speakers with more than five years of teaching experience. The three teachers had experience of teaching a variety of levels (beginning through advanced) and learners (from preschool learners to adults).

3.2 The chatbot Xiaoming

The chatbot demo Xiaoming was developed using the free plan of the platform Collect-Chat[1]. The chatbots built on this platform are basically rule-based, which rely on predefined responses. The main reason for selecting this platform was its user-friendly interface, which allowed teachers to create their own chatbots without requiring any programming expertise. It employed a "drag and drop" builder system, offering the

[1] c.f., https://collect.chat

flexibility for teachers to easily customize it to align with specific teaching objectives and needs (Vázquez-Cano et al., 2021). The teachers manually input Xiaoming's responses, as illustrated in Figure 2 (Any information that could reveal the identity of students or the university was concealed.). The chatbot then interacted with users by delivering these predefined sentences.

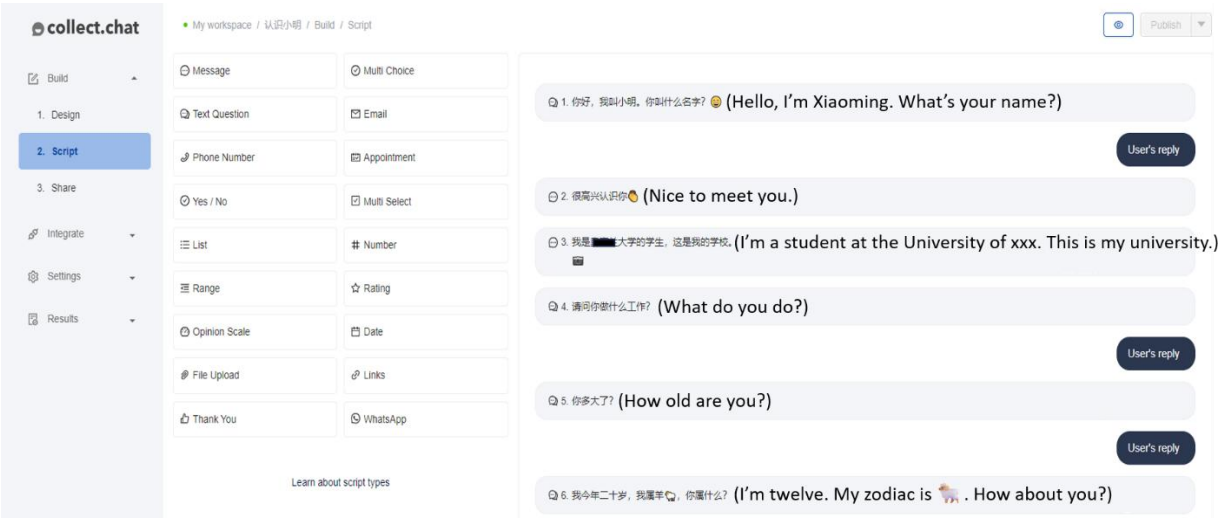
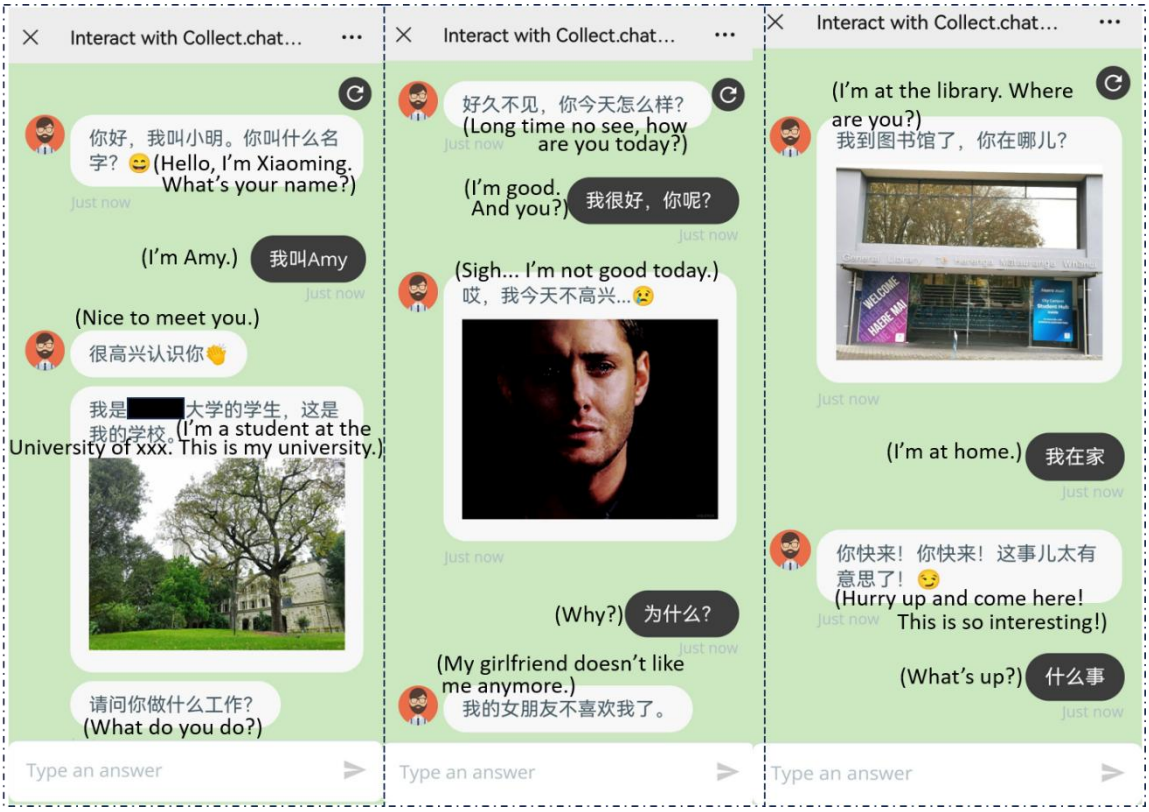


Figure 2 The editing interface of Xiaoming

Xiaoming included dialogues under three communication scenarios, which were adapted from the content of the textbook *Integrated Chinese* (Volume 1)[2]. The dialogues consisted of the topics of self-introduction, planning weekends with friends, and talking about family members. According to the content, the scenarios were named *Hello Xiaoming*, *Sad Xiaoming*, and *Gossip Xiaoming*. There was a logical relationship among them. Students conversed with *Hello Xiaoming* first. In this conversation, students and Xiaoming introduced themselves to each other like meeting a new friend. It was followed by *Sad Xiaoming*, which was about arranging some activities on weekends to cheer Xiaoming up, as he was sad because of a quarrel with his girlfriend. In the third scenario, students were supposed to study at the library with Xiaoming. However, Xiaoming was distracted by the gossip of Wenzhong Gao and Xiaoying Gao, who were part of the main characters in the textbook. Students started conversing with Xiaoming via scanning QR codes. Figure 3 a. to c. (see next page) shows samples of the conversations under each scenario. Given that Xiaoming could only respond using pre-defined sentences, it was designed to initiate and guide the conversations to maintain the logical coherence of the conversations as much as possible. The chatbot-student conversations were collected by the platform as shown in Figure 4 (see next page).

[2] Yuehua Liu, Tao-chung Yao, Nyan-Ping Bi, et al., *Integrated Chinese: Simplified Characters Textbook, Level 1,* 4th ed. (Boston: Cheng & Tsui, 2017).



a. Hello Xiaoming b. Sad Xiaoming c. Gossip Xiaoming

Figure 3 Interfaces of the three conversation scenarios

Xiaoming:	你好，我叫小明。你叫什么名字？… (Hello, I'm Xiaoming. What's your name?)	请问你做什么工作？(What do you do?)	你多大了？(How old are you?)	我今年二十岁，我属羊🐑… (I'm twelve. My zodiac is 🐐.)
Student 1:	我叫Thor (I'm Thor.)	我没有工作，所以现在我是大学的学生 (I don't have a job. I'm currently a university student.)	我八岁 (I'm eight.)	我的属是猴 (My zodiac is monkey.)
Student 2:	我叫Ho[redacted] (I'm Hozxxxxxxx.)	我是大学生 (I'm a university student.)	我今年二十岁 (I'm twelve.)	我也属羊 (My zodiac is also 🐐.)
Student 3:	你好，我叫Bella (Hello, I'm Bella.)	我也很高兴认识你。我也是[redacted]大学的一名大学生 (Nice to meet you, too. I'm also a student at xxx university.)	我是十九岁。你呢？ (I'm nineteen. And you?)	我也属羊！ (My zodiac is also 🐐.)
Student 4:	我叫李[redacted]。(I'm Zxxxx Li.)			
Student 5:	我叫issy (I'm issy.)	我也是[redacted]大学的学生 (I'm also a student at xxx university.)	我十八岁 (I'm eighteen.)	我属🐵 (My zodiac is 🐒.)
Student 6:	你好我叫H[redacted] Kim (Hello, I'm Hxxxxxx Kim.)	我在咖啡工作，也是学生 (I work at a coffee shop. I'm also a student.)	21	不知道 (I don't know.)
Student 7:	我叫 Lilith ^-^ (I'm Lilith.)	我没有工作 :((I don't have a job.)	我是十九岁 (I'm nineteen.)	我也属羊 :3 (My zodiac is also 🐐.)
Student 8:	你好，我叫丁[redacted]! (Hello, I'm Kxxxxx Ding.)	我是一个学生。(I'm a student.)	二十一岁。(Twenty-one.)	属蛇 (Snake.)

Figure 4 The records of chatbot-student conversations

3.3 Procedure

First, the research of this study designed the initial version of Xiaoming with Collect-Chat and presented it to the three participants. The researcher briefly introduced Xiaoming's functions, including how to interact with Xiaoming, adapt its content, and monitor its conversations with students. The participants were invited to test and adjust Xiaoming before applying it to their classes.

Next, the participants used Xiaoming in their classes. All three classes were conducted based on the same learning content, which was about reviewing the vocabulary and sentences from Lessons One to Five of Integrated Chinese 1. The students in the three classes shared similar learning backgrounds of Chinese, who started from the beginning level and had studied in the same Chinese course for 12 weeks.

After applying Xiaoming to their classes, semi-structured interviews were conducted individually, following a protocol of interview questions that were guided by the theory of DOI. The questions consisted of participants' general experiences with educational technology and chatbots and their specific perspectives on the adoption of Xiaoming (see Appendix). The interviews were conducted on the same day after the participants used Xiaoming in their classes, and all lasted for approximately one hour. Additional questions were asked to clarify and explore the individualized experiences of the interviewees. Informed consents were obtained from the participants to audio-record the interviews.

3.4 Analysis

The interview data were subjected to qualitative analyses. The audio recording of interviews was transcribed verbatim. The transcriptions were then segmented into episodes that "sustained focus on one topic over a longer period of time or were similar in tone" (Boschman et al., 2015, p. 253). These episodes were further segmented into topical exchanges. Each exchange referred to bigger units corresponding to one of the three main concepts of DOI, namely, adopter categories, innovation attributes, and innovation-decision process. This segmentation followed the coding scheme illustrated in Figure 1.

In an effort to minimize researcher bias and ensure interrater reliability, this study followed the approach of Huang et al. (2019), which adopted a similar approach to examine English teachers' technology acceptance. Following the initial coding by the author, the second coder, who is a doctoral student in Applied Linguistics with experience in coding interview data, conducted a thorough review to ensure the accuracy and completeness of the codes. This verification process involved cross-referencing the codes with the original statements and obtaining validation from each participant while coding their respective transcripts. Throughout this validation procedure, any discrepancies were deliberated by both coders, leading to a consensus on all the coding.

4. Results

4.1 The teachers' perspectives of chatbot use: through the lens of their adopter categories

The three teachers had not attempted to integrate chatbots into their teaching prior to this study. Their perceptions of chatbot integration categorize them as either late majority or laggards. To be more specific:

Teacher A belongs to the group of the late majority, who was characterized by her skeptical, cautious attitude and her reliance on trusted advisors. Her reluctance to use chatbots for teaching primarily stemmed from her perceived "technological immaturity" (Teacher A) of chatbots. She tested ChatGPT by inputting her Master's thesis topic to see how it would respond. The responses were quite shallow, leading her to believe that current chatbot technology was not yet intelligent enough. This impression discouraged her from further engaging with chatbots. This reluctance indicated her uncertainty about the advantages and disadvantages of using chatbots for language teaching, which is a significant obstacle to adopting a new technology (Sahin, 2006). When asked about when she would feel confident or motivated to use chatbots, she said she would rely on the suggestions of teaching experts from her family. She noted:

> I often ask my dad and cousins for suggestions about how to integrate technology into my teaching, as they are all teachers and much better at using technology than I am. They have created their own platforms, websites, and even games for their students. So, I believe they would figure out how to use chatbots before me, and I would like to listen to their sharing.

Teacher B was also categorized as part of the late majority, whose reluctance to use chatbots for teaching can be mainly attributed to her perceived mismatch between the complexity of chatbot language and her students' language proficiency. She explained:

> The language complexity of chatbots like ChatGPT is far beyond my students' proficiency in Chinese, most of whom are beginners. It is ineffective to have them practice with chatbots that use language far above their level.

Interestingly, she was proactive in adopting other technologies. She believed that interaction was essential for students to learn a language and that technology made language teaching more vivid and engaging. This belief was even more evident in her online teaching. She was in charge of the online course for this language program, where she adopted many applications and platforms, such as Padlet, Flipgrid, and Google Forms, to organize activities. Her different attitude toward chatbots indicates that teachers' perceptions of chatbot use may not be significantly related to their previous experiences with technologies. This may be attributed to the distinct, and perhaps unconscious, way in which chatbots are perceived compared to other technological tools, as noted by Liu et al. (2024).

Teacher C exhibited traits consistent with the laggards, who preferred conservative teaching methods and maintained a passive attitude toward technology use. She prioritized direct and meaningful practice and the efficient use of in-class time. This teaching belief

led her to view technological tools as unnecessary in language teaching. She only used technology when other alternatives were worse, which is one of the typical features of laggards, according to Kaminski, J. (2011). Teacher C said:

> I am usually pretty passive when it comes to using educational technology...the class is mostly about practicing. I think it is more important to have students just dive into practice instead of going for those fancy setups.

When asked about her opinions of chatbot use, although she found chatbots helpful for personal tasks, particularly for searching for information and entertainment, she was concerned about the cost and necessity of using chatbots for language teaching. She noted:

> Teachers have limited preparation time. Especially for those lacking digital literacy, the learning curve for new technologies is costly. Technology evolves rapidly, and it is neither essential nor practical for language teachers to keep up with every new advancement, as this imposes overwhelming pressure.

In summary, the reasons that have hindered teachers from using chatbots for teaching include their perceived disadvantages, the mismatch with teaching demands, the complexity of use, and the lack of how-to knowledge.

4.2 The teachers' attitude shift: the mediating role of the five attributes of innovation

Surprisingly, the three teachers, who were initially reluctant to integrate chatbots into their teaching, changed their attitudes after experiencing the process of tailoring the chatbot Xiaoming to their teaching. This shift has manifested the mediating role of the five attributes of chatbots and the alignment between chatbot use and CLT practice.

Relative advantages: Socialization-oriented communication practice

The relative advantages of chatbots highlight their potential to facilitate socialization-oriented communication practice. This potential lies in the features of chatbot-facilitated conversation practice. Firstly, chatbots engaged students in conversation flows, allowing them to practice how this language was used in communication. This can be primarily attributed to the "syntactical level" (Teacher B) practice provided by chatbots. This feature distinguishes chatbots from the technologies that provide isolated practice focusing on either vocabulary or grammar, as noted by Teacher C. Also, chatbots enable students to concentrate on conversation flow by creating an anxiety-free environment. This environment was particularly beneficial for introverted students who might hesitate to engage in competitive activities, like those provided by Kahoot, or group work. Their conversations with chatbots were only visible to the teacher, ensuring them a feeling of "security," as highlighted by Teacher A. She also underscored the capability of chatbots to provide quick and smooth responses, which smoothened conversation flows between chatbots and students.

Another feature that enhances socialization-oriented practice is to provide authentic communication practice. In this study, Xiaoming was designed to converse with students based on their campus life. Teacher A underscored the images used by the chatbot during

conversations, which depicted buildings on campus, such as the library and café, that students frequently visited. She noted:

> Images are not frequently used in other chatbots. These images make the conversation more vivid and engaging, as well as more realistic, because they are from students' familiar real-life situations.

To strengthen the authenticity of conversations, an exaggerated storytelling approach was adopted to bring Xiaoming's personality to life—he had a Chinese girlfriend and loved to gossip. Teacher B stated, "The personality of Xiaoming exhibited a degree of humor and theatricality, making the conversation interesting and engaging, and it made Xiaoming feel like a friend."

Engaging students in real social situations is also essential for effective socialization-oriented practice. Conversing with Xiaoming resembled the texting format, a common mode of communication in modern life, and the communication scenarios were based on real-life situations, as noted by Teacher B:

> The texting format of Xiaoming mirrors everyday life; it is like having a casual conversation with a friend on social media. The content, design, and feedback from Xiaoming all feel very natural and easy to understand. Especially at the start, when students were first introduced to Xiaoming, it felt like meeting a new friend introducing themselves.

Compatibility: Tailoring chatbots with a curriculum-informed approach

The compatibility of chatbots with language learning can be enhanced through a curriculum-informed approach, which eliminates teachers' perceived mismatch between chatbot use and language teaching. In this study, Xiaoming's compatibility was manifested in two aspects: its alignment with the target learning content and its alignment with teachers' needs. The teachers emphasized how well Xiaoming corresponded with the learning material, reinforcing key vocabulary and grammar from their curriculum.

> Xiaoming's conversations closely follow the topics and dialogue structures from the textbook, making it an effective review practice for students. (Teacher A)

> The practice provided by Xiaoming was closely aligned with the learning content and highlighted the frequently used words and grammar we have learned, allowing students to focus directly on essential exercises, which made it highly efficient. (Teacher C)

Chatbots can also be flexibly compatible with teachers' teaching needs. In this study, Teacher A, who emphasized interaction during the learning process, used Xiaoming to supplement her in-class activities, which provided additional practice opportunities and made the practice more engaging. Teacher B focused on students' recognition of Chinese characters. She edited the conversations to include new sentences using the characters they had learned. This edition slightly challenged the students' current knowledge and provided opportunities for them to work on the characters by themselves, which is an essential process to develop their language proficiency (Nunan, 1991). Teacher C wanted to better understand students' grasp of the content they were going to learn, which could help her

better manage the pace of the upcoming class. She suggested using the chatbot to assess students' knowledge of upcoming content at the start of a lesson.

Complexity: Less challenging than perceived

The complexity of using chatbots, which initially hindered teachers from integrating them into their teaching practice, was reduced after their hands-on experiences with chatbots. Teachers' concern about "the cost-effectiveness ratio" (Teacher C) of using chatbots was essentially tied to the perceived complexity. However, it turns out that using chatbots was not as complex as she had thought. In this study, after the authors introduced the platform Collect.Chat and provided the script of Xiaoming, the teachers were able to adjust it themselves, as this process did not require any expertise in computer science or programming. Meanwhile, the in-class implementation of Xiaoming was straightforward and user-friendly for both teachers and students. For teachers, the time and effort needed to set up Xiaoming was minimal, and there was no need to explain complex usage rules to the students. From the students' perspective, the process was equally simple—they could start conversations with the chatbot by simply scanning a QR code. This ease of use contributed to the overall efficiency of Xiaoming as a teaching tool. It changed Teacher C's attitude toward chatbot use. She noted:

> Firstly, the students have a very low cognitive load to understand this. They do not need to grasp a bunch of rules first and then explore them. They can directly scan it and understand the task they need to perform, which saves a lot of time in the classroom. I think, during practice, they can fully unleash their creativity, work on their grammar, and experiment with different sentence structures. So, it is a very efficient activity.

Trialability: Opportunities for testing and adjusting chatbots

This attribute allowed the teachers to experience chatbot use from a student's perspective and customize chatbots according to their expectations. During this testing phase, the teachers recognized Xiaoming as an effective tool for engaging students in conversation practice. However, they also identified that it was not responsive enough to student input that deviated from the current dialogue. Therefore, they made adjustments based on their experiences and expectations.

Teacher A introduced prompts that would be triggered when students intentionally used off-topic responses. These prompts were designed humorously, featuring emojis, question marks, or phrases like "别扯远了，朋友！ (Back to the topic, mate!)" to steer students back to the intended conversation. Meanwhile, she extended the conversation, which could increase more turn-taking between the chatbot and students and thus better engage the students in the communication flow.

Teacher B's adaptation of the chatbot reflected her focus on the compatibility and authenticity of the conversations. She replaced certain words with those she believed the students should pay more attention to. Meanwhile, she modified the dialogues to better align with real-life logic. For example, she changed the greeting in the second conversation from "你好" (hello) to "好久不见" (long time no see), to make the dialogue more natural as people who met their old friends. She replaced the AI-generated portrait image used in

the first conversation, which was supposed to be the photo of Xiaoming's girlfriend. According to Teacher B, the original image appeared "scary" and detracted from the students' conversation experience. By implementing a more visually appealing and approachable image, she aimed to create a more engaging and comfortable communication atmosphere for students.

Observability: Visible communication process

The process of chatbot-student conversations is observable by teachers, which enables them to monitor students' performance and reactions during conversations. This is a form of process-oriented assessment. In this study, chatbot-student communications were observed through two accesses: the records of the conversations provided by the platform Collect.Chat, and teachers' in-class observation of students' reactions to Xiaoming. The records of chatbot-student communication revealed that most students completed the conversation practice, although a few did not take it seriously. Teacher A emphasized students' engagement with the practice and closely monitored their responses. If she found a student did not take the practice seriously, she would ask them to redo the conversations. The records helped Teacher B decide when to intervene. She would explain and highlight the points where most students did not respond properly. It was also observed that Xiaoming could trigger students to talk about their own experiences, which was highlighted by Nunan (1991) as an essential step for learners to practice their communicative proficiency. This could be attributed to the conversation content, which was based on students' campus life, as well as the anxiety-free environment for communication, as previously noted by Teachers A and B.

The teachers' observation of students' reactions confirmed the effectiveness of chatbots in engaging students in conversation practice:

> The class was silent when most students concentrated on practising with the chatbot. After all, this is a new way to have conversations; they were attracted by Xiaoming. (Teacher A)

> The students enjoyed the conversations with the chatbot. I could tell it from their facial expressions and their laughter every now and then. (Teacher B)

> When I walked around to see how they interacted with the chatbot, some students stopped me and discussed with me their conversations with the chatbot. They enjoyed conversing with the chatbot, and it also increased my interaction with the students. I asked them some questions about their conversation content. (Teacher C)

4.3 Teachers' adoption of chatbots: Their innovation-decision process

Teacher A exhibited awareness but limited how-to knowledge of chatbots at the knowledge stage. She tested ChatGPT once and concluded that current chatbots were not technologically mature enough. Besides this technological issue, her reluctance to further explore their use could be attributed to her reliance on her trusted advisors. This was further evidenced in the persuasion stage. When the researcher demonstrated Xiaoming during a planning meeting for upcoming classes, Teacher A did not contribute much to the decision-making process and chose to use the chatbot only after other teachers had decided to do so.

This indicates the impact of interpersonal channels on her technology use. Her attitude toward the chatbot changed during the implementation stage. She recognized the effectiveness of Xiaoming and remarked positively about its use:

> Practicing conversations with Xiaoming is more vivid than the conventional methods. It allows students to create their own sentences rather than copying others' answers. Apparently, the students are interested in this new approach.

Teacher A also expressed a desire to continue using chatbots in the future, which indicates a positive descriptor for her confirmation stage, although she changed jobs after the semester and did not have the opportunity to implement this intention.

Compared to Teacher A, Teacher B had more knowledge and was more positive about chatbot use. She used chatbots, particularly ChatGPT, for her personal tasks such as revising her CV and searching for information. However, she was hindered from using them for teaching by her perceived mismatch between the complexity of language used by chatbots and her students' language proficiency. This concern was addressed during the persuasion stage when she learned about Xiaoming from the researcher and discovered that Xiaoming's conversations were tailored to the content she was teaching. Based on the version of Xiaoming provided by the researcher, she edited some content to help students practice the words and sentences that she deemed important. She recognized the capabilities of chatbots as conversation practice partners and noted:

> To some extent, chatbots can work as better conversation partners than teachers. As a teacher, although I am a native speaker, sometimes I worry about making language mistakes that might mislead my students. Chatbots can provide accurate sentences and avoid accidental errors.

Similar to Teacher B, Teacher C had positive experiences with chatbots in dealing with her personal tasks but lacked the knowledge of using them for language teaching. Her concern about the complexity of using chatbots, due to her preference for straightforward teaching methods, hindered her from using them for teaching. Her concern about the complexity coexisted with her uncertainty about the effectiveness of chatbots, referred to as "the cost-effectiveness ratio" in the previous section. Her attitude changed during the persuasion and implementation stages. She realized that using chatbots was not as complicated as she had perceived when the researcher introduced Xiaoming to her. This realization made her willing to use Xiaoming in her class. Her hands-on experience with the chatbot further strengthened her intention to use chatbots. She noticed Xiaoming's capability to engage students in conversation flows, which distinguished it from other technologies. In the confirmation stage, she continued to use Xiaoming and recommended it to new teachers in the language teaching program.

The teachers' decision processes were influenced by both mass media and interpersonal communications. Mass media played an important role in the knowledge stage, as evidenced by the teachers' widespread awareness of chatbots, especially prevalent ones such as ChatGPT. Interpersonal communications had a greater impact on their adoption of chatbots, as the researcher played an essential role in addressing the teachers'

perceived obstacles and changing their attitudes toward chatbots. This aligns with Rogers' (2003) explanation of these two communicative channels.

5. Discussion

The perceptions and hands-on experiences of the three teachers in this study can be viewed as a microcosm of the language teaching community's encounter with chatbots. It reveals how teachers' attitudes shifted from reluctance to willingness to integrate chatbots into their future teaching practices. This trajectory provides a snapshot of the current state of chatbot diffusion in language teaching, as well as insights into using chatbots as a new tool to facilitate CLT practice.

The discussion of advanced chatbots in mass media, particularly GenAI-empowered ChatGPT, has given teachers a contradictory impression of chatbots, characterized by both open attitudes and concerns about implementation. The discussion has brought chatbots into the spotlight, contributing to a high level of awareness among the teachers at an early stage of technology adoption. This early familiarity is uncommon compared to other technologies, which usually receive less widespread attention. Notably, all three teachers demonstrated awareness of ChatGPT and had personal experience using it. In this regard, their understanding of chatbots appears to be largely limited to ChatGPT, with no mention of other kinds of chatbots. This aligns with Rogers' (2003) assertion that the influence of mass media on the diffusion process is generally limited. This limited impact is also reflected in teachers' behaviors. When it came to practical implementation, teachers were hindered by concerns such as the complexity and compatibility of chatbots, as mentioned before. They seemed not to be motivated to address these concerns independently before the intervention of this study. Nevertheless, these reservations did not equate to a rejection of chatbot use. On the contrary, the teachers maintained open attitudes and curiosity about using chatbots for teaching, as evidenced by their rapid shift in attitudes towards chatbots later. Overall, mass media have led to a cautious yet intrigued stance among teachers, characterized by a general awareness of chatbots' potential, a limited understanding of the broader landscape of this technology, and hesitancy to take action due to perceived obstacles.

Interpersonal communication was the key factor that changed teachers' attitudes towards chatbot use. This aligns with Rorgers' (2003) statement that interpersonal communication was usually more impactful than mass media. In this case study, the authors, who were teacher-researchers, played the role of the change agent in influencing the teachers' chatbot adoption. By introducing a curriculum-informed chatbot, the teacher-researchers helped mitigate the perceived barriers that initially hindered the teachers from adopting chatbots, especially in terms of the compatibility and complexity of integrating chatbots into language teaching. Interpersonal communication also happened between peers, as seen in Teacher A's experience. She reported her reliance on her trusted colleagues, and her decision to use chatbots was influenced by Teachers B and C. These findings align with the broader concept of social intervention as an essential factor in the diffusion of innovations. Social intervention, in this context, refers to the support and communication provided by an expert or peer, which can bridge teachers' initial

perspectives and their hands-on experience with innovations. It is essential during the persuasion stage of the innovation-decision process (Grgurović, 2014; Rogers, 2003).

The presence of the teacher-researcher also demonstrated that a curriculum-informed approach is key to overcoming early-stage barriers to chatbot integration. In this study, perceived challenges related to chatbot compatibility and complexity were mitigated through these curriculum-informed solutions. By grounding the chatbot's use within the curriculum, the teacher-researchers demonstrated how the chatbot could effectively meet the teachers' teaching needs, thus making the technology more relevant and easier to integrate. This approach bridged the gap between the perceived mismatch of the technology and the teachers' instructional needs. Meanwhile, the initial version provided by the teacher-researchers allowed the teachers to test and adapt the chatbot without feeling overwhelmed by the complexity of tailoring it to their own contexts. In this regard, the role of the teacher-researchers was mainly manifested through their enactment of the curriculum-informed approach.

Moreover, chatbot-facilitated CLT introduces a new dimension that extends beyond the traditional categories of instruction-oriented and fun-oriented approaches. While instruction-oriented CLT focuses on structured learning activities and fun-oriented CLT emphasizes engaging and entertaining methods, the use of chatbots enables a socialization-oriented CLT practice. This practice is characterized by providing authentic social interactions that mirror real-life communication. This was emphasized by Teacher B, who highlighted the authentic communication scenarios and formats that simulated daily text-based digital communications. This compensates for fun-oriented CLT, which usually provides tasks or activities based on virtual and decontextualized scenarios (Huang & Liu, 2004) that cannot be directly used in real-life communications. Moreover, chatbots can effectively engage students in communication flows that are aligned with the five features of effective CLT practice listed by Nunan (1991). They provide "anxiety-free environments" (Teacher A) where students can concentrate on communication practice; use authentic materials and "social formats" (Teachers B and C) that help students talk about their own experiences with the target language; and adjust their language complexity to encourage students to work on their own sentences. These features compensate for the limitations of instruction-oriented CLT practice by providing meaningful communicative tasks (Huang & Liu, 2004; Teh, 2021). In this regard, the socialization-oriented CLT practice facilitated by chatbots is characterized by preparing students for real-life social occasions. It allows learners to practice language skills in dynamic, interactive contexts, fostering a sense of connection and engagement. This unique aspect distinguishes chatbots from other technologies used for CLT and may also be leveraged to contribute to sustainable chatbot use in language education.

6. Conclusion, implication, and limitation

To get contextualized insights into the diffusion of chatbots as virtual language partners and fill the gap in studies on teachers' perspectives based on their hands-on experiences (Ji et al., 2023; Mishan & Timmis, 2015; Timpe-Laughlin et al., 2022), this study explores Chinese language teachers' perspectives and use of chatbots as virtual

language partners. From a bottom-up perspective, it has been found that under the heated discussion on chatbots, their implementation in language teaching is still limited. The new wave of chatbots has raised widespread awareness among language teachers, which fosters curiosity and an open attitude toward chatbots. However, this awareness may not be directly translated to chatbot implementation because teachers' knowledge of chatbots is still limited, and some perceived obstacles hinder them from using chatbots for teaching practice. In this regard, social intervention, especially by change agents who can bridge general-purpose chatbots with particular teaching contexts, is essential for further integration of chatbots. This integration is meaningful for language education, given chatbots' potential to enhance CLT practice and their user-friendly feature, which requires no technological expertise. Overall, the findings indicate that it is time to focus on appropriate arrangements for chatbot use rather than technological issues, as noted by Fryer et al. (2020).

This study also provides implications to enhance chatbot integration in language education:

First, a systematic understanding of chatbots is required. It can be seen that teachers' understanding of chatbots is fragmented, mainly based on prevalent ones, such as ChatGPT, while overlooking the diversity of the chatbot family. This family includes both intelligent chatbots and simpler ones like Xiaoming in this study. In this regard, synthesizing and delivering a comprehensive understanding of the broad spectrum of chatbots to language teachers is required (Bibauw et al., 2019). This will lead to a solid understanding of chatbots, effectively addressing the perceived obstacles of teachers.

Second, addressing the lack of how-to knowledge and encouraging hands-on experiences with chatbot use might be the next step to forward chatbot integration in language education. This is significant because, as Timpe-Laughlin et al. (2022) note, "what teachers say they do may not necessarily be the same as what they actually do in the classroom" (p. 1213). They can contribute to further addressing the perceived obstacles of chatbot use, as well as the uncertainty of the effectiveness of chatbot use.

Third, this study highlights the importance of adopting a curriculum-informed approach to tailor chatbots to particular teaching contexts. This is the key step to ensure that chatbots serve as meaningful virtual language partners. It should be the core of teachers' how-to knowledge of chatbot use, which underscores teachers' expertise in guiding the purposeful and pedagogically sound implementation of chatbots.

In terms of limitations, it should be noted that the results in this study come from a small sample of teachers who were involved in a single language program. Thus, the extent to which these findings can be extrapolated to other contexts remains uncertain. Meanwhile, this study focuses on teachers' perspectives and thus lacks responses from students, whose views are also essential for exploring the role of chatbots in facilitating CLT practices. This indicates the demands for future studies to explore various contexts and stakeholder perspectives, as well as those that will provide statistical evidence based on large-scale participants.

Reference

Arnold, N., & Ducate, L. (2015). Contextualized views of practices and competencies in CALL teacher education research. *Language Learning & Technology*, *19*(1), 1–9.

Ayanwale, M. A., & Ndlovu, M. (2024). Investigating factors of students' behavioral intentions to adopt chatbot technologies in higher education: Perspective from expanded diffusion theory of innovation. *Computers in Human Behavior Reports*, *14*, 100396. https://doi.org/10.1016/j.chbr.2024.100396

Bax, S. (2003). CALL—past, present and future. *System 31*(1), 13-28. https://doi.org/10.1016/S0346- 251X(02)00071-4

Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning 32*(8): 827–877. https://doi.org/10.1080/09588221.2018.1535508

Boschman, F., McKenney, S., & Voogt, J. (2015). Exploring teachers' use of TPACK in design talk: The collaborative design of technology-rich early literacy activities. *Computers & Education*, *82*, 250-262. https://doi.org/10.1016/j.compedu.2014.11.010

Chambers, A., & Bax, S. 2006. Making CALL work: Towards normalization. *System 34* (4), 465-479. https://doi.org/10.1016/j.system.2006.08.001.

Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020). A chatbot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research, 58*(6), 1161-1189. http://dx.doi.org/10.1177/0735633120929622

Coniam, D. (2008). Evaluating the language resources of chatbots for their potential in English as a second language. *ReCALL 20*(1): 98–116.

El Shaban, A., & Egbert, J. (2018). Diffusing education technology: A model for language teacher professional development in CALL. *System*, *78*, 234-244. https://doi.org/10.1016/j.system.2018.09.002

Fatemi Jahromi, S. A., & Salimi, F. (2013). Exploring the human element of computer-assisted language learning: An Iranian context. *Computer Assisted Language Learning*, *26*(2), 158–176. https://doi.org/10.1080/09588221.2011.643411

Frei-Landau, R., Muchnik-Rozanov, Y., & Avidov-Ungar, O. (2022). Using Rogers' diffusion of innovation theory to conceptualize the mobile-learning adoption process in teacher education in the COVID-19 era. *Education and information technologies*, *27*(9), 12811-12838. https://doi.org/10.1007/s10639-022-11148-8

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior, 75*, 461-468. https://doi.org/10.1016/j.chb.2017.05.045

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.2018.12.023

Fryer, L., Coniam, D., Carpenter, R., & Lapusneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology, 24* (2), 8–22. http://hdl.handle.net/10125/44719

Gallacher, A., Thompson, A., Howarth, M., Taalas, P., Jalkanen, J., Bradley, L., &

Thouësny, S. (2018). "My robot is an idiot!"–Students' perceptions of AI in the L2 classroom. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds), *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 70-76). https://doi.org/10.14705/rpnet.2018.26.815

Grgurović, M. (2014). An application of the Diffusion of Innovations theory to the investigation of blended language learning, *Innovation in Language Learning and Teaching, 8*(2), 155-170. https://doi.org/10.1080/17501229.2013.789031

Hapsari, I. P., & Wu, T. T. (2022). AI chatbots learning model in English speaking skill: Alleviating speaking anxiety, boosting enjoyment, and fostering critical thinking. In *International Conference on Innovative Technologies and Learning*, 444–453. Springer. https://doi.org/10.1007/978-3-031-15273-3_49

Hsu, M., Chen, P., & Yu, C. (2021). Proposing a task-oriented chatbot system for EFL learners' speaking practice. *Interactive Learning Environments*, 1–12. https://doi.org/10.1080/10494820.2021.1960864

Huang, F., Teo, T., & Zhou, M. (2019). Factors affecting Chinese English as a foreign language teachers' technology acceptance: A qualitative study. *Journal of Educational Computing Research*, *57*(1), 83–105. https://doi.org/10.1177/0735633117746168

Huang, S. J., & Liu, H. F. (2004). Communicative language teaching in a multimedia language lab. *English Language Teaching*, *5*(2), 94.

Hymes, D. (1972). *On communicative competence*. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Harmondsworth: Penguin.

Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 1–16. https://doi.org/10.1080/15391523.2022.2142873

Jwaifell, M., & Gasaymeh, A. M. (2013). Using the diffusion of innovation theory to explain the degree of English teachers' adoption of interactive whiteboards in the modern systems school in Jordan: A case study. *Contemporary educational technology*, *4*(2), 138-149.

Kaminski, J. (2011). Diffusion of innovation theory. *Canadian Journal of Nursing Informatics*, *6*(2), 1-6. https://cjni.net/journal/?p=1444

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Littlewood, W. (2013). Developing a context-sensitive pedagogy for communication-oriented language teaching. *English Teaching, 68*(3), 3-25.

Liu, G. L., Darvin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): a mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 1–29. https://doi.org/10.1080/09588221.2024.2310288

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, *11*(1), Article 50. https://doi.org/10.1186/s40561-024-00342-5

Markee, N. (1992). The diffusion of innovation in language teaching. *Annual review of*

applied linguistics, *13*, 229-243.
Martins, C. B. M. J., Steil, A. V., & Todesco, J. L. (2004). Factors influencing the adoption of the Internet as a teaching tool at foreign language schools. *Computers and Education*, *42*(4), 353–374. https://doi.org/10.1016/j.compedu.2003.08.007
Mishan, F., & Timmis, I. (2015). *Materials development for TESOL*. Edinburgh University Press.
Nunan, D. (1991). Communicative tasks and the language curriculum. *TESOL Quarterly*, *25*(2), 279–295. https://doi.org/10.2307/3587464
Qian, K., Shea, R., Li, Y., Fryer, L. K., & Yu, Z. (2023, June). *User adaptive language learning chatbots with a curriculum*. In International Conference on Artificial Intelligence in Education (pp. 308-313). Springer Nature Switzerland.
Rodríguez Cardona, D., Werth, O., Schönborn, S., & Breitner, M. H. (2019). *A mixed methods analysis of the adoption and diffusion of Chatbot Technology in the German insurance sector*. In the Twenty-fifth Americas Conference on Information Systems, Cancun.
Rogers, E.M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
Sahin, I. (2006). Detailed review of Rogers' diffusion of innovations theory and educational technology-related studies based on Rogers' theory. *Turkish Online Journal of Educational Technology-TOJET*, *5*(2), 14-23.
Savignon, S. J. (1991). Communicative language teaching: State of the art. *TESOL Quarterly*, *25*(2), 261-278. https://doi.org/10.2307/3587463
Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, *149*, 102601. https://doi.org/10.1016/j.ijhcs.2021.102601
Stoller, F. L. (1994). The diffusion of innovations in intensive ESL programs. *Applied Linguistics*, *15*(3), 300-327. https://doi.org/10.1093/applin/15.3.300
Swain, M., & Suzuki, W. (2008). Interaction, output, and communicative language learning. *The handbook of educational linguistics*, 557-570. https://doi.org/10.1002/9780470694138.ch39
Teh, W. (2021). Communicative Language Teaching (CLT) in the context of online learning: A literature review. *International Journal of TESOL & Education, 1*(2), 65-71.
Timpe-Laughlin, V., Sydorenko, T., & Daurio, P. (2022). Using spoken dialogue technology for L2 speaking practice: what do teachers think? *Computer Assisted Language Learning 35*(5–6): 1194–1217. https://doi.org/10.1080/09588221.2020.1774904
Van den Branden, K., Long, M. H., & Doughty, C. J. (2009). Diffusion and implementation of innovations. In *The Handbook of Language Teaching* (pp. 659–672). Wiley-Blackwell. https://doi.org/10.1002/9781444315783.ch35
Van Horn, K. R. (2024). ChatGPT in English Language Learning: Exploring Perceptions and Promoting Autonomy in a University EFL Context. *TESL-EJ (Berkeley, Calif.)*, *28*(1), 1-26. https://doi.org/10.55593/ej.28109a8
Vázquez-Cano, E., Mengual-Andrés, S., & López-Meneses, E. (2021). Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments. *International Journal of Educational Technology in*

Higher Education 18. https://doi.org/10.1186/s41239-021-00269-8

Yin, Q. & Satar, M. (2020). English as a foreign language learner interaction with chatbots: Negotiation for meaning. *International Online Journal of Education and Teaching (IOJET)*, *7*(2), 390-410. http://iojet.org/index.php/IOJET/article/view/707

Yuan Z., Cheng X. & Duan, Y. (2024). Impact of media dependence: How emotional interactions between users and chat robots affect human socialization? *Frontiers in Psychology*. 15:1388860. http://doi.org/10.3389/fpsyg.2024.1388860

Appendix

Interview questions

1. Please introduce your teaching background in CFL, including how long you have been in this field, the institutions you have worked for, the levels you teach, and what you prioritize in your teaching.

2. What do you know about chatbots?

3. How do you use chatbots?

4. Why don't you use chatbots for your teaching?

5. What do you think about Xiaoming in terms of its relative advantages, complexity, compatibility, trialability, and observability?

6. How do you think of chatbots as virtual language partners?

Teaching or Cheating: The Dark Side of ChatGPT as a Learning Companion for Beginner Chinese Students (教学辅助还是误导：ChatGPT 在中文初学者学习中的潜在风险)

Rovira-Esteva, Sara
(罗飒岚)
Universitat Autònoma de Barcelona
(巴塞罗那自治大学)
Sara.Rovira@uab.cat

Abstract: Since its introduction in late 2022, ChatGPT has garnered significant interest among foreign language educators, who have explored its potential to enhance teaching and learning. This exploratory study examines the guided use of GenAI for two written production activities conducted by first-year university students learning Chinese. The research addresses two key questions: (1) whether the system's responses are reliable and of pedagogical quality, and (2) how students interact with the chatbot to seek guidance. The study had four phases: designing structured activities with detailed instructions and evaluation rubrics, annotating and categorizing student prompts, carrying out content analysis of the system's output, and examining student feedback collected through an anonymous survey. The results indicate that students' varying levels of Chinese proficiency and AI literacy had a substantial impact on their outcomes. While ChatGPT occasionally provided high-quality responses, its output was inconsistent, often including random errors or nonsensical results in response to similar prompts. This study offers an innovative perspective by qualitatively analyzing students' prompts and the system's output or feedback from an educator's standpoint. The findings highlight the risks associated with using ChatGPT in uncontrolled settings, where its inconsistent performance combined with students limited critical thinking skills in detecting errors pose significant challenges. Ultimately, the study emphasizes the need for cautious integration of generative AI in education. This problem can be addressed by fine-tuning the system to improve the quality of its feedback and training students to help them develop their AI literacy.

摘要: 自 2022 年末推出以来，ChatGPT 就引发了外语教育工作者的广泛关注，他们开始探索该工具在提升教学与学习效果方面的应用潜力。本探索性研究聚焦于生成式人工智能在两项书面表达活动中的引导性应用，这些活动由一组学习中文的一年级大学生参与完成。研究围绕两个关键问题展开：(1)系统响应的可靠性及其教学质量；(2) 学生如何与聊天机器人互动以获取指导。研究方法分为三个阶段：设计结构

化活动并制定详细的指导与评价标准；对学生针对系统的提问及回应的语言特征进行标注；分析通过匿名调查收集的学生反馈。研究结果表明，学生的中文水平和人工智能素养差异对学习成果有着显著影响。尽管 ChatGPT 有时能给出高质量回应，但其输出稳定性不足，面对相似提示经常出现随机错误或无意义结果。本研究从教育者视角，通过对 AI 与学生互动展开质性分析，提供了一个创新性研究视角。研究结果揭示了在非受控环境中使用 ChatGPT 的风险，其不稳定表现与学生在识别错误方面有限的批判性思维能力共同构成重大挑战。最终，本研究强调了在教育环境中谨慎整合生成式人工智能的必要性，直到开发出更为可靠的系统。

Keywords: ChatGPT, AI-student interaction, TCFL, CSL, AI literacy

关键词：ChatGPT，AI 与学生互动，对外汉语教学，中文作为第二语言，人工智能素

1. Introduction

Since ChatGPT was introduced to the public at the end of 2022, the applications and possibilities of systems based on large language models have continued to grow and evolve. Amid awe and wonder, many foreign language teachers have ventured into experimenting with these systems—not only out of fear of falling behind their students in terms of usage but also to explore how useful these tools can be for both students and educators while simultaneously evaluating their risks and limitations.

A literature review was conducted, encompassing a total of 25 contributions, primarily academic journal articles, along with a selection of master's theses and conference presentations. The review aimed to identify the main methodologies, topics, and findings to contextualize the study's starting point. The studies included in the review focus on the use of generative AI (GenAI) for learning Chinese as a second language (CSL) and were all published between 2023 and 2025. Although recent, they capture a reality that has been rapidly evolving throughout this period and continues to do so.

A large portion of the reviewed studies highlight the numerous possibilities and potential of GenAI, but they often provide limited information about their objectives and the methodologies used. Among the applications or skills studied, we find speaking (6), vocabulary (5), writing (5), reading (5), materials and course design (4), grammar (3), AI literacy (2), critical thinking (1), and intercultural communication and pragmatics (1), among others. Many scholars converge on several key themes regarding the use of ChatGPT in CSL teaching, emphasizing its disruptive impact while advocating for cautious and well-planned implementation. Below is a synthesis of the main points.

ChatGPT is widely recognized for its potential to enhance CSL teaching by offering individualized learning opportunities, generating tailored instructional materials, and supporting differentiated instruction (Gao, 2024; Li, Zhang & Cai, 2024; Meng, 2024; Ou et al., 2024; X. Wang, 2024; Xu & Ma, 2023). Its ability to create multimodal learning resources and address various language skills, including vocabulary and pragmatics, is also noted (Lee & Cook, 2024; Ou et al., 2024).

ChatGPT can generate instructional materials for listening comprehension, reading comprehension, and other language tasks, with its effectiveness varying by task type and prompt specificity (Casas-Tost et al., 2023; Casas-Tost et al., 2025; Guo, 2024; Li et al., 2023; L. Wang, 2024).

The effectiveness of ChatGPT heavily depends on prompt design, with several studies offering frameworks and examples to guide educators in maximizing its utility (Koyuturk et al., 2023; Li, 2024; Wang & Williams, 2024). Many studies stress the importance of using ChatGPT as a supplementary tool rather than a replacement for human expertise, emphasizing the need for teacher intervention to adapt and enhance AI-generated content (Gao, 2024; Meng, 2024; X. Wang, 2024; Zhao et al., 2024). Limitations such as inaccuracies, ethical concerns (e.g., plagiarism), and risks of over-reliance on technology are consistently highlighted (Gao, 2024; Hellmich et al., 2024; Liu, 2023). In this context, some authors argue that the exploration of the full potential of ChatGPT and other AI tools in CSL education must include rigorous and systematic evaluation of their capabilities and impact (Liu, 2023; Xu & Ma, 2023).

Teachers play a critical role in integrating ChatGPT into the classroom, ensuring its output aligns with curricular goals and meet students' needs (Xu & Ma, 2023; Zhao et al., 2024). Effective use requires teachers to critically review and adapt AI-generated content (Tseng & Warschauer, 2023; X. Wang, 2024). Therefore, it is essential for educators to develop strong digital competencies—particularly AI literacy—in order to use ChatGPT responsibly and effectively. Many studies highlight not only the need for professional development to equip teachers with the skills to navigate the ethical and pedagogical challenges posed by GenAI (Chen, 2023; Gao, 2024; Liu, 2023; Matthews, 2024), but also the need for teachers to foster students' critical engagement with it, helping them understand its strengths and limitations while addressing concerns about its implications for learning and ethics (Donley, 2024; Hellmich et al., 2024).

In a systematic review and meta-analysis aimed at synthesizing research findings on the impact of ChatGPT interventions on student learning, Deng et al. (2025) found that students generally exhibit positive attitudes toward ChatGPT, while instructors tend to hold more ambivalent views. However, perceptions and attitudes alone do not provide concrete evidence of ChatGPT's actual impact on learning. Furthermore, cross-sectional research identified both positive and negative correlations between ChatGPT usage and academic performance.

In this context, the author conducted an exploratory study on the guided use of GenAI in two written production activities completed by first-year university students of Chinese. The study was guided by two main research questions. First, are the system's

responses reliable and of sufficient pedagogical quality for students to interact with it autonomously. Second, to what extent do students demonstrate adequate AI literacy to engage effectively with the chatbot and obtain meaningful guidance.

AI literacy has been defined "as a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" (Long & Magerko, 2020, 2). However, this definition predates the emergence of GenAI systems and their widespread integration into foreign language teaching and learning. The author proposes that AI literacy in the context of CFL as the ability to understand, evaluate, and effectively use AI tools and systems to support language acquisition and intercultural communication, while also remaining critically aware of their limitations, potential biases, and ethical implications.

The primary aim of this study is, therefore, to assess the quality of the system's responses to student queries from a pedagogical perspective. A subsidiary objective is to examine the nature of student prompts to the chatbot. In both instances, the author sought to gain a general overview through qualitative analysis, intentionally excluding a detailed discourse analysis of the student-chatbot interaction and a quantitative assessment of the system's correct or accurate responses. This decision was made to focus on the content provided by an untrained system, as well as how inexperienced students conveyed their needs to the system through their prompts.

The study was divided into three phases. First, two written production activities were designed (one mid-semester and the other at the end), providing students with detailed instructions and an evaluation rubric. Second, students' tasks were collected, and their prompts to the system were logged in a spreadsheet, along with the system's most relevant responses related to Pinyin transcription, vocabulary, punctuation, terminology, grammar, and pragmatics. Since the author's approach was primarily qualitative, only non-repetitive outputs were recorded. Finally, an anonymous survey was administered to students to gather their feedback on the experience, and the results were triangulated with the rest of the collected data.

The rest of the paper is organized as follows. It begins with an introduction to contextualize the study, followed by a methodology section that details the context, participants, study design, and data collection and analysis. The results are divided into four sections addressing the author's two objectives. The author's analysis of ChatGPT's didactic content in an uncontrolled setting highlights areas where the system falls short of quality standards (Section 3.1) and where it excels and adds value to learning (Section 3.2), addressing the first objective. The third and fourth sections address the author's second objective, focusing on data related to students' AI literacy, derived from their prompting to the system (Section 3.3) and their subjective assessment of the experience via a questionnaire (Section 3.4). The article concludes with a discussion of the findings' implications and a conclusion summarizing the main results, outlining study limitations, and suggesting future research directions.

2. Methodology

2.1 Context and participants

The researcher has 30 years of experience teaching CSL. The study was conducted during the first half of the academic year 2023-2024 with a first-year Chinese group from the bachelor's degree in Translation and Interpreting at the Autonomous University of Barcelona. The group had a total enrolment of 35 Spanish or Catalan-speaking students, eight of whom reported having studied Chinese previously, including two heritage language learners. However, only around 20 were regularly attending classes at the time of the second activity under study.

The instructional manual used was *Lengua china para traductores* (Casas-Tost et al., 2024), which students primarily use in print, although it is also available as an open-access digital version.[1] At the time of the first activity, the first three lessons had been completed, just before the first midterm exam. The second activity took place after completing Lesson 5, right before the final exam. Both activities were graded and, together with eight other graded activities, contributed to the continuous assessment score, which accounted for 30% of the final grade.

The participation rate in the first activity was 91% (32 students), while in the second activity, it dropped to 77% (27 students). The average score was 8.6/10 in both cases, which is significantly higher than the average score (6.4/10) of the other three activities conducted without the use of GenAI tools during the semester.

Feedback collected through the form some students completed at the end of each activity revealed that 14 (41%) had little to no prior experience using GenAI systems for non-academic activities, while 19 (56%) had never or rarely used GenAI systems for academic purposes before. Additionally, 10 students (29%) reported experiencing some difficulty using the GenAI system to complete the activities.

2.2 Activity description and assessment

Both activities involved using GenAI and students were allowed to use any other GenAI system to their like to create, refine, and evaluate sentences in Chinese, following specific linguistic and grammatical guidelines based on class content. In the first activity (see Appendix 1), students were asked to independently translate given sentences into Chinese, transcribe them in Pinyin, and ensure accuracy by interacting with the chatbot to correct any errors while adhering to the vocabulary and structures from Lessons 1–3. In the second activity (see Appendix 2), students had to use the chatbot to generate five original sentences of 10–15 characters each, providing specific vocabulary and grammatical structures for the chatbot to use, ensuring no repetition of grammar points, and transcribing the results in Pinyin. In both tasks, students had to critically engage with the chatbot, ask the system to correct identified errors, and document the entire interaction

[1] The manual can be downloaded for free from https://publicacions.uab.cat/llibres/lengua-china-para-traductores-volumen-7a-ed

in a PDF file, including their final sentences in Chinese, Pinyin, and their translations, alongside their observations on vocabulary, grammar, and transcription accuracy.

The activity assessment focused on five key areas (see Appendix 3 for the rubric applied): linguistic accuracy, absence or presence of lexical and grammatical errors in the final Chinese sentences (40%); lexical richness and variety, i.e., diversity and appropriateness of vocabulary and structures used, with penalties for repetition or sub-level usage (10%); adequacy to level, i.e., alignment of constructions and vocabulary with the course content, with deductions for exceeding or not meeting the expected level (20%); Pinyin accuracy, i.e., adherence to official Pinyin orthographic rules, penalizing minor or significant transcription errors (10%); fulfilment of instructions, i.e., compliance with the task instructions and completeness of all required results (20%). Each area was scored based on performance levels, from not acceptable to excellent.

An analysis of the different aspects evaluated in the rubric individually reveals that an average score of 8.3/10 was achieved in linguistic accuracy by the students. This is a relatively high value, though not as high as might be expected given that they used a GenAI system to generate the responses under evaluation. For lexical richness and variety, students achieved an average score of 9.8/10. Regarding adequacy to level, the average score was 9.4/10, which is relatively high considering that one of the issues identified by some authors (Casas-Tost et al., 2023; Casas-Tost et al., under review) is ChatGPT's tendency to produce text at a higher level than specified by the user. In fourth place, concerning Pinyin accuracy, students received the lowest average score (8.1/10), as ChatGPT either fails to correct their mistakes or the system itself makes errors according to official Pinyin orthography. Finally, in the fulfilment of instructions category, the average score was also 8.1/10, indicating that some students did not follow all the required steps outlined in the instructions to complete the activities. This suggests that, even when teachers provide clear guidelines or example prompts to achieve optimal results, we cannot guarantee that students will effectively use the information provided to them.

2.3. Data collection and analysis

The students submitted their homework via a Moodle platform in PDF format, with lengths ranging from 20 to 30 pages, making it a relatively large analysis corpus for manual processing. The author then evaluated both the results and the students' compliance with the instructions, providing them with both qualitative and quantitative feedback. Additionally, the author recorded in a spreadsheet the parts of the interactions that emerged as noteworthy—whether due to formal or substantive aspects—considering them relevant for analysis. The author noted the system's successes as well as various types of errors or inaccuracies, including Pinyin transcription, vocabulary, punctuation, terminology, grammar, and pragmatics, alongside the prompts provided by the students to the system.

To anonymize the participating students (who were previously informed that their data would be recorded for research purposes), the author assigned each of them a unique code in a sequential manner ranging from S1 to S59. Although the analysis was conducted in two separate phases immediately after the completion of each activity, and the results were recorded in two different spreadsheets, this article treats them as a single dataset to

simplify the data presentation and analysis. It is worth noting that, even though the instructions for the two activities were not identical, the results were similar.

According to the activity instructions, students were also expected to include a direct link to their interaction with the GenAI system. However, some failed to do so, or the links provided were broken and could not be recovered. To ensure transparency and traceability, Appendix 4 provides an anonymized list of the participating students, along with the links to their submitted interactions.

Lastly, the researcher aimed to gather insights into the informants' experience, perceptions, and challenges related to using AI systems for learning Chinese. To achieve this, a survey was created using Google Forms. The survey was designed to assess participants' familiarity with GenAI systems for both academic and non-academic purposes, their interest and engagement in AI-based Chinese learning tasks, and their ability to effectively use these tools. Furthermore, it explored their willingness to participate in similar activities in the future, their interest in receiving additional training, and their likelihood of continuing to use GenAI for language learning. Open-ended feedback was also solicited to gain a deeper understanding of participants' views on the utility and practicality of GenAI in their learning experience.

3. Results

3.1 The dark side of ChatGPT as a learning assistant

To address the first research question—namely, if GenAI system responses were reliable and of sufficient pedagogical quality for students to interact with them without supervision—a content analysis of ChatGPT responses to student queries was conducted. Given the extensive nature of the analysis corpus and the repetition found in both the instructions provided by students and ChatGPT's responses, this section focuses on presenting a few representative examples of the aspects considered during the analysis, namely, terminology, Pinyin, punctuation, vocabulary, grammar, and pragmatics. Except for one Chinese student, all other interactions were conducted in either Catalan or Spanish. All excerpts from the interactions have been translated by the author of this article and are presented within inverted commas to facilitate the identification of ChatGPT's responses, although they are not direct quotations. Readers can, however, access the original interactions in Spanish or Catalan by consulting the list of links provided in Appendix 4.

3.1.1 Terminology

It is vital that students are not hindered by ad hoc or uncommon terminology, allowing them to adapt seamlessly to various teachers and textbooks. Furthermore, a core objective is to prepare them for the job market as competent professionals. For those using Chinese as their primary working language, developing robust metalinguistic skills—enabling precise discussions about the language itself—is paramount. For these reasons, the author specifically investigated ChatGPT's rigor in using consistent academic terminology.

The author has noted that ChatGPT sometimes refers to tones in Chinese as *accents* (S1) and confuses the concepts of *character* and *word*, as well as *translation* and *transcription* (S10). While characters and words are closely related, they are not strictly synonymous; many characters in Modern Standard Chinese correspond to bound morphemes and, therefore, cannot function as words on their own.

Regarding terms for parts of speech in Chinese, ChatGPT in some occasions uses *counter* or *quantifier* instead of *measure word* (S1), refers to adjectives as verbs in comparative structures (S4), classifies the negative adverbs 没 (*méi*) and 不 (*bù*) as particles (S11), and incorrectly identifies the classifier for *jar* (壶, hú) in the sentence under analysis as 一 (*yī*) (S36), which is a numeral.[2]

In terms of syntax, ChatGPT invents syntactic constituents that do not exist in the most widely recognized and authoritative Chinese grammars. For example, it analyses 为什么 (*wèishénme*) in 你为什么喜欢这个? (*"Why do you like this?"*) as a "complement of reason" rather than correctly identifying it as an adverbial (S11). Additionally, the system claims that interrogative pronouns can function as *complements of place or time*, offering the example 你去哪儿? (*"Where are you going?"*). However, according to most widespread Chinese grammar approaches, the pronoun 哪儿 (*nǎr*) in this sentence functions as an object of the verb 去 (*qù*), not as a complement (S11).

3.1.2 Pinyin transcription

The correct use of the Pinyin transcription system, as outlined in the official document *Basic Rules of the Chinese Phonetic Alphabet Orthography* (General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China & National Standardization Administration, 2012), is essential for ensuring accurate pronunciation and effective communication in teaching CSL. Adhering to the Pinyin standard is critical to prevent misunderstandings and to facilitate students' learning, as incorrect transcription can lead to errors in pronunciation and comprehension, thereby negatively affecting the educational process. Given the importance of this aspect, the author believes accurate Pinyin transcription should be emphasized during instruction and applied systematically and rigorously by teachers. To this end, ChatGPT's performance in this area was analyzed, and these are the issues identified.

As far as capitalization and punctuation is concerned, ChatGPT does not correct students' improper capitalization in Pinyin, such as failing to capitalize the first letter in a sentence or names of countries or languages (e.g., *zhōngguó* for China or *zhōngwén* for Chinese, instead of *Zhōngguó* and *Zhōngwén*, respectively) (S11). Moreover, if students submit Pinyin with incorrect capitalization, ChatGPT retains those errors in its feedback instead of correcting them.

[2] The author wonders if this error is due to the system confusing the sinogram 壶 (*hú*, teapot) with the traditional form of the numeral 一 (*yī*), which is written as 壹. The authors thanks Antonio Paoliello for bringing this to their attention.

ChatGPT occasionally suggests that a correct transcription should be amended but then provides the exact same transcription, which can be confusing for students. For example, it states that *liǎng* should be changed to *liǎng* with the third tone (S10). However, the most problematic issue is that ChatGPT often provides inaccurate or misleading explanations that completely contradict the official rules. For instance, it claims that *Zhōngguórén* (Chinese person) must be separated into *Zhōngguó rén* instead of being written as a single unit, arguing that in the Pinyin system, "each character is written as a separate word to maintain clarity" (S6). This explanation is both incorrect and misleading.

The system also fails to provide accurate corrections and explanations regarding the neutral tone and tone sandhi. For example, in an interaction with one of the students (S10), ChatGPT states:

> The *bù* (不) can be written without a tone mark when it is used as an interfix in word formation, especially when followed by words that start with the fourth tone. In such cases, it is pronounced as *bu* without the tone mark. For example, in 不错 (*bù cuò*, 'not bad'), the *bù* does not carry the tone mark because '*cuò*' is in the fourth tone. So, in summary, *bù* is marked with a tone when pronounced independently and the tone needs to be indicated, but when it is part of a compound word where the following character begins with the fourth tone, it can simply be written as *bu*.

In fact, 不 (*bù*), which is originally pronounced in the fourth tone, is pronounced in the neutral tone when used as an interfix or to form alternate questions, and it is pronounced in the second tone when it precedes another fourth tone. The system also does not seem to account for tone changes in certain morphemes in words like 妈妈 (*māma*) or 爱人 (*àiren*), which it transcribes as *māmā* and *àirén*, respectively (S15).

ChatGPT misinterprets tone rules, providing convoluted and often incorrect explanations. In Chinese, the tone change associated with the neutral tone represents a semantic shift, where a character's tone changes based on its meaning or grammatical function. It is not context-dependent like tone sandhi but rather inherent to the specific use of the word. In the explanation provided above to students S10 and S15, the system conflates two distinct phonological phenomena: neutral tone and tone sandhi. This misinterpretation demonstrates a lack of understanding of the rules governing tone behavior in Standard Chinese, further confusing students and undermining the reliability of the system's responses. Moreover, according to Pinyin orthographic rules, tone changes should generally not be reflected in the transcription.

Sometimes, the Pinyin provided does not match the Chinese characters (e.g., transcribing 你的爱人 as "*nǐ àiren*" instead of "*nǐ de àiren*") (S10). The author also found an example where it did not conform to the standard transcription practices for foreign

institution names, exemplified by rendering *Universitat de Barcelona* as *Bākèlún Dàxué* instead of *Bāsàiluónà Dàxué* (S10).[3]

Other examples of ChatGPT failing to adhere to Pinyin rules include its omission of apostrophes where required, particularly in cases where syllables need separation for clarity, as in *nǚ’ér*. When asked about the use of apostrophes, ChatGPT responded (S10):

> The apostrophe is used to separate syllables when one of them carries a different tone. In this case, *nǚ* is a syllable with a third tone, and *ér* is a syllable with a second tone. The use of the apostrophe in Pinyin applies to other similar examples, such as *wǒ'ér* (我儿) and *lǎo'ér* (老儿).

The apostrophe is used when the second syllable begins with a zero initial (a vowel sound), regardless of its tone. Therefore, in this case, the system not only fabricates a rule but also provides an example (*wǒ’ér*) using a non-existent word—something that the student, being inexperienced, was unable to detect.

When asked to revise Pinyin transcriptions, ChatGPT occasionally failed to detect errors made by either the students or the system itself. It often did not provide the correct form unless explicitly prompted, and sometimes not even then. In summary, the issues outlined above highlight significant gaps in ChatGPT’s ability to handle Pinyin accurately and its tendency to provide flawed explanations.

3.1.3 Punctuation

In six cases, issues with punctuation were identified, all of which revolved around the same problem. When translating the sentence “Your wife is five years younger than you, isn’t she?” into Chinese, the students preceded the final interrogative particle *ba* with a comma probably due to negative transfer from their native language (*Tu esposa es cinco años menor que tú, ¿verdad?*). ChatGPT not only failed to correct the error but also provided contradictory information on the matter and justified why the comma should be included in Chinese in a similar manner across different interactions. For example:

- “No, in Chinese, *ba* (吧) and *shì bù shì* (是不是) are not separated by a comma from the main clause. Here’s how it’s written correctly: Correct: *Nǐ àiren bǐ nǐ xiǎo wǔ suì, ba?* Neither *ba* nor *shì bù shì* requires a comma before them when used at the end of a sentence to form a question” (S10).

- “The comma in the phrase 你爱人比你小五岁，吧? is not strictly necessary. However, using it can help separate the two parts of the sentence and make it clearer” (S15).

[3] At the time of designing the activities, ChatGPT did not allow the option to browse the Internet; therefore, the instructions did not mention anything in this regard. It is unknown whether students had access to this option and whether they enabled it while interacting with the system. However, the author’s hypothesis is that with Internet access, this type of errors should be minimized.

- "In many cases, you can use a comma before 吧 (ba) to indicate a pause, especially if you're presenting a statement followed by a confirmation" (S18).

Any of these formulations could be genuinely confusing and misleading for students, as they contradict basic punctuation rules in Modern Standard Chinese. Unfortunately, most students accepted these explanations as accurate.

3.1.4 Vocabulary

In the area of vocabulary, ChatGPT likely demonstrates its best performance, at least at this level, as the errors detected in this domain are significantly fewer in both number and severity. However, it is worth noting that in one instance, the system informed the student that their sentence contained a Pinyin error, when in fact it was a lexical selection error caused by using the numeral *èr* (二) instead of *liǎng* (两) (S1).[4] Moreover, the explanation it provided for making this correction was incorrect. In other words, while ChatGPT successfully identified the error, it neither categorized it properly nor justified the correction appropriately.

In the remaining cases, the main issue was the use of vocabulary that was beyond the students' level or even outside the lists they had provided in their prompts, despite their repeated efforts to ensure the system adhered to those lists.

3.1.5 Grammar

In its grammar corrections and explanations, the author observed the highest number of missteps by ChatGPT. These range from telling students that a correct sentence is incorrect and then either offering the exact same sentence or making unnecessary changes to sentences that had no errors (1), failing to fix their incorrect sentences (2), creating grammatically incorrect sentences (3), providing explanations that make no sense whatsoever or clearly contradict Chinese grammar rules (4), or even offering examples that fail to illustrate the intended grammar point (5). Below, examples for each of these phenomena will be provided.

Regarding the first issue, the author identified several instances in which ChatGPT made unnecessary changes to students' homework. For example, when a student asks the system to correct their sentence, it responds: "The sentence 你爱人比你小五岁，是不是? is almost correct, but there is a small adjustment to make it sound more natural. You could say: 你爱人比你小五岁，是吗?" (S15).

In another case, when a student asks ChatGPT to remove the character 道 (*dào*) from a sentence because they do not understand it, the system replies: "I have removed the character 道, and the sentence still retains its meaning," offering *这菜多么好吃！(How

[4] The original interaction in Spanish is as follows. Student's prompt: "我妈妈比你妈妈大二岁。 *wǒ māma bǐnǐmāma dà èr sui*". ChatGPT's answer: "Tu frase es casi correcta, pero hay un pequeño error en el uso del pinyin. El correcto sería: 我妈妈比你妈妈大两岁。*wǒ māma bǐnǐmāma dà liǎng suì*. Correcciones: "二" (*èr*): Se usa para "dos" en general, pero cuando se habla de la edad, es más común usar "两" (*liǎng*)".

delicious this dish is!) as a grammatically correct sentence without further explanation of why the measure word 道 (*dào*) should be used in Modern Standard Chinese (S59).

When the supposed learning companion is expected to correct the student's sentences, numerous examples where this does not happen were encountered, and the system accepts constructions that are far from genuine in Standard Chinese. For instance, a student asks ChatGPT to do something impossible, namely, to use a numeral other than *liǎng* in a given context. In response, the system generates the sentence: *我妈妈比你妈妈大俩岁 (My mom is two years older than yours), justifying its choice by claiming that 俩 (*liǎ*) is a colloquial way of saying *two* (S9). However, the use of this word is not correct in this context. The system also accepts student-produced sentences like *我妈妈比你妈妈大二岁 (My mother is two years older than yours) as correct, even though the choice of numeral 二 (*èr*) is incorrect (S9).

Regarding ChatGPT's generation of grammatically incorrect sentences, a wide variety of instances were detected. For instance, when providing an example of a comparative of inferiority, it invented the following sentence: *这座房子没有那座房子老得不太多 (This house is not as old as that house) (S15). Other similar cases include: *我应该什么去家？ (*What should I go home?) and *孩子是大的。 (The children are big) (S1).

In one case, when a student insisted that the system offer alternative formulations to the sentences they had written— all of which were correct—ChatGPT ended up providing aberrant alternatives or so-called hallucinations[5] such as: *你女儿多大几岁？ (How old is your daughter?), which it justified by saying, "although this sounds a little strange, it retains the same vocabulary." It also suggested *我妈妈大你妈妈两岁 (My mom is two years older than yours), describing it as "a restructuring that keeps the same vocabulary", or *你爱人小五岁比你吧？(Your spouse is five years younger than you, right?), explaining it as "a less common but possible way to reorder the sentence". Fortunately, in this case the student (S32) had enough language and critical thinking skills to question these sentences, prompting the system to finally respond:

> My apologies for the earlier confusion. Interspersing the characters in the way I suggested in Section 1 is not correct in Chinese. The structure of the language is quite rigid, and characters cannot be rearranged in this way without losing the meaning of the sentence.

The activity proposed by the teacher included asking the system questions about grammar and vocabulary. In this aspect, some explanations that did not align at all with Chinese grammar were found, as the illustrative examples below.

When providing an example of how to construct a comparative of inferiority, the system writes: *他比我不高 (He is not taller than me) (S25) instead of 他没有我高 (He

[5] In the context of GenAI, a *hallucination* refers to system-generated content that appears fluent and plausible but is factually incorrect or contextually inappropriate.

is not as tall as me) or 他不比我高 (He is not taller than me). In another case, when asked how to form comparative sentences of superiority, it explains: "When you use 比 (*bǐ*) for comparisons of superiority, you can add adverbs like 非常 (*fēicháng*), 很 (*hěn*), or 特别 (*tèbié*) to intensify the adjective" (S15), which is not true.

When explaining the comparative of inferiority, it provides an incorrect formula, stating: "The structure for a comparison of inferiority with 没有 (*méiyǒu*) would be: Subject A + 没有 (*méiyǒu*) + Subject B + Adjective + 那么 (*nàme*)," and offers *这座房子没有那座房子老那么 (This house is not as old as that house) as an example (S15). In this construction, 那么 (*nàme*) should always be placed before the adjective.

Regarding the particle 了 (*le*), it states that "it can be used to ask if someone has had a certain experience, often in combination with verbs" (S18), which would more accurately describe the aspectual particle 过 (*guo*). As for 的 (*de*), another particle that is particularly challenging for students, the system explains in one instance that it "connects long phrases with nouns to provide context" (S20), rather than clarifying that it's used to connect a noun with the element modifying it. In another case, when demonstrating its various uses to a student, it provides the sentence *这的老师很有名 (This teacher is famous) as an example (S5), placing the particle where a measure word should go.

The final notable aspect of ChatGPT's performance as a virtual tutor for learning Chinese grammar is that the examples it provides often do not align with the grammar points it aims to illustrate. For instance, when asked to provide an example of a complement of degree (程度补语), ChatGPT offered the student the sentence 我非常喜欢喝啤酒，常常和朋友一起干杯 (I really like drinking beer, and I often toast with friends), analyzing it as follows: "Subject (我) + Verb (喜欢) + Complement of degree (非常) + Verb (喝) + Object (啤酒) + Circumstantial Complement (常常和朋友一起干杯)" (S55). Although the sentence provided as an example is grammatically correct, it does not illustrate the use of the complement of degree in Chinese, and the explanation, overall, does not make sense.

To another student, the chatbot provided the following incorrect sentence as an example of a complement of manner (情态补语): *老师不好，所以大家都不去上课上得认真 (The teacher is not good, so nobody attends class seriously) (S48). In another case, as an example of the use of the resultative complement (结果补语), the student was presented with the sentence 她是餐厅的老板，而且还是我的朋友 (She is not only the boss of the restaurant but also my friend) (S58). The issue with these examples is that none of them effectively illustrates the grammatical points they intend to demonstrate.

To close this section on the grammatical aspects of the analysis, the author would also like to highlight a feature that is, in principle, pedagogically beneficial: ChatGPT often presents information using formulas or schematic structures, even when not explicitly prompted to do so. However, the problem is that with each interaction, it tends to use a different format to explain the same grammatical point, which may lead to confusion for students. For example, regarding different types of comparatives, the following formulations, among others, were found:

1. [Element A] + 比 + [Element B] + [adjective/complement], which is partially incorrect as it is formulated (S2)

2. [subject 1] + 比 (*bǐ*) + [subject 2] + [adjective] (S26)

3. A + 比 (*bǐ*) + B + Adjective + (optional amount) (S28)

4. A 比 B + Adjective + C, where C refers to the quantity or measurement (S29)

5. Subject A + 比 + Subject B + Adjective+ Quantity (S37)

Note that the system uses different terms to refer to the concept of complement of quantity (数量补语), including *complement*, *optional amount*, *quantity or measurement*, and *quantity*.

3.1.6 Pragmatics

Although both activities involved writing sentences without context, some issues that could be categorized as pragmatic errors were identified, i.e., the appropriateness of language in relation to the discursive or cultural context. For example, regarding the different ways of asking about age, ChatGPT claims that it is better to say *你爸爸几岁? (*Nǐ bàba jǐ suì*?) than 你爸爸多大? (*Nǐ bàba duō dà*?) (S9) or that 几岁 (*jǐ suì*) is an informal way of asking, where the use of 您 (*nín*) makes it respectful, while 你多大? (How old are you?) is more direct and informal, suitable for children (S11). ChatGPT comments on the sentence 你爸爸多大? (How old is your father?) from another student with the following: "This is correct, but a more natural translation of 'How old is your father?' would be *Nǐ bàba jǐ suì*?" and justifies it with these words: "So, although it is more common to use 几岁 for young children, it is not limited to that age group" (S13). As a matter of fact, the usage in mainland China is quite the opposite. Moreover, if there are differences in usage across Chinese-speaking regions, the system should highlight this to students, enabling them to learn the most appropriate way to ask in each cultural context, thereby further developing their pragmatic skills.

Regarding the modal particle 了 (*le*), whose usage is better explained in terms of discourse rather than grammar, the research corpus contains a couple of noteworthy examples. In one case, ChatGPT modified a student's correct sentence about age by adding this modal particle at the end with the following argument: "The presence of 了 in questions of this type softens the sentence and makes it sound more polite and natural, as it implies that the current situation or state of the topic in question is being discussed" (S12). Meanwhile, for another student, it provided the following justification for the change: "So, while it is not mandatory, 了 often makes the question friendlier and more colloquial" (S4). From my perspective, neither explanation sufficiently describes the most plausible pragmatic function 了 (*le*) fulfils in this kind of context, which is to update the interlocutor with new information relevant to the communicative act, typically following the Chinese

topic-comment information structure. In other words, the particle 了 (*le*) can be more accurately explained or understood as a modal particle that marks the end of the comment.

Other examples within the realm of pragmatics include: *我喜欢吃书或者东西 (I like eating books and things) (S59) and *我怎么去家 (How do I go back home?) (S59). The first sentence is grammatically correct but highly implausible from a pragmatic point of view. The second does not conform to the usual way Chinese speakers express this idea, as they typically use the verb 回 (*huí*) when referring to returning to their own home.

3.2 The bright side of ChatGPT as a learning assistant

Despite the inaccuracies and errors highlighted so far, which should caution us against relying on ChatGPT excessively and without critical thinking, there are positive aspects worth mentioning. As these systems continue to evolve and improve in areas where they currently fall short, the elements that already work well will become even more valuable assets as virtual tutors in the future.

It must be acknowledged that, from a formal perspective, the system offers quite interesting features. For example, in terms of typography and formatting, it is very clear and organized. By default, it often makes use of lists, tables, formulas, and summaries to illustrate the material in a more schematic and clarifying manner.

On certain occasions, ChatGPT employs contrastive linguistics to explain grammar points. For instance, it explains to a student: "没有 is used to express that something or someone 'is not as...as' something or someone else. It is similar to 'not being as...as' in Spanish" (S19).[6] Similarly, when asked about the uses of the particle 的 (*de*), it responds: "The particle 的 (*de*) is one of the most common and versatile particles in Chinese. Its primary function is to indicate possession or connection between words, acting similarly to the apostrophe + 's' ('s) in English or 'de' in Spanish" (S20).[7]

Another example of good practice is its explanation: "The structure 是不是 (*shì bù shì*) in Chinese is primarily used to form yes/no questions and is equivalent to '¿right?' or '¿isn't?' in Spanish" (S25).[8]

However, the main issue is that the comparison language used is somewhat arbitrary and does not always consider the student's native language or the language of interaction. Students may not necessarily have sufficient knowledge of English for these explanations to be fully effective. For instance, in response to a question about the uses of the particle

[6] Original system's answer: "没有 se usa para expresar que algo o alguien 'no es tan...como' otra cosa o persona. Es similar al 'no ser tan...como' en español" (S19).
[7] Original system's answer: "La partícula 的 (*de*) es una de las partículas más comunes y versátiles en chino. Su función principal es indicar posesión o conexión entre palabras, actuando de manera similar al apóstrofo + 's' ('s) en inglés o 'de' en español" (S20).
[8] Original system's answer: "La estructura 是不是 en chino se usa principalmente para formular preguntas de tipo sí o no y equivale a '¿verdad?' o '¿es cierto que...?' en español" (S25).

de (的) posed in Catalan, the system included the following explanation: "Function: It marks possession, similar to the apostrophe 's in English" (S3).[9]

ChatGPT has also demonstrated strengths in certain explanations at the graphemic and lexical levels. For example, when a student asked about the structure of the character 追 (*zhuī*), it explained that it means "to chase" or "to follow" and is composed of two components: 辶 (*chuò*), known as the "walking radical" or "to walk", which appears in many characters related to movement or the act of walking, and 隹 (*zhuī*), which refers to a type of bird, noting that in this character it serves as the phonetic component (S49). However, while the explanation is correct, it is likely not comprehensive enough, as ChatGPT fails to mention that 自 is an allograph of 隹. To make the explanation fully understandable for first-year students, the system should supplement or expand its argument.

ChatGPT often provides accurate grammatical or usage explanations. For instance, when asked about the difference between the two possible ways to express the numeral *two* in Chinese, it explained: "两 (*liǎng*) is used when counting or measuring something specific (such as age, objects, or people) and is always followed by a measure word or a noun. (...) Here, 两 precedes 岁 (*suì*), which is a unit of measurement for age (years)" (S16).

Regarding ways to ask about age, the system explains: "To ask the age of children or young people (generally under 10 years old): Structure: Subject + 几岁? (*jǐ suì*)" (S16). To another student, it clarified: "If the daughter is older, especially if she is a teenager or an adult, it would be more appropriate to use 多大 (*duō dà*) instead of 几岁 (*jǐ suì*)" (S18).

The author also noted that in one interaction, although it did not explicitly clarify that the student's sentence was incorrect, it offered a modification that improved it, stating: "Although it's not a critical error, using a comma before 吧 could sound more natural if you replaced it with 对吧 (*duì ba*) to make it sound more fluid" (S47).

In conclusion, ChatGPT's responses vary in accuracy, necessitating constant vigilance during interactions. Consequently, to utilize ChatGPT effectively as a learning companion, users must possess sufficient AI literacy skills to mitigate the risk of misinformation.

3.3 Gauging students' AI literacy through their prompting

Although students were free to use any GenAI system, all but one chose ChatGPT, with a single student opting for Gemini. To maintain a homogeneous sample, this case was excluded from the analysis of students' prompts and system outputs. In completing the two activities according to the teacher's guidelines, students engaged with ChatGPT in various ways, which can be grouped into the following key patterns:

- Interaction and iterative refinement through prompting:

[9] Original system's answer: "Function: It marks possession, similar to the apostrophe's in English" (S3).

  - Submitting all their sentences at once—either by typing them directly or uploading images of handwritten drafts—for correction or generation. Others preferred to work on their sentences one by one until they achieved the desired outcome. No significant differences in the results were observed based on these different input methods.
  - Employing various strategies to ensure that ChatGPT adhered to the required language level. Some attached the full coursebook in PDF format, others copy-pasted the complete list of vocabulary covered up to that lesson in their initial prompt, and some provided the relevant vocabulary and grammatical structures on a sentence-by-sentence basis. No notable differences in the system's performance were observed across these different approaches.
  - Asking the system to use simpler words when unfamiliar terms appeared. In fact, most students critically evaluated ChatGPT's suggestions, checking whether its corrections aligned with classroom instruction.
  - Engaging in iterative interactions, correcting errors that ChatGPT overlooked or requesting refinements in vocabulary, grammar, and syntax.
  - Guiding ChatGPT step by step to adjust sentences, grammar, vocabulary, Pinyin, and translations until they were satisfied with the final output.

- Correction requests:
  - Asking ChatGPT to correct sentences in Chinese previously written by them or by the system in the case of the second activity, ensuring adherence to specific levels and vocabulary learned in class.
  - Emphasizing strict corrections, avoiding alternative grammatical constructions or advanced vocabulary.

- Grammar and vocabulary clarifications:
  - Requesting the system to generate example sentences using specific grammar points.
  - Asking the system to explain what grammar points had been used in given sentences.
  - Asking targeted grammar questions, e.g., usage of words like 的 (*de*), or 比 (*bǐ*).
  - Requesting vocabulary explanations and simplified examples with a focus on their learning level.
  - Querying about differences between simplified and traditional characters.
  - Inquiring about the graphic structure of specific characters.

- Transcription and translation:
  - Providing sentences for translation from Spanish or Catalan to Chinese, then refining them based on ChatGPT's responses.
  - Asking the system to provide the Pinyin transcription according to its official orthographic rules, while others did not specify this condition.
  - Requesting the system to explain specific transcription rules, such as determining when to add spaces in Pinyin or when an apostrophe is required.

- Cultural and pragmatic analysis:
  - Exploring correctness of given expressions or alternative wordings focusing on subtle differences in meaning or pragmatics, such as confirming certainty with interrogative particles like 吧 (*ba*) or 吗 (*ma*).
  - Delving into culturally specific language use, such as distinguishing between phrases used to refer to adults versus children when discussing age in Taiwan and Mainland China.
  - Asking ChatGPT if there was a way to indicate that a given Chinese name belonged to a male individual.

3.4 Student's feedback

The working definition of AI literacy in this paper goes beyond just understanding, evaluating, and effectively using AI tools and systems for language acquisition and intercultural communication. It also includes critically acknowledging their limitations, potential biases, and ethical implications. Because of this comprehensive definition, the author felt it was important to collect student feedback. This additional perspective enhanced the validity and comprehensiveness of our findings, particularly for our second research objective.

To gather feedback from students regarding the use of GenAI in their Chinese language homework, a brief survey consisting of eight Likert-scale questions (with 1 being "completely disagree" and 5 being "completely agree") and a final open-ended question was designed. The questions and the average response ratings are shown in Figure 1.

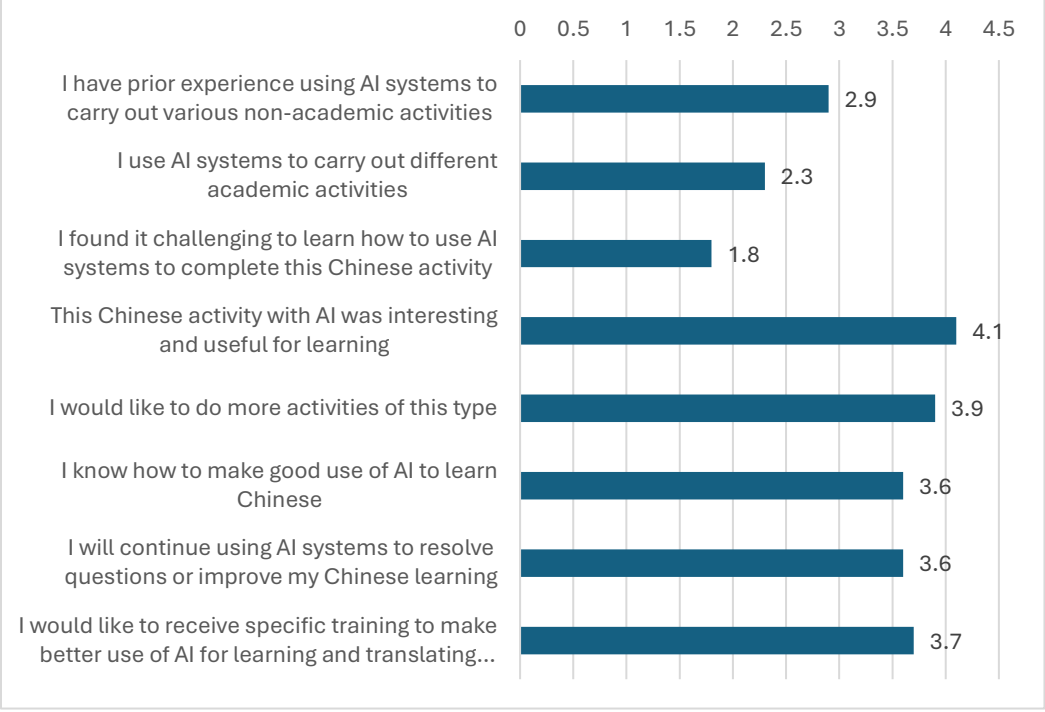


Figure 1 Students' responses to the Questionnaire on the use of GenAI

Only three students chose to answer the final open-ended question ("If you would like to add anything else about this teaching experience, you can do so below"). Their responses were as follows:

1. "I think a lot of value can be extracted from these tools; I like to input all the vocabulary I know and have it written a story with it."

2. "I was absolutely amazed. This was my first experience with ChatGPT, and it left me speechless!!!"

3. "I had never used AI for this type of activity; it was entertaining to see how it could help me in this regard."

The results of the survey reveal several key insights into students' experiences and perceptions of using GenAI for their Chinese language homework. While prior experience with AI systems for both academic (2.3) and non-academic (2.9) purposes was generally low, students found the Chinese activity with AI both interesting and useful for learning, as reflected in the relatively high average score (4.1). Additionally, a substantial portion of students expressed interest in engaging in similar activities (3.9) and recognized their fair ability to make good use of AI for learning Chinese (3.6), although they also highlighted the need for further training to optimize their use of the tool (3.7). The low difficulty score (1.8) suggests that most students did not struggle significantly with learning how to use AI in this context.

4. Discussion

4.1 Evaluating the pedagogical value of ChatGPT-generated responses

Learning and teaching styles have undergone significant changes over the past few decades. Students have not only become the central focus of the educational process, but they have also taken control of their own learning. This shift has been greatly facilitated by the increasing number of resources available online and, more recently, by the emergence of GenAI systems. These systems seem to adapt effortlessly to students' learning paces and individual needs, positioning GenAI as free, knowledgeable, always available, and ubiquitous virtual tutors in the eyes of users.

However, previous research has shown that AI-generated text is not always factual. Users must not only be aware of AI's potential to "hallucinate" but also be able to detect it. Unfortunately, less experienced students are particularly vulnerable in this regard. To address this, it is important to evaluate the quality and veracity of GenAI-generated content resulting from interactions with first-year Chinese language students, and simultaneously, gauge their prompt management strategies and to ascertain student attitudes toward these systems. To this end, two structured activities were designed for students to complete using a GenAI system of their choice, ensuring a relatively homogeneous data sample. Employing a bottom-up approach, the author categorized the identified problems into six distinct areas: terminology, Pinyin transcription, punctuation, vocabulary, grammar, and

pragmatics. These categories encompass both linguistic and metalinguistic aspects crucial to Chinese language acquisition.

Students must be able to adapt to diverse pedagogical approaches and textbooks, avoiding confusion from *ad hoc* or contextually uncommon terminology. Furthermore, for those aiming to use Chinese professionally, the development of metalinguistic skills—to articulate language concepts accurately and appropriately—is paramount. Consequently, the author specifically investigated the extent to which ChatGPT maintains academic rigor and terminological consistency in this domain. The results reveal that ChatGPT is not always consistent in its metalinguistic use of the Chinese language. Furthermore, the system not only fails to correct students when they make terminological errors but also reinforces these mistakes by automatically incorporating them into the interaction. The author hypothesizes that inaccuracies in terminology stem from biases inherent in the data used to train the system, the English's dominance in ChatGPT's training data and a lack of consensus within academia on a common terminological framework. These issues could potentially be mitigated by customizing a chatbot specifically for Chinese language learners, with training that explicitly addresses these terminological challenges.

Adhering to official transcription rules is crucial to prevent misunderstandings and support student learning, as incorrect transcription can lead to pronunciation and comprehension errors, ultimately hindering the educational process. This is especially important for students in Translation and Interpreting programs, since their future professional work will often require them to transcribe Chinese terms in Spanish texts rather than translate them. Finally, Pinyin transcription proficiency is increasingly vital as handwriting gradually disappears and reliance on Pinyin input for Chinese data and text processing continues to grow. However, the system's performance regarding Pinyin transcription is also problematic. It not only fails to correct student errors but also produces transcriptions that don't match Chinese characters, deviates from official Pinyin orthographic rules, and even fabricates rules to rationalize its output. These observations echo Rovira-Esteva's (2025) conclusion: GenAI systems perform poorly in Pinyin transcription. The author found that even with custom fine-tuning, ChatGPT-4 couldn't provide a perfect, standard-compliant transcription. This can mislead users rather than support their learning, particularly for first-year students, for whom Pinyin should serve as a supportive tool to help them grasp the meaning and pronunciation of sinograms.

In the sample of texts generated from the ChatGPT-student interaction, only one reiterative case related to text punctuation was found. Consequently, a generalization on this matter cannot be drawn. Nevertheless, the author observed that the system neither corrected students' erroneous punctuation nor offered an accurate explanation when prompted by students on the subject.

ChatGPT performs best in vocabulary. The errors detected in this area are significantly fewer and less severe. The main issue observed was the system's tendency to use words not explicitly provided by students or to offer vocabulary above their proficiency level. This finding aligns with what other researchers (Casas-Tost et al., 2023; Casas-Tost et al., 2025) found when using ChatGPT or DeepSeek to create reading comprehension activities in Chinese for beginners.

Grammar is the aspect under analysis where more pitfalls were observed. The activity proposed by the teacher also included asking the system questions about grammar. In this respect, the author found many cases where the responses were mostly accurate and appropriate for the students' level. However, some explanations were misleading or completely misaligned with Chinese grammar. For instance, ChatGPT sometimes unnecessarily corrected student writing, failed to correct grammatically incorrect sentences, or even offered grammatically flawed sentences itself. Cases were also observed where grammar explanations contradicted classroom teaching or provided examples that didn't illustrate the intended grammar point. Such cases may confuse students, as they are presented with unjustified modifications or incorrect justifications from the system. Due to the plausibility of these explanations, they could negatively impact the learning process of novice students. This aligns with the findings of Li (2024, 218), who argues that ChatGPT sometimes generates examples that do not conform to the required grammatical structures and fails to adequately identify, and correct common errors made by Spanish-speaking students, even when such errors are explicitly provided in the prompts.

The use of formulas or schematic structures to explain grammar should be a strong point for ChatGPT. However, the system changes them with each interaction and doesn't always include all the necessary information. This inconsistency isn't helpful for novice students who are specifically consulting the system because they have doubts about the subject. Although Li (2024, 128-129) concluded that most of the sentences generated by ChatGPT are grammatically correct in both Spanish and Chinese prompts, the author's findings suggest that ChatGPT struggles to meet students' needs by providing accurate and reliable grammar explanations and examples in a consistent and rigorous manner.

The pragmatic component, though often overlooked in the CFL classroom (Casas-Tost & Rovira-Esteva, 2015, 52), is crucial for providing effective linguistic correction. It helps learners develop intercultural competence, which is vital for authentic communication, especially when acquiring a foreign language outside their native culture. However, instances were also found where the system failed to provide reliable pragmatic information. This included a lack of appropriate guidance on language use in specific communicative contexts or explanations of grammar points from a discursive and pragmatic standpoint.

In sum, regarding ChatGPT's performance, it was observed that ChatGPT produced errors across all linguistic levels, including Pinyin, vocabulary, punctuation, grammar, and pragmatics. However, these errors were random and highly inconsistent. In fact, when presented with very similar prompts, the system occasionally delivered acceptable, relatively high-quality responses, while at other times, it generated nonsensical hallucinations. In other words, ChatGPT's responses are a mix of accurate and inaccurate. This means we must remain highly attentive during every interaction. To effectively use it as a learning companion, students absolutely need strong AI literacy skills to avoid being misled by the system. This leads us to our second objective: assessing students' AI literacy.

4.2 Students prompts as indicators of AI literacy

With powerful AI technologies now widespread in modern society, many are arguing that AI literacy is one of the most crucial literacies of the 21st century, standing alongside traditional reading, writing, mathematical, and general digital skills (Ng et al., 2021, 2; Krüger, 2024, 14). In addition to knowing and using AI ethically, AI literacy serves as a set of competencies that enables individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI (Long & Magerko, 2020).

Most students in this study crafted detailed prompts to maximize the educational value of ChatGPT's responses, often steering the AI back to beginner-level content when it produced more advanced answers. This process helped them engage critically with both the tool and their learning material. However, this strategy did not always work because the system had several pitfalls when interacting with students, such as language inconsistencies, over-accommodating incorrect feedback, failing to meet their learning needs, and not following prescriptive instructions. For instance, when students questioned certain correct translations or usages provided by ChatGPT, the system often altered its response to match the student's misunderstanding, even providing incorrect answers instead of explaining why its initial response was correct.

Another problem arose when students requested sentences with a specific number of characters (as required by the activity), ChatGPT often failed to meet this requirement. This issue aligns with findings in the literature indicating that ChatGPT struggles with accurately counting characters or words (Casas-Tost et al., 2023). Interestingly, most students overlooked this pitfall.

The data also revealed interaction cycles where more proficient students engaged in an iterative debate with the system when they distrusted its responses. This behavior was likely influenced by the activity instructions and evaluation rubric, which not only expected students to ensure vocabulary and grammatical structures were adequate but also encouraged critical engagement with the system, prompting them to challenge it when they identified mistakes. Therefore, explicitly including these items in the assessment seemed to motivate some students to question the system's outputs. Although these repeated interactions may have become tiresome, the author believes they empowered those students, encouraging them to take ownership of their learning and approach AI-generated content with a critical mindset. This process ultimately contributed to the development of their AI literacy and critical analysis skills. In this sense, ChatGPT's dynamic adaptation to individual needs, offering more personalized and tailored support, stands out as one of its best features. However, ChatGPT's excessive compliance and lack of subject knowledge pose significant risks, particularly when left in the hands of students with a low level of Chinese and insufficient critical thinking skills.

Given that some researchers (Ng et al., 2021, 4) associate AI literacy with perceived abilities, confidence, and readiness in learning AI, the author also prepared a survey. This survey aimed to assess students' subjective opinions on their prior GenAI knowledge, their comfort levels completing the activities, their willingness to continue using these tools in the future, and their interest in receiving training to enhance their Chinese language

learning. Their responses seem to indicate that they overestimate their AI literacy while simultaneously being open to further training on its use. They also seem to over-rely on AI technology, as none of them complained about its inaccuracies, mistakes, or the time invested in the activity. The open-ended responses, although scarce, further reinforce these findings, showcasing enthusiasm for the activity, a sense of discovery, and an acknowledgment of AI's potential to enhance learning. However, the variability in prior experience and the expressed desire for specific training indicate that while GenAI has clear pedagogical potential, careful guidance and scaffolding are necessary to ensure all students can fully benefit from its use.

While the use of GenAI in this experiment appears to have helped students improve their performance, the submitted assignments still contained some errors and showed room for improvement, despite students being allowed to use the system to review their work. This means current GenAI systems must be used in controlled environments, ensuring they truly complement and reinforce classroom learning. At this point, teachers' role is key, since they can train or guide students on how to correctly and ethically use it, as well as critically evaluate the tool's output through specific activities. As noted by Huang & Cassany (2025, 24), ChatGPT lacks the ability to provide a structured curriculum or holistic and sequential language learning content. In this regard, they suggest that "rather than focusing solely on teaching specific knowledge, the role of teachers in the AI era should perhaps shift towards fostering students' critical thinking and self-directed learning skills" (Huang & Cassany, 2025, 26).

5. Conclusions

In this study the author aimed at both assessing the quality of the system's responses as a virtual tutor and analyzing the nature of the students' interactions with the chatbot. To this end, students' overall performance with the help of ChatGPT was considered, followed by the analysis of some key aspects for CSL instruction, namely the system's use of terminology, its accuracy with Pinyin transcription, its awareness of punctuation issues, as well as its performance regarding vocabulary, grammar and pragmatics. The author also took note of ChatGPT's strengths, analyzed the nature of student-system interactions, and gathered student feedback regarding their experience with ChatGPT during these two activities.

This was not a quantitative study but rather an exploratory one, taking mainly a qualitative approach to data. However, the problems detected are numerous and serious enough to raise all alarms. The main problem with ChatGPT is that it does not genuinely understand what is being said and, therefore, does not acknowledge its inability to answer a question. Instead, it consistently provides a plausible response, which can give users a false sense of confidence. Therefore, it is essential to actively counteract the potential impact of uncontrolled use in the hands of students or educators with insufficient AI literacy to prevent greater issues. While some students were highly demanding and persistent, others accepted responses at face value. In other words, the analysis of the interactions revealed that students have varying levels of Chinese proficiency and AI literacy, both of which directly influence the quality of their outcomes.

The common denominator among previous studies is the conclusion that chatbots have significant potential as tools for learning various aspects of CSL. However, a critical issue is how to integrate them into teaching practices and, especially, how to manage their use by students. The first objective of this study identified the system's weaknesses and areas needing special attention; these findings offer practical guidance for instructors. They can use this information to supervise student tool usage and to know what to prioritize if they customize or fine-tune a chatbot for their courses. Meanwhile, the results from the second objective, detailing the nature and content of GenAI-student interactions, provide a basis for designing specific training to cultivate students' AI literacy.

Research so far has been characterized by isolated initiatives driven by the personal curiosity of the involved teachers and the need to adapt to a technology that, to some extent, surpasses educators. The results of this study partially align with previous findings. However, it is innovative as it examines AI-generated text from the perspective of the system's interaction with students, performing a qualitative analysis through the lens of an experienced instructor familiar with the course content and the students' skill levels. Moreover, based on the analyzed prompts, the author claims that using ChatGPT in an uncontrolled environment presents more risks than benefits. This is not only because most students lack sufficient AI literacy and foreign language proficiency to identify errors, but also because the system currently fails to deliver a consistent minimum threshold of quality across interactions.

The transformative potential of GenAI in the context of language learning and education is undeniable. Its purported ability to provide interactive and personalized support—whether for practicing conversations, learning vocabulary, improving grammar, or exploring cultural nuances in language use—has led to its widespread perception as an ideal language coach or partner. The innovative and constantly evolving nature of GenAI, particularly since the advent of ChatGPT, is one of the main challenges researchers and educators face. These challenges extend beyond AI literacy and ethics, as the stress from feeling perpetually behind rapidly changing technology creates a medium- to long-term psychological impact we are likely underestimating. To overcome this difficulty, researching in multidisciplinary teams could be highly effective. Such collaboration would allow for a joint study of the technology, combining diverse interests, backgrounds, and experiences. This approach fosters a more holistic and empirical understanding, moving beyond the widely discussed "dazzling potential" to rigorously and systematically evaluate these systems in a global and scientific manner.

Every study has its limitations, and this one is no exception. The primary limitation is the sample size and the format of the interaction texts, which in some cases were limited to screenshots from the students' mobile phones, making analysis more challenging. Another difficulty was that the textual analysis was conducted manually, whereas using a specific discourse analysis tool might have also allowed for the collection of some quantitative data. Thirdly, although students were free to use any GenAI system, all but one chose ChatGPT, which undeniably has biases against Chinese. Therefore, a similar study conducted with China-based systems could yield different outcomes, particularly in terms of the quality of the system's outputs. Fourth, another interesting approach would be to conduct a longitudinal study or one with a control group to determine whether the guided

use of ChatGPT contributes to improving students' Chinese proficiency in the medium or long term. The research design enabled the author to assess the outcomes but provided limited insight into the students' learning processes and what they actually gained from interacting with the system. Finally, neither the type of activity nor the student profile is, nor was intended to be, representative enough to generalize the results. Future lines of research could expand the study by including activities that evaluate additional linguistic skills of students, involving learners from higher levels, testing analytical tools that enable more objective and systematic error tagging, examining whether the language used for interactions influences the quality of the results, or adopting a more quantitative approach to assess the chatbot's accuracy in providing feedback. In any case, the author hopes that this study will be of interest to readers and encourage further research into the integration of GenAI systems in the teaching of CSL.

Acknowledgements: The author extends her deepest gratitude to their research group colleagues for being a constant source of inspiration and academic growth. Their invaluable feedback and corrections were crucial, both in designing the activities that led to this study and in their careful review of the initial draft of this article.

References

Casas-Tost, H., González-Torrents, I., Rovira Esteva, Sara, & Vargas-Urpí, M. (2025). *Using GenAI in the teaching and learning of additional languages*. [Manuscript submitted for publication].

Casas-Tost, H., Olalla-Soler, C., Paoliello, A., Rovira-Esteva, S., Sánchez-Gijón, P., & Vargas-Urpí, M. (2023). The use of ChatGPT for the creation of activities in the Chinese as a foreign language classroom: A case study on reading competence. Instituto Confucio de León. *East and West Contact and Dialogue: International Conference of Contemporary China Studies, online.* https://ddd.uab.cat/record/284643

Casas-Tost, H., & Rovira-Esteva, S. 2015. Mapping Chinese Language Pedagogy from 1966 to 2013: A Bibliometric Study of the Journal of Chinese Language Teachers Association. *Journal of Chinese Language Teachers Association 50(2)*, 31–58.

Casas-Tost, H., Rovira-Esteva, S., & Suárez, A-H. (2024). *Lengua china para traductores*. 学中文，做翻译 (7th ed., Vol. 1). Servei de Publicacions de la Universitat Autònoma de Barcelona. https://doi.org/10.5565/lib/9788410202108

Chen, Y. (2023). Discussion on the digitalization of international Chinese teaching resources in the era of artificial intelligence. *University Education, October,* 20, 22–25. [陈悦. (2023). 人工智能时代国际中文教学资源数字化探析. 大学教育, 20, 22–25.]

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Donley, K. (2024). Teaching with ChatGPT as a linguistically responsive tool for multilingual learners. *Technology in Language Teaching & Learning, 6(3),* 1719. https://doi.org/10.29140/tltl.v6n3.1719

Gao, A. (2024). Self-narrative exploration of ChatGPT assisting international Chinese teaching: Opportunities and challenges. *The Journal of Modernization of Chinese Language Education, 13(1),* 52–63. [高安娜. (2024). ChatGPT 辅助国际中文教学的自我叙事探究：机遇和挑战. 中文教学现代化学报, 13(1), 52–63.]

General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China & National Standardization Administration. (2012). Basic rules of the Chinese phonetic alphabet orthography (Standard GB_T 16159-2012). http://www.moe.gov.cn/ewebeditor/uploadfile/2015/01/13/20150113091717604.pdf [中华人民共和国国家质量监督检验检疫总局, & 中国国家标准化管理委员会. (2012). 汉语拼音正词法基本规则. GB_T 16159-2012.]

Guo, L. (2024). The use of generative artificial intelligence for preparing reading comprehension activities in Chinese as a foreign Language. [Master's thesis. Universitat Autònoma de Barcelona.] [Guo Lele. (2024). El uso de inteligencia artificial generativa para preparar actividades de comprensión escrita en chino como lengua extranjera. Tesis de máster. Universitat Autònoma de Barcelona.]

Hellmich, E. A., Vinall, K., Brandt, Z. M., Chen, S., & Sparks, M.M. (2024). ChatGPT in language education: Centering learner voices. *Technology in Language Teaching & Learning, 6(3),* 1741. https://doi.org/10.29140/tltl.v6n3.1741

Huang, S., & Cassany, D. (2025). Spanish language learning in the AI era: AI as a scaffolding tool. *Journal of China Computer-Assisted Language Learning*. https://doi.org/10.1515/jccall-2024-0026

Koyuturk, C., Yavari, M., Theophilou, E., Bursic, S., Telari, A., Testa, A., Boiano, R., Hernandez-Leo, D., Ruskov, M., & Ognibene, D. (2023). Developing effective educational chatbots with ChatGPT prompts: Insights from preliminary tests in a case study on social media literacy. *31st International Conference on Computers in Education (ICCE),* Matsue, Japan.

Krüger, Ralph. (2024). Outline of an Artificial Intelligence Literacy Framework for Translation, Interpreting and Specialised Communication. *Lublin Studies in Modern Languages and Literature 48(3)*, 11–23. https://doi.org/10.17951/lsmll.2024.48.3.11-23

Lee, B. J., & Cook, D. R. (2024). Exploring the potential of AI for pragmatics instruction. *Technology in Language Teaching & Learning, 6(3),* 1521. https://doi.org/10.29140/tltl.v6n3.1521

Li, J., Ren, X., Jiang, X., & Chen, C.-H. (2023). Exploring the use of ChatGPT in Chinese language classrooms. *International Journal of Chinese Language Teaching, 4(3),* 36–55. https://doi.org/10.46451/ijclt.20230303

Li, Q., Zhang, J., & Cai, W. (2024). Utilizing ChatGPT to implement differentiated instruction. *International Journal of Chinese Language Teaching, 5(1),* 74–89. https://doi.org/10.46451/ijclt.20240106

Li, W. (2024). ChatGPT as an assistant for Chinese teachers: Its role in grammar teaching. [Master's thesis. Universitat Autònoma de Barcelona]. [Li Weijia.

(2024). El desempeño de ChatGPT como asistente del profesor de chino: un estudio de su potencial y limitaciones para la enseñanza de la gramática. Tesis de máster. Universitat Autònoma de Barcelona.]

Liu, L., Zhou, X., Gao, X., Pan, H., Liu, X., Rao, G., Shen, S., Chen, K., Xin, P., Zhang, H., Lin, Z., Liu, H., Yu, Z., Zhou, Z., Chen, Q., Chen, M., Ren, S., Yang, X., Ma, Y., & Han, X. (2023). ChatGPT is here: New opportunities and challenges for international Chinese language education. A Discussion [Part 1]. *Language Teaching and Research, 3,* 74–89. [刘利, 周小兵, 高雪松, 潘海峰, 刘晓海, 饶高琦, 沈索超, 陈肯, 辛平, 张辉, 林筠, 刘华, 俞玮奇, 周斐, 陈青, 陈默, 任思潼, 杨绪明, 马一鸣, & 韩晓明. (2023). ＣｈａｔＧＰＴ来了：国际中文教育的 新机遇与新挑战”大家谈[上]. 语言教学与研究, *3*(221), 1–14].

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3313831.3376727

Matthews, B. (2024). Finding the human in an era of machine intelligence: A flat ontological analysis of generative AI and language learning. *Technology in Language Teaching & Learning, 6(3),* 1612. https://doi.org/10.29140/tltl.v6n3.1612

Meng, K. (2024). Performance and application strategies of ChatGPT in international Chinese vocabulary teaching. *International Journal of Chinese Language Teaching, 5(1),* 3–14. [孟凯. (2024). ChatGPT 在国际中文词汇教学中的表现与应用策略. 国际汉语教学学报, 5(1), 3–14.]

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ou, Z., Liu, Y., Qin, K., Wang, Y., & Li, X. (2024). Generation and evaluation of artificial intelligence multimodal teaching resources. *Modern Educational Technology, 34(9),* 37–47. [欧志刚, 刘玉屏, 覃可, 王艺, & 李欣燚. (2024). 人工智能多模态教学资源的生成与评价. *现代教育技术, 34*(9), 37–47.]

Rovira-Esteva, S. (2025). *Evaluating current tools for Pinyin transcription: Can customising a chatbot lead the way forward?* [Manuscript submitted for publication].

Tseng, W., & Warschauer, M. (2023). AI-writing tools in education: If you can't beat them, join them. *Journal of China Computer-Assisted Language Learning, 3(2),* 258–262. https://doi.org/10.1515/jccall-2023-0008

Wang, C., & Williams, J. (2024). Utilizing ChatGPT in the Development of Science and Technology Chinese Textbooks: A Case Study of Teacher-student Collaboration in a One-on-one Classroom. In C.-H. Lin, H. Tao, J. Da, & S. Liu (Eds.), *Proceedings of the 12th international conference and workshops on technology and chinese language teaching* (pp. 81–87). [王楚蓁, &韋志安. (2024). ChatGPT 輔助科技華語教材編寫——以單班課師生協作模式為例. 林金锡, 陶红印, 笪骏, & 刘士娟 (主编)， *第十二届国际汉语电脑教学研讨会论文集* (pp. 81–87). http://www.tclt.us/tclt12/proceedings.php].

Wang, L. (2024). *The use of ChatGPT and Gemini for preparing CLE listening comprehension activities: A comparative study*. [Master's thesis. Universitat

Autònoma de Barcelona]. https://ddd.uab.cat/record/300353 [Wang Lu. (2024). *El uso de ChatGPT y Gemini para la preparación de actividades de comprensión oral de CLE: un estudio comparativo. Tesis de máster.* Universitat Autònoma de Barcelona.]

Wang, X. (2024). ChatGPT as a conversational agent in Chinese language learning for Spanish speakers. *China Studies Review, 18(1),* 71–98. https://doi.org/10.18002/sin.v18i1.8428 [Wang Xiaoling. El ChatGPT como agente conversacional en el aprendizaje de la lengua china para los hispanohablantes. *Sinología Hispánica, 18(1),* 71–98.]

Xu, J., & Ma, R. (2023). Technological changes in international Chinese education in the ChatGPT era. *Research in International Chinese Language Teaching, 38(2),* 41–52. [徐娟, & 马瑞祾. (2023). ChatGPT 浪潮下国际中文教育的技术变革. *国际汉语教学研究*, 38(2), 41–52.]

Zhao, Q., Hsu, Y., & Huang, C.-R. (2024). *Make LLMs your portable teachers of Chinese as a foreign language—Prompt patterns of an AI-teacher.* In C.-H. Lin, H. Tao, J. Da, & S. Liu (Eds.), *Proceedings of the 12th international conference and workshops on technology and Chinese language teaching* (pp. 131–139). [肇群, 許又尹, & 黃居仁. 让大语言模型成为你的随身汉语老师——AI 教师的提示范式研究. 林金锡, 陶红印, 笪骏, & 刘士娟 (主编)， *第十二届国际汉语电脑教学研讨会论文集* (pp. 131–139). http://www.tclt.us/tclt12/proceedings.php].

Appendix 1

Translation of the instructions for the first activity

Summative Assessment Activity - Lesson 3

Instructions:

Translate the following sentences into Chinese and transcribe them correctly into Pinyin following its orthographic rules:

1. How old is your daughter?
2. How old is your father?
3. My mother is two years older than yours.
4. Your wife is five years younger than you, isn't she?
5. Xiao Qing is Chinese, isn't he?

Ask ChatGPT (or another AI system) to correct your translations, ensuring they adhere to the vocabulary and structures covered in class (Lessons 1–3). Do not use vocabulary or grammar you haven't learned. You may communicate with the chatbot in any language, and you must interact until you achieve the desired results.

Ask the chatbot at least two questions about vocabulary and three about the grammar in the resulting sentences. Note any errors you identify in vocabulary, syntax, or transcription and ask the chatbot to correct them.

Submit a complete copy of the entire conversation, from start to finish, as a PDF file via the virtual campus, including:

- Your name
- The name of the system used
- A link to the conversation
- Complete the following survey.

Appendix 2

Translation of the instructions for the second activity

Summative Assessment Activity - Lesson 5

Instructions:

Write 5 sentences, each consisting of at least 10–15 characters, with the help of ChatGPT (or another AI system). In each case, provide the vocabulary and structures you want the system to use and include the corresponding Chinese translation and Pinyin. Specify the grammatical points being addressed in each sentence, ensuring no repetition of grammar points. Examples of initial prompts in Spanish:[10]

- Write a sentence of at least 10–15 characters that includes a resultative complement and the following words: 客, 吃饭, 饭馆.

- Translate into Chinese: "The teachers ate to their heart's content at that restaurant," using a resultative complement, beginner-level vocabulary, and providing its transcription in Pinyin according to official orthographic rules.

Ensure the sentences align with the vocabulary and structures covered in class (Lessons 1–5).

You may interact with the chatbot until you achieve the desired results. Identify and highlight any issues with vocabulary, syntax, translation, or transcription, and request corrections as necessary.

Submit a complete copy of the entire conversation, from start to finish, as a PDF file via the virtual campus, including:

- Your name
- The name of the system used
- A link to the conversation
- The final 5 sentences in Chinese, correct Pinyin, and their translation
- Complete the following survey[11]

[10] The original wording in Spanish was:

- Escribe una oración de al menos 10-15 sinogramas que incluya un complemento resultativo y por lo menos las palabras siguientes: 客，吃饭， 饭馆。
- Traduce al chino: "Los profesores comieron hasta hartarse en ese restaurante" usando un complemento resultativo y vocabulario de un nivel inicial y proporcionando también su transcripción al pinyin de acuerdo con sus normas oficiales de ortografía.

[11] c.f. https://docs.google.com/forms/d/e/1FAIpQLScL3suq2-Bww9t07EPNAGePUpTp15MvZK8V1uD4pQjlPdIk9A/viewform

Appendix 3

Translation of the rubric for both activities

Assessment Criteria

Category	**Excellent (3–4)**	**Good (2–2.9)**	**Pass (1–1.9)**	**Not Acceptable (0–0.9)**
Linguistic accuracy of the final Chinese version (40%)	No lexical or grammatical errors.	Few lexical or grammatical errors.	Some lexical or grammatical errors.	Many lexical or grammatical errors.
Lexical richness and variety of structures (10%)	A diverse range of structures and vocabulary is used.	A diverse range of structures and vocabulary is used but below the required level or repetitive.	Vocabulary and structures are below the required level and repetitive.	Vocabulary and structures are below the required level, repetitive, or sentences are missing.
Adequacy of sentences to the level (20%)	All constructions and vocabulary align with the textbook's content.	Some constructions or vocabulary exceed the level.	Several constructions or vocabulary exceed the level.	Constructions and vocabulary significantly exceed the level.
Accuracy of *Pinyin* transcription (10%)	*Pinyin* transcription fully adheres to official rules.	*Pinyin* transcription contains minor errors.	*Pinyin* transcription contains several errors.	*Pinyin* transcription does not follow basic rules.
Fulfilment of instructions (20%)	All instructions are followed, and all required results are presented.	Not all instructions are followed, or some results are missing.	Not all instructions are followed, and several results are missing.	No instructions are followed, and the activity is incomplete.

Appendix 4

List of informants with the link to their interactions with ChatGPT

Informant	Link
S1	https://chatgpt.com/share/671e6cac-2924-8003-8122-ad6d2abc355d
S2	https://chatgpt.com/share/671d0769-1e54-8003-b04a-64bb4b218950
S3	https://chatgpt.com/share/671d0c5e-083c-800d-88b6-bec0b106ccc3
S4	https://chatgpt.com/share/671e91b6-3fc0-8001-a527-e92c98929301
S5	https://chatgpt.com/share/671d6fa0-cae0-8003-aa46-1cbd5e04cd05
S6	https://chatgpt.com/share/671d526b-9420-800a-afb4-361fee580bd2
S7	https://chatgpt.com/share/671eb9de-ab1c-800b-92c6-478ddaff52ef
S8	https://chatgpt.com/share/671e3a65-a684-8005-9045-3b1a070298e8
S9	Not available
S10	https://chatgpt.com/share/671e26b1-b260-8003-8308-c6853c093abc
S11	https://chatgpt.com/share/671e7bd7-8efc-800a-8b07-8d8d036b0d2c
S12	https://chatgpt.com/share/671d594c-cfec-8006-a9f5-087c187efbb7
S13	https://chatgpt.com/share/671e7e99-dc24-800c-bbdd-b165c7dfa738
S14	Not available
S15	https://chatgpt.com/share/671e8e65-c9e8-8007-8166-6d6ab1766835
S16	https://chatgpt.com/share/671ebe57-d144-8005-8030-b8f597f54ea1
S17	Not available
S18	https://chatgpt.com/share/671ccd84-970c-8007-be07-59d5244584c0
S19	Not available
S20	https://chatgpt.com/share/671ec381-01bc-8009-9528-5ac08ab99e2f
S21	Not available
S22	https://chatgpt.com/share/671baf46-2ad4-8011-a79f-7b3e0d98a2fa
S23	https://chatgpt.com/share/671b7ffa-6250-8006-9a47-12430ae8183a
S24	https://chatgpt.com/share/671e55df-c4a8-8002-82e9-5e1803ae85fa
S25	https://chatgpt.com/share/671eafc3-8fd0-800c-987c-e8262a89c96c
S26	https://chatgpt.com/share/671cca60-5b74-8012-a933-21fa448b755d
S27	https://chatgpt.com/share/671e9c0b-926c-8011-8802-d2f984f1f444
S28	Not available
S29	Not available
S30	https://chatgpt.com/share/671e741f-b470-800d-a3bb-b468a4e722d4
S31	https://chatgpt.com/share/671ce9b5-b5b8-8009-ab00-9c2658a2b290
S32	https://chatgpt.com/share/671cccb5-8200-8007-8e93-edb594281e04
S33	https://chatgpt.com/share/675dcee0-6dd8-8003-bd2b-c198f1107245
S34	https://chatgpt.com/share/675f2841-10a8-8003-9cf9-8b9f023248c4
S35	https://chatgpt.com/share/675da68b-7674-800d-bcae-7551f5739fef
S36	https://chatgpt.com/share/675f4b62-5bfc-8001-a8f7-ef50ce8f9d66
S37	https://chatgpt.com/share/675e2f3a-5774-8003-934d-e8a2730e61d7
S38	https://chatgpt.com/share/675e2f3a-5774-8003-934d-e8a2730e61d7
S39	Not available
S40	https://chatgpt.com/share/675fe85e-dd40-8003-a081-891b3b653ed0
S41	https://chatgpt.com/share/675d76c9-18fc-800a-8d88-a579091fa639

S42	Not available
S43	https://chatgpt.com/share/675f154f-65ec-800c-b6ec-240c080129cc
S44	Not available
S45	https://chatgpt.com/share/675f1e4b-c714-8007-8489-ccc3db037510
S46	https://chatgpt.com/share/675fe189-6e94-8005-8841-ce46fe0c44f6
S47	https://chatgpt.com/share/675edde8-9e24-8007-be50-05036e1288b3
S48	https://chatgpt.com/share/675f4abc-a4fc-8009-b5d3-44b5fb721ccc
S49	https://chatgpt.com/share/675f27d4-23bc-800f-88bc-8f7462cffd47
S50	Not available
S51	https://chatgpt.com/share/675ef827-bbc0-8006-9a01-713037f7674d
S52	https://chatgpt.com/share/675eb3ea-15a4-8002-bb4f-7019fced08e3
S53	Not available
S54	Not available
S55	https://chatgpt.com/share/675dd849-4b58-8011-bf13-635ac2a45142
S56	https://chatgpt.com/share/675f55cf-b0f4-8001-be81-258569165e77
S57	https://chatgpt.com/share/675efc69-87ec-800d-9025-391ee51d1f67v
S58	https://chatgpt.com/share/675d6bde-61d0-8009-9542-d08cf2a57cd1
S59	https://chatgpt.com/share/675f3398-5dc8-8007-b665-db93b252d7c3

Volume 16 Number 1, June 2025
pp. 85-117

多元视角下国际中文慕课视频设计的评价框架研究
(A Study on the Evaluation Framework for International Chinese MOOC Video Design from Multiple Perspectives)

马铖
(Ma, Cheng)
北京语言大学
(Beijing Language and Culture University)
macheng@blcu.edu.cn

徐娟
(Xu, Juan)
北京语言大学
(Beijing Language and Culture University)
xujuan@blcu.edu.cn

摘要：本研究聚焦国际中文慕课视频设计，构建兼具政策参考、学术焦点与学习者需求的评价框架。基于政策文本与学术文献，自上而下提取核心要素。通过对 16 门汉语慕课 5339 条学习者评论进行文本挖掘，自下而上提炼出学习者需求的实践关切点。最终整合多重视角，形成包含内容设计、交互设计、情感设计和创新设计 4 项一级指标，9 个二级指标、28 个观测点的系统评价框架。该框架为提升国际中文慕课视频设计质量提供科学依据，助力国际中文教育与中华文化传播。

Abstract: This study focuses on the design of International Chinese MOOC videos and aims to construct an evaluation framework that integrates language and education policies, academic perspectives, and learner needs. By analyzing policy documents published in Mainland China and academic literature, core elements were identified through a top-down approach. Simultaneously, a text mining analysis of 5,339 learner comments from 16 Chinese MOOC courses revealed practical concerns related to learner needs using a bottom-up approach. The integration of these multiple perspectives resulted in a systematic evaluation framework comprising four primary indicators—Content Design, Interaction Design, Emotional Design, and Innovative Design—along with nine secondary indicators and 28 observation points. This framework provides a scientific basis for improving the design quality of International Chinese MOOC videos, thereby contributing to the advancement of global Chinese language education and the dissemination of Chinese culture.

关键词：国际中文教育、慕课、视频设计、评价框架

Keywords: International Chinese Education, MOOC, Video Design, Evaluation Framework

1. 引言

互联网技术与教育传播方式的深度变革催生了慕课（Massive Open Online Course，MOOC）的全球化普及，其开放性与规模化特性显著拓展了国际中文教育的实践边界、中文及中国文化的传播效能（Zhang et al.，2024）。然而慕课发展的繁荣表象之下，教学质量的系统性缺陷持续制约着其可持续发展能力（Stracke & Trisolini，2021）。现有研究发现，国际中文慕课存在三重深层矛盾：一是课程建设缺乏系统性，表现为内容较为单薄（郑才华，2019）、文化元素融入不足（Wang-Szilas & Bellassen, 2017）、真实语境下的语言实践机会缺失（崔希亮，2020）；二是学习者面临交互单一（辛平，2019）、自我调节能力不足（陈晨，2021）等困境；三是教师团队面临教学信念重塑、教学理念更新（王添淼、张越，2017）、信息素养有待提升（甄刚，2019）等专业发展瓶颈。2024 年教育部全面推进“慕课出海”行动[1]，让中国慕课走出国门，积极推动中国慕课服务全球中文学习者与国际教育市场，进一步凸显教学供给侧改革与学习体验优化的紧迫性。

作为慕课课程的核心载体，教学视频承担着多重功能使命：既需实现语言知识和技能的精准传递，又要建构文化价值的认同纽带（Bayeck & Choi，2018），同时满足远程学习者的情感支持需求（Liu et al.，2024）。尽管学界已从教育学（李艳、张慕华，2015；孔艳，2024）、传播学（陈淼，2015；黄奕宇，2021）等视角展开慕课研究，但学术缝隙依然存在：一是现有评价体系多侧重宏观层面的课程整体质量评价（邱均平、欧玉芳，2015；童小素、贾军，2017），或着眼于通用性视频制作标准（李青、刘娜，2016；王雪、周围、王志军，2018；武家辉，2019），未能突显国际中文教育中语言习得、文化认知与数字媒介的三重交互特性；二是研究视角呈现单维度倾向，或依赖政策文本的规范解析（杨继龙，2024），或仅关注学习者的体验反馈（Ding，2019；周汶霏、宁继鸣，2020；Deng & Gao，2023；李雯静、骆蓉，2023），尚未建立多向联动的评价框架。

本研究采用混合研究路径突破上述局限：基于政策文本和学术研究的“自上而下”视角，结合学习者使用过程反馈的“自下而上”视角，综合采用文本挖掘法中的多种方法，以“集语言技能教学、语言要素教学、文化知识教学和交际能力训练为一体”的汉语综合课（赵金铭，2006）为例，构建并完善国际中文慕课视频设计的评价框架。框架试图在以下三个方面进行创新：第一，选择汉语综合课为观察窗口——该课型在 Coursera[2]、edX[3]、中国大学 MOOC[4]等慕课平台上的数量最多（张浩，2023），其融合语言要素、文化知识与交际能力培养的复合性特征，更能全面映射多模态教学的设计张力；第二，“政策文本分析、学术研究支撑、用户数据挖掘”的多元视角，能够消解传统评价中主客体立场割裂问题；第三，开发的评价框架，可转化为国际中文慕课视频设计、研制与评价的操作工具，提供诊断性标准。

[1] 具体参见：http://www.moe.gov.cn/fbh/live/2024/55785/sfcl/202401/t20240126_1112426.html
[2] 具体参见：https://www.coursera.org/
[3] 具体参见：https://www.edx.org/
[4] 具体参见：https://www.icourse163.org/

研究结果可为“慕课出海”行动的实施提供理论与实践双重支撑。具体围绕以下三个问题展开研究：

1. 政策与学术视角下，哪些核心维度代表着国际中文慕课视频设计的质量要求？
2. 学习者评论揭示了哪些具体的设计关注点，它们与政策和学术视角有何差异？
3. 如何整合多元视角，构建国际中文慕课视频设计的评价框架？

2. 理论基础

本研究的理论基础来源分为两个部分，一是关于设计评价的理论基础，二是关于教学视频设计的理论基础。

2.1 关于设计评价的理论基础

由于国际中文慕课视频主要涉及中文学习者即用户的认知、情感、学习活动中的交互等，因此最相关的模型即为与用户体验相关的评价模型，可用于了解学习者对视频的接受程度和满意度。

用户体验评价在一些研究中也称为用户体验质量（Quality of Experience，QoE），这类模型旨在评估设计特征是否满足用户需求，以及对用户体验的影响，发现当前产品设计上存在的问题，为之后的改善和产品的迭代提供参考（Yong，2013）。有关用户体验评价的相关理论也因用户体验定义的不同而在发展和深化中。常见的用户体验评价模型有 KANO 模型（KANO Model，Noriaki Kano，1984）、用户体验蜂巢模型（User Experience Honeycomb，Peter Morville，2004）、用户体验轮模型（UX Wheel，Magnus Revang，2007）、杰西·詹姆斯·加勒特的用户体验要素图（The Elements of User Experience，Garrett，2011）等，另有国内学者罗仕鉴等（2022）提出可从设计的三个层次进行评价，即本体评价层、行为评价层和价值评价层。

这些评价模型并非来自单一的理论源泉，而是由多位专家学者根据不同的研究对象和视角提出的。本研究将它们统一归类并称为“用户体验类评价模型”。教育领域中“以人为本”的理念，最早由 Rogers 提出（Rogers, 1969），强调将学习者的需求和体验置于教学设计的核心。用户体验类评价模型同样关注个体的感受和需求，二者在理念上高度契合。这种“以学习者/用户为中心”的共同追求，与国际中文教育中“以学习者为中心”的理念是一致的。

2.2 关于教学视频设计的理论基础

关于教学视频设计的理论源自多个学科，包括教育学、心理学、社会学等，具备明显的跨学科特点。本研究根据卜彩丽等（2021）基于国外 72 篇实证研究的系统综述，结合汉语作为第二语言学习的特点，将国际中文慕课视频设计的理论基础分为以下 4 类：

1）一般的认知学习理论

这类理论包括认知负荷理论（Cognitive Load Theory，Sweller，1994）、工作记忆理论（Working Memory Theory，Baddeley，1974）、信息加工理论（Information Processing Theory，Miller, 1956）等。主要用于研究和解释“人如何学习”“人如何有效学习”等问题，是为了使教学视频能够更好地促进学习者学习，并且能够按照人脑工作的方式去进行设计。其中引用频率最高的是认知负荷理论（卜彩丽等，2021）。

2）专门的多媒体学习认知理论

这类理论包括多媒体学习的认知理论（Cognitive Theory of Multimedia Learning，Mayer，2005）、移动学习理论（Mobile Learning Theory，Kay，1972）、双重编码理论（Dual Coding Theory，Paivio，1971）等。这类理论主要用于研究和解释“多媒体如何起作用”“多媒体是否起作用”以及“多媒体什么时候起作用”。其中最受研究者重视和引用最多的是 Mayer 的多媒体学习认知理论。他的研究最重要的贡献是人们如何选择、组织和加工语词与画面信息的，因此成为教学视频设计最重要的理论依据。

3）社会学习类和情感理论

这类理论主要包括社会存在感理论（Social Presence Theory，Short, Williams & Christie，1976）、社会代理理论（Social Agency Theory, Mayer, Sobko, & Mautone，2003）、具身认知理论（Embodied Cognition Theory，Barsalou，2008）等。这类理论是随着互联网的飞速发展兴起的，主要用于研究和解释教学视频中，尤其是在线环境下，教师是否需要出镜，即教师是否需要呈现，为什么需要教师形象的呈现等问题。它们都强调，在互联网学习环境下，教师形象作为社会线索可以增强师生的社会性联系，因而可以促进学生采用更深层次的信息加工方式，激发学生积极的学习情绪和学习态度，从而促进学习。其中，社会存在感理论的引用频率最高。

本研究的国际中文慕课视频，教学内容是中文语言。语言是人类沟通的桥梁，其本身就具备社会性功能。基于此理论，教学视频中影响学习者社会存在感主要涉及以下几个因素：师生图像的呈现以及呈现的方式，营造真实的语境，人机交互以及人际交互等。

4）语言学习类理论

这类理论主要包括二语习得领域的输入假说（Input Hypothesis，Krashen，1985）和注意假说（Noticing Hypothesis，Schmidt, 1990），用于解释语言学习者

是如何学习语言的。因学习者对外界语言信息的注意是有选择的，因此这类理论主要用于研究和解释在语言学习过程中，什么样的信息输入、如何输入才能促进学习，概括地说，输入在很大程度上是由有选择的注意决定的。

这两个假说关注学习者如何接收和处理信息，以及注意力在语言学习过程中的重要性。对于教学视频设计而言，本研究认为“输入”（Input）与“注意”（Noticing）间存在密切关系，是分不开的。中文慕课视频，不仅涉及视频作为媒介载体的因素，还涉及语言的方方面面，综合运用各种能够引起学习者“注意”的要素，输入的效果就更好。所以输入假说和注意假说也应该是中文慕课视频设计非常重要的理论基础。这两个理论的启示是，应灵活地综合运用有利于引起学习者“注意”的各种要素，依据教学内容进行更有效地输入，提升内容吸引力。

3. 研究设计

3.1 数据来源

本研究的数据来源主要有三个方面：

一是政策文本：匡昕、冯丽萍（2023）年指出，政策是实践的基本依据和价值遵循。在开展资源建设与研究之前，参考政策文本，理清建设思路，规范要素标准，是提高资源建设针对性、适用性及有效性的重要手段。中外语言合作交流中心出台的《国际中文教育教学资源建设行动计划（2021—2025 年）》等 4 个政策文本（如表 1 所示），既是国际中文教育领域的总体建设规划，也是对相关教学资源和学习材料的质量要求，为中文慕课视频设计评价的要素提取提供了政策参考，是中国大陆地区国际中文教育领域关于资源建设的权威性和代表性文件。因此本研究从这 4 个政策文本中提取中文慕课视频设计评价指标体系中应纳入的核心维度或指标。

表 1 国际中文教育资源建设政策文本

序号	文件名称	发布时间
1	国际中文教育教学资源建设行动计划（2021—2025 年）[5]	2021
2	国际中文教育数字资源建设指南（试行）[6]	2021
3	国际中文在线教育行动计划（2021—2025 年）[7]	2021
4	国际中文教育教学资源建设项目管理办法[8]	2021

二是学术研究：本研究将总结该领域已有的研究成果，梳理出评价指标体系的相关要求，并在此基础上，整合评价要素。以中国知网 CNKI 为主要文献数据来

[5] 具体参见：https://www.chinese.cn/uploads/file/20220125-1643091013876953.pdf
[6] 具体参见：https://www.chinese.cn/uploads/file/20220125-1643090953255465.pdf
[7] 具体参见：https://www.chinese.cn/uploads/file/20220125-1643091053961452.pdf
[8] 具体参见：https://www.chinese.cn/uploads/file/20220824-1661335012957560.pdf

源，以“中文教学视频”“汉语二语教学视频”“对外汉语教学视频”“外语教学视频”“语言教学视频”“教学视频”“在线视频课程”等为主题词进行检索，同时通过文献资料中的参考文献来获取相关文献。通过阅读文献题目和摘要剔除了与本研究不相关或相关性较弱的文献，再通过阅读文献全文，最终确定共有 39 篇文章涉及教学视频设计的具体维度或观测点研究。

三是学习者评论文本。本研究认为，在慕课平台上学习后，学习者留下的评论隐含了很多潜在的学习者需求和情感信息，值得充分挖掘。本研究选取具有较强代表性的，且能够获取学习者评论文本的“中国大学 MOOC”和“Coursera”为主要平台，爬取这两个平台上面向中文学习者的汉语综合课评论文本进行分析。本研究在“中国大学 MOOC”和“Coursera”平台上搜索到的面向中文学习者的汉语综合课共 16 门，如表 2 所示。数据爬取日期为 2024 年 4 月 26 日，最终共获得 6520 条评论，其中 6308 条来自“Coursera”，212 条来自“中国大学 MOOC”。经数据清洗后，最终得到有效数据 5339 条。

表 2 “Coursera”“中国大学 MOOC”平台汉语综合课程一览表

MOOC 平台	课程名称	开课单位	评论数量
Coursera 10 门	Chinese for beginners	北京大学	5243
	More Chinese for Beginners	北京大学	391
	Mandarin Chinese 1 Chinese for Beginners	上海交通大学	322
	Mandarin Chinese 2 Chinese for Beginners	上海交通大学	62
	Mandarin Chinese 3 Chinese for Beginners	上海交通大学	36
	Learn Mandarin Chinese Capstone Project	上海交通大学	27
	Mandarin Chinese for Intermediate Learners Part 1	上海交通大学	44
	Mandarin Chinese for Intermediate Learners Part 2	上海交通大学	11
	Mandarin Chinese for Intermediate Learners: Part 3	上海交通大学	8
	Learning Chinese: Start From Scratch（零到一学中文）	台湾大学	164
中国大学 MOOC 6 门	初级汉语综合	浙江科技大学	32
	初级综合汉语	北京语言大学	16
	功能汉语速成	北京语言大学	35
	汉语 upup	武汉大学	48
	汉语初级入门	上海外国语大学	13
	你好，中文（中级） Intermediate Chinese	上海交通大学	68

3.2 研究路径

本研究的研究路径如图 1 所示。

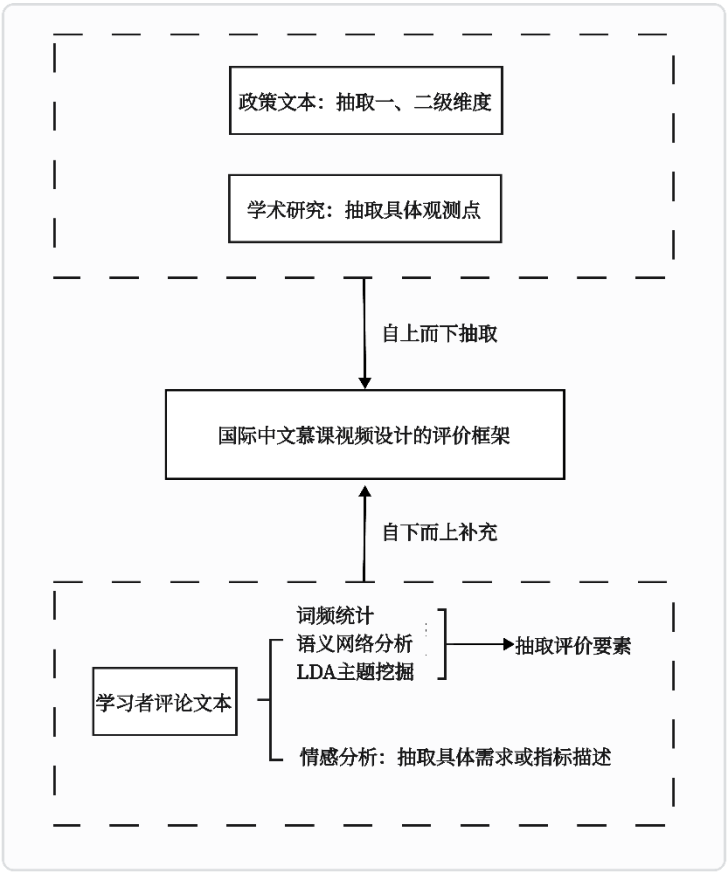


图 1　本研究路径图

4. 基于政策与学术视角构建的要素提取

4.1 政策文本的内容分析

这些文件共同强调了国际中文教育的现代化和数字化发展，包括利用新技术提升教育质量，构建多元化的教学资源体系，以及规范化的项目管理。这些政策的核心范畴旨在通过技术创新和资源整合，推动国际中文教育的持续发展和中文走向世界的进程。对于中文慕课视频设计来说，是其作为教学资源建设的最初环节之一。

为更准确地挖掘或反映政策文本中关于评价的核心要素，同时为更加系统和量化地对文件进行分析，本文采用内容分析法对以上 4 个文件进行分析，使用 MAXQDA 2022[9]软件对这些文件进行编码。整个过程如下：

[9] 具体参见：https://www.maxqda.com/zh-cn

首先，对 4 份文件进行初步分析，根据本文的研究问题，初步确定类目。这是一个预设的类别列表，用于指导后续的编码过程，本研究的预设类目为内容设计、交互设计、情感设计、区域国别化设计、创新设计 5 个。

其次，详细阅读每一段文本，以有完整意义的小句为单位，标记出与教学资源或教学视频设计相关的分析单元，共获得 173 个编码片段。然后将标记的编码片段分类到相应的类别中。每个编码片段只能归到一类，但不可出现无类可归的情况。如有标记的片段不属于先前预设类别列表中的任何一个，需要建立一个新的类别，这种情况可视为在政策文本中挖掘到了另外的潜在一级指标。本研究经过仔细阅读和梳理，并未出现这种情况。总之，整个过程为编码和归类分析的过程。最终结果如表 3 所示。

表 3 国际中文教育资源建设政策文本编码情况

类目	编码片段数量	出现文件数量	编码片段示例
内容设计	70	4	内容积极向上，丰富多彩 反映中华优秀传统文化
交互设计	34	4	丰富多样、互动性强的活动和任务 版式和界面设计清晰美观、方便使用
情感设计	34	3	便于教师发挥教学能动性 从学习者的实际需求和兴趣出发，激发并保持学习兴趣
区域国别化设计	19	3	推进资源的国别化和区域化发展 提供本土化、个性化、系统化在线中文教育解决方案
创新设计	16	3	引入虚拟现实、增强现实、扩展现实等技术 鼓励新技术融合

预设类目在政策文本中出现的比例如图 2 所示：

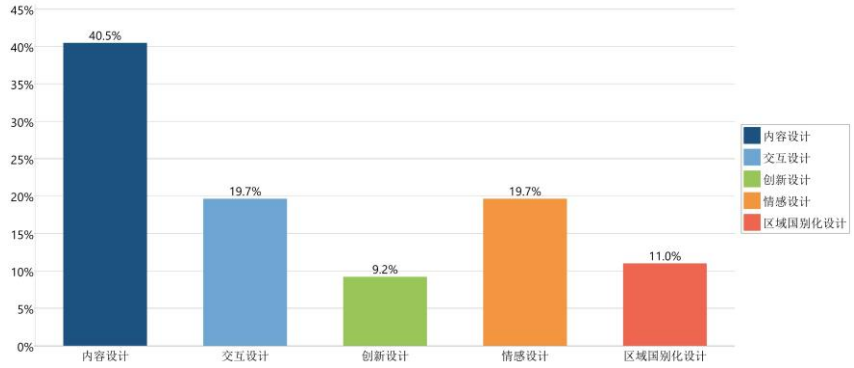


图 2 国际中文教育资源建设政策文本中预设类目占比

从表 3 和图 2 可知，所有 4 份文件均涉及内容设计和交互设计，而情感设计、创新设计、区域国别化设计也出现在了 3 份文件中。因此本文将一级指标暂定为内容设计、交互设计、情感设计、区域国别化设计、创新设计 5 个维度。提取到的国际中文慕课视频设计评价框架的一级指标和二级指标如表 4 所示。

表 4 国际中文教育资源建设政策文本中提取的一级指标和二级指标

一级指标	二级指标	在政策文本中标注的编码要素
内容设计	教学内容	教学目标、文化国情、教学方式
	文本内容	文本材料
	听觉内容	教学媒介语、声音
交互设计	人机交互	练习测试、板块设计
	交互界面	操控导航
情感设计	师生形象	教师、学习者
	视觉呈现	视觉线索、视频制作质量
区域国别化设计	未提取到	无
创新设计	创新元素	游戏设计、激励设计
	创新技术	新型信息技术

4.2 学术研究文献的要素形成

具体观测点是对国际中文慕课视频设计评价实现的具体方案，包含具体观测点及其要求，必须具备可操作性和可执行性。本文考察了 39 篇文章（见附录一）涉及教学视频设计的具体维度或观测点研究。这些设计元素或观测点能够反映目前学术界对教学视频设计的关注点所在。本文将其中与本研究不相关的具体技术参数以及视频拍摄/录制的形式等元素删除，对相似概念说法进行归并后，共得到 23 个具体观测点，如表 5 所示。

表 5 学术研究中提取到的教学视频设计的具体观测点

具体观测点	指标要求
教学目标	提供教学目标的文本呈现
教学材料	主题明确且单一
先行组织者	呈现适当的引导性材料
字幕呈现方式	呈现字幕，中外对照可选
教学媒介语	中文或适量的中外结合
背景音乐	适当使用背景音乐
视频时长	视频长度较短
控制视频进度	允许学习者自己控制（包括播放、暂停、快进、后退、倍速等），根据教学环节可视化时间轴

具体观测点	指标要求
嵌入问题或测试	视频中应有嵌入问题或测试，供学习者多练习
控制字幕	允许学习者选择字幕是否呈现以及如何呈现
弹幕	允许或控制弹幕呈现，并能发送弹幕
界面类型	课堂实录式、教师与 PPT 结合式、研讨式和动画式
学习者呈现	外国人可出镜
教师在屏	呈现教师形象或动画人物代理
教师画面大小	教师画面占屏幕面积比例适中
教师面部表情	面部表情丰富
教师目光	教师有目光引导
教师手势	教师使用指示性手势
教学语言风格	教师采用对话式风格、幽默
视频整体颜色	与教学内容相适宜
视觉线索	使用图片、动画、颜色、闪烁、箭头、线框、加粗、加大字号等提示重要信息
视频制作质量	视频分辨率高、音质清晰
游戏化设计	有游戏等激励形式

此外，本研究还根据前文所述理论基础、国际中文教育的学科专业特点以及政策文本中标注的编码要素，梳理出具体观测点共 4 项，如表 6 所示。

表 6 理论基础和政策文本中提取到的具体观测点

具体观测点	指标要求	增加项理由或文献出处
教学方式	选择合适的教学方式	政策文本中的编码要素
文化国情	融入中华文化及当代中国国情	政策文本中的编码要素
视频切块	在重要的信息后停顿，学习者可点击继续播放	梅耶的分段原则，卜彩丽等（2021）
新型信息技术	适时适当使用虚拟现实、增强现实、扩展现实等技术，模拟真实交际场景；融入 AI 智能伴学等	政策文本中的编码要素

将表 6 中 23 个具体观测点和表 7 中 4 个观测点取并集，得到 27 个具体观测点。因此本研究的初始版评价框架共包含 5 个一级指标、9 个二级指标、以及 27 个具体观测点。

5. 基于学习者体验视角的要素提取

5.1 高频词分析

本研究使用武汉大学沈阳（2008）开发的 ROST-CM6 软件导入停用词词表[10]，对最终获得的数据条目进行了分析和处理，并进行了词频统计。前 50 个高频特征词如表 7 所示。

表 7 学习者评论文本中前 50 高频词及词频

排名	词语	词频	排名	词语	词频	排名	词语	词频	排名	词语	词频	排名	词语	词频
1	课程	2986	**11**	简单	326	**21**	汉字	223	**31**	语法	136	**41**	学生	113
2	学习	1516	**12**	知识	300	**22**	教学	185	**32**	清楚	136	**42**	材料	112
3	中文	975	**13**	基础	296	**23**	日常	156	**33**	对话	136	**43**	英语	111
4	老师	880	**14**	中国	268	**24**	方法	155	**34**	完成	134	**44**	水平	104
5	汉语	846	**15**	发音	263	**25**	时间	154	**35**	句子	125	**45**	机会	104
6	初学者	783	**16**	解释	259	**26**	清晰	154	**36**	结构	124	**46**	复习	95
7	理解	570	**17**	词汇	253	**27**	教授	153	**37**	文化	124	**47**	易懂	92
8	语言	451	**18**	单词	232	**28**	使用	148	**38**	问题	120	**48**	精彩	92
9	适合	352	**19**	普通话	230	**29**	测验	145	**39**	入门	119	**49**	讲解	90
10	有趣	345	**20**	练习	230	**30**	拼音	143	**40**	高兴	116	**50**	口语	86

从词频统计情况来看，中文学习者评论文本中出现频率较高的词汇主要集中在以下五个方面：

一是与课程相关的通用术语。“课程”是出现频率最高的词，但需要注意，这类词汇（如“学习”“中文”“汉语”等）很可能是评论时采用的必要通用术语，它们构建了基本的交流语境，而非必然代表学习者的特定关注点或价值判断。要真正理解学习者的关注焦点，需要进一步分析这些通用术语的具体语境和与之相关联的词汇。

二是教师与教学方式。高频词“老师”反映了学习者对教师的重视。“教授”一词在汉语中有两种含义：一种是指教师的职称，用来指代教师；另一种是讲解传授知识、技能。由于外文翻译的问题，为了更准确地分析，本研究对评论文本中涉及“教授”的条目进行了逐一审阅，确认学习者通常使用“professor”来指代教师，这进一步突显了他们对教师的关注。“教学”“方法”等词则表明学习者对教学方式的关注。因此，可以提取出教师和教学方式两个要素。对于教师而言，应选择经

[10] 具体参见：https://gitcode.com/open-source-toolkit/595be/?utm_source=tools_gitcode&index=top&type=card&

验丰富且教学风格受欢迎的教师来授课，以提升视频及整体课程的吸引力和教学效果。同时，鼓励教师在授课过程中保持积极、热情的态度感染学习者。对于教学方式，需采用多样化的教学手段，如讲解、演示和练习等，以满足不同学习者的需求。此外，应注重教学方式的清晰度和易理解程度，确保学习者能够轻松跟随并理解课程内容。

三是教学内容与难度。“基础”“汉字”“词汇”“单词”“语法”“拼音”表明学习者关注中文的语言要素。同时，“句子”“对话”“日常”等高频词也提示我们，学习者也在意教学视频中是否呈现了例句和日常对话。“对话”“句子”“口语”等也表明学习者关注语言技能和应用，特别是口语和对话能力。而“简单”“适合”“初学者”等词表明学习者关注课程的难易程度。“文化”表明学习者对中华文化的兴趣，也反映了他们的学习需求。这个方面可抽取出教学内容这个要素。教学视频设计需确保课程内容丰富、有趣，覆盖学习者关注的高频词汇所代表的知识点，如基础词汇、语法、日常对话等。设计结构清晰、逻辑连贯，使学习者能够循序渐进地学习，从简单到复杂，逐步提升中文水平。在课程中融入文化元素，以提升学习者对中华文化的了解和兴趣。

四是教学效果与反馈。“理解”“清楚”“清晰”“易懂”等词表明学习者关注教学视频本身和呈现内容的清晰度以及易理解程度。“完成”“测验”“复习”则反映了学习者对学习进度和复习测验等反馈机制的重视。因此，可抽取出视觉呈现和交互体验两个要素。在视觉呈现方面，应利用现代技术优化教学视频的制作和呈现效果，支持学习者随时随地访问和学习。在交互体验方面，应提供及时的反馈机制，如测验、练习等，以帮助学习者巩固所学知识并跟踪学习进度。

五是学习者情感体验。“有趣”“高兴”等词表明学习者对教学内容趣味性的关注以及在观看教学视频过程中的情感体验。为增强学习者的情感体验，教学内容应注重趣味性和互动性。教师在授课时保持积极、热情的态度，也能进一步激发学习者的情感共鸣，使他们在学习过程中感受到乐趣和满足感。

基于这些高频词及相应的分析，对应教学视频设计的评价要素可抽取出教师、教学方式、教学内容、视觉呈现和交互体验等维度。

5.2 语义网络分析

本研究采用语义网络分析（Semantic Network Analysis；Carley & Kaufer，1993；Carley，1997）对评论文本进行解构，识别文本中词汇之间的关联情况，分析得出中文学习者对教学视频的整体认知和评价情况，如图 3 所示。语义网络分析把文本中的词语或概念看作网络中的“节点”，把它们在同一句话或同一条评论中共现的关系视作“连线”（Quillian, 1968；Collins & Quillian, 1969）。由此生成的图示可以通过节点的共现频率和中心度等指标衡量各概念的重要性（Carley & Kaufer, 1993），帮助快速捕捉学习者最关注的评价要点。

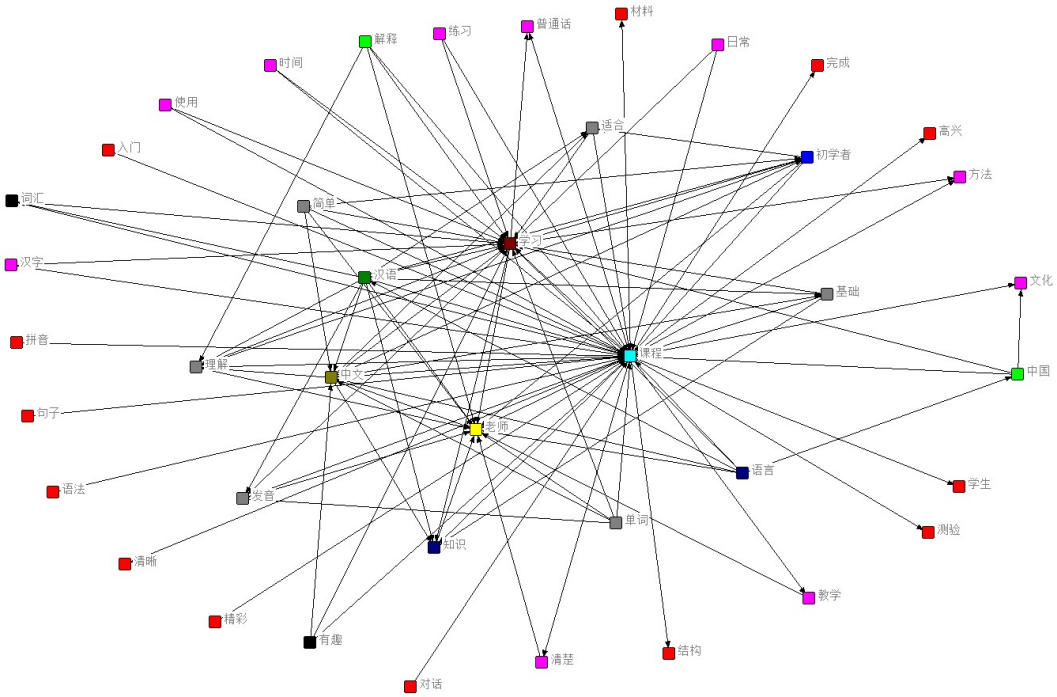


图 3 学习者评论中高频词语义网络关系

图 3 从内到外，按照节点的大小可以分为 3 个层次：

一是核心层，与课程和教师相关。这些节点的中心度（degree）值均大于 10，共有 5 个，分别为“课程”“中文”“汉语”“老师”“学习”，它们处于中心位置。这些节点与其他节点有很多连接，表明它们在评论中被频繁提及，是学习者关注的核心。

二是中间层，与教学方式相关。这些节点度在 5—10 之间（蓝色和灰色部分），分别为“初学者”“基础”“语言”“单词”“知识”“发音”“理解”“简单”“适合”。这些节点表明从学习者的角度来说，关注教学内容是否与学习对象所处的学习阶段匹配，教学方式和教学内容是否合适及易于理解。

三是边缘层，与教学内容、交互体验和视频本身属性相关。这些节点度在 5 以下（红色、粉色、绿色、黑色部分）。例如“词汇”“汉字”“语法”等表达的是学习者对语言要素学习的期待和体验，属于教学内容。“清晰”“清楚”“解释”指向“老师”，可以理解为学习者对授课教师的发音和讲解方式的要求。

总体来看，以上语义网络分析图揭示了 MOOC 评论文本中学习者的主要关注点，包括教学内容、教学方式、教学活动的两个主体、交互体验等方面。

5.3 主题挖掘分析

本研究采用 LDA 主题分析总结中文学习者关注的教学视频设计维度，这些维度也可以归纳为评价框架的要素。

LDA（Latent Dirichlet Allocation）主题分析，又称隐含狄利克雷分布，是一种无监督学习方法，最早由 Blei 等人在 2003 年提出。作为一种常用的文本挖掘技术，LDA 能够将评论文本按照概率分布归纳为若干潜在主题。通过 LDA 主题分析，可以揭示评论文本中隐藏的主题结构，并了解不同主题的关键词及其分布情况（Blei et al.，2003）。确定主题数量的常用方法主要有以下三种：第一种是根据经验进行反复调试，观察聚类的数量和效果，经过人工判断后最终确定。第二种是进行困惑度（Perplexity）计算。困惑度是一种常用的评估主题模型质量的指标。困惑度越低，模型越好，由此来确定最佳主题数量。第三种是进行一致性（Coherence）计算。通常，一致性数值越高越好。因人工判断的主观性较强，因此本研究采用一致性和困惑度计算相结合的方式来寻求最佳主题数量，具体使用 R 语言来实现。当主题数量从 2—20 个时，学习者评论的 LDA 一致性和困惑性检验情况，分别如图 4 和图 5 所示。

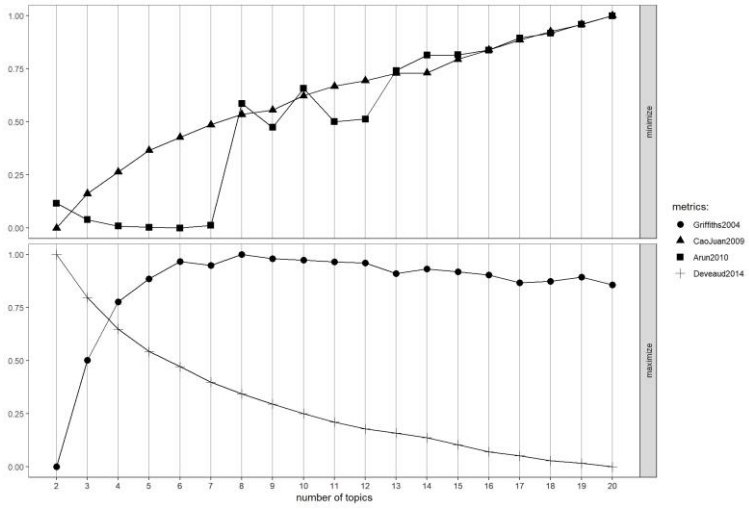


图 4 学习者评论 LDA 一致性检验结果

图 4 展示了 LDA 一致性检验结果。图中包括两部分，分别是需要最小化和最大化的主题数量。Griffiths2004（圆点）和 Deveaud2014（加号）都表示需要最大化的指标。数值越小时表示数量越好。两者交汇处在 4 附近。CaoJuan2009 三角线随着主题数量呈上升趋势。Arun2010 方块线在 4—6 之间数值均比较小。综合考虑这些指标，4、5、6 是三个比较明显的主题数量候选者。

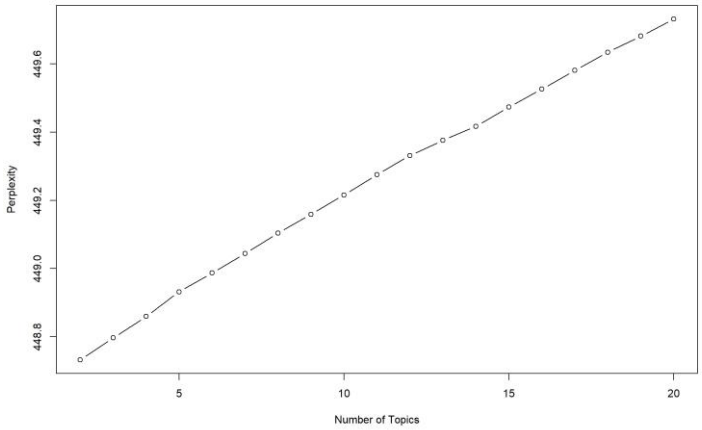


图 5 学习者评论 LDA 困惑度检验结果

图 5 显示了困惑度（Perplexity）随主题数量变化的情况。困惑度整体趋势是上升的。在该检验中，4 的困惑度更小，因此结合一致性检验结果，可以确定学习者评论文本的最佳主题数量为 4 个。使用 R 语言进行分析后，主题分布如图 6 所示。

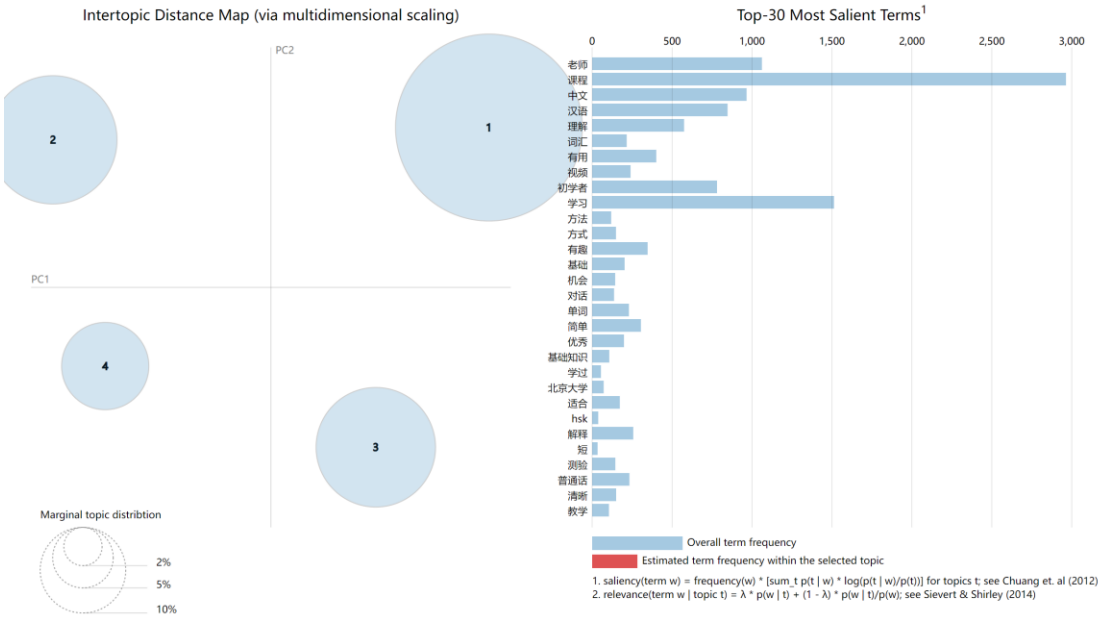


图 6 学习者评论 LDA 主题分布

由图 6 可知，4 个主题圈没有重叠，这意味着这些主题在学习者评论中具有较为明显的区分度，可赋予不同的概念。图 7 至图 10 分别是 4 个主题及其高频特征词的展示。

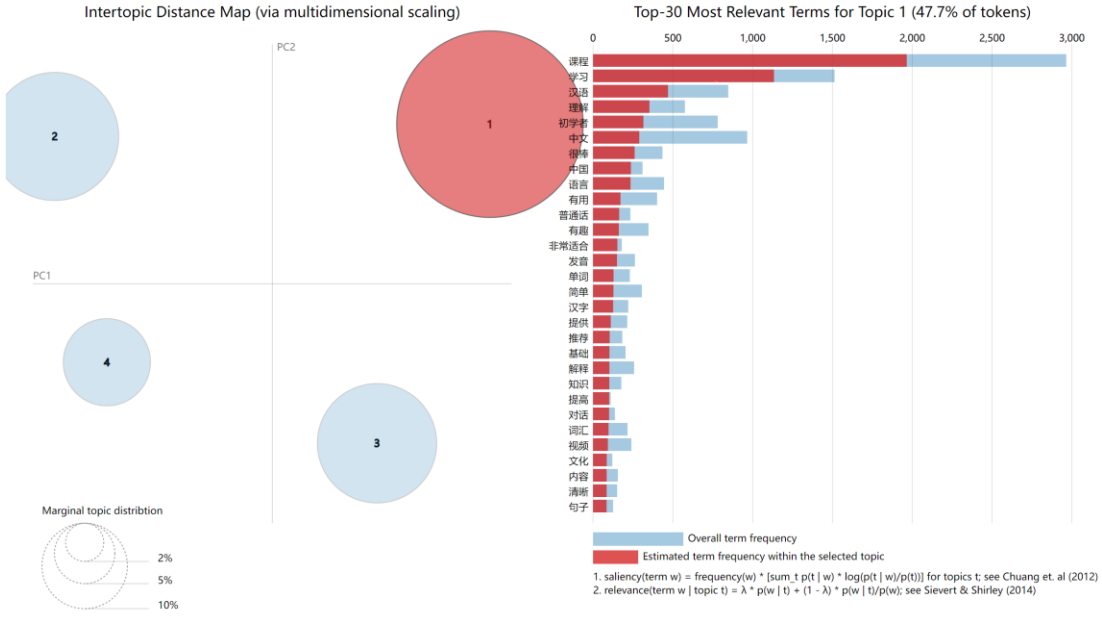


图 7 主题一前 30 个高频特征词展示

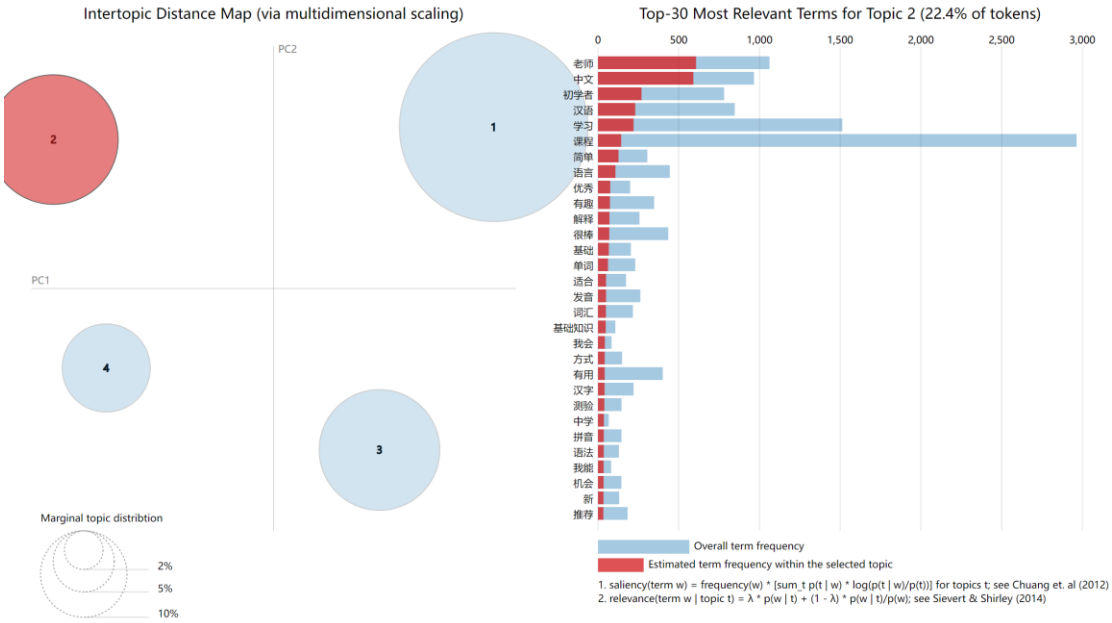


图 8 主题二前 30 个高频特征词展示

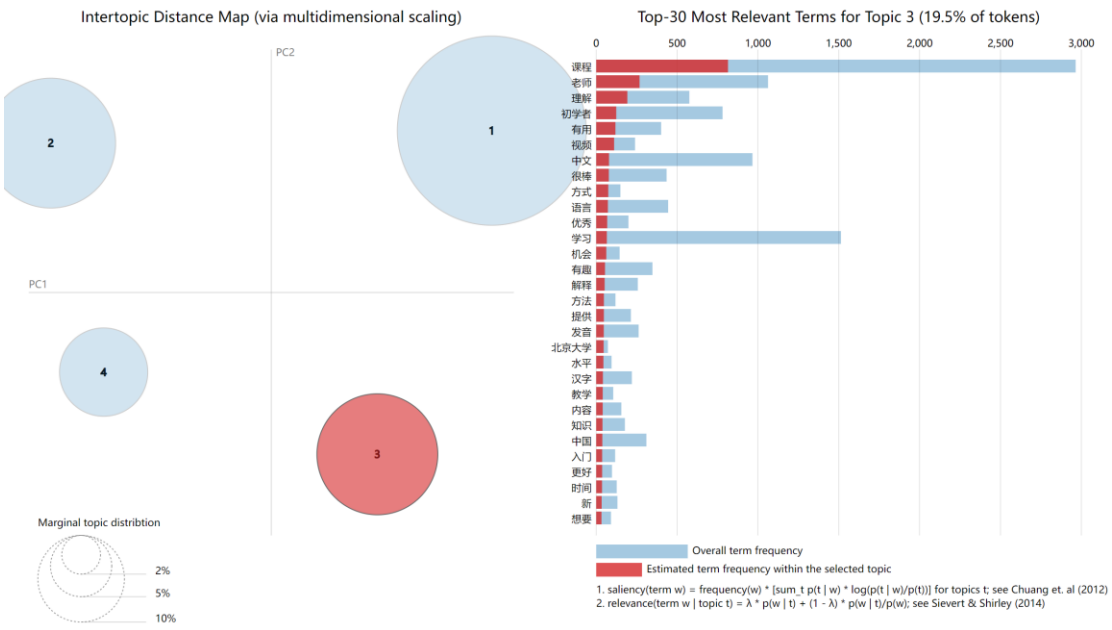


图 9 主题三前 30 个高频特征词展示

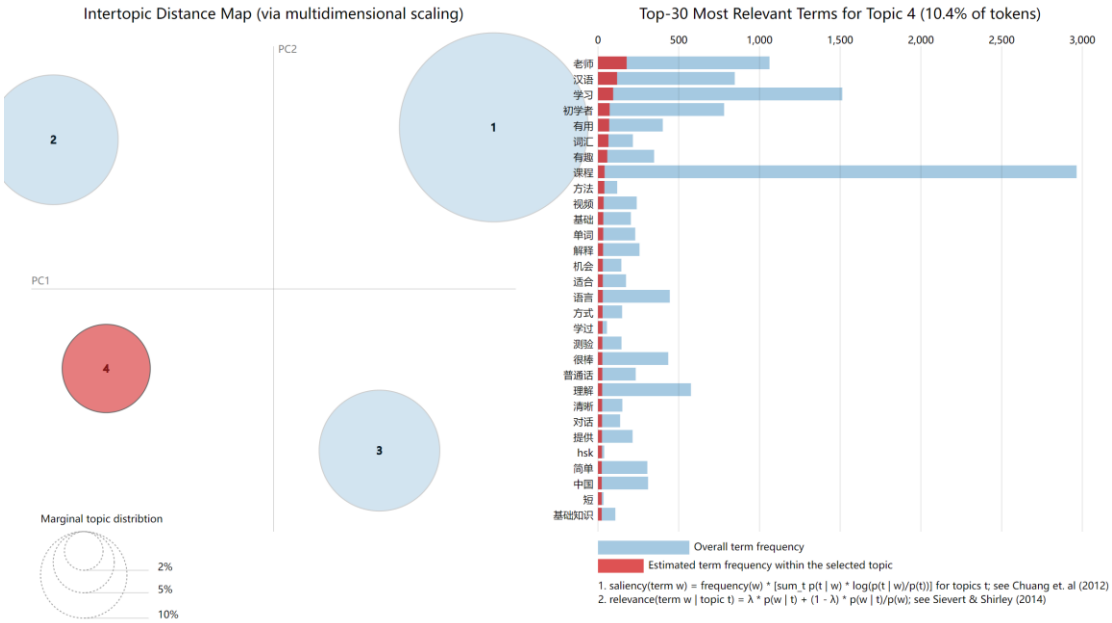


图 10 主题四前 30 个高频特征词展示

本文根据这 4 个主题类及其排名前 30 的特征词而赋予每个主题一个概念，如表 8 所示：

表 8 LDA 主题、特征词及权重展示

主题	特征词及权重
教师及教学方式	课程 0.118, 老师 0.039, 理解 0.028, 初学者 0.018, 有用 0.017, 视频 0.016, 中文 0.011, 很棒 0.011, 方式 0.011, 语言 0.010, 优秀 0.010, 学习 0.009, 机会 0.009, 有趣 0.008, 解释 0.008, 方法 0.007, 提供 0.007, 发音 0.007, 北京大学 0.007, 水平 0.006, 汉字 0.006, 教学 0.006, 内容 0.006, 知识 0.006, 中国 0.005, 入门 0.005, 更好 0.005, 时间 0.005, 新 0.005, 想要 0.005
教学内容	课程 0.116, 学习 0.067, 汉语 0.028, 理解 0.021, 初学者 0.019, 中文 0.017, 很棒 0.015, 中国 0.014, 语言 0.014, 有用 0.010, 普通话 0.010, 有趣 0.010, 非常适合 0.009, 发音 0.009, 单词 0.008, 简单 0.008, 汉字 0.007, 提供 0.007, 推荐 0.006, 基础 0.006, 解释 0.006, 知识 0.006, 提高 0.006, 对话 0.006, 词汇 0.006, 视频 0.005, 文化 0.005, 内容 0.005, 清晰 0.005, 句子 0.005
交互和情感体验	老师 0.048, 汉语 0.032, 学习 0.025, 初学者 0.020, 有用 0.019, 词汇 0.017, 有趣 0.016, 课程 0.011, 方法 0.011, 视频 0.009, 基础 0.009, 单词 0.009, 解释 0.008, 机会 0.008, 适合 0.008, 语言 0.008, 方式 0.008, 学过 0.008, 测验 0.007, 很棒 0.007, 普通话 0.007, 理解 0.007, 清晰 0.007, 对话 0.007, 提供 0.007, hsk0.007, 简单 0.006, 中国 0.006, 短 0.006, 基础知识 0.006
学习材料	老师 0.076, 中文 0.074, 初学者 0.034, 汉语 0.029, 学习 0.028, 课程 0.018, 简单 0.016, 语言 0.014, 优秀 0.010, 有趣 0.009, 解释 0.009, 很棒 0.009, 基础 0.008, 单词 0.008, 适合 0.006, 发音 0.006, 词汇 0.006, 基础知识 0.006, 我会 0.005, 方式 0.005, 有用 0.005, 汉字 0.005, 测验 0.005, 中学 0.005, 拼音 0.004, 语法 0.004, 我能 0.004, 机会 0.004, 新 0.004, 推荐 0.004

5.4 情感分析

本研究首先将不同评论归属到上文中四个主题下，然后对评论文本进行情感分析，计算积极、消极和中立这三种情感的分类情况，以便挖掘中文学习者对设计维度的好感度情况，为之后的设计评价提供参考。

综合考虑本研究的数据量规模以及几种常见的文本情感分析方法的优缺点和使用场景，最终选择采用深度学习中的 DistilBERT 模型（见下页图 11）。该模型是原始 BERT（Bidirectional Encoder Representations from Transformers）模型的精简版，具有轻量化和高效的特点，模型参数数量约为原始 BERT 的 60%，但基本完全保持了原模型的语言理解能力，并且在推理时的速度显著提升。

本研究使用公开可用的 distilbert-base-multilingual-cased-sentiments-student 模型[11]，该模型基于 tyqiangz/multilingual-sentiments 数据集进行训练和评估，在预测结果与其教师模型相比，达到了 88.29%的一致性（Yuan，2023）。这表明该模型在训练过程中有效地学到了教师模型的预测规律。本研究在 5339 条数据中随机标注了 600 条数据的情感标签，以测试该模型在本研究数据集上的性能。为便于理解，举例如表 9 所示：

[11] 具体参见：https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student

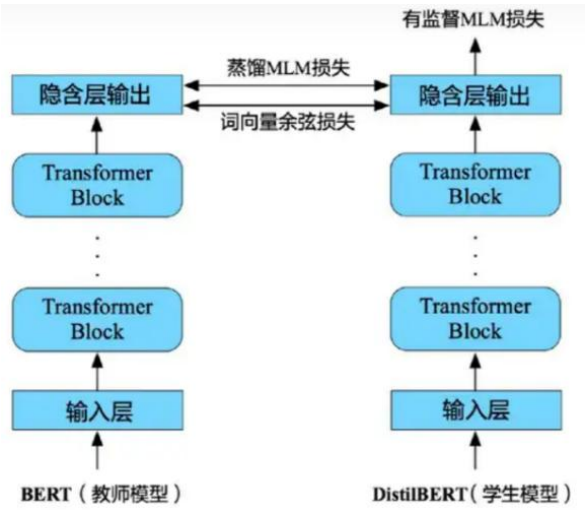


图 11 DistilBERT 模型的知识蒸馏过程

表 9 学习者评论文本情感标签分类举例

情感标签	评论内容
positive	真的学到了很多东西。我一开始对中文一无所知，但这门课程确实为我指明了学习中文的正确方向。
neutral	本课程非常适合初学者学习中文。遗憾的是，中文里没有关于结构语法的解释。
negative	西班牙语字幕有时与老师所讲得不一致（西班牙语字幕不好）。

结果显示准确率（Accuracy）为 0.93，召回率（Recall）为 0.93，F1 值（F1 Score）为 0.92，表明模型整体上表现很好，不仅识别准确，且在各个类别上的表现均衡且良好。因此，本研究直接用来对慕课评论文本进行情感分析，得到每个主题的情感分布情况如图 12 所示。

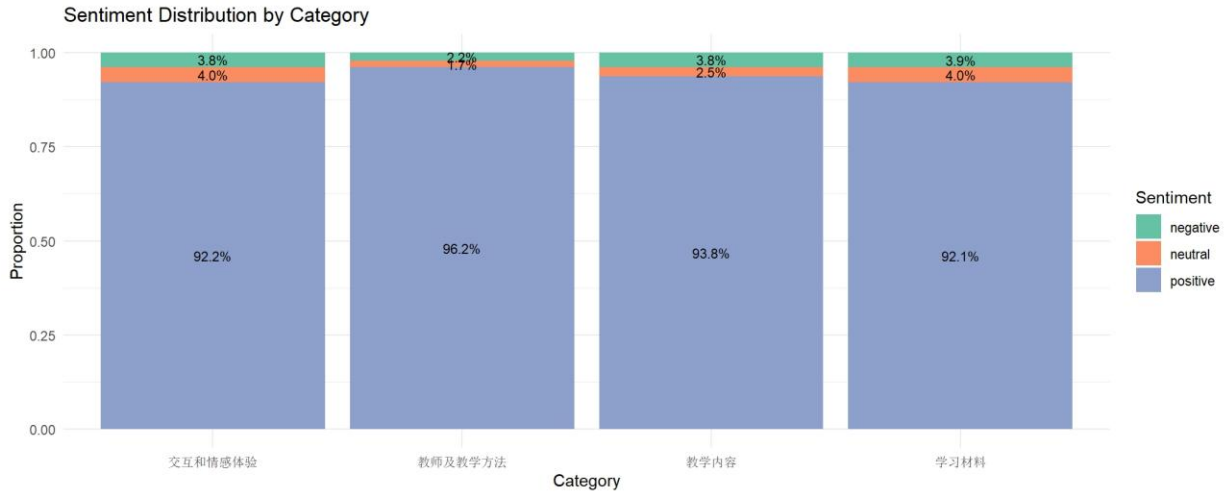


图 12 四个主题的情感分布情况

由图 12 可知，每个主题类别下的积极情感占比均超过了 90%，说明中文学习者整体评价是积极的。其中对教师及教学方式的积极情感最高，达到 96.2%，表明中文学习者对这个类别的体验是最佳的。在中立和消极情感方面，四个主题的分布情况差别不大。此外，虽然学习材料的积极情感占比也到了 92.1%，但是相较于其他三个主题是最低的，而且这个主题下消极情感占比最大，可以推断中文学习者对学习材料需要改进的需求最大。

学习者对现有的汉语综合课慕课程的积极情感占据了大多数，因此，以往基于经验制作的教学视频在大多数方面为学习者提供了良好的体验。本研究认为，中性和消极情绪的评论更能够体现学习者的反馈思考和意见建议，因此重点对中性和消极情绪的 324 条评论做进一步细致的人工分析。这些评论都关注一些具体细节的体验方面，可以归结到评价指标体系的具体观测点中。经分析总结，学习者在以下五个方面的体验不佳或需要进一步提升，具体情况如下：

第一，学习者反映教学视频语法讲解不充分。

第二，部分教学视频只展示了拼音，未展示汉字，或未教授汉字如何书写。

第三，学习者希望能提供其母语的字幕，尤其以西班牙语为母语的学习者需求最多，而现有的教学视频经常无法准确显示字幕。

第四，学习者对教学视频中听觉质量提出了要求，涉及教学媒介语、发音和背景音乐等。例如有学习者对教学媒介语的意见是，使用外文解释的时候应尽量精准，而不增加学习者的认知负荷，令其产生困惑。关于对话的发音部分，尤其是有中文学习者出镜时，相较于学习者真实的发音，部分学习者希望能够听到的是母语水平的对话。虽然背景音乐不能盖过对话声音是常识性问题，然而在实际的视频制作过程中可能并未得到充分重视，而导致学习者体验不佳。

第五，有学习者注意到文化差异，以及视频中涉及的某些敏感话题，例如与国家或性别相关的内容。因此，抽取出对区域国别化设计的具体指标要求：考虑不同文化特征和文化背景，适时适当加入文化对比，避免文化误解和敏感话题。

5.5 评价指标要素归纳

从学习者体验视角来看，评价指标的颗粒度较细，基本可对应具体观测点，相应的评论文本可对应为指标要求或描述，在自上而下的政策文本和学术研究中均可找到相应的名称，总体来说不如其全面。本文将从学习者体验视角提取到的要素与前文中的名称相统一，结合情感分析，其对应指标的要求或描述共 12 项，如表 10 所示。

表 10 学习者体验视角的结构要素及其描述

序号	指标要素	指标要求或描述
1	教学材料	提供语音、语法（提供语法讲练）、词汇、汉字（提供汉字书写指导和呈现，而不仅仅是拼音）的综合讲练
2	教学方式	体现精讲多练，在需要练习的地方，提供充足的时间确保学习者跟练；提供大量例句，并能体现日常对话
3	中国文化	融入中国文化
4	字幕	提供准确的、多语种字幕
5	教学媒介语	中文为主，可辅以适量精准的外文解释
6	背景音乐	最好不要有背景音乐，如果有，音量不能盖过对话
7	学习者呈现	学习者可出镜，发音须接近母语者，而非洋腔洋调
8	教师在屏	呈现教师形象
9	教师面部表情	亲切、面带微笑、积极、热情的态度
10	视频制作质量	画质高清、播放流畅、稳定易用
11	嵌入问题或测试	适当嵌入交互性反馈机制
12	区域国别化设计	生词或对话的文本翻译匹配目标国文字；考虑不同文化特征和文化背景，适时适当加入文化对比，避免文化误解和敏感话题

6. 多元视角的要素对比与分析

将表 10 与前文中的具体观测点及要求进行对比和分析后，做如下处理：

第一，对只有政策文本和学术研究视角，或只有学习者体验视角的指标要求，保留该视角的描述，以确保该具体观测点能够被观察和评价。

第二，对语义不冲突的、但又能表达不同意思的指标要求进行整合，以提供更全面的视角或优化表述。

第三，对描述相似的指标要求进行归并，以减少冗余，提高评价指标体系的简洁性和效率。

此外，这里需要对区域国别化设计这个指标进行讨论。在政策文本部分，该指标被提取为一级指标，但在相关政策文本中未能找到其他相关要素，同时在与教学视频设计相关的学术研究中也未提取出可供观察的具体观测点。而在情感分析环节，成功提取出了与该指标密切相关的具体要求或描述。因此，为确保这些具体要求在指标体系中得到体现，本研究认为在保留该指标的前提下，应对其层级结构及归属进行必要的调整。首先讨论层级结构，因无法再抽取出更下位的指标或观测点，因此应将该指标归到具体观测点中。其次讨论其归属，即应该归属到哪个二级指标下。从对区域国别化设计的理解来看，首要任务是在视频中确保文本翻译的准确性，使目标国用户能够准确理解信息，这与教学内容关注信息传达的核心理念相符。此外，考虑不同文化特征和文化背景，适时适当加入文化对比，是为了使内容更加适切，避免误解，这也属于教学内容范畴内对适切的追求。而避免文化敏感话题，确

保教学视频可被广泛接受，也是教学内容中不可或缺的一部分。综上所述，本研究将区域国别化设计归入教学内容这个二级指标下，成为其下属的具体观测点，处理动作也为保留。具体观测点的数量也因此由 27 个变为 28 个。

根据上文对评价指标要素的归纳和调整情况，形成了经学习者体验视角补充后的国际中文慕课视频设计评价框架。此版本包含 4 个一级指标、9 个二级指标和 28 个具体观测点，结果见附录二。

7. 框架的应用价值与实践指引

本研究构建的评价框架既保证了框架的系统性，又兼顾了一线可操作性，为教学实践提供了可量化、可追踪的评价和诊断工具。其实际应用价值与实践指引体现在以下三个方面：

第一，设计诊断，明确改进方向。一线教师可将本框架作为国际中文慕课教学视频的“体检表”，一线教师可在“备课——录制——发布”三个环节进行循环诊断。在备课阶段，对照 28 个具体观测点制作视频脚本时，可采用“红黄绿”三色标注法——绿色代表已充分准备，黄色代表需要加强，红色代表尚未涉及。例如，在展示汉字时，教师可逐项检查：是否设计了笔顺动画，是否留有学生跟写时间，是否融入了文化背景等。在录制阶段，每完成一个片段，立即对照框架进行快速自检。如发现背景音乐已盖过教师声音或讲解语音时，可立即调整音轨；若“交互设计——嵌入问题或测试”缺位，可补充字幕或图标提示等。在视频发布后，可持续收集学习者评论，将反馈与相应观测点一一对应，快速定位并记录问题，为后续迭代奠定基础。

第二，课程改进，形成迭代清单。慕课开发团队可将 28 个具体观测点转化为开发设计清单，按照“内容—交互—情感—创新”四大维度，根据专家建议设置“必选项”和“加分项”。例如，“必选项”包括基础的内容准确性、画面清晰度等；加分项涵盖创新元素等。在项目启动时，可根据一级或二级维度明确分工和具体责任人。在中期监控的时候，团队对照框架逐项汇报进度。未达标项自动进入“问题清单”，限期整改。在后期优化阶段，例如课程上线 3 个月后，可结合平台数据（课程完成率、互动题的正确率等）和框架评分，制定下一版本改进计划。建立“版本对比表”，直观展示每次迭代在各维度的提升幅度。

第三，质量评估，支撑多主体决策。教学管理者可基于该框架为每个维度设置权重，每个具体观测点可采用 5 分制进行评分，总分转化为星级，便于横向比较和纵向追踪。在教师自评时，课程上线前可完成自评表，反思薄弱环节。在同行互评时，可邀请几位同领域教师交叉评审，聚焦专业准确性。在学生评价时，可根据此框架，设计简化版问卷，聚焦学习体验相关指标。在专家诊断时，邀请教学设计专家进行深度诊断，提供改进建议。评估结果可用于平台准入（设最低星级门槛）、

资源分配（高分课程优先推广与经费支持）、教师培训（将薄弱维度纳入研修主题）及精品课认证（达到最高标准方可申报）。同时，结合观看率、参与度等学习行为数据，可进一步提升质量管理的科学性与透明度。

8. 结语

本研究立足国际中文慕课视频设计的实践需求，以政策文本和学术研究为切入点，结合从实际学习者评论中得到的体验反馈，构建并完善了中文慕课视频设计评价框架。相较于以往研究仅关注宏观课程或单一视角的做法，本研究综合运用多种文本挖掘方法，兼顾宏观政策参考、中观学术研究与微观学习体验，挖掘的指标层次更为细化，使得评价框架更易落实到慕课视频设计和开发之中。与此同时，本研究仍有不足：（1）学习者评论样本主要采自“中国大学 MOOC”和“Coursera”两家平台，结果的外推性需进一步验证。如果未来能够进行更系统、更广泛的问卷调查，获取到足够数量的有效问卷，将为研究提供更加全面的基础数据。（2）评价框架尚未在更多元化的教学场景和学习者群体中进行应用与迭代修正。未来，研究可进一步扩展采样范围，纳入不同平台的学习者评论，结合机器学习、情感分析等技术提升自动化分析能力；同时，框架指标可在实际视频开发与教学中持续验证和完善，并特别关注不同文化背景下学习者对设计要素的差异化需求；此外，还可以探索与生成式人工智能结合，实现实时反馈和脚本迭代，不断提升国际中文慕课视频的教学效果与学习体验，从而推动国际中文慕课视频评价框架的持续优化与创新。

致谢：本文由（中国大陆）国家社科基金项目“国际中文教师数字素养智慧研修平台构建与应用研究”（24BYY044）资助。

参考文献

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645. DOI: 10.1146/annurev.psych.59.103006.093639

Bayeck, R.Y., & Choi, J. (2018). The Influence of National Culture on Educational Videos: The Case of MOOCs. *International Review of Research in Open and Distributed Learning, 19*(1),186-201. DOI: 10.19173/irrodl.v19i1.2729

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 03, 993-1022.

Bu, C.L., Cao, W.Y., Guo, J.J., & Shen X.J. (2021). Research on Principles, Theoretical Basis and Evaluation Tools of Instructional Video Design: A Systematic Review based on 72 Empirical Studies Abroad. *Modern Distance Education,*02, 35-44.

[卜彩丽, 曹婉迎, 靳宇, 郭佳佳, & 沈霞娟. (2021). 教学视频设计的原则、理论基础与评价工具研究——基于国外 72 篇实证研究的系统综述.*现代远距离教育*, 02, 35-44.]

Carley, K. M. (1997). Network Text Analysis: The Network Position of Concepts. In C. W. Roberts (Ed.), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts* (pp. 79–100). Routledge. DOI: 10.4324/9781003064060-5

Carley, K. M., & Kaufer, D. S. (1993). Semantic Connectivity: An Approach for Analyzing Symbols in Semantic Networks. *Communication Theory*, *3*(3), 183–213. DOI: 10.1111/j.1468-2885.1993.tb00072.x

Chen, C. (2021). On Self-regulated Learning of Chinese Language Learners in MOOCs. *Journal of International Chinese Teaching*, 01, 50-58. [陈晨. (2021). 汉语慕课学习者的自我调节研究. *国际汉语教学研究*, 01, 50-58.]

Chen, M. (2015). *A study on university MOOCs from the perspective of communication studies*. [Master's Thesis, Hunan Normal University]. [陈淼. (2015). *传播学视角下的大学 MOOC 研究*. 湖南师范大学硕士学位论文.]

Collins, A. M., & Quillian, M. R. (1969). Retrieval Time from Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. DOI: 10.1016/S0022-5371(69)80069-1

Cui, X.L. (2020). Chinese Teaching against the Background of Global Public Health Emergency. *Chinese Teaching in the World*, 03, 291-299. [崔希亮. (2020). 全球突发公共卫生事件背景下的汉语教学. *世界汉语教学*, 03, 291-299.]

Deng, R.Q., & Gao, Y.F. (2023). Using Learner Reviews to Inform Instructional Video Design in MOOCs. *Behav Sci (Basel)*, *13*(4), 330.

Ding, Y. (2019). What constitutes an effective instructional video: Perspectives from Chinese EFL learners. In B. Zou & M. Thomas (Eds.), *Recent developments in technology-enhanced and computer-assisted language learning* (pp. 236–256). Information Science Reference. DOI: 10.4018/978-1-7998-1282-1.ch011

Garrett, J. J. (2011). *The elements of user experience: User-centered design for the web and beyond* (2nd ed.). New Riders.

Huang, Y.Y. (2021). The Problems and Optimization Strategies of MOOC Education in China from the Perspective of Communication. *Higher Education Exploration*, 04, 72-77. [黄奕宇. (2021). 传播学视阈下我国慕课教育问题与优化策略. *高教探索*, 04, 72-77.]

Kay, A. C. (1972). A personal computer for children of all ages. *Proceedings of the ACM Annual Conference - Volume 1*. Association for Computing Machinery.

Kong, Y. (2024). A Probe into MOOC Teaching in International Chinese Language Education. *Theory and Practice of Education*, *44*(21), 61-64.[孔艳. (2024). 国际中文教育慕课教学研究. *教育理论与实践*, *44*(21), 61-64.]

Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. Longman.

Kuang, X., & Feng, L.P. (2023). The Policy Logic and Practical Approach to the Construction of Teaching Resources in International Chinese Education. *Journal of Tianjin Normal University (Social Sciences)*, 05, 10-21. [匡昕, & 冯丽萍. (2023). 国际中文教育教学资源建设理路与进路. *天津师范大学学报(社会科学版)*, 05, 10-21.]

Li, Q., & Liu, N. (2016). Design and Making Method of Teaching Videos in MOOC——An Empirical Study of Course based on Coursera and ed X Platforms. *Modern Educational Technology*, 07, 64-70. [李青, & 刘娜. (2016). MOOC 中教学视频的设计及制作方法——基于 Coursera 及 edX 平台课程的实证研究，*现代教育技术*, 07, 64-70.]

Li W.J., & Luo, R. (2023). Research on the Quality Evaluation of International Chinese MOOCs from the Perspective of Overseas Learners. *Research on Chinese as a Second Language*, 02, 61-77. [李雯静, & 骆蓉. (2023). 海外学习者视角下的国际中文慕课质量评价研究. *对外汉语研究*, 02, 61-77.

Li, Y. & Zhang, M.H. (2015). Study on College Students' Experience About MOOC-Based Flipped Classroom Practice——Based on the Analysis of 231 Online Learning Diaries. *Modern Distance Education Research*, (05):73-84+93. [李艳,张慕华. (2015). 高校学生慕课和翻转课堂体验实证研究——基于 231 条在线学习日志分析, *现代远程教育研究*, (05), 73-84+93.

Liu, F., Ziden, A., & Liu, B. (2024). Emotional support in online teaching and learning environment: A systematic literature review (2014–2023). *Journal of Curriculum and Teaching, 13*(4), 209-218. DOI:10.5430/jct.v13n4p209

Luo, S.J., Zhu, S.S., Shen, C.Y. (Eds.) (2022). *User Experience Design*. Higher Education Press. [罗仕鉴, 朱上上, & 沈诚仪编著. (2022). *用户体验设计*. 高等教育出版社.]

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp.31-48). Cambridge University Press. DOI: 10.1017/CBO9780511816819.004

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: The role of the speaker's voice. *Journal of Educational Psychology, 95*(2), 419–425. DOI: 10.1037/0022-0663.95.2.419

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review, 63*(2), 81–97.

Morville, P. (2004, June 21). *User experience design*. Semantic Studios. https://semanticstudios.com/user_experience_design/

Noriaki Kano. (1984). Attractive Quality and Must-Be Quality. *The Journal of the Japanese Society for Quality Control*, *14*(2), 39-48.

Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart and Winston.

Qiu, J.P., & Ou, Y.F. (2015). The Construction and Application Research of MOOC Quality Evaluation Indicators System. *Higher Education Development and Evaluation*, 05, 72-81+100. [邱均平, & 欧玉芳. (2015). 慕课质量评价指标体系构建及应用研究. *高教发展与评估*, 05, 72-81+100.]

Quillian, M. R. (1968). Semantic Memory. In M. Minsky (Ed.), *Semantic Information Processing* (pp. 227–270). MIT Press.

Revang, M. (2007, April 17). *The user experience wheel*. User Experience Project. https://userexperienceproject.blogspot.com/2007/04/user-experience-wheel.html

Rogers, C. R. (1969). *Freedom to learn*. Merrill.

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics, 11*(2), 129–158. DOI: 10.1093/applin/11.2.129

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. John Wiley & Sons.

Shen, Y. (2008). *ROST ContentMining System: Software for content mining and analysis*. Wuhan University. [沈阳. (2008). *ROST ContentMining System: Software for content mining and analysis*. Wuhan University.]

Stracke, C. M., & Trisolini, G. (2021). A Systematic Literature Review on the Quality of MOOCs. *Sustainability*, *13*(11), 5817. DOI: 10.3390/su13115817

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312. DOI:10.1016/0959-4752(94)90003-5

Tong, X.S., & Jia, X.J. (2017). Constructing quality evaluation system for MOOCs. *Chinese Journal of Distance Education*, 05, 63-71+80. [童小素, & 贾小军. (2017). MOOC 质量评价体系的构建探究. *中国远程教育*, 05, 63-71+80.]

Wang, T.M., & Zhang, Y. (2017). The Transition of Teacher's Roles in the Context of MOOC Education Based on a Narrative Research. *Curriculum, Teaching Material and Method*, 03, 110-115. [王添淼, & 张越. (2017). 慕课教学中教师角色转换的叙事研究. *课程.教材.教法*, 03, 110-115.]

Wang, X., Zhou, W., Wang, Z.J., & Han, M.Q. (2018). Optimum design in instructional videos in MOOCs: the case of the top 20 MOOCs in Class Central. *Chinese Journal of Distance Education*, 05, 45-54. [王雪, 周围, 王志军, & 韩美琪. (2018). MOOC 教学视频的优化设计研究——以美国课程中央网站 Top20MOOC 为案例. *中国远程教育*, 05, 45-54.]

Wang-Szilas, J., & Bellassen, J. (2017). Dualism-based design of the Introductory Chinese MOOC 'Kit de contact en langue chinoise'. *In Q. Kan & S. Bax (Eds), Beyond the language classroom: researching MOOCs and other innovations*, 43-57. DOI:10.14705/rpnet.2017.mooc2016.670

Wu, J.H. (2019). Research on the Teaching Evaluation System for MOOCs. *Education Modernization*, 03, 38-40. [武家辉. (2019). 慕课教学评价体系的研究. *教育现代化*, 03, 38-40.]

Xin, P. (2019). Analysis of Teaching Chinese as a Second Language MOOCs from the Perspective of Teaching Philosophy. *Journal of Higher Education*, 16, 6-9. [辛平. (2019). 教学理念视域下的对外汉语教学慕课分析. *高教学刊*, 16, 6-9.]

Yang, J.L. (2024). Analysis of Development Context and Problems of MOOC Policy in China. *Adult Education*, 05, 49-57. [杨继龙. (2024). 我国慕课政策发展脉络及问题分析. *成人教育*, 05, 49-57.]

Yong, L.T. (2013). User experience evaluation methods for mobile devices. *Third International Conference on Innovative Computing Technology (INTECH 2013)*, 281-286. DOI: 10.1109/INTECH.2013.6653647

Yuan, L. X. (2023, June 24). *distilbert-base-multilingual-cased-sentiments-student (Revision 2e33845)*. Hugging Face. https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student

Zhang, H. (2023). *Studies on MOOC for Teaching Chinese to Speakers of Other Languages*. Current Affairs Press. [张浩. (2023). *汉语国际教育慕课研究*. 时事出版社.]

Zhang, Y., Luo, Y., & Li, Y. (2024). China's outward-oriented internationalization pathways to MOOCs. *Media and Communication Research, 5*(1), 105-110. DIO: 10.23977/mediacr.2024.050117

Zhao, J.M. (Editor) (2006). *Teaching Chinese in This Way: Language Skills*. The Commercial Press. [赵金铭主编. (2006). *汉语可以这样教——语言技能篇*. 北京：商务印书馆.]

Zhen, G. (2019). The Current Situation and Strategies of the Construction of Course Moving Platform for Chinese International Education. *Education and Teaching Forum,* 46, 218-219. [甄刚. (2019). 汉语国际教育慕课平台建设现状及策略探究. *教育教学论坛*, 46, 218-219.]

Zheng, C.H. (2019). Standard Academic Translation: Current Situation Analysis and Countermeasure Research on Online Education for Teaching Chinese as a Foreign Language. *Journalism Communication*, 08, 34-35. [郑才华. (2019). 对外汉语在线教育的现状分析及对策研究. *新闻传播*, 08, 34-35.]

Zhou, W.F., & Ning, J.M. (2020). Construction of International Chinese MOOCs from the Perspective of Learners: A Comparative Approach. *Journal of International Chinese Teaching*, 03, 82-89. [周汶霏, & 宁继鸣. (2020). 学习者视角下的国际中文慕课建设：一种比较的路径. *国际汉语教学研究*, 03, 82-89.]

附录一 教学视频设计具体观测点的学术研究文献来源

序号	文献来源
1	Cao, R., & Liu, S. Y. (2017). Present Situation and Reflections of MOOC for Teaching Chinese as a Foreign Language. *Journal of Liaoning Normal University (Social Science Edition), 40*(06), 114-121. [曹儒, 刘思远. (2017). 对外汉语教学慕课的发展现状及思考. *辽宁师范大学学报(社会科学版)*, *40*(06): 114-121.]
2	Chen, M. N. (2020). *The Impact of Instructor's Facial Expressions and Eye Gaze on Learners' Learning in Instructional Videos*. Master's Thesis, Huazhong Normal University. [陈闽楠. (2020). *教学视频中教师面部表情和眼睛注视对学习者学习的影响*. 华中师范大学硕士学位论文.]
3	Chen, Z. R. (2021). *Study of Chinese Learning Video on Tiktok (Mobile Application Software)* [Master's Thesis, Shanghai Normal University]. [陈子柔. (2021). *移动端应用软件 Tiktok 中汉语学习短视频研究*. 上海师范大学硕士学位论文.]
4	Jin, H. G., Kim, S. E., & He, W. C. (2021). Pedagogical and Acquisitional Implications from Recent Video Caption Research in SLA. *Chinese Teaching in the World,* (01), 98-114. [靳洪刚, 金善娥, 何文潮. (2021). 视频字幕研究及其对二语习得和教学的启示. *世界汉语教学*, (01): 98-114.]
5	Li, J. X. (2020). *Effects on Different Subtitling Modes on Chinese Vocabulary Incidental Acquisition and Video Comprehension* [Master's Thesis, Lanzhou Jiaotong University]. [李剑雄. (2020).*字幕的不同呈现方式对汉语词汇附带习得和视频理解的影响*. 兰州交通大学硕士学位论文.]
6	Li, L. (2021). Principles of Screencast Video Design for Elementary Chinese Listening Teaching——Analysis Based on the Cognitive Load. *Journal of International Chinese Education*, (01), 69-84. [李琳. (2021). 初级阶段听力录播课教学视频设计原则——基于认知负荷的分析. *汉语国际教育学报*, (01): 69-84.]
7	Li, S. H. (2021). *Research On the Design of Chinese Micro-class Based on Principles of Multimedia Instruction Design—Take the Award-winning Works of the National Postgraduate Chinese Teaching Mico-class Competition as an Example* [Master's Thesis, Dalian University of Foreign Languages]. [李思涵. (2021). *基于多媒体教学设计原则的汉语微课设计研究——以全国研究生汉语教学微课大赛获奖作品为例*. 大连外国语大学硕士学位论文.]
8	Meng, X. D., & Xu, E. Q. (2020). Does Animated Online Teaching Videos Require Presenting Subtitles? ——Experimental Analysis based on the Comparison of College Students' Academic Performances. *Modern Educational Technology*, *30*(09), 42-48. [孟祥栋, 徐恩芹. (2020). 动画式在线教学视频需要呈现字幕吗？——基于大学生学习成绩对比的实验分析. *现代教育技术*, 30(09): 42-48.]
9	Peng, X. Y., Zhang, L. Y., & Fu, L. (2019). How MOOC Video Affects Students' Participation——Enhancing the Attractiveness of MOOC Video. *Digital Education, 5*(05), 42-47. [彭小艳, 张林燕, 傅霖. (2019). MOOC 视频如何影响学生学习参与度——增强 MOOC 视频吸引力研究. *数字教育*, 5(05): 42-47.]
10	Ren, Y. F. (2020). *Research on Chinese Teaching Video Based on Seamless Learning Theory* ——Taking YouTube Chinese Teaching Video as an Example [Master's Thesis, Fujian Normal University]. [任怡斐. (2020). *基于无缝学习理论的汉语教学视频研究——以 Youtube 汉语教学视频为例*. 福建师范大学硕士学位论文.]
11	Wang, C. Y., & Luo, H. M. (2017). Comparative Study on Students' Learning Performance of Two Different Teaching Videos: A Perspective from Video Production Cost. *E-Education Research, 38*(10), 68-73. [汪存友, 罗慧敏. (2017). 两种不同类型教学视频的学习效果比较研究——视频制作成本的视角. *电化教育研究*, 38(10): 68-73.]
12	Wang, C. Y., Wang, Y., & Chang, W. J. (2018). Teaching Video Design from The Perspective of Factors Affecting Autonomous Learning Effectiveness. *Chinese Journal of ICT in Education, 2018* (24), 5-9. [汪存友, 王彦, 常文静. (2018). 从影响自主学习效果的因素看教学视频设计. *中国教育*

序号	文献来源
	信息化, (24): 5-9.]
13	Wang, C. Y. (2015). An investigation of narrative subtitle usefulness in online instructional video clips. *Chinese Journal of Distance Education*, (12), 61-66+80. [汪存友. (2015). 对在线教学视频中讲解字幕有用性的调查与思考. *中国远程教育*, (12): 61-66+80.]
14	Wang, C. Y. (2016). Are Subtitles Necessary in Online Learning Video? The Modification of Redundant Effect. *E-Education Research, 37*(03), 59-65+85. [汪存友. (2016). 在线教学视频呈现讲解字幕有必要吗?——兼论对冗余效应边界条件的修正. *电化教育研究*, 37(03): 59-65+85.]
15	Wang, D. J. (2012). The Impact of Different Subtitle Presentation Methods on Foreign Language Learners' Video Comprehension. *E-Education Research, 33*(05), 85-90. [王电建. (2012). 字幕的不同呈现方式对外语学习者视频理解效果的影响. *电化教育研究*, 33(05): 85-90.]
16	Wang, H. Y., Hu, W. P., & Pi, Z. L., Et Al. (2018). Effects of the Instructor's Behaviors on Learning Performance in Teaching Video Lectures based on Eye Tracking. *Journal of Distance Education, 36*(05), 103-112. [王红艳, 胡卫平, 皮忠玲等. (2018).教师行为对教学视频学习效果影响的眼动研究. *远程教育杂志*, *36*(05): 103-112.]
17	Wang, J., Hao, Y. H., & Lu, J. L. (2014). An Experimental Study on The Impact of Teaching Video Presentation Methods on Independent Learning Effectiveness. *E-Education Research,* (03), 93-99+105. [王健, 郝银华, 卢吉龙. (2014). 教学视频呈现方式对自主学习效果的实验研究. *电化教育研究*, (03): 93-99+105.]
18	Wang, X. Y., Wu, X. M., & Hao, J. C. (2022). Analysis And Optimization Strategies for Teaching Videos in Design-Related Online Courses. *Digital Education, 8*(04), 17-23. [王小越, 吴筱萌, 郝佳畅. (2022). 设计类在线课程教学视频的分析及优化策略研究. *数字教育, 8*(04): 17-23.]
19	Wang, X., Han, M. Q., Gao, Z. H., Et Al. (2020). Research on the Mechanism and Optimization Strategy of Visual and Auditory Emotional Design in Instructional Videos. *Journal of Distance Education*, *38*(06), 50-61. [王雪, 韩美琪, 高泽红, 等. (2020).教学视频中视听觉情绪设计的作用机制与优化策略研究. *远程教育杂志*, *38*(06): 50-61.]
20	Wang, X., Wang, Z. J., & Hou, A. Z. (2016). The Eye Movement Study on the Design of Subtitles in Network Teaching Videos. *Modern Educational Technology, 26*(02), 45-51. [王雪, 王志军, 候岸泽. (2016). 网络教学视频字幕设计的眼动实验研究. *现代教育技术, 26*(02): 45-51.]
21	Wang, X. (2015). An experimental study on design and application of micro-video in colleges and universities. *Experimental Technology and Management, 32*(03), 219-222+226. [王雪. (2015). 高校微课视频设计与应用的实验研究. *实验技术与管理, 32*(03): 219-222+226.]
22	Wang, Y., & Dong, L. L. (2018). An International Comparative Study on The Design Principles of MOOC Teaching Videos. *Chinese Journal of ICT in Education,* (22), 10-15. [王岳, 董丽丽. (2018). MOOC 教学视频设计原则的国际比较研究. *中国教育信息化*, (22): 10-15.]
23	Xie, Y. H., Yang, J. M., & Pi, Z. L., Et Al. (2021). Effects of in-video questions and answers on learning performance. *Chinese Journal of Distance Education*, (12), 63-71+77. [谢耀辉, 杨九民, 皮忠玲等. (2021). 教学视频中前嵌问题与反馈对学习的影响研究. *中国远程教育*, (12): 63-71+77.]
24	Yang, J. M., Chen, H., Yang, W. D., Et Al. (2022). How to Design Subtitles in Instructional Videos? —— A Meta-analysis Based on 44 Experimental and Quasi-experimental Research Literatures from 2000 to 2021. *E-Education Research*, (08), 59-67. [杨九民, 陈辉, 杨文蝶, 等. (2022). 教学视频中的字幕应该如何设计？——基于 2000—2021 年 44 篇实验和准实验研究文献的元分析. *电化教育研究*, (08): 59-67.]
25	Yang, J. M., Liu, C. X., Chen, L. Q., Et Al. (2023). How Do Instructor's Depictive Gestures Take Effect in Instructional Videos?——The Influence of Gesture Learning Strategies and Learner's Spatial Ability. *Modern Distance Education Research, 35*(04), 92-101. [杨九民, 刘彩霞, 陈楼琪, 等. (2023). 教学视频中的描述性手势何以有效——手势学习策略及学习者空间能力的影响. *现代远程教育*

序号	文献来源
	研究, *35*(04): 92-101.]
26	Yang, J. M., Tao, Y., & Luo, L. J. (2015). The Presence of Teacher in MOOCs' Videos: Current Status and Future Research Agenda. *China Educational Technology,* (06), 59-63. [杨九民, 陶彦, 罗丽君. (2015). 在线开放课程教学视频中的教师图像分析:现实状况与未来课题. *中国电化教育*, (06): 59-63.]
27	Yang, J. M., Wu, C. C., Pi, Z. L., Et Al. (2019). Do Bullet Comments Promote or Interfere with Learning - A Meta-Analysis of The Impact of Bullet Comments on Learning. *Electrified Education Research*, *40*(06), 84-90+120. [杨九民, 吴长城, 皮忠玲, 等. (2019). 促进学习还是干扰学习——弹幕对学习影响的元分析. *电化教育研究*, *40*(06): 84-90+120.]
28	Yang, J. M., Xu, K., Han, J. X., Et Al. (2020). The Interaction Effects of the Type of Cues and Learners' Prior Knowledge on Learning in Video Lectures. *Modern Distance Education Research, 32*(01), 93-101. [杨九民, 徐珂, 韩佳雪, 等. (2020). 教学视频中线索类型与学习者先前知识经验对学习的交互影响. *现代远程教育研究*, *32*(01): 93-101.]
29	Yang, J. M., Yang, W. D., Chen, H., Et Al. (2022). Does Instructor's Gesture Take Effect in Instructional Videos?——Based on the Meta-Analysis of 40 Experimental and Quasi-Experimental Studies from 2000 to 2021. *Modern Distance Education Research*, (01), 92-103. [杨九民, 杨文蝶, 陈辉, 等. (2022). 教学视频中的教师手势起作用了吗？——基于 2000—2021 年 40 篇实验和准实验研究的元分析. *现代远程教育研究*, (01): 92-103.]
30	Yang, J. M., Zhang, Y., Li, L., Et Al. (2019). The Interaction of an Instructor's Guided Behaviors and Learners' Prior Knowledge in Predicting Learning from Video. *China Educational Technology*, (07), 74-81. [杨九民, 章仪, 李丽, 等. (2019). 教师引导行为与学习者先前知识水平对视频学习的交互影响. *中国电化教育*, (07): 74-81.]
31	Yang, J. M. (2014). *The Effect of the Instructor on Learning Processes and Effectiveness in Online Video Courses* [Doctoral Dissertation, Huazhong Normal University]. [杨九民. (2014). *在线视频课程中教师对学习过程与效果的影响*. 华中师范大学博士学位论文.]
32	Yang, X. Z. (2016). *Influence of Teacher's Presentation in Instructional Videos on Learner's Learning* : an Example of Declarative Knowledge and Procedural Knowledge Instructional Video [Master's Thesis, Huazhong Normal University]. [杨显忠. (2016). *教学视频中教师形象呈现方式对大学生学习的影响研究——以陈述性知识和程序性知识教学视频为例*. 华中师范大学硕士学位论文.]
33	You, Y. (2020). Research on *The Relationship Between Visual Comfort of Teaching Video and Cognitive Load of Learners* [Doctoral Dissertation, Northeast Normal University]. [尤洋. (2020). *教学视频视觉舒适度与学习者认知负荷关系研究*. 东北师范大学博士学位论文.]
34	Zhang, B. (2015). Video Shooting in Chinese Language Course, Based on the Module of Stories of the Chinese. *Frontiers of International Chinese Language Education*, (00), 85-94. [张斌. (2015). 语言课程教学视频拍摄手法探讨——以微课《中国人的故事》为例. *对外汉语教学与研究*, (00): 85-94.]
35	Zhang, J. H., Zhang, J. P., Huang, L. Y., Et Al. (2009). An Eye-tracking Experiment Research on Three-part-separated Screen Web Course. *Journal of Distance Education*, *17*(06), 74-78. [张家华, 张剑平, 黄丽英等. (2009). "三分屏"网络课程界面的眼动实验研究. *远程教育杂志*, *17*(06): 74-78.]
36	Zhang, R. Z. (2018). Research on the Influence Factors of Teacher's Video Language Presentation Skills of Massive Open Online Courses. *E-Education Research*, *39*(05), 33-39. [张润芝. (2018). 大规模开放在线课程教师视频语言表达技巧影响因素研究. *电化教育研究, 39*(05): 33-39.]
37	Zhang, S. (2020). A Qualitative Study on The Impact of Foreign Language Teaching Video Design on College Students' Foreign Language Learning. *Overseas English*, (17), 158-159+163. [张硕. (2020). 外语教学视频设计对大学生外语学习影响的质性研究. *海外英语*, (17): 158-159+163.]

序号	文献来源
38	Zheng, J., Wang, Y. N., Wang, K. L., Et Al. (2012). A Study on The Design of Micro-Learning Videos. *China Educational Technology*, (04), 21-24. [郑军, 王以宁, 王凯玲, 等. (2012). 微型学习视频的设计研究. *中国电化教育*, (04): 21-24.]
39	Zheng, J., Zhao, H. H., Yan, Z. Q., Et Al. (2012). The Effect of Teacher's Image on the Multimedia Video Learning. *Psychological Research, 5*(05), 85-90. [郑俊, 赵欢欢, 颜志强等. (2012). 多媒体视频学习中的教师角色. *心理研究, 5*(05): 85-90.]

附录二 补充版国际中文慕课视频设计的评价框架

一级指标	二级指标	具体观测点	指标要求
内容设计	教学内容	教学目标	提供教学目标的文本呈现
		教学材料	主题明确且单一，并提供与主题相匹配的语音、语法（提供语法讲练）、词汇、汉字（提供汉字书写指导和呈现，而不仅仅是拼音）的综合讲练
		教学方式	选择合适的教学方式；体现精讲多练，在需要练习的地方，提供充足的时间确保学习者跟练；提供大量例句，并能体现日常对话
		文化国情	融入中华文化及当代中国国情
		区域国别化设计	生词或对话的文本翻译匹配目标国文字；考虑不同文化特征和文化背景，适时适当加入文化对比，避免文化误解和敏感话题
	文本内容	先行组织者	呈现适当的引导性材料
		字幕呈现方式	有教师讲解或对话的部分需提供准确的、多语种字幕
	听觉内容	教学媒介语	中文为主，可辅以适量精准的外文解释
		背景音乐	适时出现，如有，音量不能盖过对话
交互设计	人机交互	视频时长	视频长度较短
		控制视频进度	允许学习者自己控制（包括播放、暂停、快进、后退、倍速等），根据教学环节可视化时间轴
		嵌入问题或测试	视频中应有嵌入问题或测试，供学习者多练习
		控制字幕	允许学习者选择字幕是否呈现以及如何呈现
		弹幕	允许或控制弹幕呈现，并能发送弹幕
	交互界面	界面类型	课堂实录式、教师与 PPT 结合式、研讨式和动画式
		视频切块	在重要的信息后停顿，学习者可点击继续播放
情感设计	师生形象	学习者呈现	学习者可出镜，发音须接近母语者，而非洋腔洋调
		教师在屏	呈现教师形象或动画人物代理

一级指标	二级指标	具体观测点	指标要求
		教师画面大小	教师画面占屏幕面积比例适中
		教师面部表情	亲切、面带微笑、积极、热情的态度
		教师目光	教师有目光引导
		教师手势	教师使用指示性手势
		教学语言风格	教师采用对话式风格、幽默
	视觉呈现	视频整体颜色	与教学内容相适宜
		视觉线索	使用图片、动画、颜色、闪烁、箭头、线框、加粗、加大字号等提示重要信息
		视频制作质量	视频分辨率高、音质清晰，兼容移动设备；播放流畅、稳定易用
创新设计	创新元素	游戏化设计	有游戏等激励形式
	创新技术	新型信息技术	适时适当使用虚拟现实、增强现实、扩展现实等技术，模拟真实交际场景；融入 AI 智能伴学等

Podcasting and L2 Listening: Impacts on Behavior and Attitude in Chinese Language Learners
(播客与二语听力：对中文学习者行为和态度的影响)

Shi, Zhongqi
(史中琦)
Columbia University
(哥伦比亚大学)
zs2132@columbia.edu

Du, Yiran
(杜伊然)
Columbia University
(哥伦比亚大学)
yd2331@tc.columbia.edu

Yu, Wenqian
(于文芊)
Columbia University
(哥伦比亚大学)
wy2360@tc.columbia.edu

Abstract: This study explores the impact of pedagogical podcasts on second language (L2) listening behaviors and attitudes among Chinese language learners in a U.S. university context. Employing a quasi-experimental pretest-posttest design, 38 undergraduate students from intermediate and advanced Chinese classes engaged with a custom-designed podcast series over a 14-week semester. The intervention included both textbook-aligned recordings and authentic talk episodes, with weekly access provided via mobile platforms. Quantitative results indicated a significant increase in both the frequency and duration of listening practice outside the classroom, alongside a notable diversification of listening sources. Despite a slight but significant rise in perceived listening difficulty—likely attributable to increased metacognitive awareness—confidence levels remained stable, suggesting the development of cognitive resilience. Correlation analyses identified perceived difficulty as the strongest predictor of sustained podcast engagement, while confidence and perceived usefulness were less influential. Students attributed distinct benefits to the two podcast types: textbook recordings were seen as supportive of classroom learning, while authentic talks enriched vocabulary, facilitated practical comprehension, and enhanced learner engagement. Qualitative feedback highlighted a strong preference for supplementary scaffolding (e.g., vocabulary lists and transcripts), particularly for authentic content. These findings emphasize the value of podcasts as accessible, motivating tools for promoting autonomous listening practice and extensive exposure to the target language. The study concludes with pedagogical recommendations and calls for future research incorporating objective usage data, control group comparisons, and proficiency-specific podcast design to further optimize podcast-based L2 instruction.

摘要：本研究探讨了教学型播客对美国大学中文学习者在第二语言（L2）听力行为与态度方面的影响。采用准实验性的前测—后测设计，38 名分别来自中级和高级中文课程的本科生在为期 14 周的学期中参与了本研究，并定期收听为其语言水平量身定制的播客系列。该干预包含课本配套录音与真实语料播客，每周通过移动设备平台发布。

量化结果显示，学生在课外听力练习的频率与时长均显著提升，听力材料来源也呈现出多样化趋势。尽管听力任务的感知难度略有上升——这可能源于元认知意识的增强——但学生的信心水平保持稳定，表明其在面对语言挑战时表现出认知韧性。相关性分析显示，感知难度是持续使用播客的最显著预测因素，而信心与感知实用性则影响较小。学生认为两类播客各有优势：课本音频有助于课堂学习，而真实谈话类播客则增强词汇掌握、实用理解能力及学习兴趣。质性反馈显示，学生普遍期望增加词汇表、文本稿等辅助材料，尤其是在面对真实语料时更为明显。研究结果强调了播客在提升学习动机、促进自主听力练习以及扩大语言输入方面的价值。文章最后提出教学建议，并呼吁未来研究引入客观使用数据、对照组设计及分级播客内容优化策略，以进一步提升播客在第二语言教学中的有效性。

Keywords: podcasts, L2 listening, Chinese language learning, learner behavior, attitudes, technology-enhanced language learning

关键词：播客，第二语言听力，中文学习，学习者行为，态度，技术辅助语言学习

1. Introduction

The importance of listening in foreign language acquisition is well established, yet students consistently receive insufficient practice both in and out of the classroom (Graham, 2006, 2016; Long, 1987; McCaughey, 2015). This discrepancy between the recognized significance of listening skills and the actual practice students engage in presents a persistent challenge in language education. A 2021 survey of over 100 college students enrolled in Chinese courses at Columbia University revealed that more than half had not listened to any Chinese audio materials in the past month, and 25% listened less than once per week. These findings, consistent with previous research (Goh, 2008; Karlin & Karlin, 2023), clearly illustrate a common challenge in foreign language acquisition: the lack of consistent listening input outside the classroom. Addressing this disconnect between the acknowledged importance of listening and students' limited engagement requires more accessible and motivating approaches to listening practice.

In recent years, podcasts have emerged as an innovative and increasingly popular tool in second language learning. Combining accessibility, flexibility, and engagement, podcasts offer a dynamic medium through which students can enhance their listening comprehension and sustain motivation (Abdous et al., 2012; Cai, 2022; Naidionova & Ponomarenko, 2018). The widespread adoption of smartphones and other portable technologies has further facilitated the integration of podcasts into educational contexts, enabling learners to access authentic audio content anytime and anywhere.

This study aims to explore the potential of podcasts to address the common challenge of limited listening practice in language education. Specifically, it examines their impact on students' motivation and perceptions of listening to Chinese-language materials. By evaluating the effectiveness of podcasts in enriching listening practices, the research seeks to contribute to the development of more engaging and effective language learning strategies and resources.

2. Literature Review

2.1 The role of listening input in L2 learning

Listening is fundamental to second language (L2) acquisition, serving as the primary channel through which learners are exposed to the target language (Goh & Vandergrift, 2022; Mohsen & Almudawis, 2021; Nunan, 2002; Vandergrift, 2004). A robust body of research has consistently demonstrated that both the quantity and quality of listening input significantly influence the development of L2 listening comprehension skills (Chang & Millett, 2013; Goh, 2008; Karlin & Karlin, 2023; Vandergrift, 1999).

Despite its recognized importance, listening remains one of the most demanding skills for L2 learners to master (Cai, 2022; Vandergrift, 2004). Several interrelated factors contribute to this challenge. The rapid pace of native speech often exceeds learners' processing abilities, hindering real-time comprehension (McBride, 2011). Pronunciation and intonation variations, whether individual speakers or regional dialects, can further complicate understanding (Cheng, 2018). Additionally, syntactic complexity and the frequent use of idiomatic expressions can overwhelm learners, particularly those with limited proficiency (Shang, 2008). These difficulties are compounded by the need to simultaneously decode sounds, recognize vocabulary, and interpret meaning, all of which place a considerable cognitive load on the listener. Such demands can erode learners' confidence, diminish motivation, and reduce their willingness to engage in further listening practice.

Moreover, a recurring theme in L2 research is that learners often receive inadequate exposure to the target language through listening activities (Khalili Sabet, 2012; Yin, 2015). Several factors contribute to this deficiency, including limited time for practice (Siegel, 2013), difficulty locating content matched to learners' proficiency levels (McCaughey, 2015), and a lack of structured support or pedagogical guidance for autonomous listening (Bozorgian, 2014). This gap in listening input poses a significant barrier to the development of effective comprehension skills.

2.2 Pedagogical applications of podcasts in L2 listening instruction

Podcasting has increasingly been recognized as a valuable pedagogical tool for addressing the challenges inherent in L2 listening instruction. Unlike general-purpose podcasts, pedagogical podcasts are intentionally designed with language acquisition goals in mind. These resources often incorporate controlled vocabulary, modified speaking rates, and content aligned with learners' curricular needs (Rosell-Aguilar, 2013; Thorne & Payne,

2005). This distinction is significant, as research in educational technology emphasizes that the efficacy of digital tools lies not in the technology itself, but in how it is pedagogically integrated into instruction (Levy, 2009; Stockwell, 2012).

Empirical studies have consistently demonstrated the positive impact of well-designed podcasts on L2 listening development. For instance, O'Bryan and Hegelheimer (2007) integrated 14 instructor-created podcasts into an ESL listening strategies course at Iowa State University. The podcasts were specifically designed to complement classroom instruction by summarizing concepts, bridging between lessons, and introducing new material. The findings showed highly positive responses, with all six students listening to each podcast multiple times. The researchers demonstrated successful integration where podcasts became a normalized component of the curriculum. Cross (2014) conducted a 9-week case study with a Japanese EFL learner, teaching her to autonomously use BBC podcasts for listening development through metatextual skills and metacognitive instruction. The learner developed a structured 10-step listening approach and showed improvement in self-evaluation ratings from 50% to 93% comprehension. Through weekly guidance and journal keeping, she progressed from basic note-taking to sophisticated analysis including discourse construction and self-directed strategy development. The study demonstrated that learners can develop autonomous podcast exploitation skills with appropriate initial guidance.

In the context of Chinese language education, Yang (2014) conducted a comprehensive study examining the role of podcasts in enhancing listening comprehension. The 60-day intervention followed a structured three-phase design: selecting suitable audio materials, identifying key linguistic features (e.g., vocabulary and grammar), and implementing listening tasks alongside pre- and post-assessments. Results showed that consistent exposure to curated podcast episodes significantly improved learners' listening skills. Yang further proposed an instructional model for classroom integration, emphasizing content selection, strategy instruction, and promotion of autonomous listening outside class settings.

In summary, these findings underscore the pedagogical potential of podcasts when they are purposefully incorporated into language instruction. Their flexible, learner-centered design can support both structured classroom learning and self-directed practice, bridging a key gap in L2 listening input.

2.3 Impact of podcasts on listening: Behavior and attitude

Beyond measurable gains in listening comprehension, podcasts have been shown to influence learners' affective and behavioral responses to L2 listening. Hasan and Hoon (2013), in a systematic review of podcast applications in language education, found that several studies reported positive shifts in learners' attitudes. These included reduced anxiety, greater willingness to engage with authentic audio content, and increased confidence in handling listening tasks. Such attitudinal changes are particularly important, as positive emotions and beliefs have been consistently linked to higher levels of engagement and persistence in language learning (Dörnyei & Ushioda, 2011).

A central psychological construct related to learner attitude is self-efficacy, defined as an individual's belief in their capacity to perform a specific task (Bandura, 1977). In the domain of L2 listening, Rahimi and Abedi (2014) examined the impact of regular podcast use among 60 EFL learners and found substantial improvements in listening self-efficacy scores. They attributed this enhancement to the autonomy that podcasts afford, allowing learners to regulate listening difficulty and revisit content as needed, thereby aligning input with personal proficiency and fostering a sense of control over learning.

However, the relationship between podcast use and self-efficacy is not uniformly linear. Pei et al. (2023) found that while podcast-based instruction led to improvements in listening performance, these gains did not consistently result in higher self-efficacy. This divergence underscores the complex, multifactorial nature of self-efficacy development and suggests that exposure alone may be insufficient. Explicit strategy instruction, timely feedback, and scaffolded practice may be essential to translate performance gains into stronger self-beliefs.

From a behavioral perspective, podcasts offer unparalleled flexibility, enabling learners to embed language exposure into daily routines via mobile devices. Yet, empirical research has not thoroughly investigated the behavioral patterns associated with sustained podcast use. Key questions remain regarding how the availability of podcasts influences the frequency and duration of listening sessions, the consistency of learner engagement over time, and the contextual factors that support long-term use of these resources.

Moreover, the attitudinal impact of podcast listening on learners' perceptions of L2 listening materials remains underexplored. It is unclear whether long-term podcast engagement alters students' perceptions of task difficulty, enhances their willingness to attempt more challenging content, or increases intrinsic motivation. Gaining insight into these psychological and behavioral dimensions is essential for optimizing podcast integration in language instruction.

To address these gaps, this study explores both the behavioral engagement and attitudinal shifts associated with pedagogical podcast use across varying proficiency levels. It seeks to answer the following research questions:

1. How do pedagogical podcasts impact students' time investment and frequency of listening practice outside the classroom?
2. How does the use of pedagogical podcasts influence students' listening confidence and their perception of listening difficulty over time?
3. What factors influence students' motivation to continue listening to podcasts?
4. How do students evaluate different types of podcast content, and what improvements do they suggest?

3. Methods

This study employed a quasi-experimental pretest-posttest design to investigate the impact of pedagogical podcasts on Chinese language learners. The mixed-methods

approach combined quantitative measures to assess changes in students' out-of-classroom listening behaviors (RQ1) and attitudes toward Chinese listening materials (RQ2), with qualitative data exploring students' willingness to continue using podcasts after the intervention (RQ3). Additionally, we gathered their suggestions for improvement (RQ4).

3.1 Participants

The study involved 38 undergraduate students enrolled in two Chinese language classes at a private research university on the East Coast. Participants ranged from 19 to 28 years old, with a mean age of 22. The gender distribution was 42% male and 58% female. The group comprised 47% White, 34% Asian, and 19% Hispanic students. Prior to the study, all participants completed an in-house Chinese placement test developed and administered by the university's Chinese program to assess language proficiency. Corresponding to the ACTFL proficiency scale, 21 students in the Advanced Chinese class were classified as Advanced Low, while 17 students in the Intermediate Chinese class were categorized as Intermediate High.

3.2 Procedure

During the first week of classes, the instructor led a discussion session on strategies for improving listening skills, during which students were introduced to a Chinese podcast series specifically developed for this study, and students were required to subscribe to the channel. Detailed information about the podcast content, structure, and accessibility is provided in the following section. Students were informed that participation in the study was entirely voluntary and unrelated to their course grades, and all students consented to participate. A pretest survey was conducted immediately following the discussion session, before engaging with the podcast series. This initial assessment documented their prior experience with Chinese audio materials and baseline perceptions regarding Chinese listening content.

Throughout the semester, participants received automatic notifications from the podcast platform when new podcast episodes became available, with no additional reminders from either the researchers or the instructors. The course instructors did not incorporate the podcasts into classroom activities or discussions, maintaining separation between the research intervention and regular coursework. At the end of the semester, participants completed a posttest survey that evaluated their current listening habits and gathered feedback on specific research areas of interest, including their experiences with different podcast content types and their willingness to continue using podcasts for language learning.

3.3 Podcast design

Students subscribed to podcast series created specifically for their Chinese course, receiving one or two episodes weekly. Two podcast series were developed to match the different proficiency levels of the classes: one for Intermediate Chinese and one for Advanced Chinese.

Each podcast series comprised two subcategories:

(a) **Textbook Audio**: Recordings of the main texts from the course textbook, providing students with supplementary listening practice of familiar content.

(b) **Real Talks**: Authentic narratives related to textbook themes but featuring natural speech patterns and authentic language use. These episodes were adapted from real-world sources such as online articles, business publications, and news reports, with content edited for appropriate length and complexity while maintaining authentic discourse characteristics. New word frequency was carefully maintained to ensure appropriate difficulty while remaining comprehensible (see Table 1 for details).

Table 1 Key differences between the two podcast series

Feature	Intermediate Level Podcast	Advanced Level Podcast
Textbook Audio Length	1-3 minutes	2-5 minutes
Real Talks Length	2-6 minutes	5-10 minutes
Speaking Speed	170-180 words per minute (wpm)	210-220 words per minute (wpm)
New Vocabulary Ratio	30-33 words per 1000	30-35 words per 1000

All podcast content was developed by a member of the research team who also served as the course instructor to ensure alignment between curriculum goals and appropriate difficulty calibration based on his knowledge of student abilities and needs. Topics were carefully selected to correspond with the chapter themes covered in each class (see Appendix 1 for sample topics). Over the 14-week semester, 21 episodes for each series were created and released. The podcasts were hosted on Spotify and Apple Podcasts, providing students with free and convenient access. New episodes were primarily released on Thursday nights, with automatic notifications sent directly to students' mobile devices.

3.4. Data collection

Data was gathered through pretest and posttest surveys administered via the Qualtrics platform. The surveys included a combination of Likert scale and open-ended questions (see Appendix 2 for key survey questions). To facilitate meaningful comparisons, certain questions were designed to be parallel across both surveys. The questions focused primarily on participants' actual Chinese audio listening practice and how they perceived the activity of listening to Chinese.

3.5 Measures

The study employed several measures to assess participants' engagement with Chinese audio materials, their attitude toward Chinese listening abilities. Moreover, participants were asked to provide specific feedback regarding the two podcast categories

(i.e., textbook audio recordings and authentic talk recordings), including their perceived usefulness and suggested improvement for these two kinds of Chinese podcasts. Lastly, participants self-reported their anticipated frequency of engagement with Chinese audio listening in the future as an indicator of continued motivation to engage with Chinese audio content.

Engagement with Chinese audio materials

Listening frequency: Participants reported their frequency of listening to Chinese audio materials on a 5-point scale. The scale ranged from 1 to 5 (1 = *never*, 2 = *once or twice a month*, 3 = *once or twice a week*, 4 = *three or four times a week*, and 5 = *five times a week or more*), reflecting different levels of engagement with the Chinese audio materials.

Listening duration: Participants reported their daily listening duration by choosing from a set of four options: *Fewer than 30 minutes, 30 minutes to 1 hour,1 to 2 hours, and More than 2 hours*.

Listening sources: Participants were asked to specify the audio sources they used for Chinese listening materials apart from the podcasts. Options included *music*, *movies*, and *social media*, with the count of choices recorded.

Attitude towards Chinese listening: Perceived difficulty in Chinese audio materials: Participants assessed their perceived difficulty in listening to Chinese audio using a single item on a 5-point scale, where 1 = *very easy* and 5 = *very difficult*.

Confidence: Participants rated their confidence in Chinese listening ability on a 5-point Likert scale, where 1 = *not at all confident* and 5 = *very confident*.

Attitudes towards two specific categories of Podcasts

Perceived Usefulness: Participants rated the usefulness of two types of podcasts—textbook audio recordings and authentic talk recordings—on a 6-point scale, where 1 = *not useful at all* and 6 = *extremely useful*. Moreover, participants reported in what aspects the podcasts were useful in enhancing their listening ability.

Suggestions for improvement: Participants also offered specific suggestions for how to improve these two types of podcasts to enhance their effectiveness in assisting with Chinese listening.

Motivation to continue listening to Chinese Audio materials

Participants indicated their anticipated frequency of interaction with Chinese audio materials using the following options: 1 = *never*, 2 = *once or twice a month*, 3 = *once or twice a week*, 4 = *three or four times a week*, and 5 = *five times a week or more*.

4. Results

4.1 Listening practice frequency and duration

To answer RQ1, we investigated three dimensions of students' engagement with Chinese audio materials: frequency of listening, duration of listening sessions, and diversification of listening sources. To examine the effect of our podcast intervention on students' listening behaviors, we compared participants' responses at the beginning of the semester (pretest, before we launched the podcast intervention) and at the end of the semester (post the podcast intervention) using dependent t-tests.

Listening frequency

Analysis of listening frequency revealed a notable increase over the course of the semester. At the beginning of the study, participants reported an average listening frequency of 2.63 (SD = 1.15) on the 5-point scale, indicating that most students were listening to Chinese audio materials between "once or twice a month" and "once or twice a week". By the semester's conclusion, this average had increased to 3.45 (SD = 1.01), suggesting that students were now listening closer to "three or four times a week". To determine the significance of this change, a dependent t-test was conducted. The results showed a statistically significant difference between pretest and posttest reported frequencies, $t(37) = 3.89$, $p < .001$. The magnitude of this difference was moderate, as indicated by Cohen's $d = 0.49$. This effect size suggests that the podcast intervention had a practically meaningful impact on students' listening habits, with participants engaging more frequently with Chinese audio materials by the end of the semester.

Listening duration

At the beginning of the study, before subscribing to the podcast channel, a substantial majority of participants (65.5%) reported spending fewer than 30 minutes daily on Chinese audio. This indicates that most students had limited engagement with Chinese listening materials prior to the intervention. The remaining participants were distributed as follows: 20.7% spent between 30 minutes to 1 hour, 3.4% listened for 1 to 2 hours, and 10.3% engaged for more than 2 hours daily.

By the end of the semester, after exposure to the podcast series, there was a notable redistribution of listening durations. The percentage of participants spending fewer than 30 minutes daily decreased to 55.2%, suggesting that some students had increased their listening time. Correspondingly, there was a substantial increase in the proportion of students listening for 30 minutes to 1 hour, rising from 20.7% to 34.5%. This shift indicates that a significant number of participants had moved from the shortest duration category to a moderate listening duration.

Interestingly, the percentage of participants listening for 1 to 2 hours remained constant at 3.4%, suggesting that the intervention did not impact this particular duration category. However, there was a slight decrease in the proportion of students listening for more than 2 hours daily, dropping from 10.3% to 6.9%.

Listening sources

At the beginning of the semester, movies were the most popular additional listening resource, with 14 out of 33 students (42%) reporting their use (Figure 1). By the end of the semester, this number had increased to 22 students (72%), suggesting that students sought out more diverse and authentic Chinese language content along with increased exposure to the podcasts.

The popularity of Chinese songs as a listening resource increased notably over the course of the study. At the outset, 11 students (22%) reported using Chinese songs for language practice; by the posttest, this number had risen to 19 students (50%), reflecting a growing interest in musical content as a medium for developing listening skills.

Social media emerged as one of the most rapidly growing sources of Chinese listening input among participants. At the beginning of the semester, only six students reported using social media platforms for Chinese listening practice. By the end of the study, this figure had more than doubled to 16 students, suggesting an increased comfort with spontaneous, conversational Chinese as encountered in authentic digital contexts.

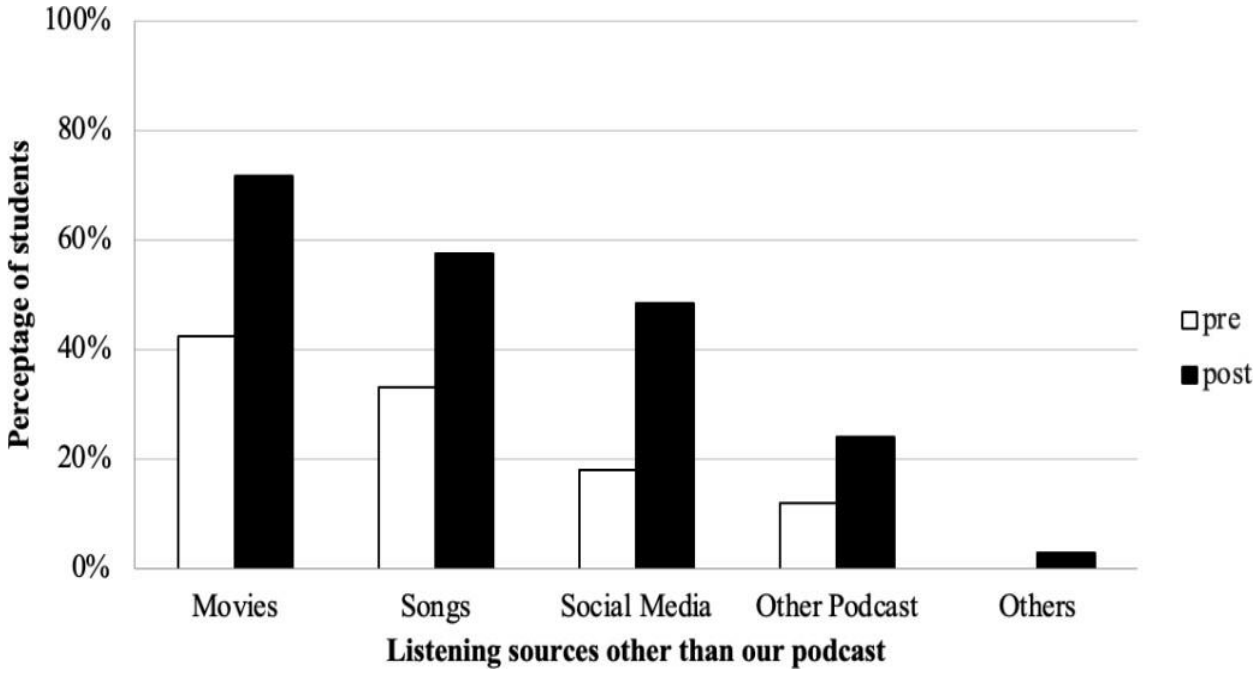


Figure 1 Listening resources in the pre- and posttest

Similarly, the use of external Chinese podcasts—those not included in the study's intervention—also demonstrated notable growth. While only four students reported engaging with other Chinese podcast content at the outset, this number increased to eight by the posttest. This upward trend may indicate that structured exposure to the study's pedagogical podcasts stimulated interest and confidence in independently exploring additional podcast-based resources.

Interestingly, while no students reported using other Chinese audio resources initially, one student reported listening to additional resources, such as TV shows, by the posttest. This indicates a slight diversification in listening sources beyond the categories explicitly measured.

To quantify the overall change in the diversity of listening sources, a dependent *t*-test was conducted. The results showed a significant increase in the average number of Chinese audio resources used by participants. In the pretest, participants reported using an average of 1.06 ($SD = 0.75$) other Chinese audio resources besides the study podcasts. By the posttest, this average had increased significantly to 2.06 ($SD = 1.02$) additional resources. The dependent t-test results ($t(32) = 4.81$, $p < .001$) indicate that this increase was statistically significant. Furthermore, the large effect size (Cohen's $d = 0.84$) suggests that this change was not only statistically significant but also practically meaningful.

4.2 Changes in listening attitudes and perceptions

To answer this research question, we examined whether there is a change between the pretest and posttest in students' attitudes toward Chinese language listening along two dimensions: their perception of listening difficulty and their confidence levels when engaging with Chinese audio materials. We also analyzed the relationships between these attitudinal factors and students' listening behaviors over the course of intervention.

Perceived difficulty in listening to Chinese audio materials

The study examined changes in participants' perceived difficulty of listening to Chinese audio materials over the course of the semester (Table 2). On a 5-point scale, where higher scores indicate greater perceived difficulty, the average score in the pretest was 3.07 ($SD = 0.84$). This initial score suggests that at the beginning of the semester, students perceived listening to Chinese audio as moderately difficult. By the end of the semester, after exposure to the podcast series, the average score for perceived difficulty increased to 3.38 ($SD = 1.02$). This increase indicates that, on average, students found listening to Chinese audio slightly more challenging by the end of the study period.

Table 2 Pre- and post-intervention means and t-test results (N=38)

	Pretest		Posttest		Dependent *t*-test	
Variable	*M*	*SD*	*M*	*SD*	*t*	*Cohen's d*
Listening frequency	2.63	1.15	3.45	1.01	3.89***	0.63
Perceived difficulty	3.07	0.84	3.38	1.02	2.07*	0.39
Confidence level	2.72	0.84	2.83	0.85	0.90	0.20

Note. *$p < .05$. **$p < .01$. ***$p < .001$

To determine whether this change was statistically significant, a dependent *t*-test was conducted. The results showed a significant difference between pretest and posttest scores, $t(28) = 2.07$, $p = .048$. The effect size, as measured by Cohen's *d*, was 0.39, indicating a small to medium effect. This finding suggests that the increase in perceived difficulty, while statistically significant, represents a moderate practical change.

Confidence in listening to Chinese audio materials

Confidence was rated on a 5-point scale, with higher scores indicating greater confidence. In the pretest, the average confidence score was 2.72 (*SD* = 0.84), suggesting a moderate level of confidence at the beginning of the semester. By the posttest, this average had increased slightly to 2.83 (*SD* = 0.85).

To assess whether this change in confidence was statistically significant, a dependent *t*-test (Table 2) was conducted. The results showed no significant difference between pretest and posttest scores, $t(28) = 0.902$, $p = .38$. The effect size (Cohen's $d = 0.17$) was small, indicating that any change in confidence levels was minimal from a practical standpoint.

Correlations among listening practice and perception factors

To better understand the relationships between various aspects of Chinese listening practice and perceptions, correlation analyses were conducted. These analyses revealed interesting patterns both at the beginning and end of the semester (Table 3).

Table 3 Correlations among Chinese listening behavior, Perceived difficulty, and confidence levels

	Listening Frequency	Listening Duration	Perceived Difficulty	Confidence in Listening
Listening Frequency	--	.62*** ($p < .001$)	-.42** ($p = .004$)	.38** ($p = .01$)
Listening Duration	.63*** ($p < .001$)	--	-.44** ($p = .006$)	.58*** ($p < .001$)
Perceived Difficulty	-.80*** ($p < .001$)	-.58*** ($p < .001$)	--	-.68 *** ($p < .001$)
Confidence in Listening	.49** ($p = .004$)	.25 ($p = .16$)	-.58** ($p < .001$)	--

Note: Pretest correlations are shown above the diagonal; posttest correlations are presented below the diagonal.
*$p < .05$. **$p < .01$. ***$p < .001$

In the pretest, both dimensions of students' listening behavior, frequency and duration, showed significant correlations with their attitudes: negative correlations with perceived difficulty (r = -.42 and -.44, respectively) and positive correlations with confidence levels (r = .38 and .58, respectively). This indicates that at the beginning of the semester, students who listened more frequently and for longer periods found Chinese audio less challenging and felt more confident in their listening abilities.

In the pretest, listening frequency showed moderate and significant correlations with several variables:

- A positive correlation with listening duration *(r* = .62), suggesting that students who listened more frequently also tended to listen for longer periods.
- A positive correlation with confidence in listening (*r* = .38), indicating that more frequent listeners tended to feel more confident in their listening abilities.
- A negative correlation with perceived difficulty (*r* = -.42), suggesting that students who listened more frequently tended to perceive Chinese audio as less difficult.

Interestingly, these patterns were not only maintained but strengthened in the posttest correlations:

- The correlation between listening frequency and duration remained stable (*r* = .63).
- The correlation between listening frequency and confidence increased (*r* = .49), suggesting a stronger relationship between how often students listened and how confident they felt.
- The negative correlation between listening frequency and perceived difficulty became much stronger (*r* = -.80), indicating that by the end of the semester, there was a very strong relationship between how often students listened and how difficult they perceived Chinese audio to be.

In the posttest, the correlations between listening behaviors and attitudes strengthened in several aspects. Listening frequency exhibited a substantially stronger negative correlation with perceived difficulty ($r = -.80$, $p < .001$) and maintained its positive relationship with confidence ($r = .49$, $p = .004$). For listening duration, the negative correlation with perceived difficulty increased ($r = -.58$, $p < .001$), while its correlation with confidence became non-significant ($r = .25$, $p = .16$). These patterns suggest that by the semester's end, students who listened more frequently and longer were much more likely to perceive Chinese listening as less difficult.However, the non-significant relationship between listening duration and confidence indicates that students' willingness to engage in extended listening sessions was independent of their perceived confidence in Chinese listening comprehension.

4.3 Factors influencing continued podcast engagement

To evaluate participants' willingness to continue engaging with Chinese listening materials, we analyzed students' answers to the predicted frequency of their future engagement with listening to Chinese audio materials, including podcasts. Results showed that a substantial portion (39.3%) predicted daily listening to Chinese audios, including podcasts, while six participants (21.4%) predicted to be engaged at least three times weekly. Nine participants (32.1%) predicted listening once or twice weekly, and a small minority (7.1%) did so once or twice monthly. These findings highlight participants' strong motivation and willingness to keep these listening habits and a tendency toward regular engagement among most participants, with daily or near-daily listening emerging as the most common practice.

When examining the factors that influence students' motivation to keep listening to podcasts, several key relationships emerge from the correlation analysis (Table 4). Results show a significant negative correlation between perceived difficulty and motivation (r = -.46, p = .015). In other words, students who find listening activities more challenging tend to be less motivated to continue using podcasts for language learning.

Table 4 Correlations among perceived difficulty, perceived usefulness, confidence level and continued motivation to listen to Chinese materials

	Perceived Difficulty	Perceived Usefulness	Confidence Level	Continued Motivation
Perceived Difficulty	--	.15 (p = .407)	-.58** (p < .001)	-.46** (p = .015)
Perceived Usefulness		--	-.10 (p =.607)	.15 (p = .433)
Confidence Level			--	.36 (p = .061)
Continued Motivation				--

Note. *p < .05. **p < .01. ***p < .001

The relationship between confidence and motivation is approaching statistical significance (r = .36, p =.061), suggesting a possible positive connection. While this correlation does not quite meet the conventional threshold for significance, the trend indicates that students who feel more confident in their listening skills may be more likely to continue using podcasts.

Surprisingly, the perceived usefulness of podcasts shows only a weak, non-significant correlation with future motivation (r = .15, p = .433). One might assume that students who see podcasts as useful would be more inclined to keep listening, yet the data does not support this assumption. This suggests that perceived usefulness alone may not be enough to drive continued engagement, especially when difficulty perception appears to have a much stronger influence.

4.4 Student evaluation of podcast content types and suggestions

The study compared two types of podcasts: textbook audio recordings and authentic talk recordings. For overall usefulness, students rated both types favorably. The authentic talks received a slightly higher mean rating (M = 4.61, SD = 0.91) compared to the textbook audio (M = 4.17, SD = 1.64). However, the dependent t-test results indicate that this difference was not statistically significant, $t(37) = 1.26$, p = 0.22, Cohen's d = 0.204. This suggests that students found both podcast types useful, albeit in different ways.

The qualitative analysis of student responses reveals distinct perceived benefits for each podcast type (see Table 5). Textbook audio recordings were primarily valued for

enhancing listening comprehension (10 mentions) and aiding class preparation (9 mentions), with additional benefits for pronunciation and tone (4 mentions). In contrast, authentic talks were appreciated for their contribution to vocabulary and expressions (6 mentions), practical and supplementary nature (6 mentions), listening comprehension (6 mentions), and engagement and interest (3 mentions).

This pattern suggests that students perceive textbook audio recordings as more directly supportive of coursework and foundational language skills, while authentic talks provide enrichment, real-world application, and greater engagement with the language. The complementary nature of these benefits indicates that a combination of both podcast types might best serve students' language learning needs.

Table 5 Frequency of reported Chinese podcast learning benefits by types.

Category	Authentic Talks	Textbook Audio
Listening Comprehension	6	10
Pronunciation & Tone	0	4
Fluency & Natural Speech	1	2
Vocabulary & Expressions	6	4
Engagement & Interest	3	0
Practical & Supplementary	6	0
Class Preparation	2	9

The data on students' suggestions for improvement provides valuable insights into how pedagogical podcasts could be enhanced to better serve language learners (see Table 6). A detailed quantitative analysis of these recommendations reveals several key themes, with a notable observation that students offered suggestions exclusively for the Authentic Talks series and not for the Textbook Audio series.

Table 6 Student suggestions for improving the authentic talk series podcasts

Category	Frequency
Supplementary Support	
Provide vocabulary lists	7
Provide transcripts	3
Content and Delivery	
Adjust difficulty level	3
Adjust speeding speed	2
Include more Episodes	1
Incorporate different voices	1
Instructional Integration	
Integrate podcast listening as course assignment	2

As shown in Table 6, the most common recommendation from students was the inclusion of supplementary materials, with vocabulary lists being the most frequently requested (7 mentions). Many students also expressed a desire for transcripts (3 mentions), indicating a strong preference for textual support alongside audio content. These requests

suggest that learners may struggle with unfamiliar vocabulary and complex structures in authentic talks and see supplementary materials as a way to reinforce their understanding and maximize learning outside of listening sessions.

In terms of adjustments to the podcasts themselves, students highlighted concerns about speaking speed (2 mentions) and difficulty level (3 mentions). These suggestions indicate that some learners found the content challenging, either due to its pace or complexity. Notably, the call for difficulty adjustments aligns with the correlation data, which shows that perceived difficulty significantly affects students' motivation to continue using podcasts for language learning.

Although less frequently mentioned, students also suggested incorporating different voices (1 mention) and increasing the number of episodes (1 mention). These recommendations reflect an interest in exposure to a wider range of speakers and expanded learning opportunities, underscoring students' awareness of the benefits of diverse linguistic input in language acquisition.

One particularly interesting recommendation was integrating podcasts into formal assignments (2 mentions). This suggests that some students believe structured use of podcasts in coursework could enhance their engagement and commitment to listening practice. From a pedagogical standpoint, this highlights the potential value of embedding podcasts into the curriculum to encourage consistent and meaningful use.

5. Discussion

5.1 Impact of pedagogical podcasts on Chinese listening practice

A key finding of this study is the marked increase in listening frequency among students who subscribed to pedagogical podcasts. The shift from an average frequency of 2.63 (between "once or twice a month" and "once or twice a week") to 3.45 (approaching "three or four times a week") highlights the potential of podcasts to foster more regular and sustained engagement with listening materials. This trend reflects the motivational and practical affordances of podcasts—namely, their accessibility, portability, and integration into daily routines. These findings echo McBride's (2009) study on Spanish language learners, which demonstrated that podcast availability significantly boosted the amount of time learners spent on listening practice.

Equally important is the observed redistribution in listening duration. Fewer students reported spending less than 30 minutes on daily listening, while a greater proportion engaged for 30 minutes to an hour. This increase in both frequency and duration suggest a meaningful expansion of learners' total listening exposure—an essential component for the development of listening proficiency, as supported by previous research (Chang & Millett, 2013; Goh, 2008).

In addition to quantitative improvements, a qualitative shift was also evident in the variety of listening sources students engaged with. The increased use of movies, music,

social media, and other non-course podcasts suggests that pedagogical podcasts serve not only as a primary resource but also as a catalyst for broader exploration of Chinese audio content. This behavioral change aligns with Siegel's (2011) advocacy for extensive listening, which emphasizes the pedagogical value of diverse and authentic listening materials in promoting listening fluency and adaptability.

Crucially, the data also revealed that students engaged with podcasts voluntarily, without institutional or curricular compulsion. This pattern points to the emergence of intrinsic motivation, an essential element in autonomous language learning. According to Ryan and Deci's (2000) self-determination theory, such self-motivated behaviors are often more enduring and impactful. The voluntary and repeated use of pedagogical podcasts thus indicates not only learner interest but also the capacity of well-designed content to stimulate meaningful, self-directed engagement with the target language.

In sum, these findings underscore the multifaceted benefits of pedagogical podcasts in enhancing both the quantity and quality of L2 Chinese listening practice. Their role in fostering autonomy, expanding exposure, and encouraging engagement with diverse materials positions them as a valuable asset in contemporary language learning environments.

5.2 Impact on perceptions of Chinese listening materials

An intriguing and somewhat paradoxical finding of this study is the slight yet statistically significant increase in learners' perceived difficulty of Chinese listening tasks, despite regular exposure to pedagogical podcasts. At face value, one might expect that increased exposure would correspond with reduced perceived difficulty. However, this counterintuitive result aligns with well-documented patterns in metacognitive development within second language acquisition.

This phenomenon can be understood through the lens of flawed self-assessment, as described by Dunning, Heath, and Kruger (2004). In the early stages of skill acquisition, learners often overestimate their competence due to a limited grasp of the domain's complexity. As learners gain more experience, particularly through frequent engagement with authentic or semi-authentic materials, they are exposed to subtler linguistic features and structural complexities previously unnoticed. This heightened awareness fosters a more accurate, albeit sometimes harsher, self-appraisal of their listening ability. Thus, the increase in perceived difficulty may not reflect declining skill but rather a maturing understanding of the demands of L2 listening.

Importantly, while perceived difficulty rose, learners' self-reported confidence remained stable or showed slight improvement. This stability suggests a meaningful form of cognitive resilience. Rather than discouraging learners, the increased challenge appears to have been internalized as a natural part of the learning process. This finding is consistent with motivational theories emphasizing the importance of competence and self-efficacy in sustaining learner engagement (Dörnyei & Ushioda, 2011; Ryan & Deci, 2000).

The coexistence of increased perceived difficulty and stable confidence thus offers a promising insight: learners may become more metacognitively aware without losing motivation. Such a development suggests that pedagogical podcasts not only enhance linguistic exposure but may also foster a more sophisticated and realistic learner mindset—an outcome beneficial for long-term language development.

5.3 Continued use and influencing factors

The study's findings regarding learners' intentions to continue using podcasts for Chinese listening practice post-intervention are particularly promising. A notable 39.3% of participants reported plans to engage with podcast content daily, while an additional 21.4% indicated intentions to listen at least three times per week. These figures suggest that the podcast intervention not only enhanced short-term listening engagement but also cultivated sustainable listening habits for a substantial proportion of learners.

To understand the factors influencing continued podcast use, a correlation analysis was conducted. The results revealed that perceived difficulty emerged as the strongest and most significant predictor of continued motivation. Learners who perceived the listening tasks as more challenging were less likely to express their intentions to continue using podcasts. While the relationship between confidence and continued motivation approached statistical significance, perceived usefulness exhibited only a weak and non-significant correlation with learners' willingness to persist.

These findings underscore the pivotal role of difficulty perception in sustaining learner engagement with podcast materials. Among the variables examined, it is not the perceived utility of the content, but rather the learner's subjective experience of challenge, that most strongly determines ongoing motivation. This insight carries important pedagogical implications: careful calibration of difficulty levels may be more impactful for promoting long-term engagement than simply ensuring content relevance or usefulness.

This interpretation is well-aligned with Krashen's (1985) Input Hypothesis, which posits that optimal language acquisition occurs when learners are exposed to input that is just beyond their current level of competence—commonly referred to as "i+1." If the input is too difficult (i+2 or beyond), learners may become frustrated or demotivated; if too easy, progress may stagnate. Thus, podcast content that consistently challenges learners at an appropriate level may not only enhance listening skills but also sustain motivation over time.

In summary, the study highlights the importance of balancing linguistic challenge with learner capacity in podcast-based instruction. Achieving this balance is essential for maximizing the motivational and pedagogical potential of podcasts in second language education.

5.4. Evaluation of podcast content types and suggestions for improvement

Students' evaluations of the two podcast types—textbook audio recordings and authentic talks—revealed a complementary relationship in terms of perceived usefulness.

While both formats were rated similarly for overall utility, qualitative feedback highlighted distinct pedagogical functions for each.

Textbook audio recordings were primarily valued for their role in reinforcing classroom learning. Students reported that these materials enhanced listening comprehension, supported class preparation, and improved pronunciation accuracy. These benefits underscore the importance of structured, level-appropriate input in consolidating linguistic foundations and aligning with formal instruction. As such, textbook-based content appears particularly effective in building core listening skills through repetition and review.

In contrast, authentic talks were praised for fostering vocabulary development, promoting practical language use, and increasing learner engagement. Students noted that this format exposed them to natural speech patterns, colloquial expressions, and culturally relevant topics—elements that enriched their learning experience and provided valuable real-world application. The engagement driven by authentic content also suggests an affective benefit, potentially increasing motivation and interest in the language.

Together, these findings support a complementary instructional model that integrates both controlled and authentic audio input. Textbook podcasts offer structure and alignment with curricular goals, while authentic talks provide challenge, enrichment, and exposure to diverse linguistic forms. The simultaneous use of both types appears to address a broader range of learner needs and learning styles.

Students also offered several suggestions for improving the podcast experience. Most notably, they requested additional scaffolding, such as vocabulary lists, transcripts, and glossaries. These supports are particularly important for authentic materials, which often include unpredictable vocabulary and fast-paced delivery. As Vandergrift and Tafaghodtari (2010) emphasize, strategic scaffolding significantly enhances comprehension and learning outcomes when engaging with challenging audio texts.

Furthermore, several participants recommended modifying the speed and difficulty level of the podcasts, reinforcing earlier findings that perceived difficulty strongly influences continued motivation. These suggestions align with established principles in input design, highlighting the importance of calibrating content to match learners' evolving proficiency levels in order to maintain optimal challenge and sustain engagement.

In summary, student evaluations affirm the pedagogical value of both podcast types, while also identifying key areas for enhancement. Implementing targeted scaffolding and differentiated content can maximize the instructional impact of podcasts and better support diverse learner profiles.

6. Conclusion and future directions

This study contributes to the expanding literature on podcast integration in L2 listening instruction by examining its impact on Chinese language learners. The findings

demonstrate that pedagogical podcasts can substantially increase both the frequency and duration of out-of-class listening practice. Notably, many students also broadened their exposure by engaging with additional Chinese-language media, including films, songs, and social platforms, suggesting that podcast use can serve as a catalyst for extensive listening beyond the classroom.

Although exposure to podcast content was associated with a slight increase in perceived difficulty, learners' confidence levels remained stable, indicating that they were not discouraged by the added challenge. This combination of growing self-awareness and resilience highlights the developmental benefits of structured yet authentic listening input. However, the significant negative correlation between perceived difficulty and continued motivation reinforces the importance of carefully calibrating podcast content to learners' proficiency levels to maintain engagement over time.

Students identified distinct yet complementary benefits from the two podcast formats examined. Textbook audio recordings were seen as instrumental for supporting class preparation and reinforcing foundational skills, while authentic talks enhanced vocabulary acquisition, real-world application, and learner engagement. These findings suggest that a balanced approach—integrating both structured and authentic content—best supports diverse learner needs and promotes comprehensive listening development.

Several limitations merit consideration. First, the study relied on self-reported data, which may be subject to bias or inaccuracies. Future research should incorporate objective behavioral measures (e.g., app analytics or log data) to better capture listening patterns and engagement. Second, the sample focused on intermediate and advanced learners; further studies should explore how learners at varying proficiency levels respond to different podcast features, enabling more differentiated instructional design. Third, the use of objective listening comprehension assessments and inclusion of control groups would help strengthen causal interpretations and offer more robust evidence of podcast effectiveness.

In sum, this study underscores the pedagogical value of podcasts as an accessible and engaging tool for enhancing L2 listening practice. The findings offer practical implications for educators and curriculum designers and provide a foundation for future research aimed at optimizing podcast-based interventions across varied language learning contexts.

References

Abdous, M., Facer, B. R., & Yen, C.-J. (2012). Academic effectiveness of podcasting: A comparative study of integrated versus supplemental use of podcasting in second language classes. *Computers & Education*, *58*(1), 43–52. https://doi.org/10.1016/j.compedu.2011.08.021

An, I. S. (2013). Integrating technology-enhanced student self-regulated tasks into university Chinese language course. *International Journal of Computer-Assisted*

Language Learning and Teaching, *3*(1), 1–15. https://doi.org/10.4018/ijcallt.2013010101

Azmee, A. (2022). The impact of the podcast (BBC sounds) to raise metacognitive awareness in developing L2 listening comprehension among Bangladeshi undergraduate learners. *Computer Assisted Language Learning Electronic Journal*, *23*(4), 103–119.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037//0033-295x.84.2.191

Bozorgian, H. (2014). The role of metacognition in the development of EFL learners' listening skill. *International Journal of Listening*, *28*(3), 149–161. https://doi.org/10.1080/10904018.2013.861303

Brandl, K. (2021). *Communicative language teaching in action: Putting principles to work* (2nd ed.). Pearson Prentice Hall.

Cai, W. (2022). *Teaching and researching Chinese second language listening*. Taylor & Francis.

Chang, A., & Millett, S. (2013). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal*, *68*(1), 31–40. https://doi.org/10.1093/elt/cct052

Cheng, Y.-L. (2018). Unfamiliar accented English negatively affects EFL listening comprehension: It helps to be a more able accent mimic. *Journal of Psycholinguistic Research*, *47*(4), 899–911. https://doi.org/10.1007/s10936-018-9562-y

Clark, A. J. (1989). Communication confidence and listening competence: An investigation of the relationships of willingness to communicate, communication apprehension, and receiver apprehension to comprehension of content and emotional meaning in spoken messages. *Communication Education*, *38*(3), 237–248. https://doi.org/10.1080/03634528909378760

Cross, J. (2014). Promoting autonomous listening to podcasts: A case study. *Language Teaching Research*, *18*(1), 8–32. https://doi.org/10.1177/1362168813505394

Goh, C. (2008). Metacognitive instruction for second language listening development. *RELC Journal*, *39*(2), 188–213. https://doi.org/10.1177/0033688208092184

Goh, C., & Vandergrift, L. (2022). *Teaching and learning second language listening: Metacognition in action* (2nd ed.). Routledge.

Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, *34*(2), 165–182. https://doi.org/10.1016/j.system.2005.11.001

Graham, S. (2016). Research into practice: Listening strategies in an instructed classroom setting. *Language Teaching*, *50*(1), 107–119.

Graham, S., & Macaro, E. (2008). Strategy instruction in listening for lower-intermediate learners of French. *Language Learning, 58*(4), 747–783.

Karlin, O., & Karlin, S. (2023). Developing L2 listening comprehension through extensive and intensive listening. *Aila Review*, *36*(1), 91–111. https://doi.org/10.1075/aila.22015.kar

Kelly, W. Q., & Klein, J. D. (2016). The effect of type of podcasts and learning styles on language proficiency and confidence. *Journal of Educational Technology Systems*, *44*(4), 421–429. https://doi.org/10.1177/0047239515617159

Khalili Sabet, M. (2012). The impact of authentic listening materials on elementary EFL learners' listening skills. *International Journal of Applied Linguistics & English Literature*, *1*(4), 216–229. https://doi.org/10.7575/ijalel.v.1n.4p.216

Long, D. R. (1987). Listening comprehension: Need and neglect. *Hispania*, *70*(4), 921-928. https://doi.org/10.2307/342570

McBride, K. (2011). The effect of rate of speech and distributed practice on the development of listening comprehension. *Computer Assisted Language Learning*, *24*(2), 131–154. https://doi.org/10.1080/09588221.2010.528777

McCaughey, K. (2015). Practical tips for increasing listening practice time. *English Teaching Forum*, *53*(1), 2–13. https://files.eric.ed.gov/fulltext/EJ1060240.pdf

Milliner, B., & Dimoski, B. (2021). The effects of a metacognitive intervention on lower-proficiency EFL learners' listening comprehension and listening self-efficacy. *Language Teaching Research*, 1–35. https://doi.org/10.1177/13621688211004646

Mohsen, M. A., & Almudawis, S. (2021). Second language vocabulary gains from listening versus reading comprehension input: A comparative study. *Journal of Psycholinguistic Research*, *50*, 543–562. https://doi.org/10.1007/s10936-020-09690-y

Moyer, A. (2008). Language contact and confidence in second language listening comprehension: A pilot study of advanced' learners of German. *Foreign Language Annals, 39*(2), 255–275. https://doi.org/10.1111/j.1944-9720.2006.tb02265.x

Naidionova, A. V., & Ponomarenko, O. G. (2018). Use of podcasting technology to develop students' listening skills. *Information Technologies and Learning Tools*, *63*(1), 177-185. https://doi.org/10.33407/itlt.v63i1.1962

Nunan, D. (2002). Listening in language learning. In W. A. Renandya (Ed.), *Methodology in Language Teaching: An Anthology of Current Practice* (pp. 238–241). Cambridge University Press.

O'Brien, A., & Hegelheimer, V. (2007). Integrating CALL into the classroom: The role of podcasting in an ESL listening strategies course. *ReCALL*, *19*(2), 162–180. https://doi.org/10.1017/s0958344007000523

Oanh, P. T. K., & Anh, N. T. N. (2022). Using podcast to enhance pronunciation competence for high school students. *International Journal of Scientific and Research Publications (IJSRP)*, *12*(2), 520-527. https://doi.org/10.29322/ijsrp.12.02.2022.p12265

Pei, T., Suwanthep, J., & Lu, H. (2023). The effect of self-directed online metacognitive listening practice on Chinese EFL learners' listening ability, metacognition, and self-efficacy. *Frontiers in Psychology*, *14*, 1285059. https://doi.org/10.3389/fpsyg.2023

Rahimi, M., & Abedi, S. (2014). The relationship between listening self-efficacy and metacognitive awareness of listening strategies. *Procedia - Social and Behavioral Sciences*, *98*, 1454–1460. https://doi.org/10.1016/j.sbspro.2014.03.565

Rosell-Aguilar, F. (2013). Podcasting for language learning through iTunes U: The learner's view. *Language Learning & Technology*, *17*(3), 74–93.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, *78*(2), 199–221. https://doi.org/10.1111/j.1540-4781.1994.tb02034.x

Shang, H. F. (2008). Listening strategy use and linguistic patterns in listening comprehension by EFL learners. *International Journal of Listening*, *22*(1), 29–45. https://doi.org/10.1080/10904010701802147

Siegel, J. (2011). Thoughts on L2 listening pedagogy. *ELT Journal*, *65*(3), 318–321. https://doi.org/10.1093/elt/ccr029

Siegel, J. (2013). Exploring L2 listening instruction: Examinations of practice. *ELT Journal*, *68*(1), 22–30. https://doi.org/10.1093/elt/cct058

Stockwell, G. (2012). *Computer-Assisted language learning*. Cambridge University Press.

Tabatabaee, G. A., & Rezvani, E. (2019). Effects of using English podcasts on idioms and vocabulary learning of Iranian EFL learners. *Indonesian Journal of Learning and Instruction*, *2*(01). https://doi.org/10.25134/ijli.v2i01.1680

Tomé Díez, M., & Richters, M. A. (2020). Podcasting as a tool to develop speaking skills in the foreign language classroom. *The EuroCALL Review*, *28*(1), 40-56. https://doi.org/10.4995/eurocall.2020.13191

Vandergrift, L. (1999). Facilitating second language listening comprehension: Acquiring successful strategies. *ELT Journal*, *53*(3), 168–176. https://doi.org/10.1093/elt/53.3.168

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, *24*(24), 3–25. https://doi.org/10.1017/s0267190504000017

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning*, *60*(2), 470–497. https://doi.org/10.1111/j.1467-9922.2009.00559.x

Yang, X. (2014).. (*Case analysis of Podcasting's influence on Chinese listening comprehension.* [Unpublished Master's thesis]. Beijing Foreign Studies University. [杨晓彤. (2014). *Podcast 对汉语听力理解能力影响的个案分析* [未出版硕士论文]. 北京外国语大学.]

Yin, M. (2015). The effect and importance of authentic language exposure in improving listening comprehension. *Culminating Projects in English*, 33. https://repository.stcloudstate.edu/engl_etds/33

Appendix 1

I. Samples of the podcast series topics

	Textbook Audio	Real Talks
Intermediate Level Podcast	初到公司 会见客户 中国人的称呼 握手的讲究 改革和开放 国企、私企和外企	王小鲁谈改革 民营企业家 肯德基的中国化 龙行天下—开放的中国 外资撤离中国 数字广告
Advanced Level Podcast	肯德基全力打造中国人的快餐 可口可乐的中国功夫 必胜客：胜在起点 爱她就请她吃哈根达斯 耐克广告遭遇禁播	肯德基和麦当劳谁更胜一筹 美国中餐女王 云办公巨头 Zoom 跟着 Zara 去开店 不同于星巴克的蓝瓶咖啡 直播带货：崛起的新“风口”

II. Sample of podcast transcripts

Textbook Audio《改革和开放》Intermediate Level

从 1978 年到现在的几十年时间里，中国的经济取得了巨大的发展。中国的 GDP 年均增长 9.7%，经济总量的世界排名从第十位提高到第四位，外汇储备世界第一。随着经济的发展，中国人的生活水平也有了很大的提高。13 亿人的温饱问题已经基本上解决，农村的贫困人口从 1978 年的 2.5 亿减少到 2006 年的 2148 万。这些成绩都应该归功

于邓小平提出的改革开放政策。

改革的重点是转变经济体制。改革以前，中国一直是计划经济，市场的作用很小。1978 年以后，中国政府决定向市场经济转变，建立“有中国特色的社会主义市场经济体制”。事实证明，30 年前开始的这次改革是成功的。它从农村到城市，从沿海到内地，调动了人们工作的积极性，使中国经济取得了巨大的发展。

中国对外开放，是为了融入世界经济。利用巨大的市场，中国吸引了大量外商和外资，他们对中国经济的发展帮助很大。中国也利用这个机会，向世界展示自己发展的成绩，改善自己的形象。现在，中国已经是很多发展中国家学习的典范。还有人说，中国是第三次全球化的主角。

Real Talks《王小鲁谈改革》Intermediate Level

今天的内容，取材于（中国改革基金会国民经济研究所副所长）王小鲁先生的讲话。原文比较长，今天我们先谈其中的两个问题。

第一个问题：这 40 年改革让中国经济发生了哪些变化？

王小鲁先生说，1977 年，改革的前一年，中国的人均 GDP 折算成美元是 198 美元；2017 年是 8836 美元。如果用人民币不变价/衡量的话，人均 GDP 大概涨了 30 倍。用美元的不变价格衡量，大概是 15 倍。

还有其他的一些数字也可以说明问题，比如城镇化率，1977 年中国的城镇化率是 17.5%，也就是说，80%以上的中国居民是农民，当时农民的生活状况是很差的。2017 年城镇化率 58.5%，农村居民下降到 40%左右。所以，总体上看，改革让中国经济经历了翻天覆地的变化。

王小鲁先生认为，这四十年，主要有四个方面：

第一，从过去的计划经济转向了市场经济。效率更高，现在，总体上看，中国算是市场经济国家了。

第二，从一个封闭的国家转向了开放。改革前，中国对外贸易非常少，既没有对外投资，也没有外来投资。现在变成了一个总体上开放的经济，中国目前已经是世界第一大贸易国了。

第三，从过去的国有经济一统天下，到现在实际上民营经济已经占了一大半了。

改革开放以前，除了农业，基本上都是国有经济，在工业中，国有企业占的比重接近 80%，剩下的 20%全部是集体所有制企业。现在非国有经济占工业产出的比重已经超过了四分之三，国有和国有控股企业占不到四分之一。

Textbook Audio**《肯德基全力打造中国人的快餐》节选** Advanced Level

肯德基在中国已经走过了二十多个年头。回顾这二十多年的发展历史，我们不难发现,肯德基在中国的发展简直可以用“飞跃”两个字来形容。1987 年 11 月 12 日，肯德基落户北京，北京肯德基有限公司也是当时北京第一家经营快餐的中外合资企业。从此，肯德基在中国的扩张就一发而不可收。1992 年 10 家，1996 年 100 家，2004 年 1000 家，2007 年 2000 家，截至 2009 年 2 月，肯德基在中国的餐厅数量达到 2500 多家，创下国际快餐连锁业在中国开店数量之最。各种数据显示，肯德基在中国已经取得了巨大的成功，其发展速度和规模，让它成为在华最成功的西式快餐连锁企业。

Real Talks**《肯德基和麦当劳谁更胜一筹》**Advanced Level

这两大“洋快餐”品牌，在中国都拥有着不相上下的人气，但是，他们的策略和路线，似乎又存在着很多不同。今天，我们就对比一下，看看他们到底谁更胜一筹？

第一、无差异 还是 差异化

麦当劳刚到中国，就出现了“水土不服”的现状。这是因为在进入中国的最初几年，麦当劳采取的是“无差异市场”的战略，也就是将其在西方的经验直接搬到了中国的市场上，比如在加油站附近开设门店，这种做法虽然在美国为麦当劳带来了相当高的利润，但在中国却根本行不通。同样，在产品方面，麦当劳以汉堡和牛肉为主，这非常符合欧美消费者的饮食习惯，但在中国却并没有那么受欢迎。

而肯德基则表现出了更强的适应能力，它针对不同的地域和饮食习惯，采用了差异化的市场战略，在以鸡肉为主的同时，推出了更符合中国人口味的汤、粥、饭等各种产品，并且不断地推出有地方特色的新品，比如老北京鸡肉卷、川香辣子鸡等等。

正因为两者在最初进入市场时，采取的战略不同，肯德基一直走在麦当劳的前面。后来，麦当劳也意识到了这一点，很快地调整了市场定位，做出了差异化的改变。

Appendix 2

Survey questions

Listening Frequency and Habits

1. In the past month, how many times in a week do you usually listen to our podcasts?
2. In the past month, how many times in a week have you listened to Chinese audio materials other than our podcast?
3. How do you usually listen to our podcast?

Podcast Usage Patterns

5. How many episodes of each subseries have you listened to?
    - Authentic talks
    - Textbook audios
6. For those episodes you have listened to, how many times do you usually listen?
    - Authentic talks
    - Textbook audios

Podcast Design Evaluation

7. What do you think of the speaking speed of the podcast?
8. What do you think of the timing of our weekly podcasts (i.e., released each Thursday)?
9. Please rate the length of each subseries:
    - Authentic talks
    - Textbook audios
10. Please rate the difficulty of each subseries:
    - Authentic talks
    - Textbook audios

Perceived Value and Impact

11. Overall, how useful are the podcasts to your Chinese learning?
12. Please rate the overall usefulness of each sub-series:
    - Authentic talks
    - Textbook audios

Note: These questions represent key items from the pre- and post-intervention surveys administered to participants. Response options included Likert scales for agreement/evaluation questions and multiple-choice or open-ended responses for other items.

URL (网址): http://www.tclt.us/journal
Email (电子邮件): editor@tclt.us