

Integrating Task-Based Language Teaching and Generative AI: Design, Implementation, and Evaluation of the *CFLingo* Platform for Chinese Learning

(任务型语言教学与生成式人工智能的融合：*CFLingo* 中文学习平台的设计、实施与评估)

Shan, Limei
(单丽梅)
Lehigh University
(理海大学)
lis208@lehigh.edu

Pan, Zilong
(潘子龙)
Lehigh University
(理海大学)
zip322@lehigh.edu

Weidman, Rob
Lehigh University
(理海大学)
row3@lehigh.edu

Abstract: Recent advancements in generative artificial intelligence (GAI) have led to the development of GAI-integrated platforms to enhance foreign language learning. However, such platforms' effective design, development, and evaluation require a robust theoretical framework. This design-based study applies task-based language teaching (TBLT)—specifically the cognition hypothesis (CH) and the triadic componential framework (TCF)—to inform the design and implementation of *CFLingo*, a GAI-integrated Chinese language learning platform. The study addresses three key inquiries. First, the study examines how the cognition hypothesis can inform task sequencing within the platform. By progressively increasing task complexity, the platform scaffolds learners' cognitive load, guiding them from simpler to more challenging tasks in a structured and supportive way. Second, it explores the role of the triadic componential framework in enhancing the platform's adaptability through prompt engineering techniques, which optimize task conditions to address learners' varying proficiency levels and provide tailored feedback, creating opportunities for meaningful language practice. Third, the study evaluates the platform's effectiveness through open-ended responses and interviews with 26 college students who used *CFLingo* over a semester. The findings reveal that task sequencing and adaptive feedback enhanced task authenticity, improved performance, and enriched the learning experience. These insights offer valuable design and instructional implications for future GAI-integrated language learning platforms.

摘要: 近年来,生成式人工智能(GAI)的快速发展催生了多种旨在提升外语学习效果的GAI集成平台。然而,这类平台的有效设计、开发与评估需要一个坚实的理论框架作为支撑。本项设计型研究运用任务型语言教学(TBLT),特别是认知假说(CH)和三元成分框架(TCF),为GAI集成语言学习平台——智语学伴(*CFLingo*)的设计与实施提供理论指导。研究围绕三个核心问题展开探讨。首先,研究

探讨了认知假说如何指导平台中的任务序列设计。通过逐步增加任务复杂性，平台帮助学习者合理分配认知负荷，从简单任务逐渐过渡到更具挑战性的任务，实现结构化、支持性的学习进程。其次，研究探索了三元成分框架在平台适应性方面的作用。通过提示工程技术，平台优化任务条件，以适应不同水平学习者的需求，提供个性化反馈，为学习者创造有意义的语言练习机会。最后，研究通过对 26 名在一学期内使用 *CFLingo* 的大学生进行开放式问卷和访谈，评估了平台的有效性。结果显示，任务序列设计与适应性反馈提升了任务的真实性，改善了任务表现，并丰富了整体学习体验。本研究的结果为未来 GAI 集成语言学习平台的设计与教学提供了宝贵的启示与实践指导。

Keywords: Generative artificial intelligence (GAI), OpenAI API, human-centered platform, Task-Based Language Teaching (TBLT), Cognition Hypothesis (CH), Triadic Componential Framework (TCF)

关键词: 生成式人工智能、应用程序接口、人本主义平台、任务型语言教学、认知假设理论、三元要素框架

1. Introduction

With the rapid advancement of artificial intelligence, the integration of large language model (LLM)-supported generative AI tools (GAI), such as OpenAI's ChatGPT or Google's Gemini, into foreign language education has gained significant attention. Key features like "robust NLP capabilities, adaptability, and interactive elements" (Li, 2024, p. 26) hold great potential for enhancing foreign language pedagogy and second language acquisition (SLA) research.

As researchers have begun to consider using different types of AI tools in foreign language teaching, such as chatbots as conversational partners or writing assistants, concerns about "inaccuracies, bias, and plagiarism" (van Dis et al., 2023, p.224) persist. In the SLA process (Han, 2007), particularly during interactive tasks where much of the "negotiation and feedback" takes place, students often receive limited guidance from instructors while obtaining individualized feedback from AI. This is especially problematic when assigning interaction-based homework using tools like ChatGPT Web Application. It has become increasingly difficult for teachers to monitor students' interactions with AI, track their progress, and address potential issues such as misinformation or ethical concerns. Instructors need a secure platform to safely store student data, enabling effective follow-up feedback and assessment. Building on this need, a key question emerges: How can a secure, 'human-centered' platform (Yan, 2024) be designed to foster collaboration between teachers and AI in achieving teaching goals? OpenAI's Application Programming Interface (API) offers a promising solution. Compared to the ChatGPT web app, the API is more adaptable to classroom use, offering greater customization, support for multiple accounts, enhanced security, and better prompt control. These capabilities provided the researchers

with the possibility and feasibility of developing a GAI-integrated language learning platform using OpenAI's API, which was subsequently named *CFLingo*.

The next challenge that needs to be addressed is aligning the technical design of a GAI-integrated language learning platform with foreign language curriculum that adopts Task-Based Language Teaching (TBLT). TBLT is a research-based pedagogical framework. It is functionally oriented but has psycholinguistic validity. It emphasizes form, meaning, and use, aiming to help learners perform meaningful, real-world tasks, known as target tasks. Pedagogic tasks serve as foundational steps to develop language skills required for target tasks (East, 2021). Task-Based Language Teaching (TBLT) emphasizes the importance of sequencing tasks to progressively develop learners' cognitive abilities, thereby supporting both language acquisition and performance. In response to this need, Robinson introduced the Cognition Hypothesis (CH) and later the Triadic Componential Framework (TCF) (Robinson, 2001b, 2003b), which provides a detailed taxonomy of task characteristics. Both CH and TCF play a pivotal role in TBLT by illustrating how increasing task complexity—while accounting for cognitive, interactive, and learner-related factors—can enhance second language development. Incorporating pedagogical theories such as CH and TCF is essential for guiding the design and development of *CFLingo*—a generative AI-enhanced Chinese language learning platform that seamlessly integrates these principles throughout its development process.

This paper explores how CH and TCF can be applied to design and develop a GAI-integrated Chinese language learning platform. Specifically, this exploration intends to: (a) present how sequence tasks informed by CH can be utilized to design a GAI-integrated learning platform, (b) explore how TCF can be applied via prompt engineering on the GAI-integrated learning platform to create adaptive learning experiences, and ultimately address the research question of (c) how effective the overall design of the GAI-integrated language learning platform is.

2. Literature Review

2.1 The Growing Role of Generative AI in Language Teaching

Since GAI tools, like ChatGPT, were first released to the public, language instructors and researchers have moved through stages of concern, from attempts to ban GAI applications and eventually toward recognizing the benefits of these changes. They are now acknowledging and embracing the opportunities presented by this exciting new technology, particularly conversational agents like ChatGPT (Hong, 2023). In the past two years, researchers have taken initiatives to conduct empirical and theoretical studies in foreign language teaching, including Chinese language instruction. These studies range from examining students' experiences and perceptions of ChatGPT 3.5 (Xiao & Zhi, 2023), exploring pre-service teachers' insights into AI conversational chatbots (Belda-Medina & Calvo-Ferrer, 2022), and evaluating the suitability of ChatGPT-generated dialogue materials for EFL learners (Young & Shishido, 2023), to more extensive research on

ChatGPT's capabilities in assessing CFL learners' writing performance, generating teaching materials, and creating teaching tasks and assessments (Li et al., 2024).

As Han (2024) summarized, three research areas are emerging concerning ChatGPT's affordances for language learning: (1) ChatGPT's capabilities, such as adaptability and personalization in delivering tailored learning experiences, though the specific affordances that optimize learning outcomes remain underexplored (Han, 2024); (2) how learners leverage these affordances to enhance language skills, with a research emphasis needed on "the interaction between ChatGPT and language learners" (Han, 2024, p. 303); and (3) the role of human agency in guiding and mediating the learning process during AI chatbot interactions (Han, 2024). These research areas provided implications for future GAI integrated platform development and implementation.

2.2 Diverse Voices and Debates in TBLT

A task-based syllabus is an instructional method that structures learning around practical, real-world tasks, such as ordering a meal, for students to complete in the classroom. It consists of a sequence of tasks to foster learners' communicative skills (Pica et al., 1993). By embedding systematic language learning within practical and communicative tasks it supports structured SLA. Through fostering interaction and cognitive engagement, it aligns with SLA's focus on iterative and meaningful language use (Han, 2018). This approach enhances structured SLA by integrating real-world practice with theory-driven language instruction.

The classification of tasks in TBLT has evolved significantly over the years. Pica et al. (1993) established an early typology of communication tasks. Skehan (1998, 2003) expanded this categorization by introducing a taxonomy identifying task characteristics influencing linguistic demands, such as code complexity, cognitive complexity, and communicative stress. This framework highlighted how varying task features could affect learners' language performance (Skehan & Foster, 2001). Building upon these earlier models, Robinson developed the Triadic Componential Framework (TCF; 2007a) that offers a vital taxonomy for understanding task characteristics in TBLT by categorizing task demands into three main areas: complexity, condition, and difficulty.

The most extensive and active exploration of cognitive task complexity over the past few decades has taken place within the TBLT domain. Among these discussions, the debate surrounding Skehan's Limited Capacity Hypothesis (LCH) and Robinson's Cognition Hypothesis (CH) offers important insights into TBLT and its impact on L2 production. Skehan (1998) posited that learners face a trade-off among complexity, accuracy, and fluency (CAF) when performing tasks due to limitations in attentional resources and working memory. He argues that fluency often takes precedence during meaning-based tasks, potentially compromising complexity and accuracy. This suggests that task design must promote balanced language development, ensuring learners can effectively engage with all three aspects of language use (Skehan, 1998).

Conversely, Robinson's CH emphasizes that increasing cognitive demands within pedagogic tasks can enhance interlanguage development by encouraging learners to focus on L2 features necessary for expressing new cognitive distinctions. His study shows a correlation between task complexity and the syntactic complexity of task-doers' or L2 learners' speech production (Han, 2018). Robinson (2001a, 2003b) argues that this heightened focus leads to increased uptake and accelerates grammatical development. Furthermore, manipulating task complexity—such as reducing planning time—can improve learners' access to their current L2 abilities and enhance the likelihood of transferring learned skills to real-world tasks (Long, 2014).

Despite the theoretical foundations of both Skehan's Limited Capacity Hypothesis and Robinson's Cognition Hypothesis, empirical studies have yielded inconsistent results regarding the interplay of accuracy, complexity, and fluency in task performance. Many findings do not achieve statistical significance or reveal unexpected trends (Long, 2014). To address these gaps, this study proposes a GAI-integrated language learning platform that sequences tasks based on their complexity. By leveraging the adaptive capabilities of generative AI tools, the platform tailors task conditions to accommodate learners' perceived difficulty levels, ultimately fostering a more effective and personalized language learning experience.

2.3 The Role of GAI in TBLT

Prior to GAI, SLA literature “barely scratches the surface” (Han, 2024, p.302) of conversational agents (CAs) used in TBLT. Previous studies primarily concentrated on learners' perceptions of experiences with AI tools, revealing that students enjoyed task-based interactions with CAs. Learners favored CAs that offered both audio and visual feedback, but the research mainly examined L2 speaking and listening skills without exploring broader learning impacts, such as reading and writing, for Chinese learners. Tasks were selected from readily available CAs like Google Home Hub and Amazon Echo, leading to insufficient development of tailored tasks that meet learners' specific needs. The researcher indicated “only a handful of studies have redesigned the dialogue system to fit into specific learning contexts or redesigned the tasks for specific learning purposes” (Xiao et al., 2023, p.12).

With the emergence of GAI, several researchers have conducted empirical studies on task generation within TBLT. It is capable of generating materials across various types and difficulty levels, including a wide range of tasks (Li et al., 2024). Since GAI tools provide a “Human-machine” interface for each language learner, how can tasks be sequenced and designed so that all learners can do the same task while getting individualized interaction with AI. It becomes essential to have a deeper exploration of GAI's role in Instructed Second Language Acquisition (ISLA) which TBLT rooted in (Han, 2024). Existing literature has not yet systematically investigated the integration of GAI tools within TBLT curricula. There is a research gap on whether TBLT theories, specifically Robinson's Cognition Hypothesis and Triadic Componential Framework, can still guide the Chinese curriculum design and implementation in this new GAI-integrated language learning environment. This study will address this research gap by applying CH

and TCF to GAI-integrated language learning platform design, emphasizing the pedagogical task classifications and sequencing.

2.4 Theoretical Foundation for GAI-Integrated Language Learning Platform

To explore how TBLT can guide the design of GAI-integrated Chinese language learning platform, this study focuses on Robinson’s Cognition Hypothesis (CH) and Triadic Componential Framework (TCF), which underpin its rationale and research design.

The CH in TBLT suggests that “sequencing tasks from simple to complex creates optimal conditions for practice” (Robinson, 2003b, p.55), as it maximizes the real-world target task requirements (Robinson, 2001a). Robinson clarifies task complexity along two dimensions (Figure 1): resource-directing and resource-depleting. As shown in Figure 2, for both individual and interactive tasks, moving from simple to complex along the resource-directing dimension may reduce fluency but enhance accuracy and complexity in learners’ language use. On the other hand, along the resource-depleting dimension, when learners face less planning time, unfamiliar tasks, or multiple simultaneous tasks, the skills they develop are more likely to transfer effectively to real-world performance (Long, 2014). Thus, the Cognition Hypothesis argues that pedagogic tasks should be sequenced based on increasing cognitive complexity (Robinson, 2005). This principle serves as the first foundation for designing task sequences for the GAI platform within TBLT contexts.

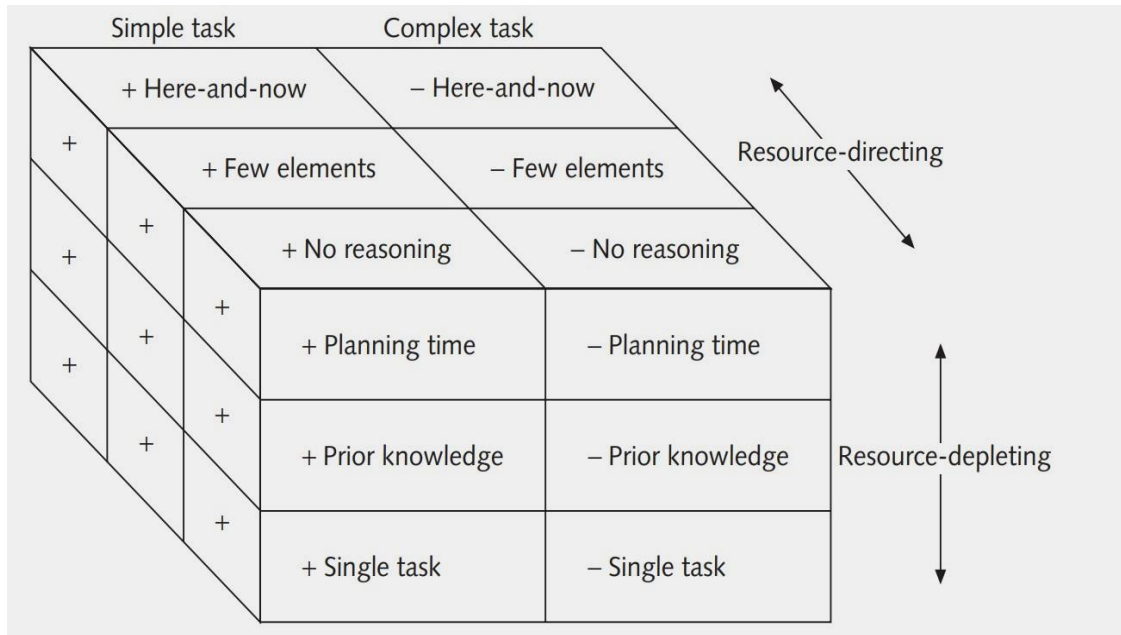


Figure 1 Resource-Directing and Resource-Depleting Dimensions of Task Complexity (Robinson, 2003a, p648)

monologic tasks	
simple	complex
+ fluency, - complexity, - accuracy	- fluency, + accuracy, + complexity
interactive tasks	
simple	complex
+ fluency, - accuracy, - comprehension checks/ clarification requests	- fluency, + accuracy, + comprehension checks/ + clarification requests

Figure 2 Proposed Effects of Task Complexity on Accuracy, Fluency and Complexity along Resource-Directing Dimensions (Robinson, 2001a)

Robinson's (2001b) Triadic Componential Framework (TCF) is a practical categorization of task characteristics. As outlined in Appendix 1 (Robinson, 2006), he identifies three categories of task demands crucial for real-world task performance: task complexity, task conditions, and task difficulty (Robinson, 2011). While Robinson emphasized sequencing pedagogic tasks based on cognitive task complexity, he also highlighted that "successful learning and performance result from the interaction of different aspects of task demands" (e.g., complexity level and task conditions) with learners' ability profiles, which influence their perceptions of task difficulty (Robinson, 2001c, 2002b).

TCF taxonomy forms the second foundation of this study. Within a single unit or lesson, the same task may produce varying outcomes depending on the conditions under which it is performed. This underscores the importance of considering both task conditions and task difficulty during task implementation. Each learner has a unique "aptitude profile"—a combination of natural abilities or capacities influencing their success in language learning. According to Robinson's Aptitude Complex/Ability Differentiation framework (Robinson, 2007a), these individual differences can be linked to the Cognition Hypothesis, which suggests that tasks can be designed with varying complexity levels to align with or challenge learners' aptitude profiles. By adjusting task conditions—such as learners' proficiency levels, available resources, and the equality of participant roles—educators can tailor tasks to match learners' aptitudes better. This approach enhances cognitive engagement and optimizes learning outcomes. These principles guide the present study's task design.

A prompt-engineering approach was employed in each task design to achieve this, enabling the generative AI to produce outputs aligned with learners' aptitude profiles. This approach establishes optimal task conditions that support second language (L2) acquisition by promoting fluency, accuracy, and complexity in learners' language performance.

2.5 Purpose of the Study

The challenge of aligning learners of varying proficiency levels with tasks that best match their abilities—thereby maximizing learning and performance—remains underexplored and infrequently applied (Robinson, 2001c). Building on the CH and the TCF, this study introduces a GAI-integrated foreign language learning environment, *CFLingo*, designed to address the core issues of effective task sequencing for acquisition and instructional adaptation to meet learners' needs. This exploration focuses on:

- a) (a). Understanding how the Cognition Hypothesis (CH) can guide task sequencing design within *CFLingo*;
- b) (b). Investigating the role of the Triadic Componential Framework (TCF) in optimizing the *CFLingo*'s adaptability.

Finally, based on the integration of CH and TCF, the study intends to address the following research question:

- c) How effective is *CFLingo* in improving task performance and enriching the learning experience?

3. Development of the GAI-integrated Language Learning Platform—*CFLingo*

3.1 Platform Background and Overview

CFLingo, 智语学伴 in Chinese, was designed and developed at Lehigh University to provide a virtual language partner that supports reading and writing practice while connecting to in-class Task-Based Language Teaching (TBLT) tasks. The platform was designed for a diverse group of learners, including heritage speakers and non-native students, all of whom have passed HSK Level 3 but exhibit varying strengths and weaknesses—particularly in reading and writing—due to limited opportunities for real-time conversations with native speakers. The platform leverages OpenAI's GPT large language model to create adaptive and authentic learning experiences tailored to each student's proficiency level and individual learner profile. By aligning AI-driven tasks with the chapter-based syllabus, *CFLingo* ensures meaningful engagement that complements classroom instruction. Additionally, the platform's adaptability allows students to practice skills in a way that reflects real-world language use, bridging the divide between structured learning and authentic application while fostering individualized growth.

Figure 3 presents the homepage of *CFLingo*: There are two to three pedagogic tasks designed for each chapter scaffolded to align with the CH, covering a total of ten chapters in *Integrated Chinese* (Volume 3), which is the leading textbook adopted at college-level 3rd-year class across the US. During the five-day teaching cycle for each chapter, students are given designated time to interact with *CFLingo* during class. Each task typically takes about half an hour to complete, ensuring that all students engage with the same pedagogical task related to the chapter. The instructor can monitor progress in real-time, either in the

classroom or by reviewing chat logs in the "My Chats" section. While instructors have access to all chat records, students can only view their own chat history for future reference.



Looking for the Perfect Language Partner? Meet CFLingo AI!

智语学伴 (CFLingo) 秋季 (Fall) 春季 (Spring) 我的聊天记录 (My Chats)

智语学伴 - CFLingo

[中文听说读写 第三册 秋季](#)
[Exercises for Integrated Chinese Volume 3, Fall CHIN 115]

[中文听说读写 第三册 春季](#)
[Exercises for Integrated Chinese Volume 3, Spring CHIN 116]

[我的聊天记录](#)
[My Chats - Archive of your past chats with CFLingo]

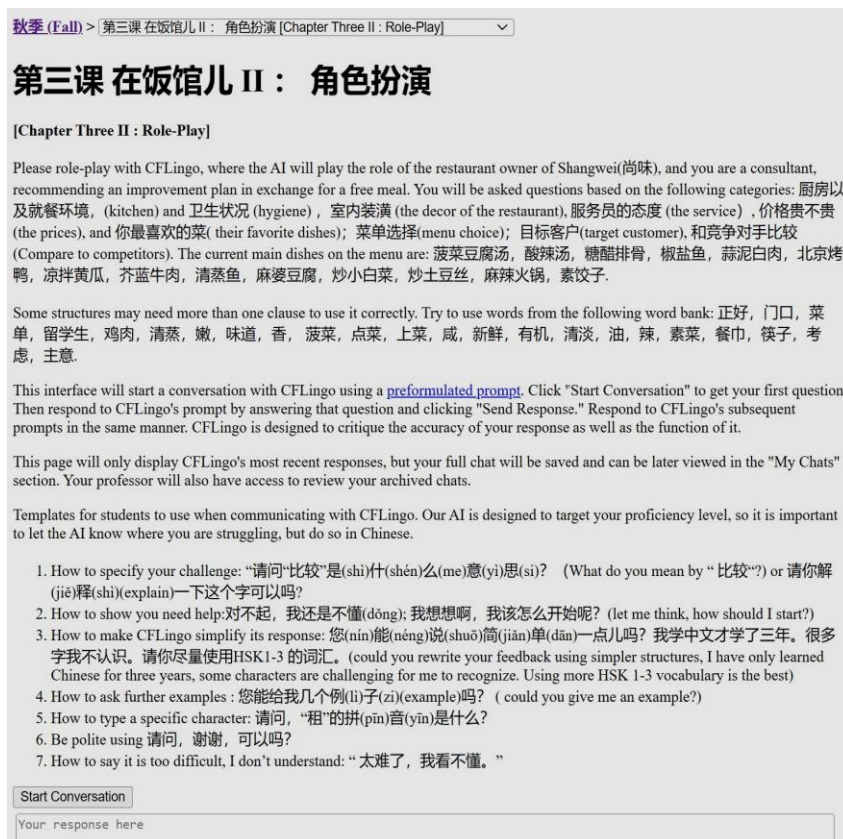
你是否还在苦苦寻找合适的语伴，希望有人能陪你一同练习、探索华语的魅力？「智能华语学伴」让一切变为可能！我们为您提供24小时在线中文练习，让你随时随地都能学习中文，快速提高。我们的个性化方案让学习更简单、更有趣。赶快加入，开启轻松的中文学习之旅！

Master Mandarin with CFLingo AI! Struggling to find a Mandarin partner? Look no further! Enjoy 24/7 online practice and personalized learning plans tailored to your pace. With CFLingo AI, mastering Mandarin is effortless and fun. Join us today and start your Mandarin journey with ease!

Figure 3 CFLingo Homepage 智语学伴

CFLingo offers task-based instructional support activities (see Figure 4), including sentence generation, sentence rearrangement, language error detection, email responses, role-playing or debating with a virtual language partner, and describing or narrating events. As students progress, the focus gradually shifts from developing interpretive reading and interpersonal communication skills to enhancing presentational writing abilities, such as essay drafting and revision. Students engage with the platform weekly, completing tasks aligned with the chapters they are studying. A dropdown menu on the interface allows students to easily switch between tasks, ensuring smooth navigation and flexibility.

For example, in Chapter Three, themed “At the Restaurant,” students begin with a “sentence generation” task, where they construct complete and meaningful sentences using words or phrases provided by *CFLingo*. This initial activity focuses on language forms by identifying student errors and prompting revisions if the sentence is incorrect. Students are encouraged to draw from a word bank, incorporate more complex structures, or create sentences with multiple clauses. Once students demonstrate linguistic readiness, they advance to a more situational task—engaging in a conversation with *CFLingo*. In this scenario, the AI takes on the role of a restaurant owner, while the student acts as a consultant recommending improvement plans in exchange for free meals. Recommendations may address areas such as kitchen hygiene, restaurant décor, service quality, menu options, or target customers. As shown in Figure 5, students are provided with word banks and templates to assist in generating further prompts in Chinese.

Figure 4 Sample Tasks List of *CFLingo* 智语学件Figure 5 Sample Task Interface of *CFLingo* 智语学件

A key requirement for these tasks is that students must communicate with AI entirely in the target language, Chinese. After completing the two scaffolded activities with *CFLingo*, students present a live restaurant review in class, focusing on descriptive and narrative language to comment on the same categories covered in the conversation. Finally, students write a detailed restaurant review in paragraph form as part of their chapter quiz. Overall, *CFLingo's* design for each chapter, in alignment with classroom instruction, guides students' language development from dynamic, interpersonal communication to structured, presentational skills, both in spoken and written Chinese. This AI integrated curriculum reflects the different linguistic demands of each mode of communication—interpersonal, role-play tasks emphasize spontaneous, interactive language, requiring features such as question forms, negotiation strategies, and turn-taking, while presentational tasks like narrating or describing rely on structured, cohesive language for extended discourse.

Figure 6 shows an example of a task interaction in *CFLingo*—News Generation and Discussion. When students access this activity, instructors guide them through the task explanation at the top of the screen, above the dialogue box. Once students understand the task requirements, they can click Start Conversation to interact with *CFLingo* and complete all the subtasks on the page. As shown in this sample, while all students work on the same task, each receives a unique AI-generated news piece and engages in an individualized discussion about it with the AI.

The screenshot displays the CFLingo interface for a task titled "第七课 电脑和网络 II: 新闻生成和讨论" (Chapter Seven II: News Generation & Discussion). The main content area shows a news article in Chinese: "中国学校推出新教育计划" (Chinese schools launch new education plan), dated 2024年1月15日. The article discusses a new plan by a Beijing school to include coding, art, and environmental courses alongside traditional subjects like math and English. Below the article, there are instructions for the AI to generate a sample news article and a "Send Response" button. The chat log shows a conversation where the AI offers help, the user asks for simpler sentences, and the AI provides safety instructions for a chemistry experiment.

Figure 6 Sample Task Chat Logs of *CFLingo* 智语学伴

3.2 Integration of Cognition Hypothesis (CH) on *CFLingo* for Task Sequencing Design

This section explicitly demonstrates the integration of the Cognition Hypothesis (CH) framework in supporting task sequencing design. In line with Robinson's Cognition

Hypothesis (2001b, 2005, 2007a), pedagogic tasks should be sequenced incrementally based on their cognitive complexity. In this study, the *CFLingo* platform comprises 10 chapters and 4 essay assignments delivered across the Fall and Spring semesters, aligning with the curriculum design of *Integrated Chinese, Volume 3*. The selected pedagogic tasks illustrated in Figure 7 represent typical task types for each semester. These tasks are structured and scaffolded to support learners in achieving the target task of “writing descriptive or argumentative essays.” As shown in Figure 7, cognitive complexity progression sequences these tasks accordingly. Figure 8 further depicts cognitive development based on the revised Bloom’s Taxonomy (Anderson & Krathwohl, 2001) and reinforces the Second Language Acquisition (SLA) process, encompassing Input, Negotiation, Feedback, and Output (INFO) as outlined by Han (2007).

The first task, situated at the bottom of the pyramid in Figure 6, involves recalling facts from the text or vocabulary provided by AI. This task corresponds to the “Remember” stage and is considered **input**.

Task 2 progresses to sentence generation and error detection, which falls under the “Understand and Apply” stages and are part of **intake**.

In Task 3, students focus on comprehending emails or other readings, engaging in understanding, applying, and analyzing the material. At this stage, comprehensive input and intake take place. Task 4 introduces role-playing or debating with AI on topics of interest. Here, students practice applying their knowledge and negotiating with AI to clarify linguistic features or **feedback**, ensuring mutual understanding. This task begins to incorporate “reasoning.”

Task 4 also includes describing or narrating a series of pictures in paragraphs, which involves applying, analyzing, and evaluating the material. When AI provides feedback on students’ descriptions or narrations, **negotiation** occurs. Students must further explain or adjust their responses to reach mutual comprehension and facilitate learning.

The subsequent task involves essay outlining and draft critique. This multi-step process supports learners in writing essays, moving from constructing an outline to producing a first draft and eventually a final draft. Through the stored chat logs on *CFLingo*, instructors collaborate with AI to confirm feedback (see Figure 8), emphasize or further explain AI suggestions, and facilitate communication. This process involves extensive clarification, confirmation, and **negotiation** until learners produce the final **output**—well-crafted essays—thus achieving the target task.

As indicated in Figure 7 (Han, 2007) and Figure 8, throughout this guided language acquisition process, task conditions are optimized through negotiation and feedback.

The progression of these pedagogic tasks demonstrates a gradual increase in complexity, preparing students to advance from sentence generation to unprepared role-play or debates with AI, then to composing longer discourse, and ultimately producing well-developed essays. As shown in Figure 7, both the resource-directing and resource-

depleting dimensions involve an increasing number of components. With each level up, more prior knowledge, additional steps, and more elements are needed. The progression transitions from tasks requiring minimal reasoning to those involving extensive reasoning, culminating in essay writing. In the final stage, students must adopt various perspectives and articulate their opinions to compose their essays. This step reflects the "full complexity level of the target task(s)" (Long, 2014, p. 226).

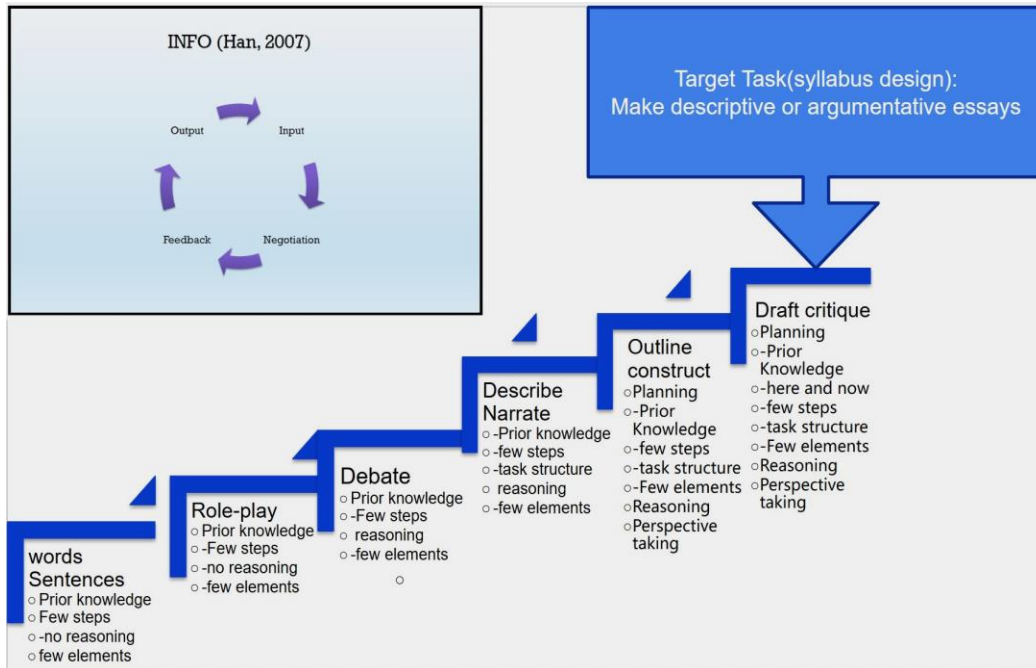


Figure 7 Cognitive Complexity Progression in the Sequencing of Pedagogic Tasks adapted from CH and TCF

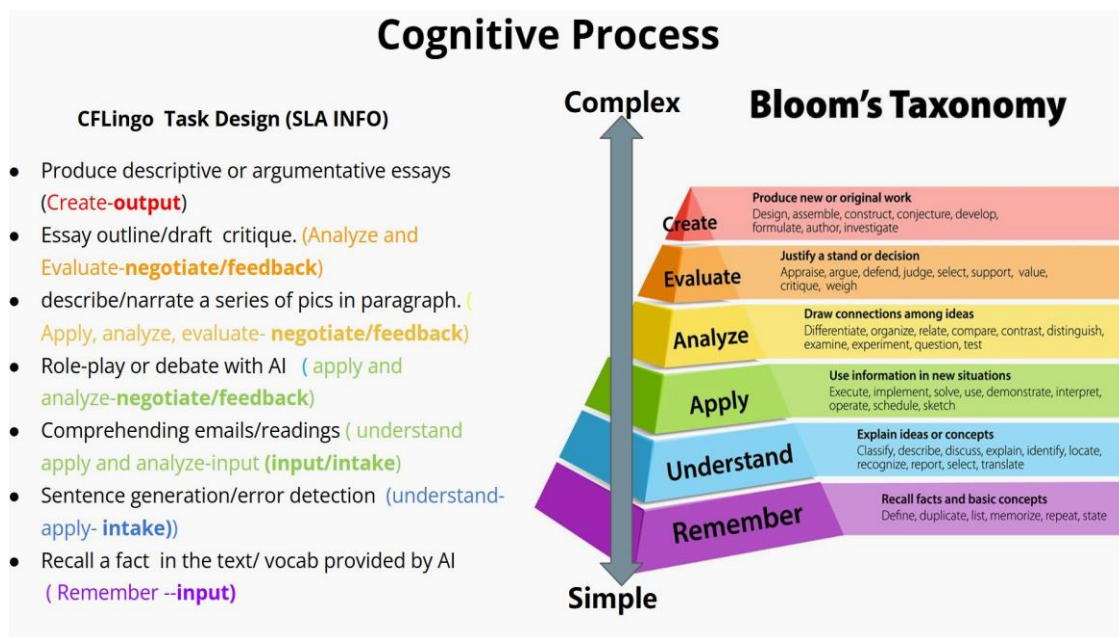


Figure 8 Cognitive Taxonomy in CFLingo Task Syllabus Design Adapted from “Revised Bloom’s Taxonomy (Anderson & Krathwohl, 2001)

An implication for overall task-based syllabus design is that instructors need to gradually increase the complexity of pedagogic tasks, not linguistic complexity. The *CFLingo* task syllabus aligns with the *Integrated Chinese* (IC) curriculum, making it easier for instructors who use the same curriculum or textbook to adopt the *CFLingo* model and effectively reach more Chinese language learners. This demonstrates the generalizability of *CFLingo*.

The CH in task-based learning indicates that as tasks increase in complexity, individual variations in cognitive abilities, such as aptitude and available cognitive resources, play a more crucial role in shaping performance and learning outcomes. This suggests that learners with higher cognitive resources or greater aptitude are better equipped to handle more demanding tasks, leading to improved performance and more effective learning (Robinson, 2003b). The following section further discusses how to leverage task conditions and consider learners' individual variables to reduce the perceived difficulty and enable success in the target tasks.

3.3 Integration of Triadic Componential Framework (TCF) into *CFLingo* for Adaptivity Design Through Prompt Engineering

3.3.1 *CFLingo* Platform System Development

In Fall 2023, the Center for Innovation in Teaching and Learning (CITL) within Lehigh University's Library and Technology Services (LTS) began developing local web interfaces using the application programming interface (API) provided by OpenAI, the developer of the ChatGPT web application and the underlying GPT large language model (LLM) that powers it. This API allows developers to directly interact with the various versions of the GPT LLM. At first, this development primarily sought to solve access issues. The most recent versions of the GPT model were only available through the ChatGPT web application for users paying a monthly subscription fee. Providing a subscription to this service for every student in a class for an entire semester presented a significant barrier. In contrast, a local application could use the most recent GPT model via OpenAI's API at a fraction of the cost. Also, integrating various GPT models, like GPT-3.5, GPT-4.0, and most recently, GPT-4 Omni, allows for tailored solutions that meet diverse teaching needs.

Focused on solely providing access to the latest versions of the GPT LLM, the initial interfaces were relatively simple and provided an open-ended, sandbox-like environment for students to explore. These local interfaces used client-side JavaScript to store, process, and format student-entered text into a format that could be sent via the API to the GPT LLM. This message was then sent to a local server-side PHP backend that would add the API Key credentials that tied the request to CITL's account with OpenAI for billing. This PHP backend could also store the outgoing and ingoing messages in a local MySQL database and associate them with the current user's account, allowing for detailed analysis and progress tracking without OpenAI retaining data, thus ensuring student privacy.

Despite the open-ended focus of these initial local interfaces, the development team also recognized the potential for special-purpose interfaces that could supply additional processing of text sent to and from the large language model to provide scaffolding to students for specific tasks. As students progress to the third year of the Chinese program, the need for focused practice in reading and writing becomes increasingly critical, providing a strong use case for developing a more scaffolded, localized interface to support their learning. Therefore, the development team extended the scope of their work to include the application that would come to be known as *CFLingo*. The researchers and developers used the previously mentioned non-scaffolded sandbox interface to design prompts that would elicit an interaction with the AI that met the needs of a specific exercise. By trial and error, an initial prompt would emerge from this process to seed appropriate interaction between the AI and a student. Then, the developer would create a customized scaffolded interface that would provide the GPT model with this seed prompt before the student began directly interacting with the AI. This ensured all students worked on the same tasks and seamlessly integrated activities into lessons, allowing them to simultaneously engage in unique, individualized conversations. Furthermore, as previously stated, the resulting interactions are saved to a local MySQL database, allowing the instructor to access all conversations and the individual students to review their own, streamlining data collection and allowing the monitoring of student progress.

3.3.2 The Role of Prompt Engineering Within the TCF Framework

For each initial seeded prompt integrated into the *CFLingo* interface and sent to the GPT LLM using OpenAI's API, researchers developed a structured template (see Figure 9) to systematically organize content. This design enabled *CFLingo* to provide students with clear instructions, respond to inquiries, correct responses, offer constructive feedback, and critically adapt to varying proficiency levels while ensuring that students could progress effectively. Figure 9 provides an example prompt, color-coded to correspond with the different components shown in Figure 7 (Cai, 2024). For instance, the prompt informs *CFLingo* that it is assuming the role of a CFL teacher, corresponding to the "Roles" component, highlighted in blue.

Roles	Settings	Task/outcome	Procedure	Tailor	Feedback	Follow-up
<p>Example: You are a Chinese as Foreign Language (CFL) teacher who will prompt English speaking students in a 3rd year of CFL class to make a complete and meaningful sentence in Chinese from a word or phrase that you provide them... provided word list. Restrict all of the words in your chats to this explicit list of 600 HSK 1-3 words.....Critique students' errors and ask them to redo the sentence if it is incorrect, give further feedback and then move on to the next word or phrase to make a sentence from. Encourage them to use words from the following word bank...</p>						

Figure 9 Prompt Engineering Color-Coded Template

Prompt engineering refers to the practice of crafting meticulously designed, text-based prompts to interact with chatbots and other generative AI tools, aiming to generate desired outputs across text, images, audio, video, or a combination of digital media (Knoth et al., 2024; Velásquez-Henao et al., 2023). Each pre-formulated prompt on *CFLingo* represents a specific task design. Comparing these to traditional TBLT task design features (see Table 1), *CFLingo* not only sequences tasks in order to increase cognitive complexity but also emphasizes interactive factors. For instance, AI-human interaction on *CFLingo* is a dynamic two-way process where the AI adjusts its proficiency slightly above or below the learner's current HSK 3 level (i+/-1), focuses on implicit feedback, and maintains equal status with the learner to foster engagement.

Table 1 Key Features comparison between *CFLingo* Task and Traditional task (Candlin, 1987)

Key features of TBLT tasks (traditional)	Each <i>CFLingo</i> Prompt/task design	Interactive factors Task condition	Affective Factors/ability variables Task difficulty
Goal/expected outcomes:	Goal/outcome: authentic language in use in real-life situations Real-life functional/communicative goals such as role-play with restaurant owner	Situational Authenticity Open solution Divergent solution	Aptitude: practical solutions and strategic language use.
Condition/input	Role and setting/context Contextual environment closest to real-life Two sensory inputs (visual on characters and pinyin recall typing) Teaching about creating conditions	Interactional authenticity spontaneous two-way flow, Equal Status	Interest and willingness to communicate (“Novelty”)
Procedure/actions	Procedure (content of the task, scaffolding step by step): enable learners to develop implicit, functional knowledge for communication	Negotiation strategy (prompt engineering for both designers and students)	Encouragement including: Compliment Encourage to Proceed
	Adaption (on proficiency level)	Learners were given instruction by <i>CFLingo</i> that is slightly more advanced than the current HSK 3 level or lower (i+/-1)	Analytical Ability: The skill to identify linguistic patterns and rules, and make an adaptation to AI or make further request

Feedback	Feedback	Feedback Strategy (mainly implicit some explicit)	Encouragement (linguistically) Politeness (pragmatics) Confirmation
	Follow up		Field independence (Individualized follow up, Learner autonomy) Challenging learners with i +1
Monitoring in the classroom	Monitoring through stored data	For Teacher and AI collaborative feedback	Processing anxiety (no interruption) Field independence

CFLingo's task design also accounts for affective variables influencing learners' perceived difficulty. It employs motivational strategies such as showing empathy, giving compliments, reducing processing anxiety, and switching tasks when learners feel bored, thus enhancing the learning experience. *CFLingo* leverages TCF to enhance adaptability through prompt engineering, focusing on "task conditions" and "task difficulty." As this is a design-based study, the refinement process is ongoing and informed by learner feedback. Each year, prompts are adjusted, and different ChatGPT models are tested for improvements.

4. Evaluation of the *CFLingo* Platform

4.1 Participants and Contexts

Throughout Fall 2023 and Fall 2024, the *CFLingo* platform was utilized in three third-year Chinese-as-a-Foreign-Language (CFL) classes, comprising a total of 27 students. Among the 27 students, 26 consented to participate in the study. The HSK Level 3 proficiency test was administered at the beginning of each semester to ensure that students shared a similar proficiency level. The results indicated that 90% of the participants scored 85 or higher in Reading and Writing, whereas 10% scored 65, which is also a passing score. By the third year of CFL study, students aim to achieve Intermediate-Mid to Intermediate-Advanced proficiency on the ACTFL scale, developing balanced listening, speaking, reading, and writing skills. The platform addresses the critical need for reading and writing practice with a language partner, acknowledging the unique linguistic characteristics of Chinese, where spoken and written forms are acquired separately. On this platform, interacting with GAI in real-life contexts helps bridge the gap by connecting the form, sound, and meaning of each character. During the refining and implementation process, "human-centered" is the key element being emphasized (Bhutoria, 2022; Yang et al., 2021).

4.2 Data Sources

Three types of data were collected to evaluate *CFLingo*'s effectiveness: students' responses to open-ended questions on their user feedback on *CFLingo*, interview feedback exploring their experiences and platform improvement, and conversational interaction data to understand the adaptivity of *CFLingo*.

4.2.1 TBLT Items and Open-Ended Questions

At the end of the semester, all students were asked to complete five open-ended questions to articulate their experiences engaging with *CFLingo*. These included questions such as: "How do you like the 'Student Follow-up Prompt Template' added to the language task, which you can refer to when you do not know how to ask questions or seek clarifications?" and "Have you experienced any noteworthy or challenging moments in your interactions with *CFLingo*?" The primary objective of these questions was to gain a comprehensive understanding of user experiences and to assess the overall self-reported effectiveness of *CFLingo* as a language learning tool.

4.2.2 Interview Protocols

Students were invited to participate in voluntary interviews at the end of the semester to discuss their experiences after using *CFLingo* in class. The interview protocol focused on their experiences and perceptions of the *CFLingo* platform, exploring various engagement and task effectiveness themes. Key questions included their general impressions of *CFLingo*, their typical interaction patterns, and the tasks they found most enjoyable or beneficial. For example, participants were asked to identify tasks they enjoyed, such as sentence generation or role-playing with it, and to suggest improvements for less effective tasks. The protocol also examined perceptions of *CFLingo* as a partner in language learning and solicited suggestions for enhancing the platform's instructional and technical design. The interview aimed to provide an in-depth understanding of students' perceptions, supplement the data obtained from the open-ended questions, and help offer insights into the potential of generative AI in supporting language acquisition.

4.2.3 Conversational Data

Participants' conversational data were collected as they engaged with *CFLingo* throughout the semesters. In total, participants generated 114 dialogue threads while interacting with *CFLingo* on various tasks, ranging from sentence generation to interviews with restaurant owners. Each dialogue thread consisted of conversations ranging from 2 to 18 exchanges as students and *CFLingo* communicated back and forth. All sentences were exported from *CFLingo* and grouped by students and tasks for data analysis. The conversational data was used to examine the adaptability of *CFLingo* in responding to students' prompts.

4.3 Data Analysis

Qualitative data from open-ended questions and interview transcripts were analyzed to address students' feedback on task sequencing and adaptivity. Additionally, student-*CFLingo* conversational data was collected and used to understand their experiences with the platform's adaptability.

An inductive coding process was applied for data from open-ended questions and interview transcripts. Researchers collaboratively analyzed the data to generate overarching themes and examined codes within each theme. It was double-coded if a data point aligned with the definitions of multiple themes or codes. Informed by constant comparative methods (Creswell & Creswell, 2017), researchers met regularly to discuss and finalize the codebook to ensure consistency. Subsequently, researchers reviewed each other's codes, refining them until 100% agreement was achieved (see Table 2).

Table 2 Codebook for Student's Open-ended and Interview Data

Theme	Codes	Definition	Example Quotes
Students' Feedback on Task Sequencing	Feedback on the progression of complexity of the tasks	Students' feedback on the design of <i>CFLingo</i> 's tasks gradually increases in complexity.	I did notice a difference, and it was still difficult towards the end, but it felt easier.
	Feedback on conceptual connections of the tasks	Students' feedback on the logical flow of <i>CFLingo</i> 's task design across all activities.	The debate task and essay feedback were interconnected, helping me see how arguments develop in different formats.
	Other comments or suggestions	Students' relevant comments or suggestions relevant to <i>CFLingo</i> 's task sequencing design	One thing I would suggest is for the 'Send Response' button to be made a little bigger
Students' Feedback on Adaptivity	Feedback on personalized conversational experiences	Students' feedback and perception of the personalized conversational experiences with <i>CFLingo</i>	I appreciated how it was like willing to work with me in that way, like a real person would.
	Feedback on vocabulary adaptivity	Students' feedback on <i>CFLingo</i> 's adaptivity on their vocabulary levels	The vocabulary would be beyond my level. I tried asking <i>CFLingo</i> to simplify its responses using prompts, and it

			helped me understand better.
	Feedback on difficulty adaptivity	Students' feedback on <i>CFLingo</i> 's adaptivity on their conversation's difficulty level	I think sometimes it's a little bit too challenging. I don't think it's ever too simple. But if I ask it to write a little simpler, it's usually pretty good about making it more understandable.
	Other comments or suggestions	Students' relevant comments or suggestions relevant to <i>CFLingo</i> 's task adaptivity design	It sometimes oversimplifies when switching to easier vocabulary, and I understand that's a tricky balance to strike

For conversational data, a deductive coding process was used, guided by CH (Peter, 2001), which informed the coding framework. Initially, the same dataset was collaboratively coded to develop a shared codebook. A selective coding approach was used to identify a central theme, such as 'Interactive Factors,' which included codes or sub-codes like 'Implicit Hints.' Each theme or code represented a unique idea, with double coding applied when data aligned with multiple codes or sub-codes. After consolidating their individual codebooks into a single, unified version, this version was used to code the remaining interview transcripts. The researchers then reviewed each other's codes until they achieved 100% inter-rater reliability. Table 3 below presents the coding structure for the conversational data. Since *CFLingo* responds to participants in Chinese, corresponding English translations are provided in the "Quotes" column. In addition, percentages were provided for each theme and code. For example, the label 61% next to 'Interactive Factors' indicates that 61% of the data was coded under the theme of Interactive Factors. In comparison, 43% of the remaining data was coded under Affective Variables. Percentages for codes were calculated relative to their corresponding theme. For instance, 31% next to 'Proficiency' indicates that, within the theme of Interactive Factors, 31% of the codes pertained to Proficiency level. Similarly, within the Proficiency level code, 42% were coded as 'Plus one' and 58% as 'Minus one'.

Table 3 Codebook for Student-CFLingo Conversational Data

Theme	Codes	Sub-codes	Definition	Example Quotes
Interactive Factors (61%)	Proficiency level (31%)	Plus One (42%)	Learners were given instruction by <i>CFLingo</i> that is slightly more advanced than the current HSK 3 level.	很好！你能不能用一个更长一点儿的句子。 Great! Could you use a slightly longer sentence?
		Minus one (58%)	Learners were given instruction by <i>CFLingo</i> that is slightly less advanced than the current HSK 3 level.	"考虑"的意思是 "to consider"。你来试试回答问题。 The meaning of "考虑" is "to consider." Why don't you give it a try and answer the question?
	Negotiation strategy (10%)		The interaction between <i>CFLingo</i> and learner reaches a mutual understanding during communication.	我会用简单的词汇重新问问题。 I will ask the question again using simpler vocabulary.
	Feedback Strategy (52%)	Implicit (Demo) (18%)	<i>CFLingo</i> gives examples and lets learners correct errors on their own.	为了使句子更完整、更清楚，可以这样改写。 To make the sentence more complete and clearer, it can be rewritten like this
		Implicit (Recast) (4%)	<i>CFLingo</i> reformulates part or all of a learner's utterance by replacing non-target language items with the correct form.	对的！白明下了飞机打算坐出租车到学校。 That's correct! After getting off the plane, Bai Ming plans to take a taxi to school.
		implicit (Hint) (53%)	<i>CFLingo</i> allows learners to try again by giving them linguistic hints.	你的句子结构很好！这里有一个小的改进建议。 Your sentence structure is excellent! Here's a small suggestion for improvement.
		Explicit (25%)	<i>CFLingo</i> clearly states or corrects the errors.	我们可以把两个从句用表示原因的连接词连接起来，使句子更流畅。

				We can connect the two clauses using a conjunction that indicates reason, making the sentence flow more smoothly.
	Equal Status (6%)		The interaction between <i>CFLingo</i> and learner shows equal roles.	没问题，我来帮助你理解。 No problem, I'm here to help you understand.
Affective Variable (39%)	Encouragement (63%) both linguistically also procedural	Compliment (30%)	<i>CFLingo</i> compliments learners based on their progress.	太好了！你造的这个句子很清楚，也很流畅。继续保持！ That's great! The sentence you created is very clear and flows well. Keep it up!
		Encourage to Proceed (70%)	<i>CFLingo</i> encourages or motivates learners to progress through the tasks.	希望这样解释更清楚，你可以尝试调整一下自己的句子。 I hope this explanation makes it clearer. You can try adjusting your sentence.
	Confirmation (16%)		<i>CFLingo</i> provides constructive feedback starting with what they did well at, aimed at reducing processing anxiety.	你的句子很清楚，也表达了意思。 Your sentence is clear and conveys the meaning well.
	Politeness (pragmatics) (21%)	Response to request (23%)	<i>CFLingo</i> is responsive and respects the learner's request.	当然可以。 Of course.
		Empathy (38%)	<i>CFLingo</i> shows empathy to learners' frustration like friends. Aimed to reduce stress and anxiety.	不要灰心！请试着这样安排... 再试一次。 Don't get discouraged! Please try to arrange it this way... Give it another shot.

		Make Request (23%)	<i>CFLingo</i> politely makes requests to learners after providing hints.	你可以尝试一下这个句型吗? Can you try using this sentence pattern?
		Apologize (15%)	<i>CFLingo</i> expresses apologies to learners during conversation.	对不起, 我解释地不够清楚。 I'm sorry; I didn't explain it clearly enough.

4.4 Evaluation Results

4.4.1 Students' Feedback on Task Sequencing Design of *CFLingo*

Students noted that the sequence of tasks is thoughtfully designed to gradually increase in complexity. Many participants appreciated how initial tasks focus on foundational skills, such as sentence generation, allowing them to build confidence before tackling more challenging activities. One student remarked, "Starting with simple tasks and moving to more complex ones really help me feel prepared... I can see my progress". The design of the task sequence also reinforces learning by revisiting previously covered material in new contexts. Students expressed that this repetition aids retention and understanding. One student stated, "I love how we come back to earlier topics in different tasks... it really helps solidify what I've learned". The results showed that a gradual increase in cognitive demands (Robinson, 2003b) supports learners in developing their skills systematically.

Moreover, students also highlighted the logical flow of tasks as a significant strength of the *CFLingo* platform. Many participants appreciated how each task builds on the previous one, creating a coherent learning experience. One student noted, "the way tasks are connected makes it easy to follow along... I always know what to expect next". This coherence helps students navigate their learning progression more effectively as they achieve higher-level language skills. An instructional implication is that when designing sequential tasks, in addition to increasing complexity, establishing clear connections between consecutive tasks can provide students with a smooth transition to new content by allowing them to easily integrate new material with their existing knowledge schema (Van Kesteren et al., 2014).

Additionally, students appreciated the variety of task types within the sequence, which kept the learning experience engaging. Many participants noted that alternating between different types of activities prevents monotony and maintains interest. One student remarked, "I enjoy how we switch between debates, role-plays, and writing tasks... it keeps things fresh". This variety contributes to sustained engagement and enthusiasm for learning.

In summary, students provided positive feedback on the sequential design of tasks embedded in *CFLingo*, which justified that the Cognitive Hypothesis can be applied in designing GAI-integrated language learning platforms and can yield positive learning experiences.

4.4.2 The Evaluation of the Adaptability Design of *CFLingo* via Students' Conversations and Feedback

4.4.2.1 The Evaluation of Adaptability via Conversation

Based on the analysis of students' conversations with the language AI platform *CFLingo*, several key themes emerged, highlighting the adaptability provided by *CFLingo*. These themes reveal how interactive and affective factors influenced learners' engagement and progress.

Interactive Factors

Interactive Factors were one of the significant themes, making up 61% of the coded conversations. Within this category, the sub-code Feedback Strategies was particularly prominent (see Figure 10), which comprised implicit and explicit forms of feedback. Notably, implicit feedback, particularly through Hints, had a significant impact, with 53% of instances recorded. For example, *CFLingo* remarked, "你已经很接近了! 不过这里的结构需要调整一下。" ("You're very close! However, the structure here needs some adjustments."). This strategy not only encourages learners but also invites them to engage in self-correction, promoting autonomy and critical thinking. In contrast, explicit feedback, as in the statement "好的, 让我们更细化地解释一下。" ("Sure! Let's break it down more clearly"), although less frequent (25%), serves to clarify specific errors and reinforce learning outcomes. This combination of feedback types illustrates that *CFLingo* can flexibly adjust its responses based on learners' real-time learning needs. This indicates the effectiveness of the engineering techniques applied in its design, highlighting the importance of both guided exploration and clear, direct instruction in language learning.

Further enhancing the interactive experience was the proficiency levels. The Input Hypothesis (Krashen, 1985) suggested "i+1" strategy for Comprehensive Input, which states that language learners acquire language when they understand messages that are slightly above their current level of competence, was evident in prompts like, "请再试一次, 用更复杂的句子。" ("Please try again, using a more complex sentence"). This approach encourages students to stretch their language abilities, promoting growth and understanding. Conversely, the "Minus One" strategy allowed students to encounter language input that was slightly below the designed HSK3 level, as seen in the interaction, "顺畅"意思是"流利"、"没有阻碍的", 英文是 "smooth" 或 "fluent" ("The meaning of '顺畅' is 'fluent' or 'unobstructed'; in English, it is 'smooth' or 'fluent'). This method provides necessary scaffolding, enabling learners to build confidence while grappling with new concepts. The balance between these strategies suggests that *CFLingo* was able to provide differentiated input, which is vital for accommodating diverse learner proficiency levels.

The findings suggest several implications for instructional design. First, integrating Plus One and Minus One strategies into language learning environments can create a more

adaptable experience for students. Educators can prompt GAI tools to adjust to varying proficiency levels, allowing learners to engage with content that is both challenging and accessible. This flexibility ensures that learners can progress at their own pace while still feeling challenged. Additionally, the diverse feedback strategies observed underscore the need for training GAI in providing effective feedback. The frequent use of implicit feedback, such as hints and encouragement, alongside explicit corrections, can foster a dynamic learning atmosphere. As evidenced by the *CFLingo*'s use of feedback, "你的句子更好了！但可以更简单些。试试这样" ("Your sentence is better! But it can be simpler. Try this."), designers should aim to balance supportive guidance with explicit instruction. Designers can input more pre-designed example sentences into the prompt to guide GAI in cultivating an environment that encourages student autonomy and risk-taking in language use.

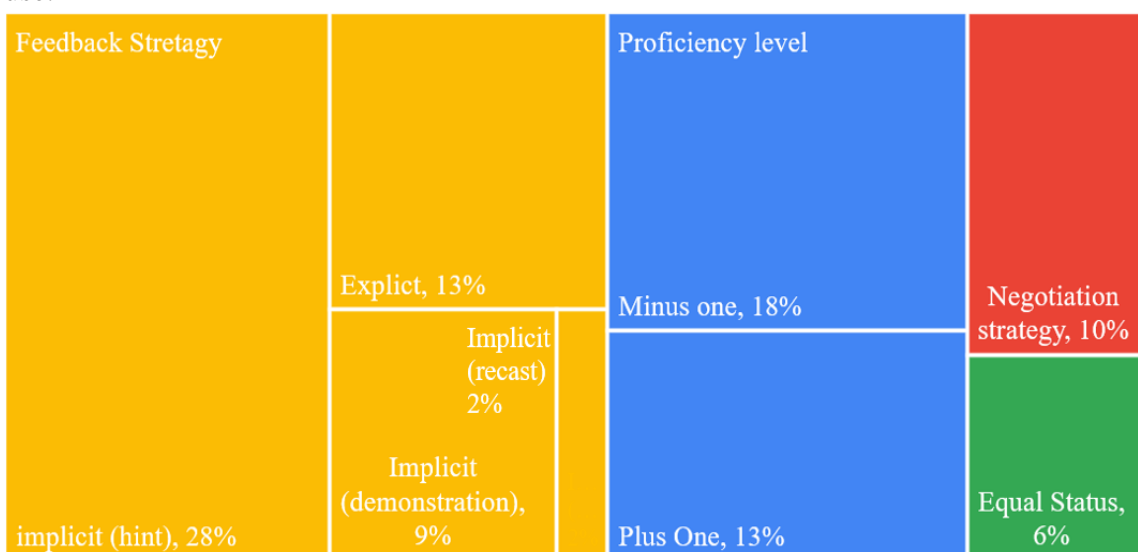


Figure 10 The Proportion of Categories under Interactive Factors

Affective Variables

The analysis also highlighted Affective Variables (see Figure 11), with Encouragement emerging as a vital factor, comprising 63% of interactions. Positive reinforcement, such as in the quote, "非常好！说得很完整！" ("Very good! You expressed it very thoroughly!"), illustrates how acknowledgment of effort can significantly enhance motivation and persistence among learners. Encouraging students to proceed, as seen in the *CFLingo*'s prompt, "接近了！但主述宾的结构不完全正确，可以稍微调整一下，请再试一次" ("You're getting close! However, the subject-verb-object structure is not entirely correct; it can be adjusted slightly. Please try again"), fosters a sense of support and empowerment. Empathy was another critical component, as the AI displayed understanding of students' frustrations, responding with phrases like, "没关系，我们一起来看看一下" ("No worries, let's take a look together."). Such empathetic engagement not only reduces anxiety but also promotes a more inviting learning environment. Educators and

designers can consider integrating empathetic responses into their instructional methods to ensure students feel valued and supported throughout the learning process.

These findings suggest that designers and instructors should prioritize the integration of affective variables in educational tools and environments. Designers of AI platforms like *CFLingo* can enhance learner engagement by incorporating features that provide regular positive feedback and encouragement, similar to the examples observed in the study. Additionally, providing options for empathetic responses can help address learners' emotional needs, creating a more supportive atmosphere. By fostering a supportive and positive environment, student engagement could be enhanced, making language learning a more enjoyable and effective experience. The following sections synthesize the specific features that enabled *CFLingo* to deliver adaptive feedback to learners.

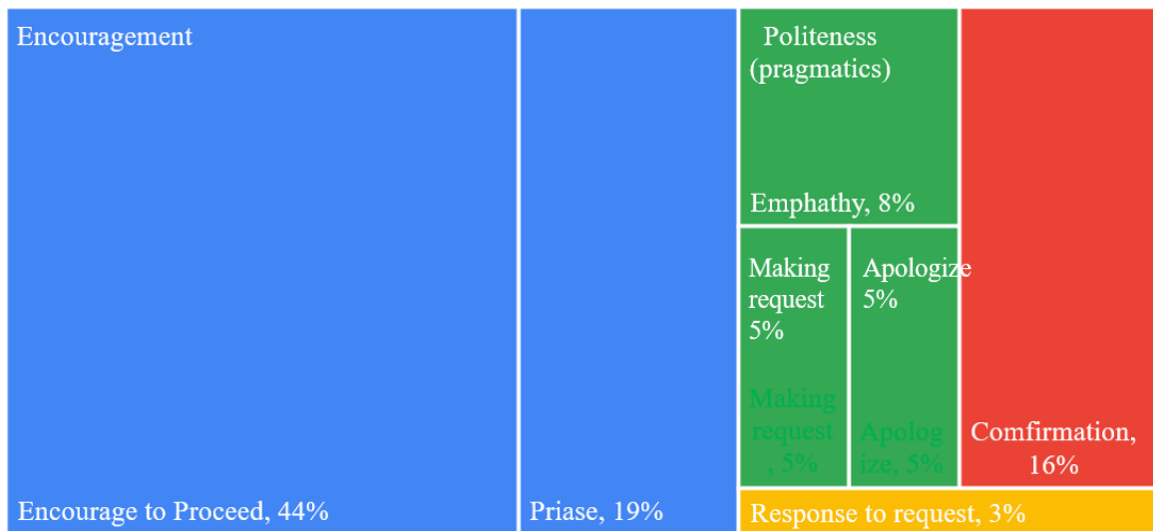


Figure 11 The Proportion of Categories under Affective Variables

4.4.2.2 Students Feedback on Adaptability

A prominent theme in adaptability was *CFLingo*'s ability to personalize learning experiences according to individual students' learning paths. Participants acknowledged that *CFLingo* offered alternative ways to articulate their thoughts while encouraging critical engagement with the content. This interaction prompted students to reflect on their writing processes and explore avenues for improvement in their language skills. The friendly and supportive tone of the feedback contributed to a positive learning atmosphere, fostering a greater sense of comfort throughout their educational journeys. In addition to immediate feedback, students valued the constructive criticism offered by *CFLingo*. Many participants mentioned that the platform not only points out errors but also provides suggestions for improvement. One student shared, "When I make a mistake, *CFLingo* doesn't just tell me I'm wrong; it shows me how to fix it, which is really helpful".

In addition, students expressed that *CFLingo* effectively adapts to their vocabulary learning needs. Many noted that the platform provides personalized vocabulary exercises

that align with their vocabulary levels. For instance, as students indicated, when they respond to *CFLingo* in Chinese, “it will evaluate the proficiency of my vocabulary”. This adaptability allows students to engage with vocabulary that is relevant and challenging for them. Another significant theme was *CFLingo*'s ability to assist with sentence comprehension. Students appreciated the platform's feedback on their sentence structures and grammar. One participant stated, “It knows what it's talking about, and it does give good feedback... I see it as a partner”. This indicates that students view *CFLingo* as a supportive companion that enhances their understanding of complex sentence structures based on their needs.

More importantly, the adaptability of task difficulty appeared to be a recurring theme in student feedback. Many students appreciated that *CFLingo* adjusts the complexity of tasks based on their performance, allowing for a personalized learning experience. One participant remarked, “I like that the tasks get harder as I improve... it keeps me challenged without being overwhelmed”. This adaptive approach ensures that learners are consistently engaged and motivated to progress in their language skills. Another aspect of *CFLingo*'s adaptability is its ability to create tailored learning paths for students. Many participants noted that the platform adjusts the difficulty of tasks based on their performance, allowing for a more customized learning experience. One student shared, “*CFLingo* can assess my progress and then suggest new topics or vocabulary that I should focus on next”. This feature not only helps students stay engaged but also ensures that they are consistently challenged at an appropriate level, facilitating steady progress in their language skills.

In summary, findings from student responses and interviews revealed that *CFLingo*'s adaptability was perceived as a key factor in enhancing their learning experiences, mainly by providing personalized learning paths. These results support the applicability of the Triadic Componential Framework (TCF) in designing adaptability for GAI-integrated language learning platforms, demonstrating its potential to yield positive learning outcomes. As Robinson (2007a) observed, learners may achieve better performance in specific learning contexts, such as a Task-Based Language Teaching (TBLT) approach, when these contexts align with their cognitive strengths. This suggests that *CFLingo* effectively leverages TCF to create an adaptive, human-centered, problem-solving environment, fostering improved L2 production and overall language performance.

5. Discussion and Implications

As shown in Table 1, *CFLingo*'s task design introduces several innovative attributes compared to traditional TBLT task design features (Candlin, 1987). Notably, *CFLingo* leverages adaptability, implicit feedback, and enhanced interactional demands to create an immersive, human-centered learning environment. Furthermore, the platform promotes teacher-AI collaboration, seamlessly integrating technology with pedagogical goals to enhance student proficiency and fluency in L2 learning. These aspects will be discussed in the following sections.

5.1 Individualized Adaptability With a Focus on Proficiency Level

In pre-designed prompts for each pedagogic task, researchers incorporate a step called “Adaptation.” Initially, *CFLingo* was given an HSK 1-3 vocabulary list and instructed to tailor its language use accordingly. Adhering to the “i+1” principle of the Input Hypothesis (Krashen, 1985), researchers prompt the *CFLingo* to present learners with language input slightly above their current level. During the “feedback or follow-up” sections, *CFLingo* is programmed to adjust its proficiency level in response to students’ signs of frustration or requests for modifications, utilizing simpler grammatical structures or rephrasing input based on the chapter’s vocabulary bank. This allows the *CFLingo*’s proficiency to dip below HSK 3, which is referred to as “i-1.” This adaptability is not limited to the “Adaptation” step; it is also evident in the “task setting” and “task follow-up” phases, where *CFLingo* provides individualized support and ensures continuity.

The data indicates that adaptability is highly rated across interactive task conditions. This feature includes dynamically adjusting proficiency levels during personalized dialogues, providing varied and constructive feedback, and facilitating negotiation. On the affective side, *CFLingo* offers both procedural and linguistic encouragement, emphasizes pragmatic language use, and motivates learners to engage with the material, revise their work, and enhance their L2 production.

5.2 Implicit Feedback: Elevating Task Performance

Constructive feedback on *CFLingo* includes both implicit and explicit forms, with data revealing (see Figure 8) that the majority of feedback provided is implicit. This feedback typically consists of examples (with measures in place to prevent learners from copying and pasting, ensuring genuine learning), hints that explain grammatical errors in Chinese without directly supplying the correct answer, recasts, and encouragement for students to try again. Research indicates that implicit feedback is particularly beneficial for learners, as highlighted by Ellis (1994). This type of feedback promotes deeper cognitive processing and helps learners cultivate an intuitive understanding of the language.

5.3 Interactional Demands Enhanced Throughout the Design

In TCF, task conditions significantly influence task performance. Within the *CFLinguo* platform, cognitive demands are carefully graded and sequenced. However, the interactional demands of pedagogic tasks are not explicitly graded and sequenced (Robinson, 2006). Instead, these demands are enhanced throughout *CFLingo* design. Most of the tasks—such as sentence generation, role play, and debates—are open-ended discussions that promote a two-way flow of communication and divergent solutions.

Robinson (2006, p. 22) emphasizes that “holding task conditions constant is important to ensuring transfer of training to real-world contexts. The more task conditions are practiced in pedagogic versions, the more elaborate and consolidated the scripts become for real-world performance, which successful transfer will draw upon outside the classroom.” Tasks that allow open solutions encourage creativity and critical thinking, while a two-way flow of communication fosters interaction between learners and the AI or

among peers, leading to richer dialogues. Additionally, tasks that promote divergent solutions support learners in developing various strategies for language use and problem-solving. *CFLingo* provides an improved context and environment for task performance, leading to enhanced outcomes in L2 production.

5.4 Immersive Authentic Chinese Environment Promotes Fluency

In the *CFLingo* platform, both AI and learners are required to communicate in the target language, enhancing both situational and interactional authenticity (ELT Concourse, n.d.). This approach also promotes learners' aptitude profiles by encouraging practical solutions and strategic language use during interactions. Often, it can be challenging for learners to engage in conversations within a target language environment during traditional pair work due to varying levels of ability or affective factors.

It is exciting to witness this seamless language flow occurring in *CFLingo* interactions, where the AI consistently provides examples and hints in the target language first, creating an experience similar to communicating with a native Chinese speaker. A thorough analysis of accuracy and complexity necessitates additional data; however, a significant improvement in writing fluency has been observed, as measured by words typed per minute and character recognition. Additional empirical data will be analyzed in the next stage of the research.

5.5 Teacher-AI Collaboration in Providing Instructional Feedback

The designer or instructor is not only able to monitor learners' progress and learning behaviors—such as whether they have completed all the questions, if they are using the target language, and whether they are applying any prompt engineering strategies—but the stored data also provides a valuable opportunity for teachers to offer follow-up feedback. As shown in Figure 12, this capability enables teachers to provide collaborative feedback during the after-task session, assisting students in revising and refining their final target tasks, such as creating descriptive or argumentative essays.

This design feature aligns with Yan's (2024) strong call for a “human-centered perspective towards the application of AI in education,” which should be encouraged and upheld in research regarding the impact of AI on education (Yan, 2024, p. 14). The design and instructional implication is that when leveraging prompt engineering through the OpenAI API to develop a platform, designers must ensure that it is structured for a CFL classroom that prioritizes a human-centered and user-friendly approach, which will significantly influence the learning experience and outcomes.

Sample Exercise Interface of CFLingo 智语学伴—Essay Critique 作文点评

Narrative Essays:outline (记叙文)

Argumentative Essays:draft (议论文)

CFLingo点评: 这个大纲很不错, 有开头、结尾和中间部分。你的主题和观点都很鲜明 (clearly)。然而, 一些部分可以更具体。比如: 1. 开头可以详细描述 (detailed explanation) 你亲人对你的生活的重要性。可以加一句, 例如: “他们在我生活中有非常重要的地位 (position)。” 2. 中间部分, 每个段落都可以增加一些细节。例如: - 家庭: 可以举例说说与每位亲人的具体回忆。例如: “和姥姥姥爷一起过年的时候, 他们总是做很可口的饭菜。” - 女朋友和朋友: 可以加入你们的一个特别的活动或者趣事: “第一次见面是在图书馆, 她帮我找到我要的书。” - 宠物: 可以描述你们平时一起做的事情, 例如: “我常常和猫一起在家看电视, 她总是喜欢钻到我的怀里。” 3. 结尾: 可以再总结一下亲人与自我成长的关系 (relationship with personal growth), 例如: “通过与他们相处, 我学会了…… (比如感恩 (gratitude))。” 请再试着修改一下这个大纲, 然后再分享一遍可以更丰富 (enrich) 你的描述。

单老师点评:

同意AI给出的大部分建议, 比如中间部分增加细节, 结尾的主题升华。请注意**标题和主题**的一致性, 比如是我爱的人, 还是我的家人? 结构上值得注意的是: 开头部分**引出话题**, 中间部分你要**介绍四类爱的人**, 要注意详略得当, **重点突出**, **故事之间的衔接**, 注意个别字词使用。推荐词汇: **不是亲人, 胜似亲人**。期待你的第一稿。

CFLingo 第一稿点评:

你的稿件给出了一个很好的开头、中间和结尾的结构, 围绕着一个明确的论点进行展开。以下是一些建议: 1. “论点的准确性”: 你明确指出了自己的观点, 即**父母不应过分介入子女的感情生活**, “门当户对”**对于婚姻和爱情很重要, 同时拥有相似但不完全一样的爱好对维持关系有好处**, **这三个观点清晰、论点鲜明**。2. “整体结构”: 文章的整体结构 (总-分-总) 设计得相对合理, 但在过渡句子上可以做得更好一些。尤其是从一个主要论点转移到下一个论点时, 可以使用一些过渡性的表达, 如**“接下来, 我们来谈谈……”**。3. “关键词使用”: 建议使用更多的关键词来增强文章的连贯性。例如, 在介绍新论点时可以使用**“首先, 其次, 最后”**等。4. “论据是否充分”: 你提供了一些生动的例子来支持自己的观点, 这是非常好的。但对于每个观点, 如果能提供更多**具体的例子或数据支持**, 会使论证更加有力。5. “论证是否合理”: 你提出的论证大体上是合理的, 但在“门当户对”这部分, 可能需要**更详细地解释为什么相似的经济和文化背景能使关系更稳定**。6. “用词选择”: 用词基本恰当, 但注意, 为了使文章更加书面化, 应尽量避免使用口语表达, 如“老公”可以换为“丈夫”, “共同朋友合不来”可以修改为“与共同朋友相处不融洽”。7. “语法结构”: 整体语法结构使用较为恰当, 但需要注意时态的一致性和主谓一致。8. “语体”: 文章的语体整体偏向书面语, 这对于论文来说是合适的。但还可以进一步规范, 尽量使用准确、正式的词汇。总的来说, 你已经完成了一篇内容丰富、结构清晰的作文。在修改时, 注意以上提到的细节, 并尝试加入更多支持你观点的证据, 使论证更加充分, 期待你的改进稿。

单老师点评: 文章可以很自然地提出问题 bring up an issue, 你试图引出论点, **AI的点评部分我并不是完全同意 (看红色强调部分)**, 你的论点可以分为两部分, 而不是“三个观点”, 但是你的确可以更加明确地阐述。分析问题 analyze 部分层次合理, 关键词使用得体。我同意AI建议, 可以增加一些具体的例子或者参考数据 (citations), 比如为什么“门当户对”更好? 解决问题 solve 总结完整, 很好地总结了你论点的两部分。其次, 针对议论文用语, **如AI建议的**, 可以增加**书面语的使用**。整体来说, 这一稿已经是一个很好的开端, 期待你进一步改进!

Figure 12 Teacher-AI collaborative feedback through Stored Data on CFLingo

6. Future Direction

The current *CFLingo* platform, guided by the Cognitive Hypothesis and Triadic Componential Framework, has paved several avenues for future research and development. First, as students interact with *CFLingo*, substantial conversational data are generated, providing researchers with valuable insights into students' conversational experiences. The next step for researchers is to associate these experiences with student characteristics, such as their perspectives, and to further evaluate how syllabus and task designs influence learners' writing fluency, accuracy, and complexity. Specifically, analyzing how students respond to *CFLingo*'s tasks and instructions—using methods such as sentiment analysis or discourse analysis—can offer significant insights into their interactions with GAI tools in language learning.

Secondly, with advancements in large language models, the researchers aim to incorporate voice input features into *CFLingo*. While the current version enables students to practice vocabulary and writing skills, speaking is fundamental to language acquisition. Integrating a voice input feature would enhance *CFLingo*'s capacity to support comprehensive language proficiency.

Finally, researchers propose creating an immersive learning environment in the target language that aligns with structured Second Language Acquisition frameworks. Moving forward, the environment will be refined by incorporating interactive elements and considering learner-specific factors. The impact on learner performance will be evaluated using empirical data collected through *CFLingo*, providing insights into the effectiveness of immersive learning designs for language acquisition outcomes.

7. Conclusion

This research introduces a pioneering platform that integrates GAI tools within task-based language teaching (TBLT) syllabus design—the *CFLingo* platform. By applying the TBLT framework's cognitive hypothesis and the Triadic Componential Framework, the design demonstrates how GAI can optimize task adaptability for individual learners, thereby enhancing the task condition: interactional factors. While the qualitative empirical data gathered showcases the adaptability of generative AI, it is important to note that the sample size is relatively small. Nonetheless, the feedback collected indicates that students have a positive perception of *CFLingo*, highlighting the benefits they experience from the progressive complexity of task sequences, and the adaptive feedback provided by the platform. As students engaged with various tasks in *CFLingo*, they acknowledged the impact of affective factors on task design, noting how these elements contributed to their Chinese learning experience, particularly the perceived difficulty level. Overall, this study suggests a promising future for GAI-integrated TBLT in further enhancing learners' accuracy, fluency, and complexity in second language acquisition.

Acknowledgment: We extend our deepest gratitude to Lehigh University's Center for Innovation in Teaching and Learning (CITL) for their generous funding and exceptional technical support, which were instrumental to the success of this project.

References

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Baleghizadeh, S. (2015). The procedural syllabus and the task syllabus: How similar, how different? *HOW*, 22(2), 104-113. <https://doi.org/10.19183/how.22.2.153>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, 2022, 12(17), 8427. <https://doi.org/10.3390/app12178427>
- Bhutoria, A. (2022). Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop model. *Computers and Education: Artificial Intelligence*, 3, 100068. <https://doi.org/10.1016/j.caeai.2022.100068>
- Candlin, C. N. (1987). Language learning tasks. In D. Murphy & C. N. Candlin (Eds.), *Language learning tasks* (pp. 5–22). Prentice Hall.
- Cai, W. (2024). *Using AI tools for Chinese language teaching and research* (Workshop). The 2024 Annual Conference of the Chinese Language Teachers Association, USA, St Louis, MO, USA.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- East, M. (2021). *Foundational principles of task-based language teaching*. Routledge.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.
- Ellis, R. (2003). *Task-based language learning and teaching* (1st ed.). Oxford University Press.

- Han, Z. (2007). *INFO: optimizing conditions for instructed learning of Chinese as a second/foreign language [invited Talk]*. The 2nd International Forum on the Teaching and Learning of Chinese, Guangzhou, China.
- Han, Z. (2018, September). Task-based learning in task-based teaching: Training teachers of Chinese as a foreign language. *Annual Review of Applied Linguistics*, 38, 162–186.
- Han, Z. (2024). Chatgpt in and for second language acquisition: A call for systematic research. *Studies in Second Language Acquisition*, 46, 301–306.
doi:10.1017/S0272263124000111
- Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1), 37-45. <https://doi.org/10.61414/jeti.v5i1.103>
- Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, 6, 100225. <https://doi.org/10.1016/j.caeai.2024.100225>.
- Krashen, S. D. (1985). *The input hypothesis: issues and implications*. Longman.
- 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, Y. (2024). Usability of ChatGPT in second language acquisition: Capabilities, effectiveness, applications, challenges, and solutions. *Studies in Applied Linguistics & TESOL at Teachers College, Columbia University*, 24(1), 24–37.
- Long, M. (2014.). *Second language acquisition and task-based language teaching*. John Wiley & Sons, Incorporated.
- Nunan, D. (2010.). *Task-based language teaching*. Cambridge University Press.
- Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communicative tasks for second language instruction. In G. Crookes, & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice*, (pp. 9-34). Multilingual Matters.
- Robinson, P. (2001b). Criteria for classifying and sequencing pedagogic tasks. *Applied Linguistics*, 22(1), 27-57.
- Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge University Press.
- Robinson, P. (2003a). Attention and Memory during SLA. In *Handbook of Second Language Acquisition* (pp. 631-678). Blackwell.
- Robinson, P. (2003 b). The cognition hypothesis, task design, and adult task-based language learning. *Studies in Second Language Acquisition*, 21(2), 45-105.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, 43(1), 1-32.
<https://doi.org/10.1515/iral.2005.43.1.1>
- Robinson, P. (2007a). Aptitudes, abilities, contexts and practice. In R. M. DeKeyser (Ed.), *Practice in second language learning: Perspectives from applied linguistics and cognitive psychology* (pp. 256–286). Cambridge University Press.

- Robinson, P. (2006). Criteria for Classifying and Sequencing Pedagogic Tasks. In *Investigating Tasks in Formal Language Learning* (pp. 7-27). Bristol, Blue Ridge Summit: Multilingual Matters. <https://doi.org/10.21832/9781853599286-004>
- Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, 61(s1), 1 - 36. DOI:10.1111/j.1467-9922.2011.00641.x
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford Applied Linguistics.
- Skehan, P. (2023). Limited attentional capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference*. (pp. 123-155). John Benjamins.
- Skehan, P. (1998). *A cognitive approach to language learning*. (n.d.). Oxford University Press.
- task-based learning*. (n.d.). ELT Concourse. Retrieved November 5, 2024, from <https://www.eltoncourse.com/training/in-service/methodology/tbl.html>
- van Dis, E. A. M., Bollen, J., Zuidema, W., Rooij, R. v., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature*. DOI: 10.1038/d41586-023-00288-7
- Velásquez-Henao, J. D., Franco-Cardona, C. J., & Cadavid-Higuita, L. (2023). Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering. *Dyna*, 90(230), 9-17.
- Xiao, F., Zhao, P., Sha, H., Yang, D., & Warschauer, M. (2024). Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning*, 4(2), 300-325. DOI:10.1515/jccall-2022-0032
- Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), 212. <https://doi.org/10.3390/languages8030212>
- Yan, D. (2024). Collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology*, 28(1), (1), 1–19.
- Yang, S. J., Ogata, H., Matsui, T., & Chen, N. S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, 2, 100008. Page number?
- Young, J. J., & Shishido, M. (2023). Investigating OpenAI's ChatGPT potentials in generating chatbot's dialogue for English as a foreign language learning. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(6), 65-72. <https://thesai.org/Publications/ViewPaper?Volume=14&Issue=6&Code=IJACSA&SerialNo=7>

Appendix 1

The Triadic Componential Framework for Task Classification – Categories, Criteria, Analytic Procedures, and Design Characteristics

Task complexity (cognitive factors) (Classification criteria: cognitive demands) (Classification procedure: information-theoretic analyses)	Task condition (interactive factors) (Classification criteria: interactional demands) (Classification procedure: behavior-descriptive analyses)	Task difficulty (learner factors) (Classification criteria: ability requirements) (Classification procedure: ability assessment analyses)
(a) Resource-directing variables making cognitive/conceptual demands ± Here and now ± Few elements ± Spatial reasoning ± Causal reasoning ± Intentional reasoning ± Perspective-taking	(a) Participation variables making interactional demands ± Open solution ± One-way flow ± Convergent solution ± Few participants ± Few contributions needed ± Negotiation not needed	(a) Ability variables and task-relevant resource differentials h/l Working memory h/l Reasoning h/l Task-switching h/l Aptitude h/l Field independence h/l Mind/intention-reading
(b) Resource-dispersing variables making performative/procedural demands ± Planning time ± Single task ± Task structure ± Few steps ± Independency of steps ± Prior knowledge	(b) Participant variables making interactant demands ± Same proficiency ± Same gender ± Familiar ± Shared content knowledge ± Equal status and role ± Shared cultural knowledge	(b) affective variables and task-relevant state-trait differentials h/l Openness to experience h/l Control of emotion h/l Task motivation h/l Processing anxiety h/l Willingness to communicate h/l Self-efficacy

Adapted from Robinson 2007b, by permission of Multilingual Matters. Reproduced with permission of the publisher, Multilingual Matters.