2024 (20th) ChinaCALL Conference & International Congress on English Language Education and Applied Linguistics

2024（第20届）语言智能教学国际会议
暨2024英语教育及应用语言学国际大会

August 23-25, 2024   Beijing, China

Conference Proceedings
(short papers)

本电子论文集收录了投稿给"2024（第 20 届）语言智能教学国际会议暨 2024 英语教育及应用语言学国际大会"的小论文。作者对其文章内容等具有全部解释权。所述观点不代表 ChinaCALL 观点。

会议论文引用格式：

Author(s) Name(s). (2024). The title of paper. In Proceedings of the 2024 (20th) ChinaCALL Conference & International Congress on English Language Education and Applied Linguistics. Retrieved from https://www.chinacall.org.cn/conference2024/index.html

Contents

CALL materials & CALL environment 技术赋能学习资源和环境建构研究

English Language Education 英语教育

Applied Linguistics 应用语言学

Translation Studies 翻译研究

CALL research

计算机辅助外语教学理论研究

Investigating Students' Perceptions of Interpersonal Interaction Strategies in Online EAP Courses: Effectiveness and Challenges

Xiaohui Guo

Xi'an Jiaotong-Liverpool University

Xiaohui.guo@xjtlu.edu.cn

Abstract

While CMC (computer-mediated communication) facilitates communication, collaboration, and reflection through diverse online platforms such as videoconferencing and forums, it also raises concerns among stakeholders. Educators are actively exploring online pedagogical methods to optimize their teaching effectiveness in this digital landscape. However, the transition to new teaching techniques sometimes falls short of expectations. A predominant issue revolves around student interaction in virtual learning environments, where instructors strive to boost participation through a range of strategies. To assess the efficacy of these strategies and the overall success of online instruction, gathering student feedback becomes indispensable. This action research aims to explore students' perceptions of the effectiveness of interpersonal strategies in online English for Academic Purposes (EAP) courses within the setting of a transnational university, drawing upon Moore's interaction framework. Employing a mixed-methods approach, the research consists of an initial quantitative phase utilizing a validated 20-item Likert scale survey, followed by qualitative data collection through semi-structured focus group interviews. Icebreaker discussion and autonomy in selecting readings that drive discussion group formation were rated as the most valued leaner-learner engagement strategies, while due date checklists and regular announcements or email reminders were considered the most valued leaner-instructor engagement strategies. The underlying reasons for students' perceptions and the challenges they faced in interacting with peers and instructors were also identified. The results of this study have implications for instructors, course designers, and policymakers in enhancing student engagement and teaching quality, ensuring instructional excellence in online courses.

Keywords: online learning, interpersonal interaction strategies, student perception

Introduction

With the growing embrace of computer-assisted language learning and the accelerated transition to online education catalyzed by the COVID-19 pandemic, language learning activities have progressively migrated to virtual learning environments, where interactions can be enriched through both asynchronous and synchronous CMC (computer-mediated communication). CMC facilitates distance education by using network technology, allowing communication between learners and instructors as well as among learners themselves (Wang, 2004). Compared to telephone conversations, computer-mediated human-human interaction has become a major phenomenon at the beginning of the 21st century. This interaction can occur in oral, visual, or written forms, or

a combination of these, involving exchanges between learners and instructors, among learners themselves, and even between native speakers in target language countries and learners in their home countries (Garrison, 2016). The fourth generation of distance education, which promotes immediate communication and interaction, offers an exceptional degree of student engagement and provides numerous opportunities for meaningful language learning (Wang & Sun, 2001). This is particularly important because interaction is central to language acquisition, given that language serves as a means of communication and interaction. Therefore, interaction is a fundamental component of the communicative language learning process (Hall, 2010; Kitade, 2000).

Interpersonal interaction, consistent with the principles of social constructivism (Vygotsky, 1997) and the social presence component of the Community of Inquiry framework (Garrison, 2016), is crucial for engaging students(Lai et al., 2019; Wang et al., 2022) and enhancing their satisfaction with e-learning (Fedynich et al., 2015; Kuo, Walker, et al., 2014). However, faculty may not be aware of strategies to encourage students to interact (Paquette, 2016). Moreover, since students' perception of interaction was a better predictor of course satisfaction than their actual measured interaction (Fulford & Zhang, 1993), it is significant to research what personal interaction strategies can best promote effective student learning perceived by students. Additionally, little research has been conducted on language teaching and learning in the Chinese context. Therefore, this research aims to evaluate the effectiveness of interpersonal interaction strategies perceived by Chinese language learners and the challenges regarding online interpersonal interaction, hoping to provide implications for instructors, course designers, and policymakers in enhancing student interpersonal interaction and teaching quality.

Literature Review

Interpersonal interaction in online education

Interaction in online education has been extensively studied. The most well-known theoretical framework is Moore's (1989) three types of interaction in distance education: learner-content interaction, learner-instructor interaction, and learner-learner interaction. The model was further expanded to include additional interaction types, such as instructor-instructor interaction and learner-interface interaction (Anderson, 2003). Among these different types of interaction, learner-learner and learner-instructor interactions can be considered interpersonal interactions (York & Richardson, 2012), which appear to be the most influential and essential components in online education (Swan et al., 2009).

Learner-learner interaction is communication between individual students or within small groups, while learner-instructor interaction refers to the two-way communication between students and instructors (Moore, 1989). Learner-learner interaction includes exchanges, such as sharing ideas with classmates, engaging in discussions, and receiving feedback from peers. Learner-instructor interaction involves asking questions,

providing support and giving encouragement to learners. Both of the two interaction types may be synchronous through videoconferencing and chats or asynchronous through e-mail and discussion boards (Bernard et al., 2009).

Significance of interpersonal interaction in online education

Interaction between students and teachers and between students and students can help to create a positive learning environment (Gilbert & Moore, 1998). This is consistent with Muilenburg & Berge's (2005) findings that lack of interpersonal interaction is the most significant barrier to online learning perceived by students. The reason for this is "When students have the opportunity to interact with one another and their instructors about the content, they have the opportunity to build within themselves, and to communicate, a shared meaning to 'make sense' of what they are learning" (Berge, 1999, P.8).

Interpersonal interaction can enhance student satisfaction by reducing transactional distance, thereby offering a more effective educational experience for learners (Moore, 1989). This is supported by a three-year study on graduate and undergraduate students' level of satisfaction with online instruction (Cole et al., 2014). The study also indicates that the lack of communication between students and the instructor is the main source of student dissatisfaction.

To achieve high course satisfaction, it is essential to recognize that simply providing students with appropriate tools for interacting with peers and faculty does not automatically guarantee enhanced interpersonal interaction. Quality instructional and interpersonal interaction opportunities need to be deliberately designed into the course in online learning environments (Berge, 1999; Bernard et al., 2009). As stated by Roblyer & Wiencke (2003, P. 87), "Highly interactive learning environments are rarely serendipitous; activities must be designed to encourage, support, and even require interaction". Therefore, the success of online courses positively relates to the quantity and quality of interpersonal interactions.

Perceptions of interpersonal interaction strategies in online education

Student's perception of interaction strategies in online education has been widely studied. Northrup et al. (2002) emphasized the importance of interpersonal interaction in online learning, particularly highlighting the high value placed on timely responses from both peers and instructors. Fedynich et al. (2015) investigated graduate students' perceptions of online learning by surveying 249 students. The study found that interaction, both among students and between students and the instructor, significantly impacts student satisfaction. Moreover, the instructor's role was identified as vitally important to student satisfaction. This is supported by Martin & Bolliger (2018) who conducted a study with 155 university students in the United States and found that learner-instructor interaction strategies were the most highly valued perceived by students.

A range of studies have been conducted to explore instructors' perceptions of interaction strategies in online education. For instance, York & Richardson (2012) examined the views of six experienced online instructors through interviews to identify the factors they believe enhance interaction among students and between students and instructors in online courses. The findings highlight factors such as group work, community, and instructor participation, underscoring the importance of interpersonal interaction. In contrast, Kuo, Belland, et al., (2014) researched K-12 teachers' perceptions of interaction type in blended learning environments and found that learner-content interaction is considered more important than interpersonal interactions. Another study by Bolliger & Martin (2018) investigated instructor and student perception of interaction strategies and found that instructors rate these strategies as more important compared to students.

Research questions

- How effective online interpersonal interaction strategies do students perceive? Why?

- What challenges regarding online interpersonal interaction do students report?

Methodology

Participant

This research involves 53 Chinese students enrolled in online English for Academic Purposes courses at an English as a Medium of Instruction(EMI) university in China. Their English language proficiency ranges from CEFR B1+ to B2. The courses were delivered through Moodle, while BigBlueButton served as the virtual classroom software, offering features such as multi-user whiteboards, breakout rooms, and polling.

Instrument

The study employs a mixed-methods approach, integrating both quantitative and qualitative research methods to provide a comprehensive understanding of the research questions. The first phase involves gathering quantitative data through a validated survey adapted from Martin & Bolliger (2018). The new survey focuses on interpersonal interaction strategies and is tailored to the teaching context of the research. The survey comprises 20 Likert-scale items, with responses ranging from 1 (very unimportant) to 5 (very important), to evaluate students' perceptions of the effectiveness of various interaction strategies. 10 items focus on learner-learner interaction strategies, while the remaining 10 address learner-instructor interaction strategies. A bilingual survey was utilized to minimize the potential for misunderstandings of specific items. Cronbach's alpha coefficients were computed to verify the internal reliability of the survey. The overall internal reliability coefficient for the instrument was 0.92, and the reliability for the two subscales was also satisfactory: a) learner-to-learner interactions ($\alpha = 0.83$), and b) learner-to-instructor interactions ($\alpha = 0.92$).

The semi-structured interviews were utilized to gain a deeper understanding of interpersonal interaction strategies and to explore reasons behind participants' perceptions, and challenges they face regarding online interpersonal interaction. Five participants were interviewed using three prepared questions, along with follow-up questions, to elicit comprehensive responses.

Data collection and analysis

Ethical approval for the research was obtained prior to data collection. The online survey was distributed to potential respondents via a URL sent through email. Participants provided their informed consent by completing the survey on the Wenjuanxin platform. To recruit interviewees, invitations were issued through the final question of the survey. Quantitative data were analyzed using SPSS, where descriptive statistics, including means and standard deviations, were computed for each Likert-scale item to identify overall trends in students' perceptions of interaction strategies. Qualitative data from the semi-structured interviews were analyzed through thematic analysis. Interview transcripts were transcribed and coded, with initial codes developed inductively and subsequently organized into themes to capture the reasons and challenges related to interaction strategies.

Results

In terms of leaner-leaner interaction strategies (Table 1), the overall mean scores range from 3.36 to 3.91, indicating that most strategies are viewed somewhat positively. Allowing students to choose readings that drive discussion group formation is highly valued (item 5, M=3.91), with icebreaker discussion as the second (item 3, M=3.85). Peer review (item 9, M=3.38) and virtual lounge (item 1, M=3.36) are the least valued learner-learner interaction strategies.

Table 1

Leaner-learner Interaction Strategies	Mean	SD
1. Students use a virtual lounge where they can meet informally to share common interests.	3.36	1.18
2. Students share their profiles on Padlet before the course starts.	3.45	0.89
3. Students introduce themselves using an icebreaker discussion.	3.85	0.86
4. Students moderate discussions (e.g. assigning a group leader).	3.72	1.07
5. Students have choices in the selection of readings (articles, books) that drive discussion group formation.	3.91	1.06
6. Students post audio and/or video files in threaded discussions on LMO forums instead of only written responses.	3.83	1.17

	Mean	SD
7. Students interact with peers through student presentations (asynchronously or synchronously).	3.68	0.83
8. Students work collaboratively using online communication tools to complete case studies, projects, reports, etc.	3.8	1.04
9. Students peer-review classmates' work.	3.38	0.97
10. Students are required to rate individual performance of team members on projects.	3.6	1.22

By evaluating learner-instructor interaction strategies (Table 2), it has been found that the instructor posting a 'due date checklist' (item 15) received the highest mean score of 4.49, indicating it is the most valued interaction strategy among students. This is followed by item 12 (M=4.34) of sending regular announcements or email reminders. The instructor referring to students by name in discussion forums (item 11) has the lowest mean score of 3.64, suggesting it is the least valued or least impactful strategy among the listed options.

Table 2

Leaner-instructor Interaction Strategies	Mean	SD
11. The instructor refers to students by name in discussion forums.	3.64	0.76
12. The instructor sends/posts regular announcements or email reminders.	4.34	1.02
13. The instructor creates a forum for students to contact the instructor with questions about the course.	4.17	1.01
14. The instructor creates a course syllabus for students.	4.08	0.9
15. The instructor posts a "due date checklist" at the end of each instructional unit.	4.49	0.89
16. The instructor creates short videos to increase instructor presence in the course.	3.94	0.91
17. The instructor provides feedback using various modalities (e.g., text, audio, video, and visuals).	4.15	0.89
18. The instructor provides students with an opportunity to reflect (e.g., via a journal or surveys).	4.06	0.95
19. The instructor posts grading rubrics for all assignments.	4.15	0.99
20. The instructor uses various features in synchronous sessions to interact with students (e.g., polls, emoticons, whiteboard, text, or audio and video chat).	4.17	0.98

The averages of the two types of interaction strategies were also compared (Table 3). It is worth mentioning that learner-instructor strategies (Subscale Mean = 4.12) are valued significantly more by students compared to

6

learner-learner strategies (Subscale Mean = 3.72), which indicates that instructor presence is much more important to students in online interpersonal interaction.

Table 3

Subscale	Mean
Learner-learner	3.72
Leaner-instructor	4.12

Reasons behind the effectiveness of learner-learner strategies perceived by students:
The quantitative data revealed that allowing students to choose readings that drive discussion group formation is highly valued. Students think that this approach aligns discussions with their interests and preferences. Specifically, giving the autonomy of selecting readings related to their field of study would enhance their contribution to group discussions. Furthermore, it appears contradictory that students value self-introduction in icebreaker discussions but dislike using a virtual lounge for informal meetings to share common interests, as both provide opportunities for social and personal interaction. However, students in interviews mentioned that, although they appreciate getting to know each other for a more comfortable learning environment, they do not find the virtual lounge as engaging or effective for fostering meaningful interaction compared to more structured and purposeful activities like icebreaker discussions. They also cited inconvenience as another reason for disliking the virtual lounge. Instead, they prefer social media platforms like WeChat for informal communication. Regarding peer-view classmates' work, while some students find feedback useful, they have concerns about their ability to provide constructive feedback to classmates.

Reasons behind the effectiveness of learner-instructor strategies perceived by students:
Student interviewees believe that sending a "due date checklist" and regular announcements or email reminders provides clear, organized information about upcoming deadlines. Additionally, They find this approach helpful as it enables them to manage their time effectively and stay on track with their learning. Students place the least value on instructors referring to them by name in discussion forums. While personal recognition is appreciated, it is considered less critical to their overall learning experience compared to other strategies.

Challenges regarding online interpersonal interaction

In examining the challenges students face with interpersonal interactions in online education, several key themes emerged from interviews:

- **Personality:** Students frequently noted that introversion, or shyness, can hinder their ability to actively connect with classmates and instructors.

- **Community:** Students often feel a sense of detachment from classmates and instructors in online learning environments. Compared to onsite learning, the rapport and sense of connection with others tend to be weaker.

- **Language Proficiency:** Students often face challenges in verbally interacting with others due to a lack of confidence in their English language skills.

- **Technical Issues:** Unstable internet connectivity and problems with personal devices can sometimes make students feel isolated from their classmates and the instructor.

Discussion and conclusion

This study aims to explore Chinese language learners' perceptions of the effectiveness of interpersonal interactions and to identify the challenges associated with online interpersonal interaction. Among all learner-learner interaction strategies, students view the option to select readings that guide discussion group formation as the most effective for enhancing their engagement in online learning. This finding is supported by Andrade & Bunker (2009), who emphasize the benefits of allowing students some choice in distance language learning. Conversely, the least valued interaction strategy is the use of a virtual lounge for informal meetings to share common interests. This aligns with Nicholson's (2002) research, which highlights the limited effectiveness of non-monitored spaces for communication, such as instant online messaging.

Regarding learner-instructor interaction strategies, the "due date checklist," which functions as a reminder system, is highly valued by students. This finding aligns with O'Neil & Buckley (2020) who reported that students have positive attitudes towards reminders. In contrast, students do not place significant importance on "the instructor referring to students by name in discussion forums" compared to other learner-instructor interaction strategies. This contradicts Bolliger & Martin's (2018) study, which emphasized the importance of using student names by instructors in distance education. Overall, learner-instructor interaction strategies are generally regarded as more valuable than learner-learner interaction strategies. This is consistent with other research that highlights the significance of the instructor's role in online interactions (Fedynich et al., 2015; Martin & Bolliger, 2018; Tanis, 2020).

Several challenges in online interpersonal interaction perceived by students were identified. Introversion, for instance, hinders students' interactions with others, which aligns with Kuo, Belland, et al. (2014), who found that extroverted students reported more frequent interactions compared to their introverted peers. Additionally, a low level of language proficiency can impede student interaction in an EMI context. This is supported by Zhang & Pladevall-Ballester (2022) who observed that when English is the sole medium of communication,

students' classroom interactions can be negatively affected. The two challenges mentioned are relevant not only to online learning but also to face-to-face teaching settings. However, the following challenges are specific to online education. Firstly, despite various activities designed to enhance social interaction, students still perceive a lack of community—a concern that has long been highlighted by researchers (Garrison, 2016; Moore & Anderson, 2003; Swan et al., 2009). Additionally, technical issues frequently arise as a challenge, as Muilenburg & Berge (2005) mentioned in their study that individuals with high confidence in using online learning technologies experience significantly fewer barriers to social interactions.

This study has limitations, such as a small sample size and a non-exhaustive list of interpersonal strategies in the survey. However, the following recommendations can be valuable for future online educators and course designers, particularly those in similar teaching contexts: 1) Instructors play a crucial role in facilitating interaction, as students still view them as the authority in learning. 2) community and rapport should be built through tools such as social media and icebreaker activities. 3) technical troubleshooting guides can be provided to reduce technical issues. 4) a relaxing learning environment needs to be created to ease pressure on students, encouraging them not to fear making mistakes and promoting autonomy in their learning choices.

Bibliography

Anderson, T. (2003). Modes of Interaction in Distance Education: Recent Developments and Research Questions. *Handbook of Distance Education*, 129-144.

Andrade, M. S., & Bunker, E. L. (2009). A model for self-regulated distance language learning. *Distance Education*, *30*(1), 47–61. https://doi.org/10.1080/01587910902845956

Berge, Z. L. (1999). Interaction in Post-Secondary Web-Based Learning. *Educational Technology*, *39*(1), 5–11.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A Meta-Analysis of Three Types of Interaction Treatments in Distance Education. *Review of Educational Research*, *79*(3), 1243–1289. https://doi.org/10.3102/0034654309333844

Bolliger, D. U., & Martin, F. (2018). Instructor and student perceptions of online student engagement strategies. *Distance Education*, *39*(4), 568–583. https://doi.org/10.1080/01587919.2018.1520041

Cole, M. T., Shelley, D. J., & Swartz, L. B. (2014). Online Instruction, E-Learning, and Student Satisfaction: A Three Year Study. *International Review of Research in Open and Distributed Learning*, *15*(6), 111–131. https://doi.org/10.19173/irrodl.v15i6.1748

Fedynich, L., Bradley, K. S., & Bradley, J. (2015). Graduate Students' Perceptions of Online Learning. *Research in Higher Education Journal*, *27*. https://eric.ed.gov/?id=EJ1056187

Fulford, C. P., & Zhang, S. (1993). Perceptions of Interaction: The Critical Predictor in Distance Education. *American Journal of Distance Education*, *7*(3), 8–21.

Garrison, D. R. (2016). *E-Learning in the 21st Century: A Community of Inquiry Framework for Research and Practice* (3rd ed.). Routledge. https://doi.org/10.4324/9781315667263

Gilbert, L., & Moore, D. R. (1998). Building Interactivity into Web Courses: Tools for Social and Instructional Interactions. *Educational Technology*, *38*(3), 29–35.

Hall, J. K. (2010). Interaction as method and result of language learning. *Language Teaching*, *43*(2), 202–215. https://doi.org/10.1017/S0261444809005722

Kitade, K. (2000). L2 Learners' Discourse and SLA Theories in CMC: Collaborative Interaction in Internet Chat. *Computer Assisted Language Learning*, *13*(2), 143–166. https://doi.org/10.1076/0958-8221(200004)13:2;1-D;FT143

Kuo, Y.-C., Belland, B. R., Schroder, K. E. E., & Walker, A. E. (2014). K-12 teachers' perceptions of and their satisfaction with interaction type in blended learning environments. *Distance Education*, *35*(3), 360–381. https://doi.org/10.1080/01587919.2015.955265

Kuo, Y.-C., Walker, A. E., Schroder, K. E. E., & Belland, B. R. (2014). Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *The Internet and Higher Education*, *20*, 35–50. https://doi.org/10.1016/j.iheduc.2013.10.001

Lai, C.-H., Lin, H.-W., Lin, R.-M., & Tho, P. D. (2019). Effect of Peer Interaction among Online Learning Community on Learning Engagement and Achievement. *International Journal of Distance Education Technologies (IJDET)*, *17*(1), 66–77. https://doi.org/10.4018/IJDET.2019010105

Martin, F., & Bolliger, D. U. (2018). Engagement Matters: Student Perceptions on the Importance of Engagement Strategies in the Online Learning Environment. *Online Learning*, *22*(1), 205–222.

Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education*, *3*(2), 1–7. https://doi.org/10.1080/08923648909526659

Moore, M. G., & Anderson, W. G. (Eds.) (2003). *Handbook of Distance Education*. L. Erlbaum Associates.

Muilenburg, L. Y., & Berge, Z. L. (2005). Student barriers to online learning: A factor analytic study. *Distance Education*, *26*(1), 29–48. https://doi.org/10.1080/01587910500081269

Nicholson, S. (2002). Socialization in the "virtual hallway": Instant messaging in the asynchronous Web-based distance education classroom. *The Internet and Higher Education*, *5*(4), 363–372. https://doi.org/10.1016/S1096-7516(02)00127-6

Northrup, P., Lee, R., & Burgess, V. (2002). *Learner Perceptions of Online Interaction*. Association for the Advancement of Computing in Education (AACE), P. 2-7 https://eric.ed.gov/?id=ED477075

O'Neil, C. A., & Buckley, K. M. (2020). Impact of Electronic Reminders on Student Grades and Attitudes in Online Courses. *Nurse Educator*, *45*(6), E66. https://doi.org/10.1097/NNE.0000000000000813

Paquette, P. (2016). Instructing the Instructors: Training Instructors to Use Social Presence Cues in Online Courses. *Journal of Educators Online*, *13*(1), 80–108.

Roblyer, M. D., & Wiencke, W. R. (2003). Design and Use of a Rubric to Assess and Encourage Interactive Qualities in Distance Courses. *American Journal of Distance Education*, *17*(2), 77–98. https://doi.org/10.1207/S15389286AJDE1702_2

Swan, K., Garrison, D. R., & Richardson, J. C. (2009). A Constructivist Approach to Online Learning: The Community of Inquiry Framework. In *Information Technology and Constructivism in Higher Education: Progressive Learning Frameworks* (pp. 43–57). IGI Global. https://doi.org/10.4018/978-1-60566-654-9.ch004

Tanis, C. J. (2020). The seven principles of online learning: Feedback from faculty and alumni on its importance for teaching and learning. *Research in Learning Technology*, *28*. https://doi.org/10.25304/rlt.v28.2319

Vygotsky, L. S. (1997). Interaction Between Learning and Development. In *Readings on the Development of Children* (pp. 34–40). Scientific American Books.

Wang, Y. (2004). Distance Language Learning: Interactivity and Fourth-generation Internet-based Videoconferencing. *CALICO Journal*, *21*(2), 373–395.

Wang, Y., Cao, Y., Gong, S., Wang, Z., Li, N., & Ai, L. (2022). Interaction and learning engagement in online learning: The mediating roles of online learning self-efficacy and academic emotions. *Learning and Individual Differences*, *94*, 102128. https://doi.org/10.1016/j.lindif.2022.102128

Wang, Y., & Sun, C. (2001). Internet-Based Real Time Language Education: Towards a Fourth Generation Distance Education. *CALICO Journal*, *18*(3), 539–561.

York, C. S., & Richardson, J. C. (2012). Interpersonal Interaction in Online Learning: Experienced Online Instructors' Perceptions of Influencing Factors. *Journal of Asynchronous Learning Networks*, *16*(4), 83–98.

Zhang, M., & Pladevall-Ballester, E. (2022). Students' attitudes and perceptions towards three EMI courses in mainland China. *Language, Culture and Curriculum*, *35*(2), 200–216. https://doi.org/10.1080/07908318.2021.1979576

Dynamic or Static Pictures? The Effects of Two Multimodal Types on L2 Verb Learning and the Role of Visual Abilities

TAN Qianyi

South China Normal University

Abstract

Built on the cognitive theory of multimedia learning and the embodied cognition hypothesis, the quasi-experimental study seeks to examine the effects of static-picture and dynamic-picture multimodal inputs on L2 verb learning and the role of visual abilities. Sixty-six Chinese EFL learners were divided into the static-picture group (SG; N=31) and the dynamic-picture group (DG; N=35). The object-spatial imagery model was adopted to determine learners' object and spatial imagery levels. It was found that the SG significantly outperformed the DG in spelling, which can be explained by the greater quantity of attention paid to the target words by the SG. No significant difference was identified in the results of the meaning tests. The spelling performances in the SG significantly correlated to learners' levels of object imagery.

Keywords: Multimodal input; L2 verb learning; object imagery; spatial imagery

Introduction

As an essential category of words, verbs involve movements that can be simulated in mind and performed with bodies, distinct from the other word classes like nouns and adjectives. Due to the fact that interpreting objects from the environment is easier than interpreting actions, verbs take a longer time for a child to acquire and they have also been considered by many EFL learners as difficult to master (Imai et al., 2008; Yang & Coxhead, 2022). Therefore, some researchers have made efforts to identify effective methods of learning verbs. For L1 learners, what facilitates verb learning the most is learners themselves performing the action described by the verb (Sidhu & Pexman, 2016). Moreover, simulating the movement in mind or watching other people performing it is a more effective way than merely learning the verb on paper, which can be explained by the embodiment cognition theory.

However, whether the embodiment cognition hypothesis underlies L2 verb learning remains unknown. Few studies on L2 vocabulary zoomed in on L2 verb learning, even though research on

L2 vocabulary as a whole is extensive. Guided by the cognitive theory of multimedia learning that learners process messages through the visual information system and the verbal information system, more and more L2 language instructors incorporate multimedia (e.g. text, audio, pictures, and videos) into vocabulary teaching to deliver multi-sensory experience to their students. It was discovered that the combination of different modes of input is more beneficial to vocabulary learning than the unitary mode (Ramezanali & Faez, 2019; Teng, 2023; Yun, 2011). Furthermore, many comparative studies have been done to understand the advantages of various multimodal inputs in terms of vocabulary learning (Al-Seghayer, 2001; Sato et al., 2022). Among the findings, the efficacy of dynamic and static visual aids was controversial. According to the embodiment theory, dynamic learning aids may be more useful for verb learning than static visual aids, as dynamic pictures allow learners to observe dynamic movements. Hence, the study seeks to examine the hypothesis that dynamic input produces better verb learning outcomes than static input.

The effects of multimodal input are moderated by various factors, one of which is individual cognitive styles. The studies that considered the importance of individuals' cognitive differences were based on the visual-verbal cognitive model and failed to differentiate participants' different visual abilities, which are not a none-or-all concept (Mohsen & Balakumar, 2011; Yeh & Wang, 2003). Kozhevnikov et al.(2002) furnished more evidence about the existence of two types of visualizers: object visualizers and spatial visualizers. The former one demonstrates a higher level of competence to deal with the pictorial attributes of still images, while the latter manipulates dynamic materials flexibly. The finer classification of visual ability has been used in the education of science-related subjects and fine art (Kozhevnikov, 2022; Pitta-Pantazi et al., 2013), but its potential in L2 multimodal teaching was still underexplored. Hence, the hypothesis is that learners with better object imagery learn better with static aids, while learners with better spatial imagery learn better with dynamic aids.

To examine the above two hypotheses, this study was guided by the two research questions:

1. Which type of multimodal input (text + static pictures VS text + dynamic pictures) contributes to better learning and retention of L2 verb form and meaning? Why?

2. Do the levels of object and spatial imagery correlate to the learning outcomes?

Methodology

Participants

Sixty-six Chinese eleventh graders aged from 16 to 17 from two intact classes in Meizhou, Guangdong Province, China, participated in the experiment. The two classes were selected because of their similar English scores in previous examinations. Their English proficiency is roughly equivalent to B1 according to the Common European Framework of Reference for Languages. The two parallel classes were randomly divided into the dynamic-picture group (DG; $N = 35$) and the static-picture group (SG; $N = 31$).

Verb Learning Materials

Two experienced English teachers chose 26 motion verbs from the word list of the Graduate Record Examination (GRE). Those words were chosen because they did not occur in the participants' textbooks and exercise books, and the animation depicting the motions was accessible. To determine their frequency, the 26 target words were examined using the VocabProfile section of Compleat Lexical Tutor (www.lextutor.ca). The frequency levels of all the target words fall between the K-5 and the K-9. To ensure these words were unknown to at least most of the participants, pre-tests were also conducted in both groups. As seen in Figure 1, the multimodal verb learning materials used in both groups were created by MS PowerPoint and each slide includes five elements: (a) an EFL motion verb (written form), (b) its phonetic transcription, (c) its Chinese meaning, (d) an audio recording of the EFL verb (spoken form), (e) a picture describing the motion of the word. During the experiment, each slide was presented for 12 seconds.

The only difference between the materials used in both groups was that the pictures were dynamic in the DG and static in the SG. To ensure the information communicated to both groups was the same, the static pictures were the screenshots of the keyframe that defines the motion in the dynamic pictures.

Figure 1

An example of the verb learning materials

dangle /ˈdæŋɡ(ə)l/
悬摆

Instruments

Vocabulary tests. In the current study, a pre-test and an immediate post-test were administered in each group. Each of the tests contained two parts (the meaning test and the spelling test) and each part covered 26 items. The items in the two tests were the same but in different orders to minimize test-retest learning effects. Referring to Nation (1990), the meaning tests adopted multiple-choice items which required the respondents to choose the Chinese equivalent to each L2 verb. For each word, there were five choices: one correct answer, three distractors, and one named "I don't know" to prevent students from guessing. The correct Chinese meaning was provided by the Oxford Chinese-English dictionary, and the distractors were created based on two criteria: (a) the distractors' length should match that of the correct answer, and (b) their part of speech should be the same as that of the correct answer (Heaton, 1988; Hughes, 2002). The spelling tests provided the Chinese translation of each word, and the participants were required to write down the corresponding target words.

Each correct answer scored 1 point and each wrong answer scored 0 point on the meaning tests. One of the researchers rated all the participants' answers. The spelling test papers were distributed after the participants submitted their meaning test papers so that they could not get any hints from the vocabulary meaning items.

The Object-Spatial Imagery Questionnaire (OSIQ). The Objective-Spatial Imagery Questionnaire (OSIQ) is a self-report questionnaire created by Blajenkova et al. (2006) to evaluate individual variations in visual imagery preferences and experiences. A series of studies have

confirmed the OSIQ's ecological, discriminant, and predictive validity (Blajenkova et al., 2006). The survey comprises a scale for object imagery and a scale for spatial imagery. The object imagery scale measures preferences for depicting and processing vivid, detailed, and high-resolution images of objects (e.g., *When reading fiction, I usually form a clear and detailed mental picture of a scene or room that has been described.*), while the spatial imagery scale assesses preferences for depicting and processing schematic images, object movement, spatial relationships between objects, and spatial transformations (e.g., *I was very good in 3-D geometry as a student.*). Each of the scales has 15 five-point Likert items and one item in the spatial image scale needs to be reversed. The questionnaire was translated into Chinese by the researchers to facilitate participants' understanding. The accuracy and clarity of the translation were confirmed by a professor specializing in psycholinguistics. In this study, a spatial imagery score and an object imagery score were created for each participant after ratings of the 15 items for each factor were averaged. The Cronbach's alpha reliability was .704 for the translated spatial imagery scale, and .786 for the translated object imagery scale, indicating high reliability of the Chinese questionnaire.

The Questionnaire on participants' behaviour. To further explore how the input materials affect their L2 verb learning, a survey on participants' behaviors during the experiment (i.e. attention allocation) was distributed. It consists of two items, concerning the amount of time the pupils focused on the pictures and textual information while viewing the slides, offering four choices (A. never, B. seldom, C. often, D. always).

Procedures

The experimental design is shown in Figure 2. The experiment took place in the students' English classes and they had the right to ask for leave for personal reasons. The experiement lasted for two weeks. In week one, they learned 26 English motion verbs with the materials. During the intervention, both groups watched the slides projected on a large screen in front of their classrooms.

Figure 2

The experimental design

5
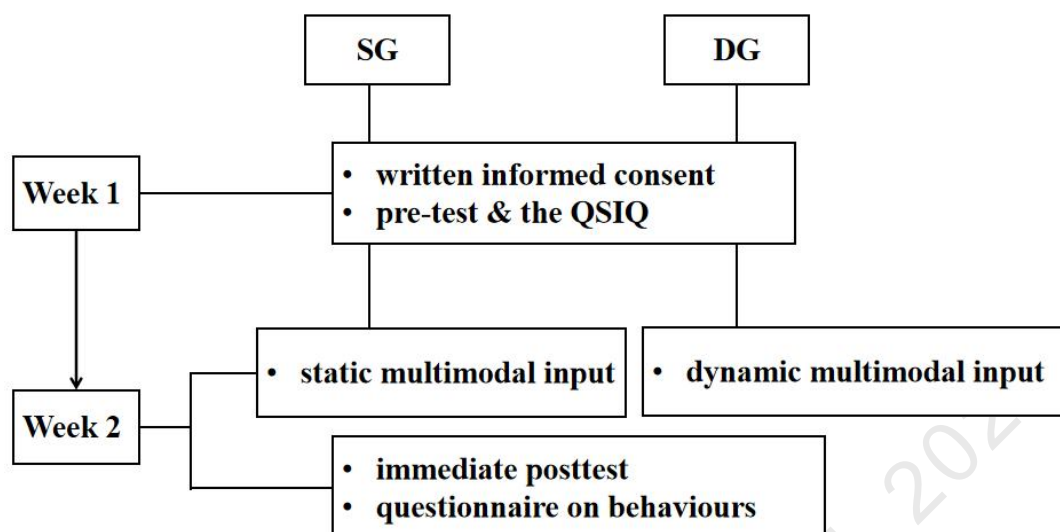
Data analysis

All the data analyses were performed in SPSS 23 in the study. To answer the first research questions, the scores of meaning test and spelling test were submitted to an independent sample t-test respectively. Additionally, Spearman's correlation was used to evaluate the relationship between the vocabulary scores and the amount of attention paid to the pictures and the text so that we can gain a deeper understanding of the effects of multimodal types on L2 verb learning. The level of significance was set at an alpha level of $p < .05$.

To answer the second research question, Pearson's correlation was employed to determine the relationships between the level of object imagery, spatial imagery, and vocabulary scores.

Results

Vocabulary tests

The Mann-Whitney U test showed that there was no significant difference between both groups in the scores of the meaning pre-test and the spelling pre-test, indicating that the two groups had the same knowledge about the target words. Table 1 illustrates the independent t-test results of the post-spelling test and the post meaning test. It shows that significant differences exist between the two conditions in the immediate($t=-4.784$, $p<.001***$), with cohen's d of 1.180. As the values of Cohen's d were greater than .80, they are considered to be large effect sizes,

indicating that the differences between the two conditions are very meaningful. On the other hand, no significant differences were identified between the two groups in the immediate meaning test, which means both groups performed similarly in the meaning vocabulary tests.

Table 1

Results of the independent sample t-test

	Levene's test for equality of variances		t-test for equality of means						95% confidence interval of the difference	
	F	Sig.	t	df	Sig.(2-tailed)	Mean difference	Std. Error difference	Cohen'd	Lower	upper
Spelling	2.978	.089	-4.784	64	.000***	-6.525	1.364	1.180	-9.250	-3.801
Meaning	.003	.956	-.643	64	.523	-.346	.538	0.159	-1.419	.728

***$p<.001$.

The Object-Spatial Imagery Questionnaire (OSIQ)

The data of the two subscales are normally distributed according to the Shapiro-Wilk test results ($p>.05$). To answer the second research question, a Pearson correlation was used to analyze the relationships between the scales' scores and the vocabulary scores. For the static-picture group, participants' object imagery ratings significantly correlated with their spelling scores (r=.439, p=.014*) in a positive way. No other significant relationships were observed.

The questionnaire on learners' behaviour

The results of the questionnaire on learners' behaviour were shown in table 2. To find out whether the vocabulary learning performances were correlated to their viewing behaviors, this study conducted a Spearman's correlation and a Mann-Whitney U, as the measure type was considered ordinal. It was found that the frequency of viewing the text significantly related to their spelling score (r=.401, p=.001**) in a positive way. This means that the more the EFL learners attended to the text in the materials, the higher spelling scores they were likely to gain. The amount of attention paid to text by the SG was significantly larger than that by the DG (U=327.500, p=.003**), while the amount of attention given to the pictures by the DG was also significantly larger than that by the SG (U=325.000, p=.003**).

Table 2

Attention allocation

		Never	Seldom	Often	Always
		n	*n*	*n*	*n*
SG(N=31)	Viewing pictures	1	16	9	5
	Viewing text	5	10	16	0
DG(N=35)	Viewing pictures	1	5	16	13
	Viewing text	11	20	3	1

Discussion

Vocabulary learning performances of both groups

Surprisingly, the results indicated that the SG outperformed the DG in the spelling test significantly, while they had similar scores in the meaning recognition test, which goes against the hypothesis that the DG will surpass the SG in verb learning. The less satisfying spelling performance of the DG may be explained by the distracting nature of dynamic aids. To furnish more evidence, this research conducted a Mann-Whitney U test and a Spearman's correlation to find out the relationship between the frequency of gazing at text and their vocabulary scores. The results showed the SG gave significantly more attention to the text, whereas the DG paid more attention to the pictures. And the participants who spent more time on textual information had higher accuracy in form learning. These findings are supported by the Noticing Hypothesis

(Schmidt, 1990), which suggests that people learn more from the things they pay attention to than the things they do not. Long et al. (2001) also emphasized the significance of noticing in the obtaining of linguistic form.

As more evidence for the advantages of static multimodal aids, Chun and Plass (1996) found that the words annotated with text and pictures were learned and retained better than the words annotated with text and videos. According to the Cognitive Load Theory (Chandler & Sweller, 1991), dynamic pictures might subject the learners to more extraneous load, which is determined by the instructional material used to convey information. In terms of the vocabulary retention, the better result in the SG can also be attributed to the less distracting nature of the static aids, as Vanderplank (1990) stressed the pivotal role of attention (form-focused) in the intake of form.

Also, no significant differences between the impact of dynamic aids and static aids on L2 word meaning acquisition emerged in many studies (Akbulut, 2007; Lin, 2009; Sato et al., 2022), which are resonant with this research. These findings are unsurprising given that animated pictures are typically thought to be particularly helpful for understanding processes or procedures, whereas static images should be sufficient for learning declarative information or simple facts (Weiss et al., 2002). In many instances, static images may also suffice to gain deeper comprehension (Höffler & Leutner, 2007). In this study, the still pictures were an appropriate carrier of meaning, as they adhere to the Congruence Principle (Tversky et al., 2002), which states that a carefully chosen or produced picture can convey the meaning of the content precisely. As a result, the still pictures provided as many cues as the dynamic pictures did for the learners to recall the meaning in this study.

This study implies that using dynamic pictures to learn words in this study may not give full play to the Embodied Cognition Theory, which asserts that cognition has roots in motor behavior and animation may help pupils develop an internal representation of the external environment they are mentally emulating. The reason could be that the use of dynamic pictures only may not fully activate the embodiment effect. It is strongly advised that instructors teach vocabulary with embodied interactive video, gestures, and other tools to create embodied effects (Huang et al., 2021; Hung et al., 2017; Xu & Ke, 2020).

Relationships between visual abilities and vocabulary learning

9

The results showed that the respondents in the SG with stronger object imagery ability scored higher in the immediate spelling test, as there were positive relationships between the object imagery ratings and the scores, which corresponds to the second hypothesis. Object imagers are more likely to have a photographic memory, as they possess an excellent capacity for processing the pictorial properties (such as shape, color, and texture) of static images (Blazhenkova & Kozhevnikov, 2010). By keeping a picture of the images in mind, they easily constructed a connection between form and meaning. The foregoing findings were in line with the Individual Differences Principle proposed by Mayer and Moreno (1998) that multimedia effects depend on individual differences.

However, the second hypothesis was not fully verified because the positive relationships between the DG's spatial imagery ratings and their learning performances did not reach a significant level. The finding was in accordance with Rouhi and Mohebbi's (2013) conclusion that the high and the low spatial ability group did not score differently on a significant level. They argued that the cognitive loads imposed by their vocabulary glosses were so heavy that even the learners with high spatial ability failed to properly manipulate them, which highlighted the importance of a balance between individuals' working memory capacity and cognitive load of tasks. The advantages of high-spatial-ability learners show only when the materials are under relatively high loads, as the knowledge may be demanding to digest profoundly for those with low spatial abilities (Plass et al., 2010). In the current study, the cognitive load imposed by the dynamic-picture materials was far below learners' capacity, which allows learners with different levels of spatial ability to manage the materials easily. This was supported by Tight (2010), who stated that specific learning preferences may neither be a booster nor hindrance to L2 vocabulary learning and that participants appear to be equally capable of learning new words regardless of their preferred perceptual learning approach.

Conclusions, implications, and limitations

To address the first research question, this study found that the static-picture group significantly outperformed the dynamic-picture group in the spelling test, as the SG participants viewed text more frequently. The result revealed the limitation of applying the Embodied Cognition Theory to L2 verb learning in the multimedia environment. As for meaning acquisition,

the two groups earned similar scores, indicating that static pictures can represent the verb meaning as well the dynamic pictures.With regards the second research question, there existed a significant positive relationship between participants' object imagery levels and their spelling performance when they were using the static visual aid.

The foregoing findings bring some implications for explicit L2 vocabulary instruction. Firstly, it is advised that static pictures should be used to introduce new words due to their greater effectiveness, and dynamic pictures, which are more motivating, should be used to review learned words. Secondly, understanding students' visual preferences is necessary, as tailoring learning materials to individuals' characteristics can strengthen learning effects. Importantly, this study suggested that the fine classification of visualizers may be of great value to L2 learning materials developers. Thirdly, teachers should apply multimedia in conformity with the cognitive load theory (Zhang & Zou, 2022). When adopting multimodal input, teachers should reduce extraneous load, moderate intrinsic load, and optimize germane load. For example, instructors should avoid using images containing seductive details, which may steer students' attention from the knowledge.

Inevitably, some limitations lie in this study. Firstly, this study highlighted the advantages of still pictures in explicit L2 word learning and the role of imagery preferences in the multimedia environment. The findings are, however, tentative. The small samle size, the overall experimental design, the participants' mental status, and some other variables may influence the conclusion. Future studies may involve more subjects and take more variables into consideration, such as L2 proficiency, L1 background, etc. Secondly, this study only utilized a self-report questionnaire to determine participants' levels of imagery. More object and spatial imagery measures could be used in the future, such as the Degraded Pictures Test for the object imagery and the Paper Folding Test for the spatial imagery. Thirdly, an eye-tracking technique could also be adopted to understand participants' attention allocation.

References

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science, 35*, 499–517.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language, Learning & Technology, 5*(1). 202–232.

Blazhenkova, O., & Kozhevnikov, M. (2010). Visual-object ability: A new dimension of non-verbal intelligence. *Cognition, 117*(3), 276-301.

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: a new self-report imagery questionnaire. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 20*(2), 239-263.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The modern language journal, 80*(2), 183-198.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction, 8*(4), 293-332.

Heaton, J. B. (1988). *Writing English language tests (Longman handbooks for language teachers)* (pp. 54–55). Longman

Höffler, T. N. & Leutner, D. (2007) Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*(6): 722–738.

Huang, T. H., & Wang, L. Z. (2021). Artificial intelligence learning approach through total physical response embodiment teaching on French vocabulary learning retention. *Computer Assisted Language Learning*, 1-25.

Hughes, A. (2002). *Testing for language teachers* (p. 181). Cambridge University Press.

Hung, I. C., Kinshuk, & Chen, N. S. (2017). Embodied interactive video lectures for improving learning comprehension and retention. *Computers & Education, 117*, 116–131.

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child development, 79*(4), 979-1000.

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and instruction, 20*(1), 47-77.

Kozhevnikov, M., Ho, S., & Koh, E. (2022). The role of visual abilities and cognitive style in artistic and scientific creativity of Singaporean secondary school students. *The Journal of Creative Behavior, 56*(2), 164-181.

Lin, C. C. (2009). Learning action verbs with animation. *The Jalt Call Journal, 5*(3), 23-40.

Long, M. H., De Bot, K., Ginsberg, R., & Kramsch, C. (2001). Focus on form: A design feature in language teaching methodology. *English language teaching in its social context*, 180-190.

Mayer, R. E., & Moreno, R. (1998). A cognitive theory of multimedia learning: Implications for design principles. *Journal of educational psychology, 91*(2), 358-368.

Mohsen, M. A., & Balakumar, M. (2011). A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. *ReCALL, 23*(2), 135-159.

Nation, P. (1990). *Teaching and learning vocabulary.* Newbury House

Pitta-Pantazi, D., Sophocleous, P., & Christou, C. (2013). Spatial visualizers, object visualizers and verbalizers: Their mathematical creative abilities. *ZDM, 45*, 199-213.

Plass, J. L., Kalyuga, S., & Leutner, D. (2010). Individual differences and cognitive load theory. In J. L. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 65–87). New York: Cambridge University Press.

Ramezanali, N., & Faez, F. (2019). Vocabulary learning and retention through multimedia glossing. *Language Learning & Technology, 23*, 105–124

Rouhi, A., & Mohebbi, H. (2013). Glosses, Spatial Intelligence, and L2 Vocabulary Learning in Multimedia Context. *3L: Southeast Asian Journal of English Language Studies, 19*(2).

Sato, T., Lai, Y., & Burden, T. (2022). The role of individual factors in L2 vocabulary learning with cognitive-linguistics-based static and dynamic visual aids. *ReCALL, 34*(2), 201-217.

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics, 11*(2), 129-158.

Sidhu, D. M., & Pexman, P. M. (2016). Is moving more memorable than proving? Effects of embodiment and imagined enactment on verb memory. *Frontiers in psychology*, 7, 1010.

Teng, M. F. (2023). The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in language learning and teaching, 17*(3), 738-754.

Tight, D. G. (2010). Perceptual learning style matching and L2 vocabulary acquisition. *Language Learning, 60*(4), 792–833.

13

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate?. *International journal of human-computer studies, 57*(4), 247-262.

Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with Uni-Lingual (CEEFAX) sub-titles. *System, 18*(2), 221-234.

Weiss, R.-E., Knowlton, D.-S., & Morrison, G.-R. (2002). Principles for using animation in computer-based instruction: theoretical heuristics for effective design. *Computers in Human Behavior, 18,* 465e477.

Winke, P., Gass, S., & Syodorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*(1), 65–86.

Xu, X., & Ke, F. (2020). Embodied interaction: Learning Chinese characters through body movements. *Language Learning & Technology, 24*(3), 136–159.

Yang, L., & Coxhead, A. (2022). A corpus-based study of vocabulary in the new concept English textbook series. *RELC Journal, 53*(3), 597-611.

Yeh, Y. & Wang, C. (2003) Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *CALICO Journal, 21*(1): 131–144.

Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning, 24*(1), 39-58.

Zhang, R., & Zou, D. (2022). A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. *Computer Assisted Language Learning, 35*(9), 2790-2816.

Emerging technologies & language education

新兴技术在外语教学中的应用

Decelerating language endangerment through artificial intelligence (AI): A proposed AI Framework

Mustapha Bala Tsakuwa

School of Foreign Languages, Yulin University, Yulin City, 719000, Shaanxi, China

mbtsakuwa@yulinu.edu.cn

Abstract

The revitalization of indigenous languages and cultures in Nigeria faces great challenges amidst the influences of English, Hausa, Pidgin, social media and technological advancements. In this article, insights were obtained from ChatGPT and published articles to explore the dual role of artificial intelligence (AI) as both a potential language preserver and a complex player in this domain. AI offers significant benefits as it facilitates the documentation and preservation of endangered languages, enhances community engagement and contributes to linguistic and cultural research. However, using AI technology in the Nigerian context seems challenging because there is a scarcity of accurate language and linguistic data, a lack of trained professionals, and technological and infrastructural constraints. The political atmosphere and socio-cultural practices further complicate the issue. This paper, therefore, examines the benefits and challenges through the lens of Nigeria's diverse linguistic and cultural landscape, proposing an AI framework for decelerating language endangerment through the use of artificial intelligence. The article also emphasizes the need for collaborative efforts among local communities, linguists, technologists, and policymakers to slow down language endangerment and/or language death.

Keywords: Decelerating language endangerment; artificial intelligence; AI Framework; benefits; challenges

Introduction

For decades languages around the globe have been fading away because of many *natural* and *artificial* factors. Language experts, linguists, educationists as well as community-development societies and governments have also been making concerted efforts to maintain and save the languages (Galla 2016; Kono 2019). Plans are made, money is spent, and policies are implemented, all in the name of revitalizing the languages. These efforts seem *fruitless*, especially in some countries or regions (Akpan 2018; Akpan 2020; Ayakoroma 2017). Some questions that might come to mind are: Are these languages maintainable, revivable and/or savable? What does humankind lose if languages disappear? How *best* languages can be revitalized in the era of artificial intelligence (AI)? In other words, how can AI technologies be used to preserve or decelerate language endangerment or extinction? We will attempt to provide answers to these questions. But, before then, let's begin with the following quotation, which motivates this study, from Geeksforgeeks (2024: para.1) which says:

> Imagine a world where ancient tongues, on the brink of fading into silence, are reborn. Where stories whispered through generations find a digital echo and cultural knowledge carried in every syllable is amplified across the internet. This is the promise of language revitalization in generative AI, a revolutionary field that seeks to leverage the power of artificial intelligence to resurrect endangered languages and empower their communities.

AI has greatly influenced the preservation and revitalization of indigenous languages. Through various techniques, AI has been able to analyze and translate these languages, making them more accessible to both native speakers and non-native speakers. This has not only preserved the cultural and linguistic heritage of the indigenous communities, but it has also aided in creating a more diverse and inclusive world. With AI, more people can understand and appreciate these languages, leading to a better understanding and appreciation of different cultures (Geeksforgeeks 2024). In addition, AI has helped in developing language learning tools and applications, further enhancing the accessibility of these languages. As AI continues to advance and improve, its role in revitalizing native languages will only continue to grow, making a significant impact in preserving the diversity of languages around the world.

However, modern technology is accused of contributing to the decline of many minor indigenous languages by compelling the use of dominant global languages like English, French, Arabic, Chinese, and the like, which today have a lot of literature and are used by search engines and social media (Irwin 2019; Pradhan & Dey 2023). Despite that, technology also offers potential avenues for language documentation, preservation, teaching and learning of the indigenous languages (Meighan 2021). For issuance, Meta and Nunavut Tunngavik collaborated to make the Inuktitut

language accessible on Facebook to more than 35,000 speakers globally. With enough linguistic data, large language models (LLM) could be used to document the world's endangered languages. The internet, AI technologies, and LLMs can be used to revive and digitize the languages and encourage more people to learn and use them in everyday life. When endangered languages die off, their unique cultures and worldviews will be lost; revitalizing the languages is like rescuing precious treasures on the verge of loss (Geeksforgeeks 2024; IBM Research 2024).

Decelerating language endangerment through technology in Nigeria and elsewhere

In today's global space, out of the 7,000 world's languages, about 50% are predicted to be extinct by 2100. This prediction may be true from the look of things in the globalized world today. To decelerate language endangerment, the UN has announced the year 2022-2032 as the International Decade of Indigenous Languages (IBM Research 2024: para. 3). AI is being harnessed to develop writing tools to strengthen languages that are at severe risk of declining in Brazil. IBM researchers worked with Brazil's University of Campinas to build a model for an AI-powered writing assistant in the Nheengatu language, which is largely spoken by parents and grandparents, but, it is not passed down to children to read or write in it. The model can translate between Nheengatu, English, and Portuguese, and also provide word definitions with complete sentence examples (IBM Research 2024; Newitz 2023). Also, UNESCO (2023b) reports that Motorola collaborates with Lenovo to incorporate native languages into smartphones. Languages like Kaingang, Nheengatu, and Cherokee are now among the 80 languages on Motorola smartphones. The integration of these languages helps in maintaining the linguistic diversity of the native languages and cultures in Brazil, the Amazon and the U.S., by giving speakers more options to engage in their local languages. Over 170 indigenous language apps have been created by Ogoki Learning, a software development company, for K-12 schools, tribal colleges and universities, and indigenous communities all over the United States, Canada, and the world (Ruf 2020). Also, to bridge digital and linguistic divides in India, AI technology is used to deliver content in local languages across the country (ITU News 2021).

Today, many African/Nigerian languages are down to less than 20,000 speakers and are on UNESCO's list of "severely endangered" languages (IBM Research 2024; Udoh & Anyanwu 2015). This implies that the languages are no longer passed down to children at home or schools. Ayakoroma (2017) believes that reviving Nigerian indigenous languages is indispensable, otherwise, the languages will gradually go into extinction. ITU News (2020) reports that in an attempt to preserve Nigerian languages, efforts have been made to advance speech recognition in Nigeria with AI and machine learning. A "Wazobia" mobile application was developed to support data collection. Newitz (2023) argues that AI-powered platforms and apps (like Google, Duolingo, Microsoft Translator, etc.) facilitate community contributions to language documentation and

revitalization projects. However, these apps don't work for every language – especially native languages in Africa, North America and New Zealand, among other regions.

UNESCO (2023b) reports that some UNESCO Member States planned to focus on digital empowerment and address the incompatibility issue between native languages and existing keyboards and software. With this, the UN Permanent Forum on Indigenous Issues encouraged big technology industries, with the active participation of indigenous communities, to support the development of digital tools to enable the use of indigenous languages. The Ubyssey (2024) believes that linguistic inclusiveness is crucial in deploying AI to safeguard, revitalize and promote native languages. However, the vast majority of these languages lack adequate technological support, depriving over a billion people from participating in global conversations. It is not easy to work on endangered languages because they lack sufficient data. But, the challenges could be less if collaboration between linguists and communities is established.

The Ubyssey (2024) shows that scholars (like Adebara, who focuses on AI incorporation in African language revitalization) underscore the significance of linguistic inclusivity, social equity and representation in AI usage for language issues and emphasize the need for transparency and accountability in how AI technologies are deployed in addressing the issues of language endangerment. Like Kono (2019), Adebara stresses the importance of community engagement in developing AI tools that cater to the specific needs of language communities and empower the communities to reclaim and revitalize their linguistic heritage.

Proposed AI framework for decelerating language endangerment in Nigerian context: Challenges and considerations

In the following parts of this article, as a form of suggestion, I outline four key challenges to decelerating language endangerment in the Nigerian context through AI technology.

Linguistic Situation: There are great challenges in using AI to revitalize Nigerian endangered languages amidst the influences of other languages (like English, Hausa, Pidgin, etc.). The expansion and daily use of English in the country, for instance, threatens the existence of not only the minority languages in Nigeria but also the big languages, particularly, Igbo and Yoruba (Aboh 2022). The Hausa language has been reported as the internal agent of language shift and endangerment in Nigeria. Many minority language speakers in northern Nigeria, including in the Nigerian capital city, Abuja are switching to the use of Hausa in their daily life (Tsakuwa et al 2023). Whereas, Pidgin is being used in daily conversations by almost all southern Nigerians instead of Igbo, Yoruba, Efik, Ishekiri, Ejagham, etc. Use of Pidgin is culturally embedded and endemic in the southern part nowadays (Callies 2024). Sometimes, it is hard to differentiate languages and dialects in Nigeria. Take for instance, languages like Bolewa and Kwamanci in the Kwami local government area of Gombe State, people take them as one language, but, they are two different languages (Aliyu

et al 2000). This kind of linguistic situation can be a challenge when using AI to document endangered languages.

Data Treatment: AI systems require large amounts of data to learn and process language information effectively. However, most indigenous Nigerian languages are either undocumented properly or under-documented in the digitized format. This could make it difficult for AI to analyze and understand the complexities of the endangered languages and their cultures. Therefore, efforts must be made to collect and digitize relevant data for AI to be effective in this context. The quality of available data to be used in training AI models should align with the dialects, accents, and orthographies of the languages (Geeksforgeeks 2024). This is because AI algorithms can be trained on biased data which can result in creating misconceptions or lead to one-sided representations of languages and cultures. Hence, AI tools should ensure inclusiveness and fair representation of the diverse linguistic and cultural landscape of Nigeria's endangered languages. Furthermore, the use of AI must respect the cultural significance and community norms of the endangered languages, ensuring that technology empowers rather than exploits speech communities. Therefore, AI tools must be developed with cultural sensitivity and in collaboration with native speakers. There should be constant application of relevant AI tools on the data in the language documentation and revitalization process under the acceptable ethics put down by the government to support the "participants' language ownership" (Kono, 2019).

Technological Advancement and Infrastructure Constraints: Some of the main challenges are the lack of access to internet connectivity and facilities, and the shortage of technical expertise to develop, deploy, and maintain AI systems for language revitalization in Nigeria. These can hinder the implementation of the necessary technology required to utilize AI tools. To effectively integrate AI into revitalizing endangered languages, modern technology and internet connectivity are needed. Local expertise must be trained to manipulate AI systems for language revitalization. The incompatibility issue between native languages and the current keyboards should be addressed by adapting the keyboards or creating compatible ones which can help in writing the languages properly. And UN's suggestion on developing digital tools for minority languages should taken seriously by large tech companies (UNESCO 2023a).

Another related challenge is social media platforms (like Facebook, WhatsApp, Twitter/X, Instagram, and so on) which seem to be more friendly to English, Hausa, Pidgin, Yoruba, and Igbo in the country. The vast majority of the Nigerian indigenous languages have no presence on social media platforms because the platforms have no language options and no accurate translation into the indigenous languages. Furthermore, the existing search engines that use AI (like Google, Microsoft Bing, etc.) have no provision for most Nigerian languages. Additionally, the technological gadgets (like phones, computers, iPads, etc.) used in accessing such platforms and search engines have no keyboards for writing indigenous languages. In a nutshell, these linguistic

expansions of English, Hausa, and Pidgin, and the technological advancements pose great challenges to the existence of more than 500 native Nigerian languages. These should be addressed by having compatible keyboards, accurate translation and well-trained personnel.

Political Atmosphere: In order to maintain Nigerian indigenous languages, the Nigerian government has good language and education policies that state clearly how the languages should be used in education and other domains. Also, the 1999 Nigerian constitution spells out and allows the use of Nigerian languages in politics and governance. However, there is poor implementation of these policies (Elugbe 2006; Udoh & Anyanwu 2015); there is a lack of political will in the implementation. This can hinder the smooth application of AI tools in preserving the dying languages and the implementation of the technology policy in the country. Also, sustainability is a big problem in Nigeria, and it is known that using AI to decelerate the rate of language endangerment can require long-term funding and support. Hence, funds and relevant AI resources should be sustainable within the complex sociopolitical Nigerian context (Udoh 2022). The government should fund research, provide scholarships for language documentation training of community members on their languages, promote literacy in the endangered languages, and help enlighten the public on language issues and policies (Adebara & Abdul-Mageed 2022).

Another challenge is the application of AI technology which requires a certain level of expertise and understanding, many linguists and language teachers in Nigeria lack such expertise in using AI facilities. There is a need for proper training of language experts and teachers to effectively utilize AI for language revitalization and education (Udoh 2022). To capture this myriad of challenges, I propose the following framework (Figure 1) summarizing the fundamental problems in reviving or decelerating language endangerment in Nigeria.
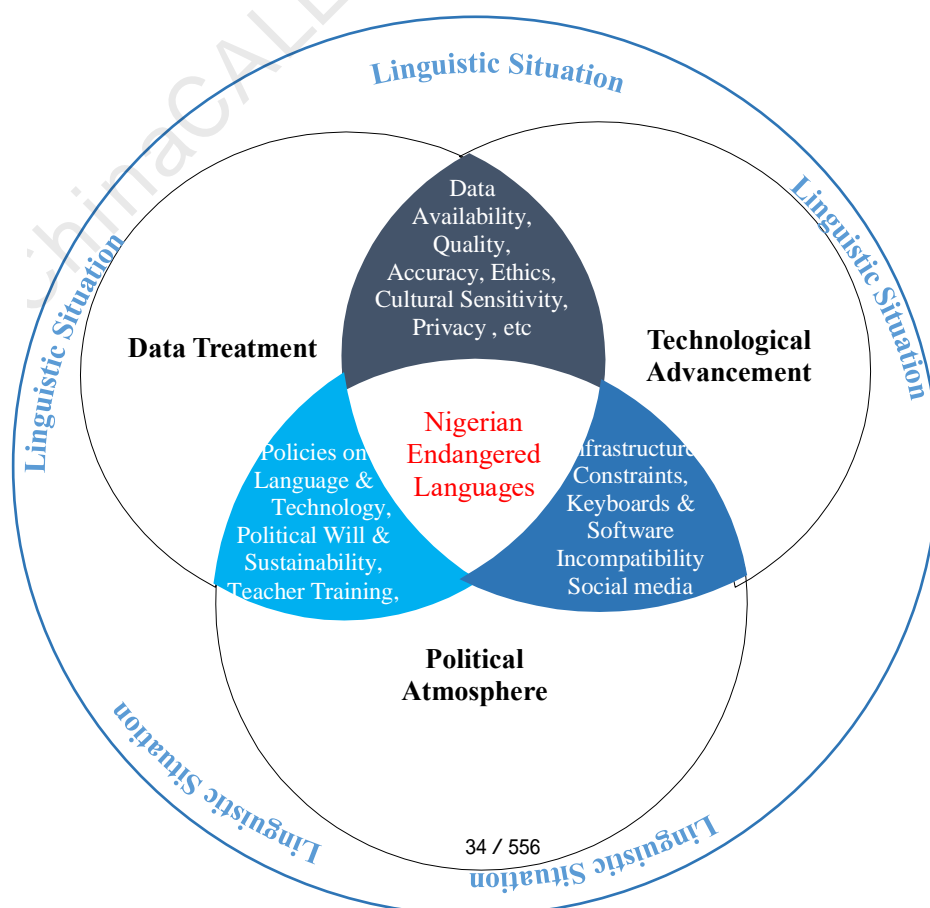
The linguistic situation, socio-political climate, infrastructure constraints, social media and technological advancements are among the potential problems that can impede the use of AI technologies to revive and preserve Nigerian endangered languages. Today, these can be argued to be among the leading factors that 'kill' many Nigerian languages gradually (Tsakuwa, et al 2023). With these, I will further propose a framework (see Figure 2) for collaboration in slowing down the problem of language endangerment in Nigeria. Elugbe (2006) affirms that there should be coordination and collaboration among stakeholders in language revitalization in Nigeria.
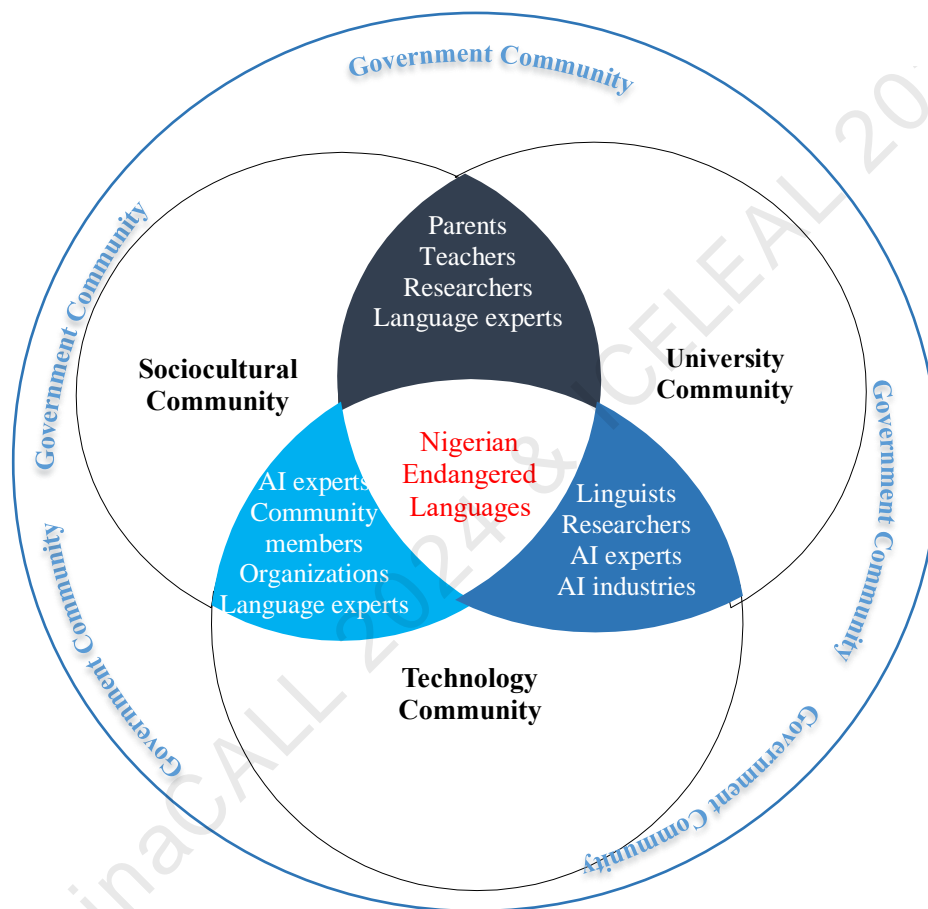
Figure 2: Community Partnership Model for Decelerating Language Endangerment

In viewing the framework clockwise, we are made to understand that the sociocultural community should have a synergy with the university community; thus, parents, community development associations, teachers, language experts and scholars should collaborate in documenting, developing and teaching the indigenous language (Adebara & Abdul-Mageed 2022; Hinton 2018). The university community, with its academics, also work with the technology community of specialists and industry personnel to develop relevant apps and software for the endangered languages. Therefore, linguists, researchers, AI experts, and so on, interact and work together to help the speech community in their language revival and preservation efforts. The technology community, simultaneously, contacts and communicates with the sociocultural community to test the developed

language apps and get feedback. Thus, community members, youth organizations, language experts, AI specialists, and so on, make concerted efforts in revitalizing indigenous languages.

All of these communities have to work in collaboration with the government community to develop and implement relevant policies. Thus, the government can be considered an additional community which is vital in the success of the endangered indigenous language revitalization and preservation which provides favorable policies and political will to be implemented in the process of the language revitalization and preservation.

This model, however, argues that minor endangered indigenous languages cannot be stopped from total extinction; their death can be decelerated through community partnerships among the sociocultural community, university community and technology community which work closely with the government policies and political wills and provide a conducive environment for endangered indigenous language maintenance and revitalization process.

Conclusion

This article presents the significant benefits of using AI in saving endangered languages, some potential challenges and critical issues in the Nigerian context. The paper believes AI can be effectively harnessed to safeguard and promote Nigeria's rich indigenous heritage, ensuring its vitality for future generations. The article also emphasizes the need for collaborative efforts among local communities, linguists, technologists, and policymakers. Since much research has not been done on Nigerian endangered languages (Udoh and Anyanwu 2015), we hope future studies will focus on this issue to practically use AI in reviving the Nigerian endangered languages.

References

Aboh, S.C. (2022). On the parameters of the power of English in Nigeria: A bottom-up approach towards revitalizing Nigerian languages, *Cogent Arts & Humanities*, 9:1, 2118447, DOI: 10.1080/23311983.2022.2118447

Adebara, I. and Abdul-Mageed, M. (2022). Towards Afrocentric NLP for African Languages: Where we are and where we can go. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers, pages 3814 - 3841, Association for Computational Linguistics

Akpan, A.D. (2018). General report on the indigenous Nigerian Sign Language Documentation Project. https://www.s-deli.org/portfolio/general-report-indigenous-nigerian-sign-language-documentation-project/

Akpan, U.J. (2020). Church diplomacy and language documentation in eastern Nigeria. A Festschrift in Honour of Prof. Imelda Icheji Lawrence Udoh. In E. Udoh, and G. Ekpo, (Eds.).

Aliyu, S.A., Shehu, A. & Abba, U. (2000). Gombe State: A history of the land and the people. Zaria: Ahmadu Bello University.

The Ubyssey (19 March, 2024). Dr. Ife Adebara incorporates African languages in AI research. https://www.ubyssey.ca/science/dr-ife-adebara-incorporates-african-languages-in-ai-research/

Ayakoroma, B.F. (2017). Reviving the use of indigenous languages in the contemporary Nigerian society: The National Institute for Cultural Orientation (NICO) Initiative. *International Journal of Language and Linguistics*, 4(4).

Bowern, C. (2022). Drivers of language loss. *Nature Ecology & Evolution* 6, 132–133 https://doi.org/10.1038/s41559-021-01621-x

Bromham, L., Dinnage, R., Skirgård, H. *et al.* (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution* 6, 163–173. https://doi.org/10.1038/s41559-021-01604-y

Callies, M. (2024). Cultural conceptualisations in Nigerian Pidgin English proverbs. World Englishes. 1-20, https://doi.org/10.1111/weng.12660

Elugbe, B. (2006). Documenting endangered Nigerian languages: Challenges and prospects. In: Ihezue, Lizzy; Osuji, Onkwugha E. (Eds.). Proceedings of the national workshop on best practices to safeguard endangered Nigerian languages. Abuja: UNESCO.

Galla, C. K. (2016). Indigenous language revitalization, promotion, and education: function of digital technology. *Computer Assisted Language Learning*, *29*(7), 1137–1151. https://doi.org/10.1080/09588221.2016.1166137

Geeksforgeeks (16 January, 2024). What is language revitalization in Generative AI? What is Language Revitalization in Generative AI? - GeeksforGeeks

Hinton, L, (2018). Approaches to and Strategies for Language Revitalization', in Kenneth L. Rehg, and Lyle Campbell (eds), *The Oxford Handbook of Endangered Languages*, 443–465, https://doi.org/10.1093/oxfordhb/9780190610029.013.22

IBM Research (16 May 2024). Can AI help to promote endangered Indigenous languages? Promoting Endangered Languages with AI - IBM Research

International Telecommunication Union (ITU News) (9 August 2022) How AI is helping revitalise indigenous languages. How AI is helping revitalise indigenous languages - ITU

Kono, N. (2019). Linguistics: Community-based Language Revitalization. In: Iphofen, R. (eds) Handbook of Research Ethics and Scientific Integrity. Springer, Cham. https://doi.org/10.1007/978-3-319-76040-7_45-1

Meighan, P. J. (2021). Decolonizing the digital landscape: the role of technology in Indigenous language revitalization. AlterNative: An *International Journal of Indigenous Peoples*, 17(3), 397-405. https://doi.org/10.1177/11771801211037672

Newitz, A.(27 September 2023) How artificial intelligence is helping keep Indigenous languages alive. How artificial intelligence is helping keep Indigenous languages alive | New Scientist

Pradhan, U., & Dey, J. (2023). Language, artificial education, and future-making in indigenous language education. *Learning, Media and Technology*, 1–14. https://doi.org/10.1080/17439884.2023.2278111

Ruf, J. (12 November, 2020) Digitizing indigenous languages: Scholars use technology to preserve their native languages. Digitizing Indigenous Languages: Scholars Use Technology To Preserve Their Native Languages | Diverse: Issues In Higher Education (diverseeducation.com)

Udoh I. and Anyanwu, O. (2015). Protection and promotion of indigenous Nigerian languages as utility vessels of indigenous knowledge systems (IKS). *Awka Journal of Linguistics and Languages* Vol. 9.

Udoh, I. (2022). Preservation and Development of Nigerian Indigenous Languages.

UNESCO (31 July 2023a) Digital Empowerment Driving the International Decade of Indigenous Languages. Digital Empowerment Driving the International Decade of Indigenous Languages | UNESCO

UNESCO (31 October 2023b) Culture and digital technologies: innovative approaches to cultural projects.Culture and digital technologies: innovative approaches to cultural projects | UNESCO

ChatGPT Applied on Writing Education -- A Systematic Review Based on International Literature

Ji Jiaojiao, Li Siying, Ren Peiling

Beijing University of Posts and Telecommunications

Abstract: ChatGPT, an extensively recognised language model created by OpenAI, has gained significant prominence across various industries, particularly in education. The integration of ChatGPT as a supplementary tool for writing instruction has gained traction. To clarify the application of ChatGPT in writing education, this paper reviewed 22 empirical studies from the Web of Science databases over the past year and a half. This study systematically reviewed these research from four dimensions, including the general research trend, research methodologies, research themes, and promises and challenges of applying ChatGPT in writing education. This study will help scholars in the field of writing better understand the utility and constraints of employing ChatGPT technology in foreign language teaching and can inform educators and researchers in developing effective teaching strategies and in designing curricula.

Keywords: writing; ChatGPT; systematic review

1. Introduction

ChatGPT was developed in 2022 as a novel chatbot rooted in Generative Pre-training Transformer architecture, and outperforms early chatbots in terms of the capability for understanding and producing human-like texts as well as providing feedback on long texts (Su *et al*. 2023). Such affordances make it a powerful writing assistant and writing tool (Barrot, 2023). Chatbots have demonstrated great potential as a writing assistant and learning partner in writing classrooms, as they can provide a broad array of language choices and feedback to students' writing process and make students feel less stressed about their writing performance in the learning process (Guo *et al.* 2022).

This study adopts the method of systemic review to comprehensively map the current status and picture the future trend of employing ChatGPT for writing education. The research is mainly guided by the following four research questions (RQ):

RQ1. What is the general research trend (including language environment, education level and target languages) indicated by the publications?

RQ2. What are the main research methodologies (including research approaches and ways of data collection) employed in the publications?

RQ3. What are the main research themes of the publications?

RQ4. What are the promises and challenges of employing ChatGPT for writing education revealed by the publications?

2. Methodology

2.1 Search strategy

The literature review search strategy must be constructed in order to return scientific publications capable of answering the research questions listed. In the elaboration of the generic string for searches on the bases of scientific publications, these terms were used in English and expanded considering the synonyms most used in the publications. The logical operators OR and AND were inserted between synonyms and keywords. The generic search string used was the following: ("ChatGPT" OR "chatbot") AND ("writing") AND ("education"). Articles published between 1 December 2022 and 30 June 2024 (i.e., within the last year and a half after the release of ChatGPT) were reviewed.

2.2 Selection criteria

This search strategy generated 127 articles that are indexed by Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index(AHCI) and Science Citation Index Expanded(SCI-E). In line with Korpershoek *et al.* (2016), we only selected articles from journals and excluded other sources (e.g., preprints and conference papers). The rigorous review processes associated with journals better assure the methodologically robust of the studies they publish (Korpershoek *et al.*, 2016). Therefore, despite its potential to exclude some relevant studies, this criterion was necessary to facilitate the initial identification of high-quality articles. However, the articles had to report on empirical studies that collected and analysed data from participants to examine the application of ChatGPT. Only English-language articles were included in the review. Table 1 summarises the inclusion and exclusion criteria applied when selecting the articles. Finally, 22 empirical studies were eventually included due to their relevance to our research purpose.

Table 1. Inclusion and exclusion criteria for article selection.

Criterion	Inclusion	Exclusion

Topic	Focused on the application of ChatGPT in writing education	Not focused on the application of ChatGPT in writing education
Study type	Empirical studies that collected and analysed data from student participants; some indicators of student engagement reported	Non-empirical studies; no indicators of student engagement reported
Source	Journals	Sources other than journals
Period	1 December 2022 and 30 June 2024	Articles outside the sample period
Language	English	Not English

2.3 Data coding and data analyses

Nvivo 12.0 was used to code and analyze the articles. We employed this software for annotating the key words, phrases or sentences and making further categorizations. During the coding and categorizing, we referred to the analytical scheme proposed by Chai, Koh and Tsai (2013), the analytical framework summarized by Macaro, Handley and Walter (2012). Four main aspects as follow are analyzed: general research trend, research methodologies, research themes, and promises and challenges of employing ChatGPT in the writing education.

3. Results and Discussions

3.1 General publication trends

Table 2 presents the language environment, education level, and target languages of the reviewed studies. Concerning the language environment, almost half of the studies (n = 9) investigated the foreign language context, followed by the L2 (n = 6) and L1 (n = 4) contexts. In addition, two studies looked into a mix of L1 and L2. In terms of the education level, most of the studies were conducted in high education settings, whereas only 2 studies were carried out in secondary education. For the last variable, the target language of 20 studies was English, and the rest of the studies (n = 2) examined Chinese. English is the dominant target language studied in the reviewed articles, and how to improve English writing effectively with the application of ChatGPT will remain as the central topic in this research field.

Table 2. Research contexts of the reviewed studies

Language environment	No. of studies	Education level	No. of studies	Target language	No. of studies
FL	9 (40.9%)	undergraduate	15 (68.2%)	English	20 (90.9%)
L2	6 (27.3%)	graduate students	2 (9.1%)	Chinese	2 (9.1%)
L1	4 (18.2%)	doctoral students	2 (9.1%)		

| L1+L2 | 2 (9.1%) | professors | 2 (9.1%) |
| FL+L2 | 1 (4.5%) | secondary student | 1 (4.5%) |

Note. FL = foreign language; L1 = first language; L2 = second language.

3.2 Research design

3.2.1 Research approaches

Based on Creswell's (2009) framework, we identified three research approaches in this review, including qualitative, quantitative, and mix-methods approach for data collection and data analysis. The largest number of studies (n =10) adopted qualitative method, followed by mixed methods (n = 8). The quantitative method was used in 4 studies, which implies that more quantitative research should be conducted to compensate the results of qualitative studies.
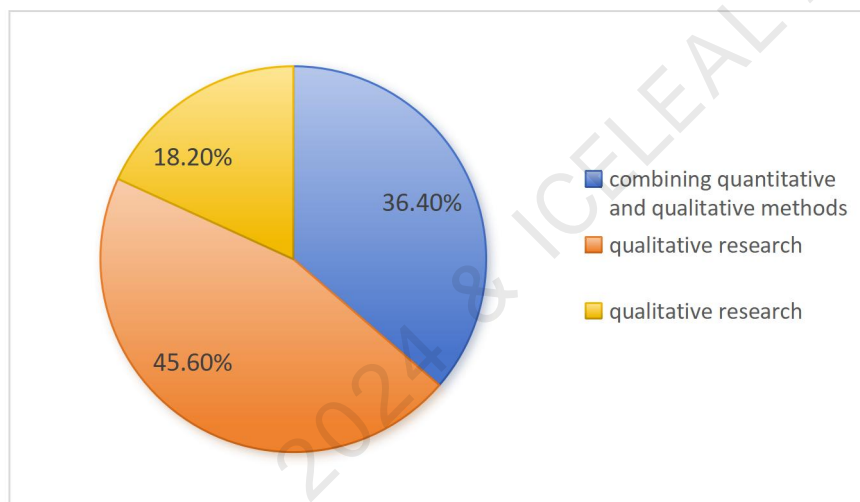
Figure 1. Research approaches of the reviewed studies

3.2.2 Ways of data collection

The data source used in the studies is presented in Figure 2, which represents a variety of data used in the reviewed studies. The largest number of studies included participants' writing texts as the only or one of the data sources (n=16), while interviews (n=12) was the next most frequently used data source. Additionally, ChatGPT-generated feedback (n=10), teacher feedback (n=6), questionnaire (n=5) and peer feedback (n=1) were other types of data used in these studies.
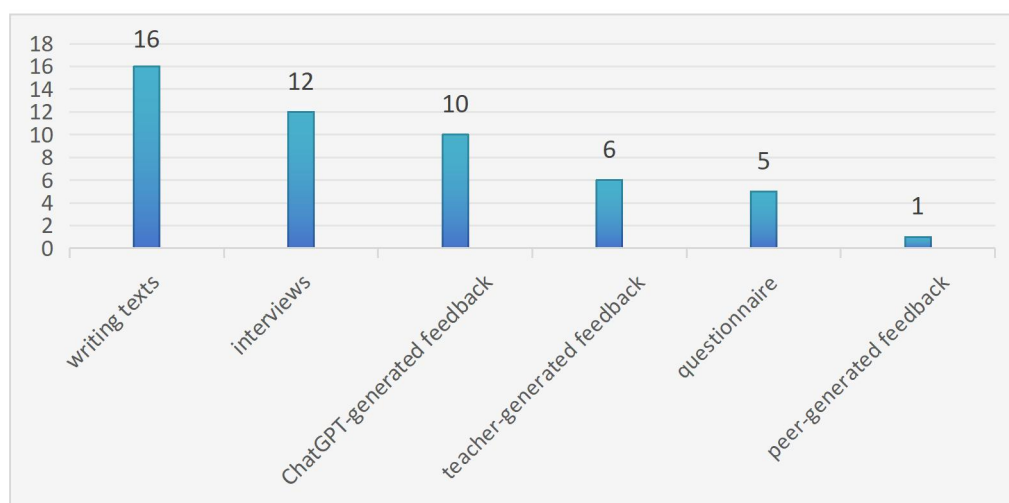
Figure 2. Frequency of different data collection approaches

3.3 Research themes

Our analysis extracted three research themes related to the application of ChatGPT on writing education, including the studies of ChatGPT-generated feedback, language learner characteristics using ChatGPT, and how ChatGPT affect teachers and learners.

Table 3. Research themes of the empirical studies

Category	Sub-categories	Number of Studies
Studies of ChatGPT-generated feedback (7 studies)	Relationships between ChatGPT-generated feedback and teacher feedback	6
	Relationships between ChatGPT-generated feedback and peer feedback	1
Learners' characteristics using ChatGPT (13 studies)	Perceived quality and usefulness	9
	Learners' perceptions/attitudes	8
	Motivation	3
	Anxiety	3
	Engagement	3
	Self-regulation	2
How ChatGPT affect teachers and learners (10 studies)	Progress and development	8
	Impact on teachers' assessment practices	4

3.3.1 Studies of ChatGPT-generated feedback

Among the reviewed articles, we found 7 studies focusing on the relationships among ChatGPT-generated feedback, peer feedback, and feedback from the human tutors, with most of them explored the relationships between ChatGPT-generated feedback and teacher feedback. For example, Qi Lu *et al.* (2023) shed light on the synergy between ChatGPT and teacher assessments.

Guo, K *et al.* (2024) found that ChatGPT produced a significantly larger amount of feedback than teachers and that compared with teacher feedback, which mainly focused on content-related and language-related issues, ChatGPT distributed its attention relatively equally among the three feedback foci (i.e., content, organisation, and language). Banihashem, S. K. *et al.*(2024) suggests a potential complementary role for ChatGPT and students in the feedback process. These studies implied that the link between AI and humans could benefit the students' writing at the forefront of educational systems.

Table 4. Studies of ChatGPT-generated feedback

Sub-categories	Number	Examples
Relationships between ChatGPT-generated feedback and teacher feedback	6	AlGhamdi, R. (2024); Escalante, J. *et al.* (2023); Lu, Q. *et al.*(2024); Steiss, J. *et al.*(2024); Wang, L. *et al.*(2024); Guo, K., & Wang, D. (2024)
Relationships between ChatGPT-generated feedback and peer feedback	1	Banihashem, S. K. *et al.*(2024)

3.3.2 Learners' characteristics using ChatGPT

Many researchers explored how ChatGPT improve learners' writing quality by influencing their psychological characteristics. Among these studies, the highly investigated topics are language learners' perceptions/attitudes towards the quality and usefulness when using ChatGPT. Researchers also put emphasis on participants' motivation, anxiety, engagement, and self-regulation. Zou, M. *et al.*(2023) indicated that attitude was a significant predictor of behavioural intention to use ChatGPT in writing and mediated the impacts of perceived usefulness and perceived ease of use on it. Tossell, C. C *et al.* (2024) explained that students recognize ChatGPT as a collaborative resource requiring human oversight, technical aptitude, subject area proficiency, and calibrated trust. Karataş, F. *et al.*(2024) suggested that ChatGPT can enhance motivation and engagement through its versatile and accessible nature in various learning activities. These studies have significant implications for leveraging ChatGPT for writing in higher education.

Table 5. Learners' characteristics using ChatGPT

Sub-categories	Number	Examples
Perceived quality and usefulness	9	AlGhamdi, R. (2024); Farazouli, A. *et al.*(2024); Zhao, X. *et al.* (2024); Zhang, R. *et al.* (2023); Zou, M. *et al.*(2023); Tossell, C. C.(2024); Barrett, A., & Pack, A. (2023); Karataş, F.(2024)

Learners' perceptions and attitudes	8	AlGhamdi, R. (2024); Escalante, J., Pack, A., & Barrett, A. (2023); Ou, A. W. *et al.* (2024); Zou, M. *et al.* (2023); Zou, M., & Huang, L. (2023); Tossell, C. C.(2024); Karataş, F.(2024); Yan, D. (2023); Yan, D. (2024)

Sub-categories	Number	Examples
Motivation	3	Song, C., & Song, Y. (2023); Guo, K. *et al.*(2024); Karataş, F.(2024); Yan, D. (2023)
Anxiety	3	Song, C., & Song, Y. (2023); Zhao, X. *et al.*(2024); Guo, K. *et al.*(2024)
Engagement	3	Song, C., & Song, Y. (2023); Karataş, F.(2024); Yan, D. (2023)
Self-regulation	2	Song, C., & Song, Y. (2023); Zhang, R. *et al.* (2023)

3.3.3 How ChatGPT affect teachers and learners

Among the articles, we found 10 studies focusing on how ChatGPT affect teachers' self-efficacy and learners' writing skills. ChatGPT exhibits capabilities in aiding the design process, generating ideas aligned with the overall purpose and focus of the paper, producing consistent and contextually relevant responses to various natural language inputs, partially assisting in literature reviews, supporting paper design in terms of both content and format, and providing reasonable editing and proofreading for articles. M. Ghafouri *et al.* (2024) found that employing ChatGPT instructions can help the students with their writing skills and increase the L2 teachers' self-efficacy. AlGhamdi, R. (2024) indicated that ChatGPT-generated feedback can be a valuable tool in enhancing the technical writing skills of computing students. Similarly, the motivational aspects of ChatGPT in language learning were further explored by Ali *et al.* (2023), who found that using AI-based instructions motivates learners and increases their reading and writing skills. In addition, Zhang, R. *et al.* (2023) identified the perceived usefulness of chatbots in consolidating target knowledge of logical fallacies in EFL writing by presenting exercises and immediate feedback. These studies contribute to a more comprehensive understanding of the role of ChatGPT within the context of a combined assessment approach.

Table. 6 How ChatGPT affect teachers and learners

Sub-categories	Number	Examples
Progress and development	8	AlGhamdi, R. (2024); Farazouli, A. *et al.*(2024); Song, C., & Song, Y. (2023); Zhao, X. *et al.* (2024); Zhang, R. *et al.* (2023); Ou, A. W. *et al.* (2024); Ghafouri, M. *et al.*(2024)
Impact on teachers' assessment practices	4	Farazouli, A. *et al.*(2024); Ghafouri, M. *et al.*(2024); Guo, K., & Wang, D. (2024); Barrett, A., & Pack, A. (2023); Matthews, J., & Volpe, C. R. (2023)

3.4 Promises and challenges of the application of ChatGPT on writing education

Using AI tools in L2 classrooms benefits students in various ways, such as providing personalized and adaptive feedback, enhancing writing quality and accuracy, facilitating self-regulation and metacognition, and increasing motivation and engagement (Ali *et al.*, 2023; Gayed *et al.*, 2022; Kohnke *et al.*, 2023). It is recommended that teachers consider incorporating ChatGPT as a supplementary tool for writing assessment. By using ChatGPT's scoring and feedback alongside their own assessments, teachers can reduce their workload and benefit from the strengths of both methods. Students should be guided on how to effectively use both ChatGPT and teacher assessments. This guidance might include teaching them how to compare their writing to assessment criteria and different feedback sources, enabling them to evaluate the feedback's relevance and identify revision strategies. Support should be tailored to students' varying proficiency levels. This approach aims to foster their independence and autonomy as learners and thinkers.

As for challenges, the application of AI-based tools might have negative impacts on the nature and quality of human learning, create more difficulties for L2 teachers, and undermine their role and identity. Thus, it becomes essential for teacher educators to equip the teachers with the knowledge and skills required to approach the forthcoming and ever-updating channels of instruction. Using ChatGPT and similar AI-powered tools requires the direct supervision of teachers as a crucial component of the education system(Ausat *et al.*, 2023).

4. Conclusion

The current study conducted a systematic review of the applications of ChatGPT for writing education. The generative artificial intelligence process is still a black box, and the content it generates lacks interpretability and a clear basis (Lu *et al.*, 2023). AI hold series of promises for language education and also encountered a variety of difficulties. We are still at the dawn of applying ChatGPT to language education in the digital age. Both technological and pedagogical innovations were called for to further maximize the potential of ChatGPT for altering the status quo of language education. Effective and ethical use of AI in education requires acknowledging its limitations, fostering AI literacy, and developing proper assessment methods.

Reference

AlGhamdi, R. (2024). Exploring the impact of ChatGPT-generated feedback on technical writing skills of computing students: A blinded study. *Education and Information Technologies*, 1-26.

Ali, J.K.M., Shamsan, M.A.A., Hezam, T.A., & Mohammed, A.A.Q. (2023). Impact of ChatGPT on learning motivation: Teachers and students' voices. *Journal of English Studies in Arabia Felix*, *2*, 41–49.

Ausat, A.M.A., Massang, B., Efendi, M., Nofirman, N., & Riady, Y. (2023). Can chat GPT replace the role of the teacher in the classroom: A fundamental analysis. *Journal on Education*, *5*, 16100–16106.

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback?. *International Journal of Educational Technology in Higher Education*, *21*(1), 23.

Barrot, J. S. (2023). Using ChatGPT for second language writing: pitfalls and potentials. *Assess. Writ.* 57:100745. doi: 10.1016/j.asw.2023.100745

Chai, C. S., Hwee Ling Koh, J., & Tsai, C. C. (2013). A review of technological pedagogical content knowledge. *Journal of Educational Technology & Society, 16*(2).

Creswell, J. W. (2009). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (3rd ed.). Thousand Oaks, CA: Sage Publications.

Gayed, J.M., Carlon, M.K.J., Oriola, A.M., & Cross, J.S. (2022). Exploring an AI-based writing Assistant's impact on English language learners. *Computers and Education: Artificial*

Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*, 13621688241239764.

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment, 8* (6).

Guo, K., Wang, J., and Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assess. Writ.* 54:100666. doi: 10.1016/j.asw.2022.100666

Guo, K., Li, Y., Li, Y., & Chu, S. K. W. (2024). Understanding EFL students' chatbot-assisted

argumentative writing: An activity theory perspective. *Education and Information Technologies*, *29*(1), 1-20. *Intelligence*, *3*, 100055.

Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, 1-24.

Kohnke, L., Moorhouse, B.L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*, 537–550.

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*(3), 643-680.

Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing?. *Assessment & Evaluation in Higher Education*, 1-18.

Macaro, E., Handley, Z., & Walter, C. (2012). A systematic review of CALL in English as a second language: Focus on primary and secondary education. *Language Teaching, 45*(1), 1-43.

Su, Y., Lin, Y., and Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assess. Writ.* 57:100752. doi: 10.1016/j.asw.2023.100752

Tossell, C. C., Tenhundfeld, N. L., Momen, A., Cooley, K., & de Visser, E. J. (2024). Student perceptions of ChatGPT use in a college essay assignment: Implications for learning, grading, and trust in artificial intelligence. *IEEE Transactions on Learning Technologies*.

Wang, L., Chen, X., Wang, C., Xu, L., Shadiev, R., & Li, Y. (2024). ChatGPT's capabilities in providing feedback on undergraduate students' argumentation: A case study. *Thinking Skills and Creativity*, *51*, 101440.

Yu, H., & Guo, Y. (2023, June). Generative artificial intelligence empowers educational reform: current status, issues, and prospects. In *Frontiers in Education* (Vol. 8, p. 1183162). Frontiers Media SA.

Zhang, R., Zou, D., & Cheng, G. (2023). Chatbot-based learning of logical fallacies in EFL writing: perceived effectiveness in improving target knowledge and learner

motivation. *Interactive Learning Environments*, 1-18.

Zou, M., & Huang, L. (2023). To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model. *Frontiers in Psychology*, *14*, 1259531.

Students' attitude to adopt AI-assisted English language learning applications for speaking learning

Author information

1. Jiajun Li; Jiajun.Li23@student.xjtlu.edu.cn; Xi' an Jiaotong-Liverpool University

2. Yujuan Luo; yujuan.luo@xjtlu.edu.cn; Xi' an Jiaotong-Liverpool University

Abstract

In recent years, with the rapid development of AI-assisted English language learning apps and their application in the education area, students' speaking skills have been improving. However, little scholarly attention has been paid to students' attitudes toward using AI-assisted English language learning apps for speaking practice in the EMI learning context in China. Against this backdrop, this study employed a quantitative research method by utilizing online questionnaires (*n*=103) to gather students' perceptions of using these apps for speaking learning. The participants were students from a Sino-British cooperative university in China. The target context features the unique characteristics of EMI. The results of this study suggested that (1) Most of the participants found the AI-assisted English language learning apps helpful, efficient, and able to improve their pronunciation; (2) In terms of gender differences, female students are more inclined to believe that AI-assisted English learning apps can increase their motivation to practice speaking as compared with male students; (3) The results of the regression analysis showed that "Helpfulness" and "Motivation to use" were significant predictors of students' perceptions of the use of AI-assisted English language learning apps for speaking learning in terms of learning experience; (4) As for the functionality of these apps, "Interface design" and "Technical problems" showed impacts on students' attitudes towards using such apps for speaking learning. This study will shed light on the potential of AI-assisted English language learning apps for improving students' speaking skills from students' perspectives, thereby contributing to the understanding of the role of AI in English speaking acquisition and informing future educational practices.

Keywords: AI-assisted language learning; speaking learning; students' attitude; EFL; higher education

1. Introduction

In recent decades, the rapid advancement of artificial intelligence (AI) has provided innovative

opportunities to improve learners' English language learning skills, especially in the context of English as a Foreign Language (EFL) (Li & Zou, 2023). In terms of language acquisition, researchers have found that AI-supported language learning has been found to be helpful for learners' listening comprehension, speaking, and reading (Wang et al., 2022a). In an English as a Medium of Instruction (EMI) environment, English speaking skills are key to the development of subject content for students. However, English speaking skills are challenging for EFL learners due to the lack of practicing environments (Chen, 2024; Heron et al., 2024).

Among the four core language skills (reading, listening, speaking, writing), speaking is considered to be the most anxiety-provoking skill for learners and can exert a negative influence on oral performance (Chen, 2024). With the popularity of MALL and the integration of AI into CALL, many AI-assisted English language learning apps have been designed to enhance EFL learners' speaking practice (Lehman et al., 2020; Litman et al., 2018; Zhang, 2021; Zou et al., 2023). Previous studies have shown that AI speaking apps can help improve speaking skills, including pronunciation, fluency, and speaking rhythm (Getman et al., 2023; Zou et al., 2023). In addition, other scholars have investigated AI-assisted English language learning apps to improve speaking skills from students' perspectives. For example, a study by Yang et al. (2022) examined Korean EFL students' perceptions of the AI-voiced chatbot 'Ellie' as an English conversation partner. The study showed that the AI voice chatbot greatly encouraged students to engage in dialogue (Yang et al., 2022). Hence, though previous research provides different insights about AI-assisted speaking apps from students' perspectives, limited scholarly attention has been paid to exploring Chinese students' attitudes toward using these kinds of apps for English speaking (Li & Zou, 2022). Besides, since many AI-assisted language learning apps are designed for young children (Getman et al., 2023), little focus has been placed on university students.

To fill the research gap, this study aims to explore students' attitudes toward AI-assisted English language learning apps for speaking learning. This study employs a quantitative research method, utilizing online questionnaires to gather Chinese students' perceptions of using these apps for speaking at a Sino-British cooperative university in China. The target context features the unique characteristic of EMI, and learning technologies are widely used both in the classroom and in students' self-directed learning processes (Tao & Zou, 2023). The aim of this study is to deepen

2

the understanding of the role of AI in English-speaking acquisitions and, thus, to provide guidance for future educational approaches.

This study is guided by the following two research questions:

1. What are students' attitudes of using AI-assisted English language learning apps for their speaking learning? Is there any gender difference?

2. What are the factors influencing students' willingness to use AI-assisted English language learning apps for speaking learning?

2. Literature review

2.1 The Role of AI in EFL Learning Context and AI-assisted ELL Apps for EFL learner's speaking Learning

As AI penetrates into all areas of human life, the application of AI in Education (AIEd) has grown dramatically in recent years (Hwang et al., 2020). The application of AI in language learning has also received much attention. In an EFL learning environment, one of the biggest challenges for EFL students is the lack of practice environments (Lin & Mubarok, 2021). Numerous studies have shown that AI can provide authentic opportunities for practice and bring benefits to EFL learning. For example, through a comparative study, Huang et al. (2023) explore students' learning performance and engagement in an AI-supported EFL classroom by offering students personalized videos. The results indicate that learners' learning outcomes and engagement levels have been improved in the AI-driven EFL classroom (Huang et al., 2023). Considering the important role of AI tools in the EFL context, it is crucial to understand how students perceive and use these tools to enhance their language proficiency.

Speaking is a core skill to communicate effectively in real-life situations (Jabber & Mahmood, 2020; Kohn & Hoffstaedter, 2017; Li & Chan, 2024; Wan & Moorhouse, 2024). Many AI-based speaking apps are widely used for EFL learners' speaking practice such as Duolingo and EAP Talk (Loewen et al., 2020; Chang et al., 2022). Numerous studies have shown that these apps can provide benefits to EFL learners for speaking practice. For example, Fathi et al. (2024) investigated the role of AI-assisted English language leaning app in supporting students' speaking skills. Their findings suggested that AI can promote student autonomy in the area of speaking (Fathi et al., 2024). However, some scholars have also pointed out challenges that AI-assisted

3

English language learning applications pose to EFL learners. In Evers and Chen's (2020) study, they suggest some AI programs (e.g. Speechnote) were less sensitive to non-native accents, resulting in less feedback on accents and demotivating learners to use the software for speaking practice. Due to these inconsistencies, it is of paramount importance to explore how students perceive AI as useful tools for speaking practice.

2.2 Students' Perceptions and Attitudes Towards AI-Assisted ELL Learning apps

The application of AI provides an effective teaching method for students' language learning and fosters their positive perceptions and attitudes towards AI (Sun et al. 2021; Zhang, 2022; Xia et al., 2022; Zhang, 2022; Zhang and Zou, 2020). This can also foster their positive perceptions and attitudes towards AI (Xia et al., 2022). For example, Zou et al.'s (2023) study investigated EFL students' perceptions of EAP Talk (an AI application for academic speaking practice), the results of the study indicated that students generally had positive attitudes towards EAP Talk, which they found valuable and easy to use, and were highly motivated to use. However, students also reported problems with the design of the interface and the lack of gamification features, which undermined their motivation to use EAP Talk (Zou et al., 2023). A similar issue is mentioned in the study by Wang et al. (2024), that they point out the interface design of AI is also a non-trivial issue, as even subtle changes can affect the learner's perception of AI. However, it remains uncertain whether and how learners' perception of AI appearance will affect their attitude of using AI-assisted English language learning apps for their speaking learning. Therefore, this study will continue to explore how AI-assisted English language learning apps will affect students' attitudes towards using such apps to practice speaking.

As stated above, while numerous researches have discussed the implication of AI-assisted English language learning apps from students or learners' perspectives, very little studies focus on university students' attitudes toward using such kind of apps for speaking practice, especially in the English as a Medium Instruction (EMI) context in China. For studies that explore university students' attitudes toward using English language learning apps for English speaking practice, the target research context focusing more on EFL learning. Zou et al.'s (2023) study is an example. In their research, totally 218 Chinese EFL university students were regulated to use EAP Talk (an artificial speech evaluation system), the results represented that students generally hold positive

attitudes towards EAP Talk. However, this study has several limitations. For instance, it only focusing on a single AI-assisted English language learning app. Based on the results of the previous studies, it is shown that speaking proficiency play an important role in the EMI leaning context. Besides, in China, it is also noted that the definition of the EMI echoes the current trends in the process of Chinese higher education, where there is a move from Chinese as the medium of instruction (CMI) toward English as the medium of instruction (EMI) at universities striving for internationalization (Rose et al., 2020; Xiao & Qiu, 2022). Nevertheless, in the area of AI-assisted English language learning, few studies have explored students' attitudes toward using AI-assisted English language learning apps for speaking practice under the EMI learning context. Therefore, this study intends to further explore this area by investigating an EMI university's students' attitudes towards using AI-assisted English language learning apps for speaking.

3. Research Methodology

3.1 Sampling and Participants

This study was administrated at a Sino-British cooperative university in China. The convenience sampling strategy allowed the researchers to collect a diverse range of perspectives, including students from different grades (both undergraduate and postgraduate), various academic backgrounds, and varying levels of English proficiency. The majority of the student participants were female (83.5%). These student participants were from a variety of disciplinary clusters, including ten different disciplinary clusters such as Education (32.04%), Business (25.24%), and Humanities and Social Sciences (11.65%) (see Table 1).

5

Table 1 Participants' demographic information

Gender	Major related to the area	Proportions for major
Female=86 (83.5%)	Education	32.04%
	Business	25.24%
	Humanity and Social Sciences	11.65%
	Advanced Technology	5.83%
	Film and Creative Technology	4.85%
	Pharmacy	4.85%
	Science	4.85%
Male=17 (16.5%)	Mathematics and Physics	3.88%
	Design	3.88%
	Language	2.91%

3.2 Instrument

In this study, a five-point Likert scale questionnaire from 1 - Strongly Agree to 5 - Strongly Disagree containing 19 items was used to investigate the attitudinal elements of students' attitudes of AI-assisted English language learning applications for speaking learning. For the designing of the questionnaire, items 1-4 aimed to collect demographic information about the participants. Items 5-13 investigated students' attitudes of using these kinds of apps for speaking practice. The questionnaire design referred to the section of *"Participant perceptions of AI apps for EFL learning"* in Zou et al. (2023) 's study. The questionnaire design effectively captures students' attitudes on various aspects of AI-assisted English language learning apps. Item 14 aimed to find out how often students use these apps. Items 15-18 focused on students' perceptions about technical functions and the personalized features of these apps, aiming to examine whether these factors would affect students' perceptions of using these kinds of apps for speaking practice. As Zou et al. (2023) 's study pointed out that students reported some problems related to the interface design, we therefore referred questions about the interface design and personalization features in our questionnaire in the hope of obtaining more general results. Item 19 used open-ended questions to collect students' suggestions.

3.3 Reliability and Validity Tests

To ensure consistency of the results obtained and to minimize measurement error, the questionnaire was tested for reliability before proceeding to other statistical analysis. As shown in

Table 2, the Cronbach's alpha value of the questionnaire was 0.911. Since the critical value for this type of cognitive questionnaire was 0.8, 0.911 was satisfactory and indicated that the questionnaire had a high level of reliability (Kline, 1999).

Table 2 Reliability analysis of the questionnaire used

Cronbach's Alpha	N of items
0.911	12

In order to test the factorizability of the data, we also performed the Bartlett's test of sphericity and calculated the Kaiser-Meyer-Olkin (KMO) value for sampling adequacy (Pallant, 2020). Table 3 showed that the KMO value was equal to 0.908, which was close to 1. According to Kaiser (1974), a KMO value higher than 0.6 implies that the sampling is adequate for factor analysis. In addition, Bartlett's p-value is less than 0.05 which indicates appropriate significance. Therefore, the questionnaire in this study had high validity and was suitable for further analysis.

Table 3 Kmo and Bartlett's test

Kmo measure of sampling adequacy		0.908
Bartlett's test of sphericity	Approx. Chi-square	668.483
	df	66
	p value	0

3.4 Data Collection and Analysis

Data from online questionnaire were analyzed by SPSS to complete quantitative data analysis. The descriptive statistics aimed to summarize students' responses and identify overall attitudes towards AI-assisted apps for speaking practice. Measures like means, standard deviations, and frequencies provided insights into the distribution of responses and the prevalence of certain perceptions. Independent samples t-tests was used to examine potential gender differences. Besides, to explore the relationships between variables, correlation analysis and multiple linear analysis were conducted to reveal potential associations between factors such as app usage frequency and willingness to use the apps.

4. Results and Discussions

4.1 Students Attitudes toward Using AI-assisted ELL Apps for Speaking

4.1.1 Result of the Descriptive Analysis

The results of the descriptive analysis (see Table 4) reveal that students generally have positive attitudes toward using AI-assisted ELL apps for speaking practice, with mean scores indicating

that students commonly express positive feelings on items like helpfulness (mean=1.87), efficiency (mean=1.87), enjoyment (mean=1.89), willingness for self-directed learning (mean=1.92), and improvement of pronunciation (mean=1.94). Additionally, students express positive feelings about items including willingness to speak English and express ideas in public (mean=1.98), motivation (mean=1.99), interest in using the apps (mean=2.01), and willingness for teachers to use them (mean=2.27).

8

Table 4 Descriptive statistics (Note: Scale ranging from 1 – strongly agree to 5 – strongly disagre

	N	Minimum	Maximum	Mean
I like to use AI-assisted English language learning apps to practice my English speaking.	103	1	5	1.89
Using AI-assisted English language learning apps is helpful for me to learn English speaking.	103	1	4	1.87
Using AI-assisted English language learning apps can improve my efficiency in learning spoken English.	103	1	3	1.87
Using AI-assisted English language learning apps makes me more willing to speak English and express my ideas in the public.	103	1	5	1.98
Using AI-assisted English language learning apps improves my spoken English pronunciation.	103	1	5	1.94
Using AI-assisted English language learning apps makes me more motivated to learn English speaking.	103	1	5	1.99
I find it interesting when using AI-assisted English language learning apps to learn English speaking.	103	1	5	2.01
I would like my teacher to use AI-assisted English language learning apps more frequently in my future classroom.	103	1	5	2.27
I am willing to use AI-assisted English language learning apps for self-directed learning outside the classroom to practice my English speaking.	103	1	4	1.92

4.1.2 Result of the Independent samples *t* tests

Interestingly, the analysis of the questionnaire data shows that male and female participants hold different attitudes toward the use of AI-assisted English language learning apps for practicing their

9

English-speaking skills (see Figure 1). An independent samples t-test is conducted to determine the gender differences in the participants' responses. The p-values for the scales range from 0.01 to 0.05 (p = 0.041), indicating a significant difference in the mean scores of the attitudes held by female students (M=1.85) and male students (M=2.29) toward the use of AI-assisted English language learning apps for speaking practice. Specifically, female students score significantly higher than male students on the scale "I would like to use AI-assisted English learning apps for independent learning outside the classroom to practice speaking English." (t=2.066, p=0.041).

Figure 1. Distribution of mean scores on the willingness to use AI-assisted ELL apps scale for men and women

■ I am willing to use AI-assisted English language learning apps for self-directed learning outside the classroom to practice my English speaking.
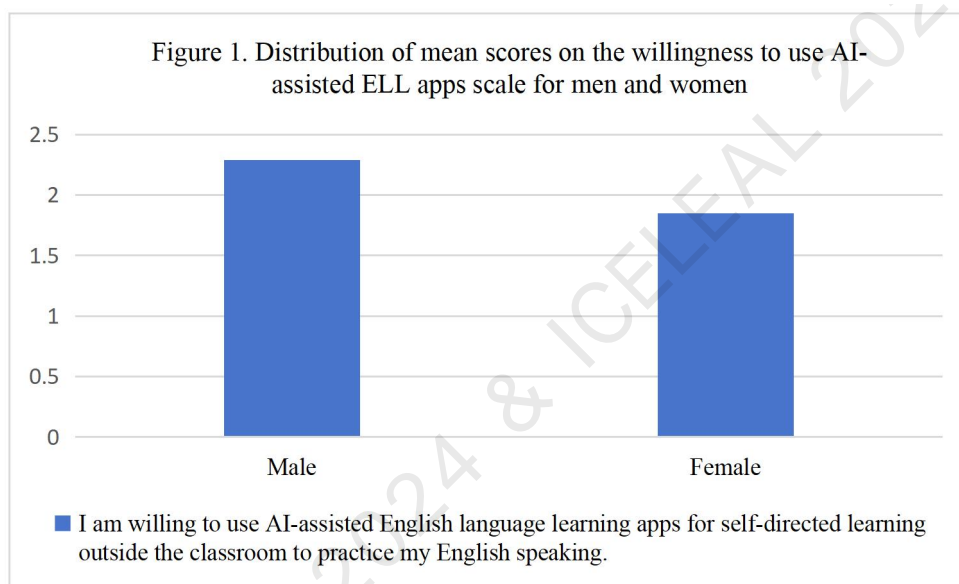
4.2 Factors Affecting Students' Attitude to Use AI-assisted ELL Applications for Speaking

To explore factors influencing students' willingness to use AI-assisted ELL applications for English speaking learning, a preliminary correlation analysis is conducted to ensure that the variables selected for the multiple linear regression analysis are not highly correlated with one another (see Table 5). The analysis reveals that variables such as helpfulness (r=0.638, p≤0.01), efficiency (r=0.571, p≤0.01), willingness (r=0.657, p≤0.01), improvement of pronunciation (r=0.480, p≤0.01), motivation (r=0.534, p≤0.01), interface design (r=0.353, p≤0.01), technical problems (r=0.156, p≤0.01), and features for social interaction (r=0.164, p≤0.01) have moderate to low correlations with each other. The correlation coefficients also indicate that these variables can be included in the regression model without significant multicollinearity concerns.

10

Table 5 Correlations among attitudes towards using AI-assisted apps and actual usage

Construct		1	2	3	4	5	6	7	8	9
Attitudes toward using AI-assisted apps	Coefficient	1								
	p-value									
Helpfulness	Coefficient	.638**	1							
	p-value	0.000								
Efficiency	Coefficient	.599**	.571**	1						
	p-value	0.000	0.000							
Willingness	Coefficient	.607**	.579**	.657**	1					
	p-value	0.000	0.000	0.000						
Improvement of pronunciation	Coefficient	.401**	.357**	.610**	.480**	1				
	p-value	0.000	0.000	0.000	0.000					
Motivation	Coefficient	.610**	.530**	.659**	.733**	.534**	1			
	p-value	0.000	0.000	0.000	0.000	0.000				
Interface design	Coefficient	.486**	.473**	.442**	.411**	.339**	.353**	1		
	p-value	0.000	0.000	0.000	0.000	0.000	0.000			
Technical problems	Coefficient	.256**	.218*	0.099	0.110	0.038	0.076	0.146	1	
	p-value	0.009	0.027	0.322	0.270	0.706	0.446	0.142		
Features for social interaction	Coefficient	.301**	.406**	.380**	.378**	.211*	.327**	.271**	0.164	1
	p-value	0.002	0.000	0.000	0.000	0.032	0.001	0.006	0.097	

Note. $* = p \leq 0.05$; $** = p \leq 0.01$

Subsequently, a multiple linear regression analysis is conducted to investigate the effects of independent variables on the dependent variable (see Table 6). The dependent variable is chosen based on students' attitudes toward using AI-assisted English language learning apps. The independent variables are chosen based on their relevance to the study's objectives and include items related to the learning experience, such as "helpfulness," "efficiency," "willingness," "improvement of pronunciation," and "motivation." For the features of AI-assisted ELL applications, "interface design," "technical problems," and "features for social interaction" are included as independent variables.

The evaluation of the regression model shows that the value of R-squared (R^2) equals 0.574, meaning the model explains 57.4% of the variance in the dependent variable. This value indicates that the model has considerable explanatory power. The F-value (F=15.855) and the level of significance (Sig < 0.001) indicate that the independent variables in the model have a significant

11

effect on the dependent variable. These statistical indicators suggest that the multiple linear regression model in this study is valid and can be used to explore the influence of learning experience and app features on learners' use of AI-assisted English learning applications. In the next two sections, we will further explore the extent and direction of the influence of each independent variable on the dependent variable.

Table 6 The linear regression analysis

	B	S.E.	β	*t*	P	R²	F and Sig
Constant	-0.213	0.257		-0.829	0.409		
Helpfulness	0.298	0.098	0.282	3.029	0.003		
Efficiency	0.181	0.123	0.159	1.472	0.144		
Willingness	0.111	0.099	0.123	1.126	0.263		
Improvement of pronunciation	-0.025	0.085	-0.025	-0.29	0.773		F=15.855
Motivation	0.227	0.105	0.232	2.154	0.034	0.574	Sig＜
Interface design	0.146	0.076	0.153	1.926	0.057		0.001
Technical problems	0.148	0.076	0.136	1.952	0.054		
Features for social interaction	-0.059	0.081	-0.055	-0.721	0.473		

4.2.1 Students' Attitude, Learning Experience and Features of AI-assisted ELL Applications

The results of the multiple linear regression analysis reveal several significant predictors of students' willingness to use AI-assisted English learning applications in their oral English learning. First, Helpfulness has a significant p-value of 0.003 (p < 0.05), indicating that the more helpful students perceive the use of AI-assisted ELL apps to be in their oral English practice, the more likely they are to use these apps. Secondly, motivation also has a significant p-value of 0.034 (p < 0.05), indicating that students who believe that AI-assisted ELL apps can enhance their motivation to practice English speaking are more likely to use these apps.

The regression analysis explored the impact of specific features of AI-assisted ELL applications on students' attitudes towards using them. Notably, while the p-values for interface design (p=0.057) and technical problems (p=0.054) approached the conventional significance level of 0.05, they did not meet the threshold for statistical significance. However, the trends indicated a positive correlation between these features and students' attitudes. This suggests that with a larger

sample size or more refined measurement, the ease of navigation in interface design and the absence of frequent technical issues could emerge as significant factors influencing students' willingness to engage with these apps for speaking practice.

4.3 Discussion

The results of this study provide valuable insights into students' attitudes toward AI-assisted English language learning apps for speaking.

Regarding the first research question, students generally hold positive attitudes towards using these apps for learning to speak. They consider helpfulness, efficiency, and enjoyment as the three most positive factors when using these apps for speaking practice. These findings align with the studies by Zou et al. (2023) and Fathi et al. (2024) on students' attitudes and acceptance of AI-assisted apps for speaking practice in EFL learning. Furthermore, there was a significant difference between male and female students' attitudes towards using these apps for independent learning outside the classroom. This aligns with previous studies suggesting that gender is a key factor in the development of speaking anxiety among EFL learners, with female learners being more likely to develop anxiety than male learners (Bensalem, 2018; Luo, 2014; Öztürk & Gürbüz, 2014).

For the second research question, we explored factors affected students' attitudes toward using AI-assisted English language learning apps for speaking practice. The regression analysis indicated that the helpfulness and motivational effects were important factors in students' attitudes towards using such apps. The more helpful students found the use of apps in their speaking practice, the more likely they were to use them. Besides, students who perceived these apps could enhance their motivation to practice their speaking learning were more likely to use them. These emphasized the importance of designing apps that were engaging and supported students' learning goals. As Wang et al. (2024) pointed out in their study, fondness for the appearance of an AI positively predicted learners' interest in learning, and whether the AI had a pleasing appearance may influence students' learning in some way. An unappealing appearance may cause anxiety and upset while learning, whereas an appealing appearance may reduce this negative emotional experience (Randall, 2019; Wu et al., 2015). Additionally, the result of regression analysis revealed a a trend where students who perceived the interface design as user-friendly and

13

encountered fewer technical difficulties tended to express a greater preference for utilizing these apps for oral English practice. However, it is important to note that these findings did not reach statistical significance, suggesting that while there may be a relationship, it is not definitive within the context of this study. This could imply the need for further research with a larger sample size or more nuanced analysis to confirm the strength and significance of these correlations. The result was in line with the findings of Zou et al. (2023). This study analyzed students' acceptance of EAP Talk for speaking practice and the findings reported issues related to user interface design, and accuracy of automated feedback (Zou et al., 2023).

5. Conclusion

The aim of this study was to investigate students' attitudes towards the use of AI-assisted English learning apps to improve speaking skills. By analyzing the responses of 103 students from a Sino-British partner university, we found that students generally have a positive attitude towards these apps. They consider AI-assisted apps to be helpful, efficient, and enjoyable, and believe that using these apps is beneficial for improving their English pronunciation, fluency, and motivation. Our findings also reveal interesting insights regarding gender differences and factors influencing students' attitudes towards using these apps. These findings suggest that by focusing on the factors that influence student attitudes and addressing areas for improvement, developers and educators can create more effective and engaging AI learning tools.

There are also some limitations to this study. Firstly, the sample size was small, including only 103 students, which may affect the generalizability of the results. Secondly, this study was limited to one Sino-British cooperative university, so the findings may not be applicable to other types of higher education institutions in China with EMI learning and teaching environments. Additionally, this study only used a quantitative research method, which may lack a deeper understanding of the issues. While this study provides valuable insights into university students' attitudes towards AI-assisted English language learning apps in an EMI environment in China, future research can address these limitations by exploring a larger and more diverse sample size and incorporating qualitative methods to further explore more feasible conclusions.

References

Bensalem, E. (2018). Foreign language anxiety of EFL students: Examining the effect of self-efficacy, self-perceived proficiency and sociobiographical variables. *Arab World English Journal, 9*(2), 38–55. https://doi.org/10.24093/awej/vol9no2.3

Chang, Y., Lee, S., Wong, S. F., & Jeong, S. P. (2022). AI-powered learning application use and gratification: an integrative model. *Information Technology & People*, *35*(7), 2115-2139.

Chen, Y. C. (2024). Effects of technology-enhanced language learning on reducing EFL learners' public speaking anxiety. *Computer Assisted Language Learning*, *37*(4), 789-813.

Chen, Y.-M. (2019). Learners' perceptions of using technology in the foreign language classroom: A case from China. *Journal of Physics: Conference Series, 1168,* 032009. https://doi.org/10.1088/1742-6596/1168/3/032009

Chen, Z. (2024). A Study of Chinese Undergraduate Students' English Language Speaking Anxiety, Expectancy-Value Beliefs and Spoken English Proficiency. *SAGE Open*, *14*(1), 21582440231219312.

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, *121*, 103254.

Getman, Y., Phan, N., Al-Ghezi, R., Voskoboinik, E., Singh, M., Grosz, T., Kurimo, M., Salvi, G., Svendsen, T., Strombergsson, S., Smolander, A., & Ylinen, S. (2023). Developing an AI-Assisted Low-Resource Spoken Language Learning App for Children. *IEEE Access, Access, IEEE, 11*, 86025–86037.

https://doi.org/10.1109/ACCESS.2023.3304274

Heron, M., Dippold, D., Aksit, N., Aksit, T., Doubleday, J., & McKeown, K. (2024). Tutors' beliefs about language and roles: Practice as language policy in EMI contexts. *Teaching in Higher Education*, *29*(1), 251-265.

Huang, A. Y., Lu, O. H., & Yang, S. J. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, *194*, 104684.

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial

15

Intelligence, 1, 100001.

Jabber, K. W., & Mahmood, A. A. (2020). Non-verbal communication between two non-native English speakers: Iraqi and Chinese. *Theory and Practice in Language Studies, 10*(2), 189–196. https://doi.org/10.17507/tpls.1002.06

Kaiser, H. (1974). A computational starting point for Rao's canonical factor analysis: Implications for computerized procedures. *Educational and Psychological Measurement, 34*(3), 691–692. https://doi.org/10.1177/001316447403400322

Kline, P. (1999). *The handbook of psychological testing (2nd ed.).* Routledge.

Kohn, K., & Hoffstaedter, P. (2017). Learner agency and non-native speaker identity in pedagogical lingua franca conversations: Insights from intercultural telecollaboration in foreign language education. *Computer Assisted Language Learning, 30*(5), 351–367. https://doi.org/10.1080/09588221.2017.1304966

Lehman, B., Gu, L., Zhao, J., Tsuprun, E., Kurzum, C., Schiano, M., Liu, Y., & Jackson, G. T. (2020). Use of adaptive feedback in an app for English language spontaneous speech. In *Lecture notes in computer science* (pp. 309–320). https://doi.org/10.1007/978-3-030-52237-7_25

Li, M., & Zou, B. (2022). Using artificial intelligence technology as a tutor for EFL speaking practice. *Smart CALL: Personalization, Contextualization, Socialization*, 82-100.

Li, Q., & Chan, K. K. (2024). Test takers' attitudes of using exam-oriented mobile application as a tool to adapt in a high-stakes speaking test. *Education and Information Technologies, 29*(1), 219–237. https://doi.org/10.1007/s10639-023-12297-0

Lin, C. J., & Mubarok, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society*, *24*(4), 16-35.

Litman, D., Strik, H., & Lim, G. S. (2018). Speech technologies and the assessment of second language speaking: Approaches, challenges, and opportunities. *Language Assessment Quarterly*, *15*(3), 294-309.

Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign*

16

Language Annals, *53*(2), 209-233.

Luo, H. (2014). Foreign language speaking anxiety: A study of Chinese language learners. *Journal of the National Council of Less Commonly Taught Languages, 15,* 99–117.

Öztürk, G., & Gürbüz, N. (2014). Speaking anxiety among Turkish EFL learners: The case at a state university. *Journal of Language and Linguistic Studies, 10*(1), 1–17.

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.

Rose, H., McKinley, J., Xu, X., & Zhou, S. (2020). Investigating policy and implementation of English medium instruction in higher education institutions in China. *British Council*.

Sun, Z., Anbarasan, M., & Praveen Kumar, D. J. C. I. (2021). Design of online intelligent English teaching platform based on artificial intelligence techniques. *Computational Intelligence*, *37*(3), 1166-1180.

Tao, Y., & Zou, B. (2023). Students' perceptions of the use of Kahoot! in English as a foreign language classroom learning context. *Computer Assisted Language Learning*, *36*(8), 1668-1687.

Wan, Y., & Moorhouse, B. L. (2024). Using Call Annie as a generative artificial intelligence speaking partner for language learners. *RELC Journal.* https://doi.org/ 10.1177/00336882231224813

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2024). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, *37*(4), 814-840.

Wu, W. C. V., Wang, R. J., & Chen, N. S. (2015). Instructional design using an in-house built teaching assistant robot to enhance elementary school English-as-a-foreign-language learning. *Interactive Learning Environments*, *23*(6), 696-714.

Xia, Q., Chiu, T. K., Lee, M., Sanusi, I. T., Dai, Y., & Chai, C. S. (2022). A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education. *Computers & Education*, *189*, 104582.

Xiao, Y., & Qiu, X. (2022). A study of the relationship between students' global perspective and willingness to communicate in english at an english medium instruction university in

17

China. *Frontiers in Psychology*, *13*, 873766.

Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, *34*(3), 327-343.

Zhang, Q. (2021). An automatic assessment method for spoken English based on multimodal feature fusion. *Wireless Communications and Mobile Computing, Wireless Communications and Mobile Computing, 2021*(6), 1–12. https://doi.org/10.1155/ 2021/1045184

Zhang, R., & Zou, D. (2022). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, *35*(4), 696-742.

Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *Sage Open*, *13*(3), 21582440231193818.

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability*, *15*(4), 2872.

Zou, B., Lyu, Q., Han, Y., Li, Z., & Zhang, W. (2023). Exploring students' acceptance of an artificial intelligence speech evaluation program for EFL speaking practice: An application of the Integrated Model of Technology Acceptance. *Computer Assisted Language Learning*, 1-26.

Research on the "San Jiao" Reform Path of College English in Polytechnics with the Background of Intelligent Education

Li Shenzhu

(Jinan Engineering Polytechnic, Jinan 250200, Shandong, PRC)

Abstract: Intelligent education is a new requirement for classroom pattern remodeling in the era of artificial intelligence, which gives new connotation to online and offline blended teaching. On the basis of the relevant research results of intelligent educational reform, this study defines the novel connotation of the "San Jiao" reform of college English in polytechnics with the background of smart education, constructs a new path to enhance virtual wisdom teaching reform, and verifies its effect through teaching practice inspection and instructing evaluation. The research results show that the smart "San Jiao" reform model combines the "intelligence" of AI teaching assistants with the "wisdom" of authentic teachers, which effectively stimulates students' enthusiasm to actively participate in class, and greatly triggers poly-technical students' autonomous consciousness of meaningful learning and active construction of knowledge system.

Key words: intelligent education; enhanced virtuality; "San Jiao" reform; collaborative teaching

1. Introduction

High quality classroom teaching is the core content of constructing high quality vocational education system. As the most active factor in classroom teaching activities, classroom interaction is an important factor affecting teaching quality and learning effect[1]. Many studies have shown that classroom discussion and innovation, peer interaction and cooperation, teacher-student interaction and communication and other interactive participation variables have an important impact on strengthening the main position of the classroom and effectively improving the quality of classroom teaching[2]. The Central Committee of the Communist Party of China and The State Council issued the Opinions on Promoting the High-quality Development of Modern Vocational Education, requiring "promoting the deep integration of modern information technology and education and teaching, and improving the quality of classroom teaching". As a new model and

new form for the application of emerging technologies in the information age, smart classroom is an ideal learning environment for implementing the education informatization 2.0 action plan[3]. Therefore, it has become an unavoidable topic of modern vocational education to explore the deep classroom interaction of intelligent classrooms in vocational colleges. In the context of smart education, we should make full use of the technical advantages of smart classrooms to enable in-depth classroom interaction in vocational colleges, so as to improve the quality of classroom teaching in vocational colleges, give full play to the main position value of classroom teaching in the process of cultivating high-quality technical and skill talents in the new era, and promote the high-quality development of modern vocational education.

In the information age, the application of educational technology and the innovation of educational ideas promote the profound change of classroom teaching mode. The emerging "AI + education" model uses big data, cloud computing, virtual reality and other technologies to have a strong impact on traditional education concepts, models, methods and content. How to integrate AI into the whole teaching process and effectively support classroom reform and innovation has become an urgent problem for foreign language education and teaching. This study attempts to build an enhanced virtual hybrid teaching model of "AI assistant teacher + authentic teacher" collaboration, and discusses the effect of dual-teacher teaching integrating wisdom elements in higher vocational colleges, in order to provide new ideas and references for the teaching practice and research of dual-teacher wisdom.

2. The connotation of intelligent dual teacher classroom and enriched virtual classroom

2.1 Intelligent dual teacher classroom

The two-teacher classroom originates from the traditional "professor + teaching assistant" mode, and its connotation becomes richer and richer with the continuous development of the "teaching assistant" function. Dual teacher classroom is the inevitable product of the development of "courier classroom" and "famous teacher classroom" to a certain stage(Zheng Xudong et al. 2021)[4]. AI teaching assistants mainly serve as mentors, teaching assistants, study partners and other roles(Wang Shichong et al. 2019)[5], whose functions cover different aspects such as intelligent perception and reasoning, planning and decision-making, control and interaction.

This paper tends to define the connotation of intelligent dual-teacher classroom as "balanced theory", that is, to bring together AI and high-quality teacher resources to the maximum extent,

2

and build a new pattern of dual-teacher classroom through iterative optimization(Zhong Shaochun 2020)[6]. Based on literature review, the author defines the intelligent dual-teacher classroom as a classroom in which authentic teachers and AI teaching assistants, based on learning big data, cooperate to build and optimize the learning network, excavate the common characteristics of individual learners and groups, and promote the intelligent development of students in the whole teaching process of deep integration of intelligent technology.

2.2 Enriched virtual classroom

Enriched virtual classroom is one of the forms of blended classroom. At present, blended classroom practice and exploration are mostly empirical practice activities, focusing only on the form of transformation, and relatively ignoring the nature and law of teaching(Zhong Shaochun 2020)[7]. Enhanced virtual classroom provides face-to-face teaching or online class functions.Teachers can carry out personalized teaching according to students' learning conditions, and students can choose suitable learning places to carry out adaptive learning at their own pace. Compared with flipped classroom mode, enhanced virtual classroom not only reverses time and space, teaching processes and course structures, but also is a flexible hybrid mode with personalized characteristics.

3. The construction of enhanced virtual intelligence dual teacher model

3.1 The feasibility of enhancing virtual classroom

How to apply intelligent technology to foreign language classrooms(Xia Xiaoyan et al. 2021)[8] has become one of the research hotspots in the field of educational technology(Zheng Chunping et al. 2021)[9]. After upgrading through embedded devices or application software, all kinds of intelligent education platforms have certain adaptive and personalized characteristics, and can basically meet the teaching requirements of enhanced virtual classrooms. Jiang Bo et al(2018)[10] believe that classroom analysis based on computer vision, emotional computing and other technologies can provide a basis for classroom upgrading. The automatic recognition algorithm of learning confusion based on facial expression provides technical support for learners' emotional modeling, and can also help explore the development path of intelligent education services. Some intelligent apps with functions such as language recognition, eye tracking and language synthesis allow machines to not only "understand" human language, but also "use"

language, and create the necessary technical environment for enhancing virtual intelligence dual-teacher classrooms.

3.2 The construction of enhanced virtual intelligence dual teacher mode

Based on blended learning theory and cognitive construction theory, the enriched virtual intelligent dual-teacher mode (EVIDTM) can be constructed. EVIDTM integrates intelligent technology into the dual-teacher classroom, incorporates learning styles and habits into the instructional design framework, and generates classroom instruction and learning reports to facilitate adaptive and personalized learning for students. EVIDTM provides ample input to the classroom with a wealth of data resources and delivers learning through a combination of text, video, online presentations or other multimodal approaches. EVIDTM's strengths are mainly reflected in three aspects: intelligent diagnosis, man-machine learning and accurate assessment. The main functions of the course platform, such as real-time push, bullet-screen interaction, random roll call and classroom red envelope, ensure smooth communication between teachers and students, man-machine and students, and run through the three stages of diagnosis, learning and evaluation, as well as the five levels of teaching, learning, practice, measurement and evaluation.

AI teaching assistants assist human teachers in personalized diagnosis of students' learning conditions, providing teaching basis for "human-machine learning"; Human teachers adjust and determine the teaching activity plan according to the diagnosis report. Students not only get more "scaffolding" support from AI teaching assistants and human teachers to realize knowledge construction through meaningful interactive inquiry, but also cultivate thinking ability in teacher-student, student-student, human-computer interaction, and enhance collaborative learning and problem-solving ability. EVIDTM focuses on the assessment of learning outcomes and builds learner profiles, personalized knowledge and ability maps based on big learning data. With the help of the intelligent platform evaluation system, human teachers provide individual guidance and general summary to help students meaningfully construct knowledge.

4. Empirical study on the effectiveness of EVIDTM

This study is based on the "English Speculative Reading" of the provincial online quality course "General English at Work", which was established by the research group in November 2022. This course is designed for first-year non-English majors in vocational college in Jinan Engineering Polytechnic. It aims to cultivate students' high-level reading and critical thinking

ability through extensive reading practice. In this study, 100 students (48 boys and 52 girls) who participated in "English speculative reading" were tested with EVIDTM and traditional instruction during the first semester of the 2022-2023 school year. The course consists of 10 modules, with the first five using EVIDTM and the last five using the traditional classroom model, all taught by the same teacher. The smart campus platform collects and analyzes students' learning data, and automatically calculates data such as video viewing time, practice, and quiz scores. Subjective data such as students' understanding of course content and attitudes towards teaching methods were obtained through questionnaires to test the practical effects of EVIDTM. The study aims to demonstrate the effect of EVIDTM on reading learning for non-English majors through descriptive statistical analysis. The test is intended to refer to the reading question type of TEM-4. The same subject takes the test twice, and the test results are analyzed through the dependent sample T-test. In addition, we will verify that there are significant improvements in student engagement, learning completion, confidence, and learning performance and positive experiences in EVIDTM, including student-student, human-computer, and student-student synchronous or asynchronous interactions.

The above studies confirmed that students' learning engagement and activity completion in EVIDTM environment were significantly higher than in traditional classrooms. Compared with traditional classes, students' independence, initiative and awareness of inquiry in pre-class preparation have also been greatly improved. Compared with reading text materials, students prefer to use diversified teaching resources such as videos for learning. EVIDTM provides more autonomy in how, where and what you learn and more freedom to learn, making it more conducive to efficient learning. The assistance of AI teaching assistants makes the guidance of human teachers more accurate and intelligent. Empowered by technologies such as knowledge graph, multi-modal presentation and learning analysis, AI teaching assistants can choose the best time to intervene in students' learning, solve problems and correct deviations in time, and collaborate with human teachers to build learning "scaffolding" for students. Students are fully engaged and immersed in learning, generating positive emotions such as pleasure, interest and sense of achievement, and obtaining psychological experience similar to "flow". The combination of enhanced virtuality and wisdom breaks through the limitations of time and space in teaching, and not only integrates virtual learning resources into real classroom situations, but also integrates

real learning resources into virtual learning situations, realizing the needs of personalized, adaptive and diversified foreign language classrooms. According to different task types, teaching activities designed in line with learners' different learning styles and styles can stimulate learners' interest and motivation and effectively improve learning results(Arbaugh 2014)[11]. EVIDTM integrates the intelligence of AI teaching assistants with the wisdom of authentic teachers.

5. Recommendations for EVIDTM

AI teaching assistants comprehensively, accurately and dynamically collect students' learning data to draw a portrait of individual knowledge and ability, while human teachers give full play to their advantages of speculative innovation, emotional adjustment and humanistic care to promote students' knowledge internalization and construction through online and offline learning activities.

EVIDTM focuses on three teaching elements: teaching model, students and teachers, highlighting the diversity of blended classroom, students' subjectivity and teachers' intelligence, and building a new pattern of intelligent dual-teacher classroom. Based on the above discussion and research results, the following three suggestions are put forward on how to better combine the enhanced virtual teaching model with the dual teacher classroom.

Firstly, keep the EVIDTM classroom flexible and adaptable. At present, foreign language classroom still attaches importance to improving language ability through a lot of memorization, mechanical exercises and repeated reinforcement, but neglects to improve the innovation and richness of teaching practice with intelligent technology. The collaboration of "AI teaching assistant + authentic teacher" can make classroom teaching more diversified, flexibly choose project-based, immersive, task-driven and other teaching modes according to the actual situation of different courses and students, and create a vivid, lively and harmonious classroom atmosphere to stimulate students' subjective initiative and creativity. Therefore, teaching should not be limited to a certain mode, but should be adjusted according to the changes of teaching environment and curriculum needs. Different regions and schools should adopt the most suitable teaching mode based on teaching and learning conditions.

Secondly, adhere to student-focused EVIDTM classes. Enhancing virtuality means subverting the traditional learning mode, and students carry out teacher-student, peer, man-machine communication and collaborative exploration in a suitable learning space at their own pace. With the application of intelligent technology in classroom teaching, AI teaching

6

assistants will become omniscient and omnipotent "smart teachers" and provide more "scaffolding" for students, but classroom teaching should always adhere to the concept of "learning-centered and student-oriented". Students are the main body of learning, AI teaching assistants are responsible for creating a ubiquitous, interactive and digital smart classroom environment, and human teachers serve as the organizers, guides and collaborators of students' learning to jointly build a "student-oriented" smart blended classroom.

Thirdly, improve teachers' comprehensive literacy of numerical intelligence. The influence of enhanced virtuality on the content and mode of classroom teaching is more and more obvious, and it plays an increasingly important role in the dual-teacher classroom. The skilled use of AI technology and intelligent devices has become a necessary skill for qualified foreign language teachers in colleges and universities, which puts forward higher requirements for teachers' professional quality. Only by constantly improving the quality of data intelligence and keeping pace with the times, can teachers effectively open EVIDTM teaching and avoid being marginalized in the human-machine collaborative teaching environment. In order to adhere to the "student-oriented", mobilize students' learning enthusiasm, meet the diversified learning needs of students, and promote students' adaptive learning and personalized learning.

6. Conclusion

As a transformation of learning and teaching methods, classroom reform is an important link in the reform and development of modern vocational education[12]. With the preliminary formation of top-level design and strategic deployment in the new stage of high-quality development of vocational education in China, vocational education is transforming from the initial scale expansion to the connotation construction that pays more attention to the improvement of talent training quality. In September 2020, the Ministry of Education and nine other departments issued the Action Plan for Improving the Quality of Vocational Education (2020-2023), which clearly put forward the key tasks of "implementing the reform of the "Three Jiao" (teachers, teaching materials and teaching ways) in vocational education, promoting the " classroom revolution" in vocational colleges and universities, and pushing the curriculum teaching reform to the depth" [13]. The reform of "Three Jiao" is the starting point to deepen the connotation construction of vocational education, highlight the type characteristics of vocational education, and optimize the mode of personnel training. Among them, classroom reform has become the "breakthrough" and

7

"focus" of the reform of "Three Jiao"[14]. In order to accurately grasp the reform and development trend of vocational education, taking the reform of "Three Jiao" as an opportunity and the construction of "colorful classroom" as the research carrier, a new type of classroom teaching model of vocational education with full elements, multi-dimension and high efficiency is constructed to provide theoretical reference for the construction of poly-technical classes.

This study clarifies the content of intelligent dual-teacher classroom and enhanced virtual classroom, constructs EVIDTM teaching model, and explores the effectiveness of the model through teaching experiments. At present, the dual-teacher classroom teaching using AI is still in the exploratory stage, and EVIDTM has not yet formed a specific teaching model framework with perfect structure and smooth process. There are many problems in the research and development of smart devices, the construction of smart platform, and the acceptance and adaptation of teachers and students, which still need to be solved jointly by different subject areas. Compared with the traditional teaching model, EVIDTM brings more opportunities and challenges to foreign language classroom teaching. In the long run, the co-teaching of authentic teachers and AI teacher assistants is expected to become the new normal of foreign language classrooms in the future, and intelligent dual-teacher teaching is worth carrying out more in-depth theoretical and practical research.

References

[1] Lv Linhai, Gong Fang. A comparative study on deep learning and its influencing mechanism of undergraduates in Chinese and American research universities: An empirical analysis based on SERU survey in eight Chinese and American universities [J]. Educational Research, 2018,39 (4) : 111-120.

[2] He Kekang. Smart classroom + Classroom Teaching Structure Reform: The fundamental way to achieve the grand goal of education informatization[J]. Education Research, 2015,36 (11) : 76-81+90.

[3] Jiang Yi, Wang Wei, Li Chenyu, et al. Research on teacher-student Interaction in smart classroom environment[J]. Modern Distance Education, 2019 (3) : 13-21. (in Chinese)

[4] Zheng Xudong, Rao Jingyang, Jia Yangyang. "Three Classrooms" Promoting high-quality and balanced development of Compulsory Education: Evolution history, strategic value, relationship analysis and conceptual framework[J]. Modern Educational Technology, 2021, (6) : 14-22.

[5] WANG Shichong, Fang Haiguang, Zhang Ge, Ma Tao. Research on a new type of "dual-teacher Classroom" supported by artificial intelligence educational robots ---- On the teaching design and future prospects of "Human-machine Collaboration"[J]. Journal of Distance Education, 2019, (2) : 25-32.

[6] Zhong Shaochun. How Artificial Intelligence Drives the Education Revolution[J]. China audio-visual education, 2020, (3) : 17-24.

[7] Zhong Shaochun. How Artificial Intelligence Drives the Education Revolution[J]. China audio-visual education, 2020, (3) : 17-24.

[8] XIA Xiaoyan, Shi Hongmin, Guo Yaling, Xie Chengkai, GUO Haoru. An Analysis of the Teaching Principle and Top-level Design of VR Curriculum in Foreign Language Teaching from the Perspective of Embodiment[J]. Foreign language audio-visual teaching, 2021, (2) : 78-84, 12.

[9] ZHENG Chunping, Lu Zhihong, Liu Hanyong, WANG Lili, Han Xiaohua. A study on College students' English Learning Outlook and learning Engagement in Virtual Reality Environment[J]. Foreign language audio-visual teaching, 2021, (2) : 85-92, 13.

[10] Jiang Bo, Li Wanjian, Li Zhixuan, Ye Yun. Automatic recognition of learning perplexity based on facial expression[J]. Open Education Research, 2018, (4) : 101-108.

[11] Arbaugh J B.What might online delivery teach us about blended management education? Prior perspectives and future direction[J]. Journal of Management Education, 2014, 38 (6) : 784-817.

[12] Zheng Feng. Classroom Revolution based on Teaching Reform Practice[J]. China University Teaching, 2020 (10) : 17-20. (in Chinese)

[13] Notice of the Ministry of Education and nine other departments on issuing the Action Plan for Improving the Quality and Training of Vocational Education (2020-2023)[EB/OL].2020-09-23[2021-11-01].http://www.moe.gov.cn/srcsite/A07/zcs_zhgg/2020 09/t20200929_ 492299.html.

[14] Yang Shunguang, Shi Weiping. The Main Principles, Key tasks and promotion Strategies of deepening reform of Higher Vocational Education during the 14th Five-Year Plan Period[J]. Educational Development Research, 21, 41 (7) : 38-43.

9

A Preliminary Investigation into Human-AI Interactive Negotiation Competence: Exploring the Usage of ChatGPT by EFL Learners

Lin Chen (林晨)   Beijing Foreign Studies University

Abstract

This study examines the utilization patterns of ChatGPT by EFL (English as a Foreign Language) learners during the English writing process. While the discourse around the application of ChatGPT in education is extensive, empirical studies are sparse, particularly those focusing on the AI literacy of language learning. Addressing this gap, the research aims to explore how students use ChatGPT to aid their English writing, contributing to the theoretical framework proposed by Wen and Liang (2024) on Human-AI Interactive Negotiation Competence (HAINC). Wen and Liang's framework delineates four key usage steps: setting goals, issuing instructions, analyzing feedback, and adjusting strategies. This qualitative study involved seven first-year English major students from a prestigious foreign language university in China. The study focused on an expository writing task assigned in their courses. Through process tracing of their expository writing processes, followed by stimulated recall sessions and interviews, the research delved into their interaction with ChatGPT. The dialogues between students and ChatGPT were collected, and data were coded in accordance with the theoretical framework. The findings reveal consistent patterns across the four usage steps. Under the overarching goal of writing the essay, the goals proposed by students were categorized into three aspects: Content, Information, and Language. In issuing instructions, prompts were categorized and counted according to these aspects. The analysis of feedback was confirmed by stimulated recall and interviews to understand students' evaluative processes. Strategy adjustments were detected as students iteratively refined their prompts or shifted to other tools, showcasing an adaptive learning process. This study provides a preliminary understanding of how students use ChatGPT during English writing tasks, offering valuable insights and augmenting the theoretical framework of HAINC. These findings contribute to a deeper understanding of AI-assisted language learning, providing a foundation for further research about AI literacy.

1. **Introduction**

Large language models (LLMs) can process and generate numerous texts instantaneously. In the past one to two years, since the release of ChatGPT, educational studies have extensively focused on its pedagogical influence, particularly in the field of language learning, where ChatGPT excels in text-based tasks (Godwin-Jones, 2022; Hu & Qi, 2023; Kasneci et al., 2023; Labadze et al., 2023; Wang, 2023).

Most of the literature, however, still focuses on whether students should use ChatGPT rather than on how to use ChatGPT appropriately (Lo, 2023). Some scholars have suggested that language teaching and learning should coexist with ChatGPT and other LLM-powered chatbots (Godwin-Jones, 2022; Qin, 2023). Following this idea, some literature has reported on pedagogical cases integrating ChatGPT in both classroom and extracurricular settings, especially in EFL (English as a Foreign Language) contexts (Kong, 2024; Xu & Zhao, 2024). Some literature has begun to examine whether ChatGPT can effectively contribute to EFL teaching and learning (Jacob et al., 2024; Mao et al., 2024; Zhao et al., 2023). Some literature has begun to examine the outcomes of students using ChatGPT as an aid in their studies, finding that ChatGPT can improve their work and enhance their abilities (Ali et al., 2024; Kim et al., 2024; Urban et al., 2024). All of these studies have been providing increasingly convincing positive answers to the question of whether students should use ChatGPT. However, there is still a lack of answers to the question of how to use ChatGPT appropriately.

To address this research gap, this study closely examined the learning process with ChatGPT. Using a qualitative approach, this study selected English major students from a prestigious university in China and traced their learning process while they completed their writing assignments. In this way, this study uncovered the usage patterns of Chinese EFL learners with the assistance of ChatGPT.

2. **Literature Review**

2.1 **Discussion on LLM-powered chatbots in Language Education**

There exist many discussions on LLM-powered chatbots in language education (Ma & Xu, 2023). Many discussions made blueprints by mentioning the potential usage and risks of utilizing LLM-powered chatbots in the context of language education (Hu & Qi, 2023; Huang, 2022; Jiao,

2023; Wang, 2023; Zawacki-Richter et al., 2019). Furthermore, some discussions proposed some potential application by cases in real situation (Godwin-Jones, 2022; Kong, 2024; Qin, 2023; Su et al., 2023; Xu & Zhao, 2024). For example, Su et al. (2023) discussed the application of chatbots integrated with LLMs within argumentative writing courses and assignments, concurrently addressing associated risks. Furthermore, Kong (2024) discussed the utilization of LLMs in intensive English reading class instruction.

As Godwin-Jones (2022) and Qin (2023) both mentioned, "Coexistence with AI" emerges as a pivotal prospective orientation within the realm of ICALL. Instead of engaging in debates regarding the utilization of ChatGPT and other LLM-powered chatbots within language education, the academic focus is shifting towards exploring more effective methodologies for implementing these technologies. However, as Jiao (2023) mentioned, there is still lack of empirical study about the application of LLM-powered chatbots in language education. The research investigating and examining the use of LLM-powered chatbots in EFL learning context focus mainly on the influence on students or the outcome of students with LLM-powered chatbots (Abdelhalim, 2024; Karataş et al., 2024; Punar Özçelik & Yangın Ekşi, 2024).

Despite the testing conducted on the efficacy of LLM-powered chatbots in assisting teachers with their instructional methods and assessing students' learning outcomes, these studies largely validate the feasibility of LLM-powered chatbots' integration into teaching practices. Nevertheless, empirical investigations into the practical application of these technologies remain notably scarce. Within the discourse on the impact of LLM-powered chatbots on language education, investigating the process of its usage could provide valuable insights into its influence on teaching methodologies.

2.2 Human-AI Interactive Negotiation Competence

It has been mentioned that there is a research gap in existing studies, which focus on the outcomes and influence of LLM-powered chatbots and ignore the process of application by users in foreign language education context. To address the research gap present in current empirical studies, this research aims to investigate the application of ChatGPT from the perspective of students. To better understand the application, a theory of Human-AI Interactive Negotiation Competence (HAINC) was introduced (Wen & Liang, 2024). This theoretical framework centers

on elucidating the dynamics of human-machine interaction to optimize machine utilization. Based on experiential study, the HAINC was divided into five essential components: understanding AI, setting goals, issuing instructions, analyzing feedback, and adjusting strategies (see Figure 1). The initial phase of cultivating proficient HAINC, termed "Understanding AI," can be construed as an endeavor to perceive the contextual situation of AI. Consequently, in assessing students' HAINC, a comprehensive investigation of their utilization stage will be conducted integrating the remaining four steps.

This theoretical framework provides guidance for approaching and analyzing students' application of ChatGPT. In accordance with Wen and Liang (2024), the study will establish the HAINC course. Consequently, this research will conduct a preliminary investigation into students' utilization of ChatGPT.

Figure 1 *The HAINC Model* (Wen & Liang, 2024).
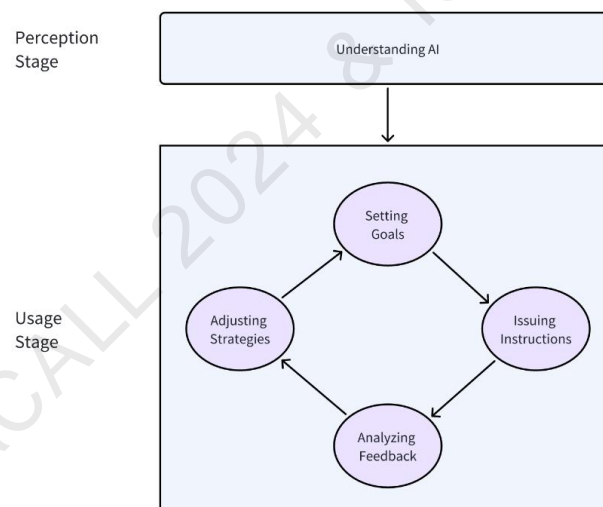
3. Method

3.1 Participants

7 participants (1 males, 6 females), aged from 18 to 19, from a top university in China, were recruited via WeChat in the qualitative research (see Table 1). All of them are Chinese native speakers. They are English majors at the first year of undergraduate. They have learned English for 9 to 14 years, and can be considered as intermediate or advanced EFL learners. All of them have experience of using AI devices in English writing assistance.

Table 1 *Interviewee Profiles*

Name (Pseudonym)	Sex	Age	Years having learnt English	Teachers in their writing courses (Pseudonym)
Chunri	Female	19	11	Lihua
Xiayu	Male	19	14	Liuli
Qiuguang	Female	19	9	Zhangwei
Dongxue	Female	19	10	Gaoshan
Mingyue	Female	19	14	Wanghong
Qingquan	Female	18	13	Chenpeng
Hehua	Female	18	12	Liuli

3.2 Setting

Participants in this research were all from the English expository writing course. As a compulsory course of freshmen in the English department, there were two periods each week (90 mins), taking up 2 credits in the training program. There were 9 teachers teaching this course. Considering the course syllabus and rules about AI assistance, participants were chosen from 6 different classes.

3.3 Instruments

Usage Log

To understand the usage of ChatGPT in English writing, usage logs were collected. Usage logs consisted of three components: process tracing, stimulated recall, and dialogue corpus.

In terms of process tracing, the writing processes of participants were screen-recorded. During the process tracing, participants were asked to write their essays using only one device. The writing task was their assignments originally. According to the actual situation, the videos submitted by the participants are fragments produced at different times of writing.

Stimulated recall method was used to validate the findings from process tracing. The

questions came from the data analysis procedure of process tracing.

The dialogues between participants and ChatGPT were collected, including prompts and generated texts.

Semi-structured Interview

To understand the thoughts of students about using ChatGPT in English writing, semi-structured interviews were conducted. Semi-structured interview guide was designed under the framework of AILL-Motivation-UTAUT model proposed and examined by An et al. (2023). Interview questions were adopted from Zou et al. (2023) and Liu et al. (2024).

3.4 Data analysis

The transcripts of interviews were written down by the researcher and one more person. The transcripts were uploaded to software NVivo. The screen-recording video fragments were connected in software Premiere, according to the actual time order. Then the joint video files were also upload to software NVivo for analysis. The dialogues between students and ChatGPT were shared through the links. Dialogues were organized diachronically. Their writing assignments were collected.

4. Findings & Discussion

4.1   Usage Patterns

Some shared usage patterns could be identified from those cases. To examine their usage patterns closely, the HAINC framework was adopted Wen and Liang (2024). It was proposed that the five elements constituting the HAINC framework are not in a non-linear and dynamic process, which means the usage patterns of ChatGPT are not fixed in a certain order. Furthermore, the process is not unidirectional but continuous and repeated in various using contexts. Examining the cases from the perspective of HAINC, key activities in students' usage processes could be identified. The model of usage pattern is shown in the Figure 2.

Figure 2 *The Usage Pattern Model*
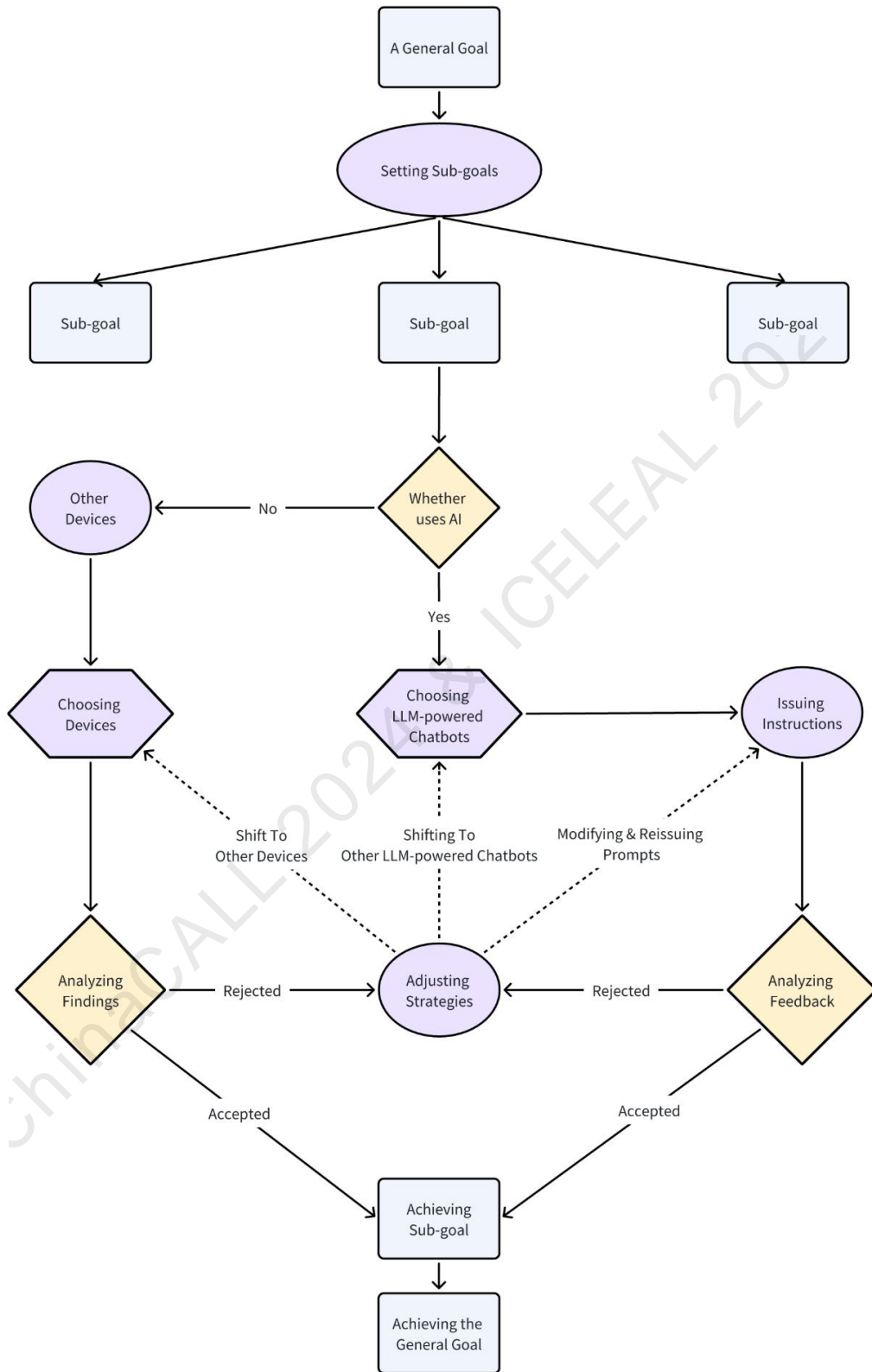
Setting Goals

Firstly, the general goal - writing an essay according to the assignment requirement – were

divided into several sub-goals by students. Students then decided whether the sub-goals can be solved by themselves. If not, they started to find assistance. Traditionally, they would use online resources such as e-dictionaries, search engines, machine translators, etc., to help them tackle problems during achieving sub-goals. Since AI devices appeared, especially LLM-powered chatbots, they have had more new choices of assistance. When students encountered a new problem, whether to use AI assistance depended on their perceptions of LLM-powered chatbots. Specifically, the relationship between perceptions and decision to use AI devices has been revealed in An et al. (2023) by method of Structural Equation Modeling.

The sub-goals could be generalized into 3 aspects: Content, Information, and Language. In terms of Content, they would like to ask for a writing structure. Participants mentioned that ChatGPT could assist them in organizing the structure of their essays by offering clearer and more logical outlines. Furthermore, it was reported that their thought process also became more organized as a result. Participants also utilized ChatGPT to synthesize ideas and provide arguments in expository writing. In terms of Information, Participants employed ChatGPT to gather data and literature. Regarding Language, the focus was on the modification and refinement of linguistic expressions.

Issuing Instructions

If they opt to leverage AI support, the subsequent action is to provide instructions, or "prompts," to inquire of the chatbot. Prompts were categorized and counted according to the intent label in Han et al. (2024) (see Figure 3). However, the result were not totally in accordance with the categorization in Han et al. (2024). Some new intentions were found. In the research, it was found that students also send ChatGPT prompts for provision. Provision refers to information is given to ChatGPT as the background as a beginning of a dialogue. In one case, *Chunri* gave ChatGPT a sample writing to be referred to. Students made provision for the dialogue suggesting that they have known ChatGPT's answers highly depend on the context. It was also found that students also inquired ChatGPT to get further information, including adding information to already given prompts, questioning closely, and giving revision of GPT's answer.

Correspondingly, based on the categorization method of Content, Information, and Language, the overall proportion can be statistically determined. Prompts for Content (including *Request for*

Generation & Request for Revision) accounted for 23.07%. Prompts for Information (including *Request for Information*) accounted for 23.08%. Prompts for Language (including *Request for Language Use & Request for Translation*) accounted for 41.03%. This indicated that in the use of chatbots, request for language use is the most frequent purpose.

Figure 3 *Percentage of Different Types of Prompts*
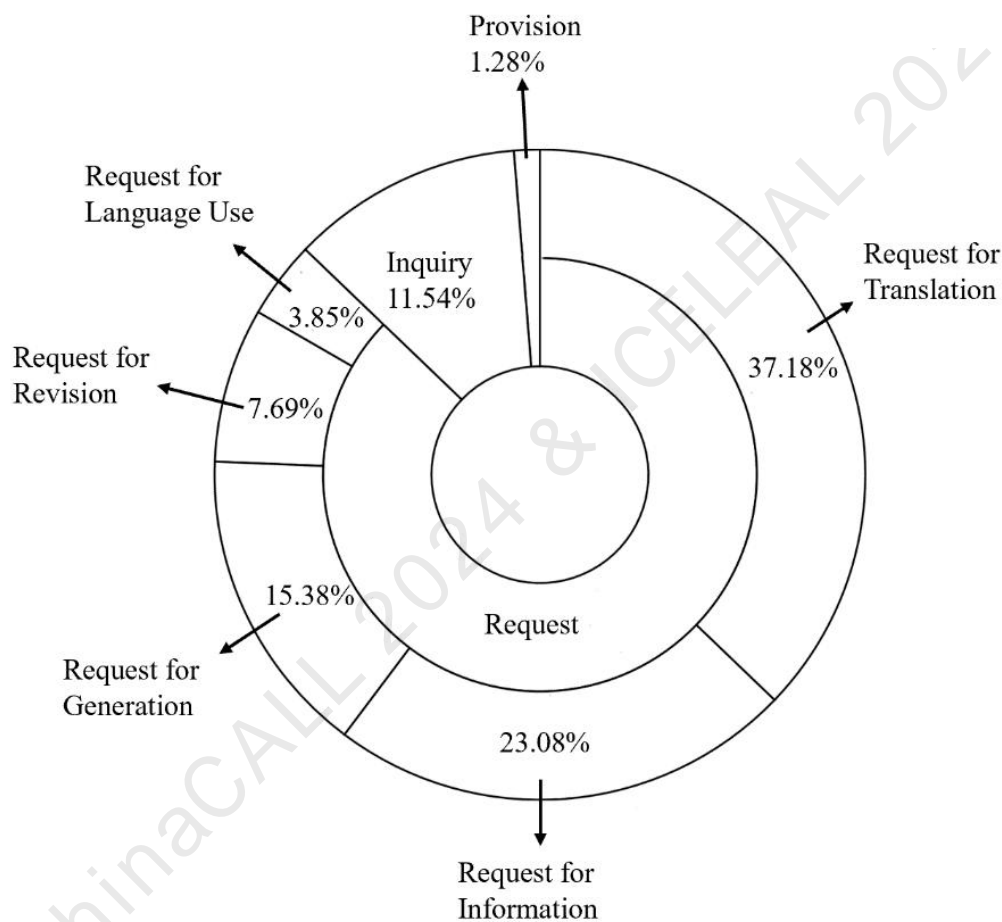
Analyzing Feedback

Whether they choose LLM-powered chatbots or traditional helpers, they will evaluate what they have found or been answered. In the simulated recalls and interviews, participants shared their processes and feelings in the analysis of feedback.

They mentioned the output quality of ChatGPT. There are both positive and negative perspectives regarding the quality of ChatGPT outputs. The positive evaluations of ChatGPT's output quality mainly revolve around two aspects: improved logical coherence and reduced

occurrence of grammatical errors. Logical coherence pertains to the flow of sentences and the overall structure. This result cohered with a majority of reviews (Darvishi et al., 2024; Hu & Qi, 2023; Qin, 2023; Xu & Zhao, 2024; Yang et al., 2023).

The negative evaluations of ChatGPT's output quality, however, covered various aspects. The accuracy of output was queried. Participants reported that ChatGPT could generate fake references of literatures with fake Digital Object Identifiers (DOIs) and International Standard Book Number (ISBNs) which could not be located on Internet. Sometimes, ChatGPT may provide DOIs or ISBNs that exist but do not correspond to the intended literature, leading to incorrect references.

In addition to accuracy, the absence of human touch in the content generated by ChatGPT was also noted. This lack was reflected in various ways, such as overly complicated diction, excessively concise expression, and rigid linking between sentences. The perception of ChatGPT as non-human was shared from both the perspectives of students and teachers.

All participants mentioned the necessity of a self-assessment phase. The first step involved judging and evaluating the generated content. Subsequently, they modified their own writing based on the generated contents, or make adjustments and revisions on the generated contents before adopting them. If they were dissatisfied with the generated content, a strategy adjustment was required.

Adjusting Strategies

If they were not satisfied with the findings or feedback, they would adjust their strategies. And then they came back to the stage of choosing types of assistance. Thus far, a cycle has emerged.

There were several methods for adjusting strategies. One was to continue using ChatGPT by adjusting the prompt words to obtain updated generated content. Alternatively, one could switch to other chatbots or transition to traditional tools such as electronic dictionaries and academic literature websites.

4.2 Usage Features

Some shared features in participants' usage patterns were identified. Firstly, all of the participants used ChatGPT at the beginning of their writing process. This indicated that ChatGPT

has helped them better open up their writing ideas. Secondly, the use of ChatGPT permeated their entire writing process. Through the observation of the process tracing, it was found that the time spent using ChatGPT constitutes the majority of their writing processes. Furthermore, their usage often involved extensive alternation or even parallel use of ChatGPT and Office Word for writing. This indicated that ChatGPT has a high level of engagement in assisting with writing.

Concurrently, the study conducted natural language processing methods to compare the text similarity between the texts generated by ChatGPT and the writing texts of the participants. The closer the similarity is to 1, the more similar the two contents are. The closer the similarity is to 0, the opposite is true. The results revealed that the student texts exhibited a high degree of similarity to the content generated by ChatGPT (see Table 2). This indicated that the participants have a high level of acceptance for the content generated by ChatGPT.

Table 2 *Text Similarities by Using Different Methods*

	Methods		
	TF-IDF	BERT	SentenceTransformers
M	.9122	.9187	.7970
SD	.0486	.0432	.1750

5. Conclusion

This study explored the usage of ChatGPT by EFL learners. Some shared features of their usage patterns were identified through process tracing. ChatGPT played a very important role throughout their writing process. Furthermore, the amount of time they spent using ChatGPT constituted a major portion of their total writing time. They used ChatGPT for multiple purposes. And it was found that they have developed strategies through their own attempts to better utilize ChatGPT.

AI literacy, which has been proposed for a long time, is a key focus in ICALL. By exploring the usage of ChatGPT, this study revealed several pedagogical implications. In the learning process, knowing how to issue good instructions is essential for better usage of ChatGPT. Here comes a key literacy in AI usage – prompt engineering, which refers to the extent to which students know how to write an appropriate prompt to receive a satisfying answer. There should be resources to teach students what a prompt is good and how to write a good prompt. Furthermore,

how to analyze the feedback from ChatGPT is also important. Analyzing exercises will still enhance students' critical thinking ability. Last but not the least, students should be also taught about how to adjust strategies when they find ChatGPT cannot tackle their issues. It requires an overall literacy in educational technologies.

Indeed, this study has limitations. Due to practical research constraints, this study had a small sample size and a relatively short tracing period. Therefore, for future research, expanding the sample size and extending the tracing duration would better support the conclusions proposed in this study. For further research, it is advisable to broaden the research context beyond writing and higher education.

References

Abdelhalim, S. M. (2024). Using ChatGPT to promote research competency: English as a Foreign Language undergraduates' perceptions and practices across varied metacognitive awareness levels. *Journal of Computer Assisted Learning*.

Ali, D., Hassan, K., Shazia, S., Dragan, G., & George, S. (2024). Impact of AI assistance on student agency. *Computers & Education*, *210*, 104967. https://doi.org/10.1016/j.compedu.2023.104967

An, X., Chai, C. S., Li, Y., Zhou, Y., & Yang, B. (2023). Modeling students' perceptions of artificial intelligence assisted language learning. *Computer Assisted Language Learning*, 1-22. https://doi.org/10.1080/09588221.2023.2246519

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, *210*, 104967. https://doi.org/10.1016/j.compedu.2023.104967

Godwin-Jones, R. (2022). Partnering with AI: Intelligent Writing Assistance and Instructed Language Learning. *Language Learning & Technology*, *26*(2), 5-24.

Han, J., Yoo, H., Myung, J., Kim, M., Lee, T. Y., Ahn, S.-Y., & Oh, A. (2024). RECIPE4U: Student-ChatGPT Interaction Dataset in EFL Writing Education. *arXiv preprint arXiv:2403.08272*.

Hu, J., & Qi, Y. (2023). ChatGPT 时代的中国外语教育：求变与应变. *外语电化教学*(01), 3-6+105.

Huang, L. (2022). 大数据时代背景下的语言智能与外语教育. *中国外语*, *19*(01), 4-9. https://doi.org/10.13564/j.cnki.issn.1672-9382.2022.01.008

Jacob, S., Tamara, T., Steve, G., Jazmin, C., Michael, H., Jiali, W., Youngsun, M., Waverly, T., Mark, W., & Carol Booth, O. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, *91*, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Jiao, J. (2023). ChatGPT 助推学校教育数字化转型——人工智能时代学什么与怎么教. *中国远程教育*, *43*(04), 16-23. https://doi.org/10.13541/j.cnki.chinade.20230226.001

Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, 1-24.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, H. K., Nayak, S., Roknaldin, A., Zhang, X., Twyman, M., & Lu, S. (2024). Exploring the Impact of ChatGPT on Student Interactions in Computer-Supported Collaborative Learning. *arXiv preprint arXiv:2403.07082*.

Kong, L. (2024). 生成式人工智能在外语专业教学中的应用：以《大学思辨英语教程·精读》教学为例. *外语教育研究前沿*, *7*(01), 11-18+90. https://doi.org/10.20083/j.cnki.fleic.2024.01.011

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1), 56. https://doi.org/10.1186/s41239-023-00426-1

Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, *13*(4), 410. https://www.mdpi.com/2227-7102/13/4/410

Ma, W., & Xu, M. (2023). 21 世纪以来我国智能语言教学研究述评. *外国语文*, *39*(06), 140-150.

Mao, Y., Wang, Y., & Xing, Y. (2024). ChatGPT 辅助高中英语写作反馈的实证研究. *教育测量与评价*(01), 3-13. https://doi.org/10.16518/j.cnki.emae.2024.01.001

Punar Özçelik, N., & Yangın Ekşi, G. (2024). Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, *11*(1), 10.

Qin, Y. (2023). 人机共生场景下的外语教学方法探索——以 ChatGPT 为例. *外语电化教学*(02), 24-29+108.

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, *215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Wang, K. (2023). 外语学科面向 AI 时代的转型与发展. *外语电化教学*(06), 3-7+105. https://doi.org/10.20139/j.issn.1001-5795.20230601

Wen, Q., & Liang, M. (2024). 人机互动协商能力：ChatGPT 与外语教育. *外语教学与研究*, *56*(02), 286-296+321. https://doi.org/10.19923/j.cnki.fltr.2024.02.006

Xu, J., & Zhao, C. (2024). 大语言模型在英语教学中的角色. *外语教育研究前沿*, *7*(01), 3-10+90. https://doi.org/10.20083/j.cnki.fleic.2024.01.003

Yang, X., Wang, Q., & Lyu, J. (2023). Assessing ChatGPT's Educational Capabilities and Application Potential. *ECNU Review of Education*, 20965311231210006. https://doi.org/10.1177/20965311231210006

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhao, X., Zhao, Z., Sun, F., & Wang, D. (2023). 面向语言文学领域的大语言模型性能评测研究. *外语电化教学*(06), 57-65+114. https://doi.org/10.20139/j.issn.1001-5795.20230610

Embodied Learning of Chinese Hand-related Verbs in Virtual Reality Scenario

Liu Yang   Cui Zejia

Beijing Institute of Technology

Abstract：Virtual reality (VR) has been applied in the field of language learning in recent years as a means to provide greater immersion and interactivity. Experimental evidence has demonstrated that VR enhances the learning of novel nouns in some dimensions (e.g. Macedonia et al., 2020; Fuhrman et al., 2021) . According to theories of embodied cognition, it can be attributed to VR providing sensorimotor experiences in the learning scenario. However, the learning target need to be further expanded, and comparison with more realistic multi-modal learning pattern need to be carried out. At the same time, more subtle dimensions suitable for testing the effect of virtual learning shall be developed. Therefore, the present study investigates whether VR promote the memory, recognition and the implicit embodied effect of learning Chinese hand-related verbs, as opposed to the non-virtual (video learning) scenario.

26 learners of Mandarin of the preliminary level learned 8 monosyllabic hand-related Chinese verbs in virtual reality, in comparison to the video learning conditions. The virtual reality scenarios were developed by a professional team to visualize and embody the interaction between participants and objects, as expressed by the target verbs carrying implicit meanings of direction, such as "outward" for "push (tui1)" or "inward" for "pull (la1)". E-prime 2.0 was used to design the learning and testing phases. Different dimensions of target verbs, including sound and spelling with the corresponding meanings, were examined by free recall, recognition and directionality tests to assess the retention, recognition and processing (embodiment effect of direction) of the verbs. No significant difference was found between virtual reality and video learning conditions in the retention, recognition and processing for individual verbs. Difference in the processing tend to be marginally significant, but only when the verbs were tested together with their noun collocations. It is suggested that the embodiment of verbs be more likely to be activated in context. The present study extends the scope of virtual learning target and expands the testing dimension for revealing the potential advantage of virtual learning scenario.

Key words：L2 verb acquisition; embodiment effect; virtual reality

1. Research Background and Significance

The Virtual Reality (VR) technology has been applied in language teaching due to its distinctive "3I" features: imagination, interactivity, and immersion. Existing research has examined the impact of VR on various aspects of language learning, including listening, speaking, reading, writing, vocabulary memorization, and comprehensive skills (Zheng Chunping, 2019). However, empirical studies need to be expanded both in breadth and depth. For instance, the dimensions of comparison are limited, with most studies contrasting VR with text, images, and actions. Yet, these learning modalities differ significantly from the more prevalent multi-modal learning, such as the use of videos in the digital age, failing to fully reflect the current status of multi-modal learning. On the other hand, research findings are not entirely consistent. Some studies support the positive effect of virtual scenarios on learning (Lan et al., 2015), while others indicate that compared to other scenarios, virtual scenarios do not always enhance learning in certain dimensions (Macedonia et al., 2020; Fuhrman et al., 2021). Researchers argue that this is primarily because the evaluation of learning is designed in traditional ways, and more targeted measurement which reflects the features of virtual learning should be devised to truly unlock the potential advantages of VR (Lan et al., 2015).

The embodied cognition provides theoretical support for evaluating learning outcomes in virtual reality scenarios from a new dimension. Embodied cognition emphasizes that cognition is multi-modal rather than relying solely on abstract, amodal symbols, highlighting the crucial role of bodily movements and sensorimotor experiences in cognition (Barsalou, 1999; Glenberg & Kaschak, 2002). Research on second language processing from an embodied cognition perspective has introduced the concept of embodied effect, which refers to the activation of body-related sensations during language processing. Empirical studies have found that understanding verbs is accompanied by the simulation of related spatial directions (Glenberg & Kaschak, 2002).

However, studies on second language processing have revealed differences in embodied effects between second language learners and native speakers (Lin Lihong et al., 2019). Some scholars attribute these differences to the learning environment (Macedonia, 2014). Learning in virtual scenarios can overcome the limitations of traditional learning environments by activating sensorimotor perceptions and enabling experiential learning (Liu Dejian et al., 2016). Therefore, measuring language learning outcomes in virtual scenarios need to extend beyond the common

dimensions of production and recognition. By incorporating processing dimensions, in this case, testing the embodied effect, it may create a more suitable assessment for virtual learning.

Previous studies of virtual learning have predominantly focused on nouns, while Chinese verbs often present obstacles for L2 learners (Cai Beiguo, 2010). This study aims to compare the learning effects of Chinese verbs between the virtual and video environments by adding processing dimension to the recognition and production tests.

2. Research Method

The study aims to investigate the effects of different learning environments on the acquisition of directional Chinese monosyllabic hand verbs (e.g., lā 'pull', tuī 'push') among 26 foreign students with a preliminary proficiency level in Chinese. The independent variable is the learning environment/scenario, specifically video-based learning versus virtual learning conditions, and the dependent variable is learning outcome, which will be measured across three dimensions: recognition, production, and processing.

There are four learning sections on E-Prime 3.0. The first three sections aim to help participants form an initial understanding of Chinese hand verbs, familiarize themselves with their meanings, and learn their pronunciations across different contexts (single verbs, verb-noun collocations, and complete sentences). The fourth section is divided into two groups, one for learning under video conditions and the other under virtual conditions.

Participants completed three tests to measure learning outcomes across diverse dimensions. 1) The memory test: a free recall and Pairing Test, including oral and written, evaluates participants' productive knowledge of the learning objectives. 2) The recognition test: a matching test encompasses pinyin-meaning, form-meaning, meaning-pinyin, meaning-form, and video-meaning matching. It measures participants' recognition abilities from Chinese pinyin and pronunciation to corresponding meanings, from the meaning of Chinese verbs to their corresponding pinyin and pronunciation, and from videos depicting verb actions to their corresponding meanings. 3) The processing test: a directionality test, investigates whether participants have acquired the embodied features of the target verbs and whether there are differences between different learning scenarios. It employs a within-subjects design featuring 2 factors (virtual learning/video learning) crossed with 2 levels (arrow direction matches verb direction/arrow direction does not match verb

direction). It is divided into two dimensions: directionality of single verbs (including pinyin and audio judgment) and directionality of verb-noun collocations (including phrases in virtual and video contexts, phrases in sentences, and unfamiliar phrases).

3. Experiment Results

3.1 Memory Test

There were no significant differences in the paired scores of oral and written free recall between the virtual and video learning groups.

Table 1: Statistical and Descriptive Data for Memory Test

Learning Scenario	Descriptive Data				Statistical Data
	VR		VD		
Output Mode	M	SD	M	SD	
Verbal	.514	.268	.510	.326	t(25)=.084, p=.934, d=.016
Written	.524	.262	.471	.313	t(25)=1.066, p=.297, d=.297

3.2 Recognition Test

In the form-meaning recognition, sound-meaning recognition, meaning-form recognition, meaning-sound recognition, and video recognition tests, there were no significant differences in the reaction time and accuracy rate between the virtual learning group and the video learning group.

Table 2: Descriptive Data for Recognition Tests

Leaning Scenario	Reaction Time				Accuracy Rate			
	VR		VD		VR		VD	
Recognition Direction	M	SD	M	SD	M	SD	M	SD
Form-Meaning	4049.15	2186.76	4036.52	1767.58	80.77	25.80	72.12	34.15
Sound-Meaning	3412.90	1592.41	3381.00	1566.05	77.89	25.81	83.65	21.15
Meaning-Form	3253.04	1671.26	2902.46	1492.75	78.85	28.89	75.00	26.46
Meaning-Sound	3732.77	1903.82	4547.22	1995.73	78.85	25.81	77.89	26.17
Video-Meaning	5209.15	1554.43	5056.92	1764.45	71.64	22.52	68.75	26.98

Table 3: Statistical Data for Recognition Tests

Recognition Direction	Test Dimension	Statistical Value	Test Dimension	Statistical Value
Form-Meaning	**RT**	t(22)=.031, p=.976, d=.006	**AR**	Z=−1.698, p=.119, r=.235
Sound-Meaning	**RT**	Z=−.335, p=.754, r=.049	**AR**	Z=−.638, p=.536, r=.088
Meaning-Form	**RT**	t(23)=.106, p=.916, d=.022	**AR**	Z=−1.231, p=.305, r=.171
Meaning-Sound	**RT**	Z=−1.800, p=.074, r=.260	**AR**	Z=−.036, p=1.000, r=.005
Video-Meaning	**RT**	t(20)=.646, p=.526, d=.141	**AR**	Z=−.647, p=.545, r=.090

3.3 Processing Test

In the directionality tests of pinyin judgment, audio judgment, phrase judgment in virtual scenarios and videos, and unfamiliar phrase judgment, the main effect of learning modality was not significant, while the main effect of direction consistency was marginally significant. The interaction between learning modality and verb-arrow direction consistency was not significant. In the directionality test of phrase judgment learned in the sentence context, the main effects of learning modality and direction consistency were not significant, but the interaction between learning modality and verb-arrow direction consistency was significant. Further simple effect analysis revealed that there were no significant differences in reaction time within the virtual (p=.338)and video (p=.123) groups respectively, when the verb and arrow directions were consistent compared with inconsistent.

Table 4: Descriptive Data for Processing Tests

Leaning Scenario		VR				VD			
Directionality		Consistent		Inconsistent		Consistent		Inconsistent	
Test Unit		M	SD	M	SD	M	SD	M	SD
Single word	**Pinyin**	536.25	194.22	538.15	249.38	543.14	304.03	551.85	224.72
	Audio	614.72	196.41	676.68	228.26	647.30	195.91	656.68	200.89
Phrase	**VR+VD**	508.79	134.57	549.76	130.54	530.55	117.49	555.35	154.13
	Unfamiliar	528.25	118.83	548.77	144.91	508.07	87.73	541.83	109.13
	Sentence	551.22	133.34	530.80	153.17	519.09	134.29	551.31	124.73

Table 5：Statistical Data for Processing Tests

		Main Effect of Learning Scenario	Main Effect of Directionality	Interaction Effect
Single word	**Pinyin**	F (1, 25)=.130, p=.722, η²=.005	F(1, 25)=.040, p=.844, η²=.002	F(1, 25)=.015, p=.904, η²=.001
	Audio	F (1, 25)=.047, p=.830, η²=.002	F(1, 25)=1.350, p=.256, η²=.051	F(1, 25)=1.011, p=.324, η²=.039
Phrase	**VR+VD**	F (1, 25)=1.031, p=.320, η²=.040	F(1, 25)=4.070, p=.055, η²=.140	F(1, 25)=.231, p=.635, η²=.009
	Unfamiliar	F (1, 25)=.100, p=.754, η²=.004	F (1, 25)=.133, p=.718, η²=.005	F(1, 25)=4.308, p=.048, η²=.147
				VR：方向一致 vs 不一致，p=.338
				VD：方向一致 vs 不一致，p=.123
	Sentence	F (1, 25)=.798, p=.380, η²=.031	F(1, 25)= 4.043, p=.055, η²=.139	F(1, 25)= .162, p=.691, η²=.006

4. Discussion and Conclusion

The main research question of this paper is whether there are differences in the learning effects of Chinese monosyllabic hand verbs by foreign students learning Chinese as a second language under virtual reality learning condition and video learning condition. The learning effects cover three dimensions, including production, recognition, and processing.

The paired free recall test results between the virtual learning group and the video learning group showed no significant differences in both oral and written dimensions. Similarly, the reaction times and accuracy rates in the recognition tests did not exhibit significant differences across the five evaluated dimensions. These findings align with the experimental results reported by Macedonia et al. (2020) and Fuhrman et al. (2021). The aforementioned results may be closely related to the duration of learning. Mayer et al. (2017) have pointed out that for learning through action, learners require at least 10 hours of practice to meet the production requirements, whereas the average learning time in this study was only 40-50 minutes. In addition, Macedonia et al. (2020) revealed that the effectiveness of virtual learning in terms of recognition only becomes evident in delayed tests.

The directionality test, which assessed whether participants had acquired the processing features of the learning targets, revealed a marginally significant main effect of directional

consistency for word phrases learned in both virtual scenes and videos, with a relatively large effect size. Similarly, the main effect of directional consistency was also marginally significant for unfamiliar word phrases, accompanied by a large effect size. These findings suggest that under both learning modalities, participants may have acquired the directional features of the target verbs, demonstrating an embodiment effect in processing and the ability to transfer this knowledge to new contexts. However, due to the small sample size, the effects were not statistically significant. Nevertheless, the observed patterns are close to the embodied processing characteristics of native speakers (see Wang Bin et al., 2019).

The lack of difference between virtual and video learning approaches indicates that videos, while not as direct as interacting with objects in virtual scenes, may still activate learners' existing embodied experiences sufficiently to perceive relevant directional features of verbs. In contrast, the lack of embodiment effect in sentence-based scenarios suggests that compared to linguistic symbols, video and virtual learning methods are more likely to stimulate kinesthetic perception and facilitate embodied learning. This underscores the potential of these modalities in enhancing learning outcomes by engaging learners' embodied experiences.

5. Limitations and Future Directions

This study examined whether embodied learning in virtual environments can facilitate the learning of Chinese verbs. The current research results indicate that there are no significant differences between video and virtual modalities in terms of memory, processing, and recognition. Considering the relatively small number of participants in the current study, as well as the insufficient learning time and frequency, future research can add to the number of participants, increase the learning time, or incorporate delayed testing to more accurately investigate potential differences.

References

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22(4), 577-660.

Fuhrman, O., Eckerling, A., Friedmann, N., et al. (2021). The moving learner: Object manipulation in virtual reality improves vocabulary learning. Journal of Computer Assisted Learning, 37(3), 672-683.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. Psychonomic Bulletin & Review, 9(3), 558-565.

Lan, Y. J., Fang, S. Y., Legault, J., et al. (2015). Second language acquisition of Mandarin Chinese vocabulary: Context of learning effects. Educational Technology Research and Development, 63, 671-690.

Macedonia, M., Lehner, A. E., & Repetto, C. (2020). Positive effects of grasping virtual objects on memory for novel words in a second language. Scientific Reports, 10(1), 10760.

Macedonia, M. (2014). Bringing back the body into the mind: Gestures enhance word learning in foreign language. Frontiers in Psychology, 5, 1467.

Mayer, K. M., Macedonia, M., & von Kriegstein, K. (2017). Recently learned foreign abstract and concrete nouns are represented in distinct cortical networks similar to the native language. Human Brain Mapping, 38(9), 4398-4412.

Cai, B. G. (2010). An investigation and analysis of the mixed use of action verbs in interlanguage. Chinese Language Teaching Worldwide, 24(4), 526-535.

Lin, L. H., Hu, C. Q., & Yu, S. Z. (2019). An empirical study on the relationship between Chinese EFL learners' verb semantic processing and verb embodiment relativity. Foreign Language Teaching and Research, 51(1), 84-95, 160.

Liu, D. J., Liu, X. L., Zhang, Y., Lu, A. F., & Huang, R. H. (2016). The potential, progress, and challenges of educational applications of virtual reality technology. Open Education Research, 22(4), 25-31.

Wang, B., Li, Z. R., Wu, L. M., & Zhang, J. J. (2019). The role of embodied simulation in the comprehension of Chinese limb action verbs. Acta Psychologica Sinica, 51(12), 1291-1305.

Zheng, C. P., Xu, L. Y., Gao, M. Y., Lu, Z. H., Cheng, Q. Q., & Yang, Z. T. (2019). A systematic literature review of the application of virtual reality technology in language teaching (2009-2018). Computer-Assisted Foreign Language Education, (4), 39-47.

A Case Study on the Application of AI Video Generation Software in English Literature Teaching

WANG Yan

Institute of Disaster Prevention

Abstract:

This research paper explores the innovative applications of AI video generation software in the realm of English literature teaching. By examining the development, current applications, and potential advantages of such technology, this study aims to evaluate its impact on teaching effectiveness. Through detailed case studies, the specific ways in which AI video generation software can be utilized in English literature teaching are investigated, with a focus on enhancing student engagement, creativity, understanding, and application of literary works. Empirical data analysis is conducted to validate the potential of AI video generation software in improving learning outcomes in English literature education. The findings provide theoretical guidance and practical insights to support the advancement of innovative teaching practices in the field of English literature.

Key words: AI Video Generators; Case Study; English Literature Teaching

1. Introduction

In the fast-evolving landscape of education, the integration of technology has become increasingly prevalent, offering new opportunities to enhance teaching and learning experiences. One such technology that has shown great promise in revolutionizing education is artificial intelligence (AI), particularly in the form of AI video generation software. Within the broader context of AI technology in education, researchers have explored the benefits and challenges of integrating AI tools into language teaching. Studies have highlighted the potential of AI to personalize learning experiences, provide instant feedback, and facilitate interactive and engaging learning environments. In the specific domain of literature education, the use of AI video generation software has emerged as a powerful tool for bringing literary works to life and immersing students in the world of storytelling.

This research paper delves into the innovative applications of AI video generation software in

the field of English literature teaching, aiming to explore the impact of this technology on teaching effectiveness. By investigating the development, current applications, and potential advantages of AI video generation software, this study seeks to shed light on how such technology can be leveraged to empower teachers, inspire creativity, enhance engagement, and improve student understanding and application of literary works.

To achieve the research objectives, a mixed-methods approach will be employed. Firstly, detailed case study and practice will be carried out in English literature class. Secondly, surveys and interviews will be conducted among teachers and students to gather their perspectives and experiences related to the use of AI video generation software in literature teaching. Thirdly, a comparative study is also conducted among the same group of students experience two different ways of teaching.

The research findings of the present study aims to provide theoretical insights and practical guidelines for educators looking to integrate AI technology into English literature education, ultimately shaping the future of innovative teaching practices in the domain of literature.

2. Literature Review:

The integration of artificial intelligence (AI) technology in education has opened up new horizons for innovative teaching practices across various disciplines. In the realm of language teaching and literature education, AI tools, particularly AI video generation software, have garnered significant attention for their potential to transform traditional teaching methods and enhance student learning experiences.

Within the broader context of AI technology in education, researchers have explored the benefits and challenges of integrating AI tools into language teaching. Studies have highlighted the potential of AI to personalize learning experiences, provide instant feedback, and facilitate interactive and engaging learning environments. In the specific domain of literature education, the use of AI video generation software has emerged as a powerful tool for bringing literary works to life and immersing students in the world of storytelling.

AI video generation software allows educators to create visually compelling content that complements literary texts, enabling students to engage with literature in a multimedia format. By

generating videos based on literary works, educators can enhance students' understanding of complex themes, characters, and narrative structures, making literature more accessible and meaningful to learners.

Moreover, AI video generation software can also stimulate students' creativity and critical thinking skills by encouraging them to create their own literary-themed videos. This interactive and hands-on approach to learning not only fosters collaborative and experiential learning experiences but also cultivates students' analytical and interpretative abilities.

While the integration of AI video generation software in English literature teaching shows promise for enriching teaching practices and improving learning outcomes, it is essential to consider the potential implications and challenges associated with adopting such technology in educational settings. Issues relating to accessibility, equity, data privacy, and ethical considerations must be carefully addressed to ensure the responsible and effective use of AI tools in literature education.

Overall, the literature on AI video generation software in English literature teaching underscores the transformative potential of AI technology in enhancing the teaching and learning of literary works. By leveraging AI tools to create engaging and interactive learning experiences, educators can inspire a deeper appreciation for literature, foster creative expression, and empower students to become active participants in the world of storytelling.

3. Innovative Applications of AI Video Generation Software in English Literature Teaching

When AI Video Generation Software is applied to the teaching of English literature, the possibilities are both exciting and transformative. The following part will explore specific application of AI video generation software in English literature teaching with *Jane Eyre*, one of the most popular English novels, as an example.

A. Specific Application Aspects

To illustrate the practical applications of AI video generation software in English literature teaching, let us consider a few specific application aspects:

Plot Presentation Videos: One important part of English literature teaching is to help the

students make clear of the plot of some literature works, which is carried out by the teacher in a traditional class, and can be replaced by AI generated videos even to a better effect at the same time of saving the labor of the teacher.

Author Biographical Sketches: Through AI video generation, teachers can create engaging biographical videos of authors, shedding light on their lives, influences, and writing styles. This approach humanizes the authors and connects their personal experiences to their literary works, fostering a more profound appreciation of the texts.

Poem Reading Videos: As for poetry works, AI generated videos can also present a better appreciation and artistic enjoyment, provoking much more interest and involvement from students with appropriate background and music.

B. A case study of application of AI video generation tools in English literature with Jane Eyre as an example.

Step 1: Preparation

Determine Video Content: Text-generation tools, such as ChatGpt, Kimi, and so on, will be used to get the brief plot of *Jane Eyre* within 500 words. Here we choose Kimi.

Choose the scenes to discuss: such as Jane's childhood, her life at Lowood School, her experiences at Thornfield Hall, her relationship with Rochester, her leaving Thornfield, and the end of the story.

Write a Narration Script: Based on the chosen scenes, write a concise narration script that covers the main plot points and character relationships.

Step 2: Using Video Generator to create

Decide a Video Generator: After lots of tries and practice, Keling is chosen as the most useful and user-friendly video generator available.

Log in to Keling Video Generator: Open the Keling Video Generator website and log in with your account. If you don't have an account, you need to register first.

Create a New Project: From the homepage, select "Create New Project" or "Create Video".

Choose a Video Template: Keling offers various templates; select one that is suitable for story narration, typically one that includes images and text.

Input the Narration Script: Enter the written narration script into the video generator in segments. You can set different scenes based on paragraph divisions.

Add Visual Elements: Choose suitable images or video clips that correspond to the storyline. For example, Scenes from young Jane's life, Life at Lowood School, Exterior views of Thornfield Hall. Keling typically provides a rich gallery, allowing you to search for related images directly.

Select Background Music: Choose appropriate background music from the audio library, ensuring that the music complements the video content and creates a suitable atmosphere.

Adjust Video Settings: Set the video duration, transition effects, subtitle styles, etc., to ensure the video flows well and is easy to understand.

Step 3: Generate and Adaptation

Preview the Video: Before generating the final video, preview it to ensure all content meets your expectations. If you're unsatisfied with any point, you can adapt it.

Generate the Video: Once confirmed, click the "Generate Video" button and wait for Keling to process it.

Download: After the video is generated, you can choose to download it locally.

C. Application of the AI Generated Videos in Class

Four teachers and ten students in our course group formed a team. Based on the teaching materials and centering on the key writers and works involved in the course of English Literature, they selected and produced a total of 20 videos introducing the plots of full-length novels, 20 videos introducing writers, and 20 videos of poetry recitations. These were put into use in the English Literature class last term. When lectures come to the explanation of the plot of a novel, an author's life, or reading of a poem, these videos are played, the lecturer's role will be shifted to help and guide students appreciate and understand the contents, instead of putting in a lot of effort doing simple repetitive tasks.

4. Empirical Study

After applying the AI generated videos in English literature class for half of last term, a

comparative study and interview are made among the target students who attend the class and the teachers who take part in this practice in hope to make an evaluation on the effectiveness of this practice. The application of the videos was made during the second half of the term, with the first half processed in the traditional way, namely, the lecturer takes the major responsibility in the class. In this way, the same group of students can experience two completely different ways of teaching, and have a deeper understanding of them. So do the teachers.

A. Research methodology

Comparative study: The application of the videos was made during the second half of the term, with the first half processed in the traditional way, namely, the lecturer takes the major responsibility in the class. In this way, the same group of students can experience two completely different ways of teaching, and have a deeper understanding of them. So do the teachers.

Interview method: After last term's practice, interviews via video conferencing are conducted among both the target students (25 volunteers from 92 students) and the teachers, 4 in number, to gather in-depth insights and perspectives from participants, which can reflect the effect of the practice for further development and adaptation.

B. Data collection and analysis

Interviews are conducted online through DingDing video conference, and recorded. The contents of the interviews are transformed into words. The following table presents the major results of the interviews with 25 volunteer students.

Major Questions	Positive Responses	Negative Responses
Initial Exposure to AI Videos	They have exposure to other form of AI usage in class, such as AI translation, AI writing.	All of the25 participants have no experience of AI videos in class.
Content Quality	The rest hold that the quality of the AI generated videos is satisfying.	4 of the 25 participants hold that there is space for improvement.
Learning Experience	Most of them have got a brand new experience and enjoy this form of presentation.	7 of the 25 participants hold that they find it sometimes difficult to catch up with the videos.

Technical Aspects	12 of the 25 think the technical aspects are acceptable.	13 of the 25 participants hold that the AI video generators are immature in their development. It'll be much better in the future.
Comparison with Traditional Teaching Methods	21 of the 25 prefer a combination of AI technology and traditional teaching methods.	4 of the 25 prefer traditional way of teaching.
Impact on Study Habits	All of the 25 participants said this experiment of using AI in class is quite inspiring for their future study and habits. They have got the consciousness of using AI.	3 of them talked about the challenges as well, such as reducing deep thinking of some problems in the quick shifting of the videos.
Suggestions for Improvement	1) Necessary pause in consideration of different levels of understanding; 2) A better combination of AI and real teacher teaching; 3) Improvement of the quality of the videos; 4) Diversity of the styles of the videos	

The second table presents the major results of the interviews with 4 teachers who take part in both the creation of the videos and the application of these videos in class.

Major Questions	Positive Responses	Negative Responses
Initial Exposure to AI Videos	1 of the 4 teachers interviewed said she had the experience of using AI in her teaching. She used AI in examination and writing.	3 of the 4 said they seldom use AI in their teaching because of their imagination of the difficulties of using it.
Difficulties and Challenges	1) The unfamiliarity with the new AI tech; 2）the extreme difficulty in choosing the most efficient video generator from so many choices; 3) Too many frustrating incidents during the video making.	
Teaching Experience	1) All of the four teachers agreed that once the videos are made and applied in class, it's quite satisfying and time and labor saving. 2) These videos are also very helpful in motivating the students and attracting their interest and attention. 3) Students' participation in class is increased greatly, too.	Almost none of them talked about negative teaching experience.

	4) The atmosphere of the class is activated.	
Comparison with Traditional Teaching Methods	1) All of the 4 teachers appreciate using AI generated videos for its efforts saving, comparing with the traditional teaching methods in which teachers will have to make a continuous output. 2) With the aid of AI videos shifting of teachers' role from a speaker to a guide becomes natural and easy.	2 of the 4 teachers also mentioned the irreplaceability of teachers. Much attention should be paid to reach a reasonable balance between AI and real teachers.
Future Using	All of the four teachers expressed their willingness to use AI even to a broader range in their teaching.	One teacher mentioned the immaturity of the AI video generators too.

C. Findings and results

The integration of AI video technology in the English literature curriculum has yielded a range of findings and results, as captured through interviews with students and teachers. The following sections delve into the detailed insights gathered from the interviewees.

Student Perspectives:

Adaptability to AI Integration:

The majority of students demonstrated adaptability to the integration of AI in their coursework, with 96% (24 out of 25) acknowledging the novelty of AI video presentations. This suggests a high level of openness to technological advancements in educational methods.

Content Quality and Satisfaction:

A significant proportion of students, 84% (21 out of 25), expressed satisfaction with the quality of AI-generated videos. However, there is a notable minority who perceive areas for enhancement, indicating a need for continuous content refinement.

Learning Experience:

The AI video format was generally well-received, offering a refreshing perspective on content delivery. Yet, 28% (7 out of 25) of students encountered difficulty in keeping pace with the video content, suggesting a need for adjustable pacing to accommodate varying comprehension levels.

Technical Proficiency and Maturation:

The technical aspects of AI videos received a mixed response, with 48% (12 out of 25) of students considering them acceptable. In contrast, 52% (13 out of 25) view the technology as immature, anticipating future improvements for a more sophisticated learning tool.

Preference for Teaching Methods:

A strong preference for a hybrid approach was evident, with 84% (21 out of 25) of students favoring a combination of AI and traditional teaching methods. This highlights a desire for a balanced educational experience that leverages technology without overshadowing human interaction.

Impact on Study Habits:

The experiment with AI in the classroom was unanimously inspiring for students' future studies and habits, with all participants (25 out of 25) reporting increased awareness and potential adoption of AI tools. However, a small subset (12%, 3 out of 25) voiced concerns over reduced deep thinking due to the rapid shifts in video content.

Suggestions for Enhancement:

Students provided constructive feedback for improvement, emphasizing the need for educationally paced content, better integration with teacher instruction, enhanced video quality, and diverse video styles to cater to different learning preferences.

Teacher Perspectives:

Familiarity with AI in Teaching:

Among the teachers, only 25% (1 out of 4) had prior experience with AI in teaching, indicating a potential for increased adoption and exploration of AI applications in pedagogy.

Challenges in AI Adoption:

The unfamiliarity with AI technology and the challenge of selecting the most efficient video generator were cited as significant barriers, alongside the frustration of technical incidents during video creation.

Teaching Experience with AI Videos:

Post-implementation, all teachers reported a satisfying experience, noting the time and labor-saving benefits of AI videos and their effectiveness in motivating students and stimulating

class participation.

Shift in Teaching Role:

The transition from being a speaker to a facilitator was seen as a natural progression with AI integration, although 50% (2 out of 4) of teachers underscored the irreplaceability of teachers and the importance of maintaining a balance with AI tools.

Future Integration of AI:

There is a strong inclination towards the future use of AI in teaching, with all teachers (4 out of 4) expressing their willingness to broaden its application. However, one teacher cautioned about the current limitations of AI video generators, indicating a need for technological maturity.

To sum up the above, the findings underscore a generally positive reception of AI video technology among students and teachers. While students appreciate the innovative presentation and interactive potential of AI, they also seek a more personalized and paced learning experience. Teachers recognize the pedagogical advantages of AI, such as time efficiency and motivational benefits, yet they also acknowledge the ongoing need for human guidance and the irreplaceable role of educators. As AI technology matures, these findings suggest a promising trajectory for its integration into educational practices, provided there is a careful and considerate balance with traditional teaching methods.

6. Conclusion

In conclusion, the integration of video generation software in the realm of English and American literature teaching has ushered in a new era of educational innovation. This study has demonstrated that such technology not only enhances the visual and auditory engagement of students but also provides a dynamic platform for the exploration of complex literary themes and narratives. The software's ability to personalize content and adapt to various learning styles has been particularly beneficial, catering to the diverse needs of the student body.

However, the findings also highlight the importance of balancing technological advancements with traditional pedagogical methods. While the software offers a wealth of interactive features, it is imperative that it serves as a complement to, rather than a replacement for,

in-person discussions and critical analysis. The technical issues and concerns about content accuracy underscore the need for continuous improvement and rigorous quality control.

Looking ahead, the potential of video generation software in literature education is vast. As the technology evolves, it is poised to offer even more immersive and interactive experiences. Educators are encouraged to embrace this tool, while also critically assessing its impact on learning outcomes and the student experience. With thoughtful implementation and ongoing refinement, video generation software can significantly enrich the teaching of English and American literature.

Walking in Different Shoes: Unveiling the Attitudes of Chinese College Teachers and Students Toward Generative AI

Key words

Generative artificial intelligence, Higher education; Technology integration; Teacher beliefs

Abstract

The rise of Generative AI (GenAI) technologies and applications, such as ChatGPT, has prompted significant changes to the landscape of Higher Education (HE). Amidst the excitement and debates about the potential of GenAI in education, it is often too easy to ignore the perspectives and attitudes of students and teachers toward GenAI and their views on how technology should be integrated into the learning environment. This study fills a gap in the literature by conducting a survey of Chinese college teachers to explore their awareness, usage patterns, and attitudes toward GenAI. The survey employed a comprehensive questionnaire that delved into various aspects of GenAI, including awareness, use patterns, and attitudes toward the technology. To offer a more holistic view, the attitudes and usage patterns of these teachers were compared with those of students using the same scale. Intriguingly, the results revealed that despite the differences in awareness and use patterns, both teachers and students exhibited similar positive attitudes toward the utility and positive impact of GenAI. However, students seemed to be more concerned about the use of GenAI tools than teachers. The study underscores the importance of providing support to both groups and considering both teacher and student perspectives in policy-making and curriculum design related to GenAI.

Authors

Ling Angela Xia, Helen Beech

Xi'an Jiaotong Liverpool University

Introduction

The integration of Artificial Intelligence (AI) technologies has become increasingly prevalent in our daily lives. In Higher Education (HE), AI-based applications such as automated grading systems and personalized learning platforms have revolutionized teaching and learning processes (Zawacki-Richter et al., 2019). However, it was not until the rise of Generative AI (GenAI) that the potential impact on academic practices truly became apparent (Farrelly & Baker, 2023; Wang et al., 2024).

Applications like ChatGPT are built on Large Language Models that have been trained on extensive datasets, allowing them to generate human-like text responses to user queries (Bozkurt, 2023). Shortly after its release, ChatGPT gained significant popularity and has been widely adopted by both students and academics in HE (Tseng & Warschauer, 2023). While it is important to assess the overall implications of GenAI on academic practices (Yang et al., 2024), it is equally crucial to understand and compare the attitudes and perceptions of teachers and students toward this emerging technology. Both teachers and students are key stakeholders in the educational landscape, and their attitudes can significantly influence the integration and usage of generative AI in teaching and learning (Tour, 2020). Previous research has pointed out that teachers' voices are often underrepresented in studies examining the adoption of AI-based solutions in HEIs (e.g., Lai & Bower, 2019). Similarly, while students are often regarded as "digital natives" who are more receptive to technological advancements (Prensky, 2011), their perceptions and attitudes toward GenAI may differ from those of teachers. Although a few studies have explored the responses of the academic community toward GenAI, they either focused entirely on teachers' perspectives (Walter, 2024) or students' views (Ding et al., 2023; Kurniati & Fithriani, 2022) in isolation and failed to make a comparison between the two groups. The current study aims to address this gap by conducting a survey of Chinese college teachers to explore their awareness, usage patterns, and attitudes toward Gen AI and comparing their responses to those gained from a group of students studying in a similar context.

Literature Review

Attitudes toward technology have been extensively studied in the literature on technology acceptance. In the extended model of the Technology Acceptance Model, for example, attitudes toward technology are directly influenced by the perceived usefulness and perceived ease of use of the technology and can, in turn, influence individuals' behavioral intention and actual usage (Kelly et al., 2023). In studies about the acceptance of AI technologies, researchers have also found that attitudes toward AI can greatly influence how people perceive and use AI systems (e.g., Bergdahl et al., 2023)

However, unlike traditional technologies, where individuals tend to hold more or less consistent attitudes, research has shown that people demonstrate diverse and divergent views on AI (Ding et al., 2023). This can be partly explained by the fact that users reported that they did not have the freedom to completely opt out of AI-powered services (Bergdahl, 2023). In addition, AI systems are deemed more complex and sophisticated, and some worry that AI may exceed or even replace humans (Cave et al., 2019). The divergent views about AI are even more pronounced with the rise of GenAI. In the HE landscape in particular, some become strong advocates of GenAI by emphasizing the inevitable transformation of education practices (Bozkurt, 2023; Tseng & Warschauer, 2023). In contrast, others voiced strong concerns about ethical and integrity issues (Liang et al., 2023), and even about the

future of certain academic disciplines (Crosthwaite & Baisa, 2023).

Amid the extensive discussions surrounding the application of generative AI in higher education, it is crucial that the attitudes of both teachers and students toward this emerging technology are examined, especially when implementing education initiatives involving GenAI. As mentioned above, most existing studies on the attitudes of the academic community toward GenAI were conducted in isolation and, therefore, failed to highlight the unique differences between teachers and students.

Another shortcoming identified in the existing research is the lack of robust GenAI attitude scales. For example, in Ding et al. (2023)'s research about students' perceptions of ChatGPT, the students were only asked to report their experiences in learning with ChatGPT, and their attitudes and concerns toward this tool were not measured in a systematic way. Moorehouse and Kohnke (2024)'s study investigated teachers' attitudes toward GenAI tools. However, their study involved in-depth interviews with thirteen participants, who may not be representative of the larger teacher population. In the current study, an adapted version of the General Attitudes Toward Artificial Intelligence Survey (GAAIS) was used. GAAIS, developed by Schepman and Rodway (2020), is a more comprehensive scale that measures both positive and negative attitudes toward AI in different domains. The original scale included 20 items (12 positive items and 8 negative items) and was initially validated. The current study plans to adapt the GAAIS survey to specifically measure attitudes toward GenAI in an educational context. The research questions (RQs) include:

RQ1: How and why are teachers utilizing generative AI tools, and how does this differ from students?

RQ2: What are the attitudes of teachers regarding the use of generative AI, and how do these differ from students?

Methods

A student version of the survey was administered to students at a large EMI university in China in April, 2023. The first author of the current paper was also the principal investigator of the student survey project and was responsible for data collection among students. The student survey included items related to their awareness, usage patterns, attitudes toward generative AI in education, as well as their training needs. Students' attitude toward GenAI was measured by an adapted version of the GAAIS, which included 10 attitude-related items (6 positive items and 4 negative items). A sample positive item is "There are many beneficial applications of Gen AI/ChatGPT that can help me with my studies". All 10 items were measured on a five-point Likert scale, with "1" being strongly disagree and "5" being strongly agree.

All undergraduate and postgraduate students registered at the university received an email with links to the ethical approval, participant information, and instructions for completing the survey. It was made very clear that participation was completely voluntary. Students were given two weeks to complete the survey. In the end, we received 716 valid and complete responses. More detailed information about the participants can be found in Table 1 below.

The teacher version of the survey followed a similar design to the student survey. In addition to survey items related to awareness, usage patterns, attitudes toward generative AI in

education and training needs, the teacher survey also included items specific to teachers' perceptions of the impact of generative AI on their teaching practices, concerns, and reservations regarding its implementation. The teachers' survey was circulated to a sample of college teachers in China in the summer of 2023, using a combination of convenience sampling and snowball sampling techniques. The participants were given ample time to complete the survey, approximately four weeks, and were assured that it was voluntary and anonymous. In the end, we were able to collect 129 complete and valid responses. The first author obtained approval from the student survey project team to use students' responses in the current study, and all survey data were analyzed using SPSS Statistics software version 27.

Table 1: Survey Participants Information

		Students (Total: 716)	Teachers (Total: 129)
Gender			
	Male	249	27
	Female	432	98
	Prefer not to answer	35	4
Year of Study			
	Year 1	117	---
	Year 2	85	---
	Year 3	47	---
	Year 4	35	---
	Master Students	368	---
	PhD Students	64	---
Academic Field			
	STEM	271	14
	Non-STEM	379	115
	Not decided	66	---
Types of Instions			
	Project 985 and Project 211 institutions	---	24
	Other Double First-Class insititutions	---	5
	EMI university	716	44
	Regular public undergraduate universities	---	10
	Regular private undergraduate universities	---	46

Results

RQ1: Awareness and use patterns of GenAI tools

Despite the time difference between the students' survey and teachers' survey (April, 2023 and summer 2023, respectively), it's worth noting that merely 1.1% of the students had not

heard about any GenAI tools when surveyed. By contrast, 11.6% of the teacher participants indicated that they had not heard of any GenAI tools. It is evident that students showed a higher awareness of the GenAI tools listed in the surveys, indicating that there was more familiarity with these tools among students compared to teachers. Among those who have heard about GenAI, the teachers and students survyed seemed to have different information channels about such tools. Although social media played a key role as the information source for both teachers (63.2%) and students (76.8%), more teachers rely on word of mouth or recommendation from colleagues (21.9%) compared to students (6.2%).

As for the use of GenAI tools, 61.0% of the student participants and 52.8% of the teacher participants had used GenAI tools prior to completing the survey. Among those who had used these tools, the survey results indicate similar frequency of use of GenAI tools among both groups. A slightly higher percentage of students (18.6%) reported using GenAI tools frequently compared to teachers (16.7%). When it comes to occasional use, the figures were also similar, with students (43.1%) only marginally more likely than teachers (42.1%) to use these tools from time to time. Among those aware of GenAI but who hadn't t tried the tools, accessibility seems to be a barrier for both groups; 13.2% of teachers and 16.0% of students were interested in trying GenAI tools but lacked access to them.

We also investigated the purposes of using the GenAI tools and the level of satisfaction among those who reported having used them. The results indicated that teachers and students were using GenAI tools for different reasons, and they also reported different levels of satisfaction with these tools. While the majority (62.7%) of teachers reported that they were using GenAI tools out of general curiosity, only 29% of students did so. The top three purposes of using GenAI tools for students included language assistance, coursework assistance, and the generation of free images and videos (10%). Teachers, on the other hand, were more likely to use the tools for professional purposes, such as research support and assessment topic generation.

Interestingly, teachers generally reported higher satisfaction with the usefulness and quality of the output generated by GenAI tools than students. For example, over 60% of teachers reported that they were either satisfied or very satisfied with the quality of the output generated for research (65.7%) and assessment generation (63.6%). The high satisfaction rates in these areas suggest that teachers saw value in applying GenAI tools for job-related tasks. In contrast, fewer students (55.4%) reported they were satisfied with the output for language assistance or coursework help. Based on the use experiences, teachers are more likely to recommend these tools to peers (87%) than students (73%).

RQ2: Attitudes and Concerns toward GenAI

As was mentioned earlier in the article, an adapted version of the GAAIS survey was used in both the teachers' survey and the students' survey to assess attitudes toward GenAI tools. Both surveys showed satisfactory reliability (Positive GAAIS scale: Cronbach's alpha = 0.88 (students) and 0.79 (teachers); Negative GAAIS scale: Cronbach's alpha= 0.75 (students) and 0.67 (teachers)).

The findings from the surveys revealed that both teachers and students generally held positive attitudes toward GenAI adoption. Although students reported slightly higher positive attitudes (average rating of 3.83) toward GenAI adoption compared to teachers (average rating of 3.74), the independent T-test revealed that the difference was not statistically significant

(t(841)= -1.256, p > 0.05).

However, when it comes to the Negative GAAIS items, it was quite surprising to find that students displayed a higher level of negative attitudes toward GenAI adoption compared to teachers. After recoding the negative items, the average rating for negative attitudes among students was 2.88, while the average rating among teachers was 3.12, indicating that teachers are generally more accepting and less concerned about GenAI adoption compared to students. The independent T-test showed that the difference in negative attitudes between students and teachers was statistically significant (t(841)=3.167, p < 0.05). This finding was surprising because age was typically associated with technology anxiety and resistance to new technologies (Tsai et al., 2020). Yet, in this case, it seems that teachers have embraced GenAI adoption more readily than the students. The reasons for this discrepancy in attitudes will be discussed further in the following sections.

We also conducted a series of tests to analyze the potential external factors, such as gender, types of institutions, and academic field, that may influence teachers' attitudes toward GenAI adoption. The analysis revealed that gender and types of institution did not have a significant influence on either teachers' positive or negative attitudes toward GenAI. Given that a significant number of participants were language teachers, it is worth examining whether language teachers and non-language teachers have different attitudes toward GenAI. In Felten et al. (2023)'s analysis about the extent to which occupations were exposed to advances in AI language modelling, English language and literature teachers and foreign language and literature teachers were found to be among the occupations with the highest exposure to AI language models. Understandably, GenAI's capabilities and potential impact on language teaching may have played a role in shaping the attitudes of language teachers toward GenAI adoption. However, our results indicated that language teachers reported very similar positive attitudes toward GenAI (3.72) compared to non-language teachers (3.74), suggesting that both groups recognize the potential benefits of GenAI in their respective fields.

Discussion and Conclusion

The current study compared how college teachers and students use GenAI tools like ChatGPT and perceive its advantages and challenges. Using an adapted version of the GAAIS survey, our study is among one of the first to carry out a direct comparison of the perception of GenAI among teachers and students. The research findings unveiled both intriguing similarities and disparities in their attitudes.

Firstly, despite the fact that the teachers' survey was administered later than the students' survey, the student participants reported higher awareness of GenAI tools like ChatGPT. In addition, the results also suggested that students were relying on different information channels to gather information on GenAI. Although both groups primarily obtained information through social media, a lot more teachers relied on recommendations from peers compared to students. The difference may be attributed to the fact that students are generally more tech-savvy and eager to explore new technologies like GenAI (Hmoud et al., 2024). By contrast, teachers may not have sufficient time to experiment with these tools; therefore, advice from professional networks probably plays a bigger role in shaping teacher's knowledge and understanding of GenAI. This also highlights the need to provide teachers with more professional training opportunities and resources to stay informed about the latest GenAI developments (Kizilcec, 2024).

Similar to the awareness level, a higher percentage of student respondents reported prior

experience using GenAI tools compared to teachers. However, the results showed among those who had used GenAI tools, the use frequency was rather similar between students and teachers. The two groups also reported different main reasons for using GenAI tools. While students primarily relied on the tools for language assistance and completing coursework, teachers used them more for testing AI capabilities, research, and developing assessment topics. This is unsurprising considering the types of tasks that students and teachers often need to complete in their daily lives, and could also be linked to confidence and competence levels in terms of the content and language knowledge that each possess. While students seem to use GenAI tools more in a supportive capacity, teachers seem to be using them more in an exploratory way.

Analysis of the adapted GAAIS results revealed both interesting similarities and differences in the attitudes of teachers and students. In terms of similarities, both groups exhibited relatively positive attitudes toward the overall potential benefits of GenAI in education, as indicated by their high mean scores on the Positive GAAIS subscale. The results are probably due to the fact that both students and teachers recognize the wide-ranging benefits and potentials of GenAI tools in their daily lives (Bower et al., 2024; Southworth et al., 2023). However, surprisingly, our survey results indicate that students were more concerned about GenAI tools than teachers, as indicated by the significant difference on the Negative GAAIS subscale for teachers compared to students. This finding contradicts the general assumption that students would be more receptive to educational technology and less apprehensive about GenAI tools (Sprenger & Schwaninger, 2021). One possible explanation for this is that students may be more sensitive to the potential issues and risks associated with the use of GenAI, such as academic misconduct. Previous studies have highlighted similar concerns of students about accuracy, transparency, and ethical considerations in the use of AI (e.g.,Chan & Hu, 2023). On the other hand, teachers may face less pressure in using GenAI compared to students and thus perceive fewer concerns. They may also possess more content and language knowledge, as well as greater awareness about ethical use in an academic context, and as such have greater confidence in their ability to utilize these tools effectively. Another factor could be that teachers may feel they have more autonomy and agency to decide whether to use GenAI in teaching, and also in what ways to utilize these tools,as opposed to students who are generally required to follow the guidance provided by their teachers which could result in increased feels of uncertainty about how to implement GenAI tools effectively. Students may also experience peer pressure and feel the need to use GenAI tools in their studies in order to keep up with their classmates rather than out of a genuine desire to utilize such tools all of which may contribute to a feeling of not being able to control the use of GenAI tools. This may contribute to students' higher level of concern. It will be interesting to explore these differential concerns through qualitative research to understand the nuances.

Lastly, our study carries important implications for the integration of GenAI tools in higher education settings, particularly in EMI universities. The fact that students were largely relying on GenAI tools to assist with language and their coursework suggests that EMI universities need to address the appropriate use of such tools proactively. This includes providing guidance on how students can use GenAI tools to enhance their learning, while also teaching students how to adhere to academic integrity policies. Related to this, the higher level of concern expressed by students compared to teachers highlights the need for institutions to develop tailored AI literacy programs to address students' apprehensions. AI literacy, defined by Long and Magerko (2020) as "a set of competencies that enable 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." (p.2), has garnered

increased interest recently and it is recognized that there is a real need to develop targeted training programs for AI literacy. Hopefully these interventions can equip students with the skills to critically evaluate and use GenAI responsibly.

In short, the differences in awareness, prior experience, and use patterns between teachers and students necessitate a coordinated effort to ensure all stakeholders are equipped with up-to-date knowledge on GenAI development. Again, carefully designed AI literacy training programs for both teachers and students, coupled with the provision of institutional policies and resources, will be key to promoting responsible adoption of GenAI in all classrooms.

References

Bergdahl, J., Latikka, R., Celuch, M., Savolainen, I., Soares Mantere, E., Savela, N., & Oksanen, A. (2023). Self-determination and attitudes toward artificial intelligence: Cross-national and longitudinal perspectives. *Telematics and Informatics*, *82*, 102013. https://doi.org/10.1016/j.tele.2023.102013

Bower, M., Torrington, J., Lai, J. W. M., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12405-0

Bozkurt, A. (2023). Unleashing the Potential of Generative AI, Conversational Agents and Chatbots in Educational Praxis: A Systematic Review and Bibliometric Analysis of GenAI in Education. *OPEN PRAXIS*, *15*(4), 261–270. https://doi.org/10.55982/openpraxis.15.4.609

Cave, S., Coughlan, K., & Dihal, K. (2019). "Scary Robots": Examining Public Responses to AI. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 331–337. https://doi.org/10.1145/3306618.3314232

Chan, C. K. Y., & Hu, W. (2023). *Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education* (No. arXiv:2305.00290). arXiv. https://doi.org/10.48550/arXiv.2305.00290

Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics*, *3*(3), 100066. https://doi.org/10.1016/j.acorp.2023.100066

Ding, L., Li, T., Jiang, S., & Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education*, *20*(1), 63. https://doi.org/10.1186/s41239-023-00434-1

Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Education Sciences*, *13*(11). https://doi.org/10.3390/educsci13111109

Felten, E. W., Raj, M., & Seamans, R. (2023). How will Language Modelers like ChatGPT Affect Occupations and Industries? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4375268

Hmoud, M., Swaity, H., Hamad, N., Karram, O., & Daher, W. (2024). Higher Education Students' Task Motivation in the Generative Artificial Intelligence Context: The Case of ChatGPT. *Information*, *15*(1), 33. https://doi.org/10.3390/info15010033

Kelly, S., Kaye, S.-A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, *77*, 101925. https://doi.org/10.1016/j.tele.2022.101925

Kizilcec, R. F. (2024). To Advance AI Use in Education, Focus on Understanding Educators. *International Journal of Artificial Intelligence in Education*, *34*(1). https://doi.org/10.1007/s40593-023-00351-4

Kurniati, E. Y., & Fithriani, R. (2022). Post-Graduate Students' Perceptions of Quillbot Utilization in English Academic Writing Class. *Journal of English Language Teaching and Linguistics*, *7*(3), 437. https://doi.org/10.21462/jeltl.v7i3.852

Lai, J. W. M., & Bower, M. (2019). How is the use of technology in education evaluated? A systematic review. *Computers & Education*, *133*, 27–42. https://doi.org/10.1016/j.compedu.2019.01.010

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, *4*(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

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

Moorhouse, B. L., & Kohnke, L. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. *Procedia - Social and Behavioral Sciences*, *122*, 1860–1868. https://doi.org/10.1016/j.system.2024.103290

Prensky, M. (2011). Digital natives, digital Immigrants. *On the Horizon*, *9*(5). https://marcprensky.com/writing/Prensky-The_Emerging_Online_Life_of_the_Digital_Native-03.pdf

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes toward Artificial Intelligence Scale. *Computers in Human Behavior Reports*, *1*, 100014. https://doi.org/10.1016/j.chbr.2020.100014

Southworth, J., Migliaccio, K., Glover, J., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, *4*, 100127. https://doi.org/10.1016/j.caeai.2023.100127

Sprenger, D. A., & Schwaninger, A. (2021). Technology acceptance of four digital learning technologies (classroom response system, classroom chat, e-lectures, and mobile virtual reality) after three months' usage. *International Journal of Educational Technology in Higher Education*, *18*(1), 8. https://doi.org/10.1186/s41239-021-00243-4

Tour, E. (2020). Technology Use in ESL: An Investigation of Students' Experiences and the Implications for Language Education. *TESOL in Context*, *20*(1), 5–21. https://doi.org/10.3316/ielapa.449297817562997

Tsai, T.-H., Lin, W.-Y., Chang, Y.-S., Chang, P.-C., & Lee, M.-Y. (2020). Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. *PLOS ONE*, *15*(1), e0227270. https://doi.org/10.1371/journal.pone.0227270

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

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00448-3

Wang, N., Wang, X., & Su, Y.-S. (2024). Critical analysis of the technological affordances, challenges and future directions of Generative AI in education: A systematic review. *Asia Pacific Journal of Education*, 1–17.

Yang, Z., Wu, J. G., & Xie, H. (2024). Taming Frankenstein's monster: Ethical considerations relating to generative artificial intelligence in education. *Asia Pacific Journal of Education*, *0*(0), 1–14. https://doi.org/10.1080/02188791.2023.2300137

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Clustering, visualising and elaborating DDL-integrated error correction process in EFL writing activities

Fangzhou Zhu, Department of Linguistics and English Language, Lancaster University, UK

Tingjun Wang, Department of Educational Research, Lancaster University, UK

Abstract

The use of language corpora for EFL (English as a foreign language) learning purposes, known as data-driven learning (DDL), has shown effectiveness from many recent studies. Although DDL seems promising, it is barely popularised in Chinese EFL context for distinct reasons, such as deficiency of process investigation and exclusivity of other available consultation resources in DDL research.

This study aims to unpack DDL-integrated EFL writing error correction in real world, where common consultation resources can be referred to along with DDL. Fifty-nine participants in a Chinese university completed six writing tasks with follow-up revisions, eleven of which then joined stimulated recalls. *TraMineR*, a sequence data analysis toolkit, was used to visualise consultation processes for error correction and cluster representative trajection by error types. Retrospective device provided insightful details to further explain sequential data in error correction processes.

Results showed DDL played a significant role in error correction activities, though it functioned variously regarding error types. DDL helped participants either retrieve prior knowledge or explore new linguistic knowledge with multiple cognitive strategies. Drawbacks of DDL also raised by participants, indicating the importance of combining other consultation resources in error correction activities for better performance.

Keywords: DDL, error correction, EFL writing, sequence analysis

Introduction

Data-driven learning (DDL) was initially proposed in early 1990s, referring to a series of exploratory learning activities conducted by learners searching in language corpora (Johns, 1990). Since then, the following three decades have witnessed the prosperity of researching DDL application in English as foreign/second language (EFL/ESL) teaching and learning. A wide range of DDL research has proved its effectiveness in EFL writing instruction, including the promotion of language pattern acquisition and error correction (e.g., Crosthwaite, 2017). Although clear evidence from rich DDL studies indicates that DDL often leads to successful learning outcomes, it is always difficult to get popular in classroom for various reasons. One of the potential explanations, as O'Keeffe (2020) pointed out, is the lack of research connecting second language acquisition (SLA) theories, DDL theories and its practice. We are eager to see more DDL research focusing on this stance.

In response, this study intended to investigate how EFL learners utilised DDL and other consultation materials when they attempted to correct specific lexico-grammatical errors in their English

writings and explore the cognitive efforts they made during their error correction behaviours, with a mixed-methods research design involving sequence analysis and think-aloud protocols. The ultimate hope for this study is that both teachers and students may have a better understanding about DDL affordances within a real EFL classroom in China.

Literature review

DDL: theories and practices

In this section, we will briefly and accessibly review the definition of DDL, its theoretical background and latest applications in EFL writings, especially in error correction. DDL was firstly mentioned at the beginning of the 1990s by Johns (1990), referring to the application of corpora in language classrooms with a series of exploratory learning activities. It is claimed that "a corpus that contains thousands of authentic text samples can greatly enhance a learner's exposure to naturally occurring language and offer a vast linguistic resource" (Quinn, 2015, p.165). The outcome of the corpus search, concordance lines, refer to a comprehensive index of the words or phrases in a language corpus searched by syntax queries, of which most common format is the KWIC (Key Word in Context) concordance (O'Keeffe et al. 2007). To successfully apply DDL in written error correction, a learner first needs to compare the language outcome and concordance lines, then the learner interprets and groups the information from the concordance lines to form generalisations and work out a solution to a language problem (Quinn, ibid).

DDL has a number of theoretical advantages to support language learning. Besides providing input authenticity, DDL uses frequency data to help learners identify language forms and meanings which are worth knowing. Rather than being fed with pre-digested rules, consulting DDL incorporates learner's autonomy, inductive learning by interaction with contextualised examples, and guidance from more experienced teachers (varied by degree of DDL curation), which reflects both constructivism and socio-cultural learning theories. Furthermore, as Boulton and Cobb (2017) stated, DDL has potentials in reflecting many theories in applied linguistics, including psycholinguistic theories, SLA theories and other learning theories, although these advantages need more research to investigate.

Regarding to practical side, to date, many DDL studies and applications lie in the field of examining its effectiveness of written error corrections with different emphases. For example, research from Yoon and Jo (2014) indicated that indirect DDL, which is curated by teachers, resulted in higher error correction rates than direct DDL (training students to use corpus tools for self-searching). Research also reported that DDL is suitable for correcting lexico-grammatical errors in general (e.g., Bridle, 2015; Crosthwaite, 2017; Tung et al., 2015), although their error type categorisation and results remain different. There are still plenty of studies focusing on perceptions of DDL, claiming that learners positively evaluated DDL as an affordable resource for improving writing accuracy and learning language use.

However, we agreed with the viewpoint from Liu and Chen (2023) that current DDL research overwhelmingly examined the effectiveness of DDL from pre- and post-treatment tests and learners' perceptions via questionnaires or interviews. The DDL-mediated error correction process and cognitive

efforts inside have been neglected, not mentioning that most relevant research treated DDL as the only resource for referencing and ignored its potential connections with other learning material. Deficiency of process investigation and exclusivity of other available consultation resources in DDL research may barely convince non-researcher teachers to use DDL in their classrooms, as well as result in the lack of theoretical underpinnings in DDL (Boulton & Vyatkina, 2021). The very few exceptions include Liou (2019) and Zhu (2021) who blended DDL with other available tools (e.g., prior knowledge, online dictionaries, search engines) in EFL writing class and examined consultation behaviours behind.

Based on the reflective review of existing DDL research, O'Keeffe shouted out "the need for greater critical engagement with the pedagogical underpinnings in the form of theories of learning and theories of language acquisition (2020, p.269)", and Mizumoto called for "pedagogical recommendations designed to enable learners to strategically utilise a wide range of language resources, from corpora to GenAI technologies, guided by their self-awareness, the specifics of the task, and relevant strategies (2023, p.1)." These proposals became the rationale of this study and encourages more DDL research to focus on process investigation in real world language class.

Written errors produced by Chinese EFL learners

Corpus linguistics and automated scoring system facilitate the error identification and description in EFL writing context. Natural language processing tool and systematic tagging system were efficiently applied in recent research in China. For example, Chang's study used an automated scoring system to identify and analyse writing errors. Lexico-grammatical errors namely preposition errors, errors of verb tense and agreement were considered as the major errors among undergraduates' English essay writing (2013). Another study which also used such online scoring system, from Zhao (2015), found that collocation errors and subject-verb agreement errors are commonly seen in students' writing. For studies adopting corpus linguistics methods, Sun's study built a learner corpus with three years' English exam writings and pointed out that wrong collocation, wrong part-of-speech, misuse of articles, wrong verb tense, agreement and voice took up nearly 80% of the writing errors (2010). Similarly, Jichun's research collected the writings from 160 non-English major freshmen in a university and used tagging tools to detect and describe errors (2015). The research finding revealed that 65.5% of the errors came from the linguistic category, where the syntactic and lexical errors, such as collocation errors, article errors, preposition errors and verb errors (tense and voice) were mostly significant in students' writings.

To conclude, as Wang stated, "the problems of Chinese learners of English lie mostly in the grammatical and lexical level (2013, p.59)." Therefore, we chose to explore these four types of errors, given that they are the most frequent which influence the accuracy of English writing produced by Chinese undergraduate students: errors of verb, article, preposition and word choice.

Research questions

This study aimed to investigate EFL learners' actual error correction processes when DDL was introduced and integrated in their revision stage of writing tasks. We intended to use both quantitative and qualitative

data to mapping what EFL learners did with DDL and other available consultation materials and how they coordinated those resources with different cognitive efforts during error correction activities. Thus, two research questions are addressed:

1. What are the patterns of correction consultation of article, preposition, verb and word choice errors with typical materials (e.g., prior knowledge, online dictionary, search engine, paper dictionary, textbook, peer discussion), with DDL only and with both?
2. How did participants cognitively interact with consultation materials for target error correction?

Research design

Participants

Fifty-nine freshers of English major in a Chinese university joined this study, as did their teacher of Integrated English module. Student participants were preparing for Test for English Major Band 4 (TEM-4) examinations. Before commencing, we confirmed that students had limited knowledge of language corpora or DDL.

Procedure

We invited student participants to experience a term-time practice of TEM-4 writing with six writing tasks and follow-up revisions. These writing tasks were further divided into 3 stages: Baseline stage (Task 1 and 2), DDL stage (Task 3 and 4), hybrid stage (Task 5 and 6). In baseline stage, participants were encouraged to use common consultation resources to help with error correction. Here belong such resources as: prior knowledge (participants' existing knowledge taught previously), online dictionary, search engine, paper dictionary, textbook and peer discussion. In DDL stage, the materials for error correction were manipulated based on error types, in the following way. During Task 3 revision, half students were provided with DDL materials for article and word choice errors, while the rest of the students were provided with DDL materials for preposition and verb errors. For Task 4, students received a reversed version of arrangement as in Task 3. DDL materials were comprised of concordance lines which were supplied as part of feedback notes provided by teacher and researcher (DDL1), and general knowledge of target error types with concordance lines as evidence in a handout (DDL2). During the error correction of Task 3 and 4, participants were required to only consult their own knowledge and/or the DDL materials provided. In hybrid stage, students received similar manipulation of DDL integration in DDL stage for Task 5 and Task 6, but they were allowed to consult other materials besides DDL materials. Additionally, guessing as error correction technique could be used at any stage for error correction.

Analyses

Each participant kept an error correction tracking spreadsheet in which they recorded the information of original error, sequences of consultation and correction decision in Microsoft Excel as each writing task was completed. Behavioural pattens of students' engagement with consultative materials had been analysed using R software with the TraMineR package, a sequence analysis toolkit (Gabadinho et al.,

2011). TraMineR package in R environment allows to manipulate, describe, visualise and cluster the longitudinal data generated by participants. To be more specific, this toolkit can automatically quantify, visualise and cluster the sequences of material consultation behaviours observed in error correction process and explore the patterns extracted from these sequences statistically.

Eleven participants were invited to stimulated recall interviews afterwards, in which they were asked to verbalise target error correction process during baseline, DDL and hybrid stages with the stimuli from error correction spreadsheet. The goal was generally for research findings to emerge from significant themes within the error correction data, without predetermined coding or analysis schemes, to support the evidence found in sequence analysis from a larger sample size.

Results

RQ1 Sequence analysis for target error corrections

This section illustrated the general sequence data for three stages of the study and then clustered the sequential representatives for each target error type (article, preposition, verb and word choice errors). Table 1 listed the frequency information of available consultation materials based on study stage. From this table, we can conclude that students' existing knowledge consistently played a vital role in error correction activities throughout the whole study. Before introducing DDL, online resources like online dictionary and search engine, were favoured by students. However, when they started receiving DDL materials, the situation changed. DDL1, which was comprised with concordance lines curated by teacher and researchers, became popular in DDL stage, and during hybrid stage, DDL1 was still the most frequently consulted material even though other common materials were allowed to use for error correction. Guessing was more preferred for error correction than traditional resources (paper dictionary, textbook, peer discussion) as well as DDL2.

Figure 1 to 3 plotted the consultation pathways that students experienced in baseline stage, DDL stage and hybrid stage respectively. The visualised data helped us explore the distribution of consultation materials among different sorts of consultation sequences and check the participants' priorities beyond the frequency. In baseline stage, prior knowledge was "prioritised", and those error correction sequences initiated with it had bigger chance to finish consultation earlier. But when students tried out DDL, DDL1 became the first material to refer to, and this trend extended to the rest of this study. The very first two attempts from participants' consultation sequences in all three stages showcased strong tendency of material preferences, while with the increase of attempts, the choice of consultation materials varied, as the colour chunks were smaller and more fragmented. It is noticeable that there was visually more space in Figure 3 than in Figure 2, indicating the application of DDL reduced the number of consultation materials needed.

Table 1. Frequency of consultation material for error correction in three stages

	Baseline	DDL	Hybrid

DDL1	N/A		239	30%	177	19%
DDL2			173	22%	72	8%
Guessing	331	14%	213	27%	99	11%
My own English knowledge	399	17%	123	28%	146	16%
Online dictionary	359	16%			104	11%
Paper dictionary	260	11%			72	8%
Peer discussion	306	13%	N/A		78	9%
Search engine	360	16%			95	10%
Textbook	292	13%			72	8%
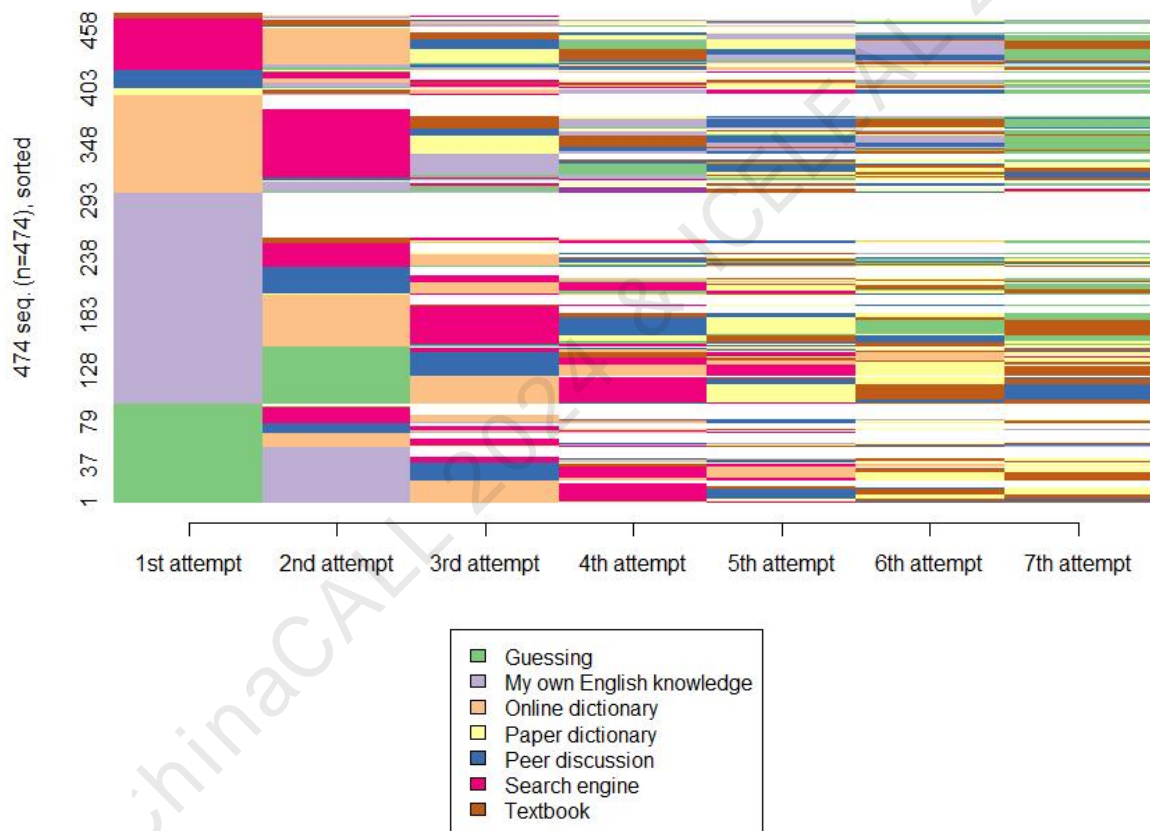
Figure 1. Sequences of consultation for target error correction by consultation attempt (Baseline stage)
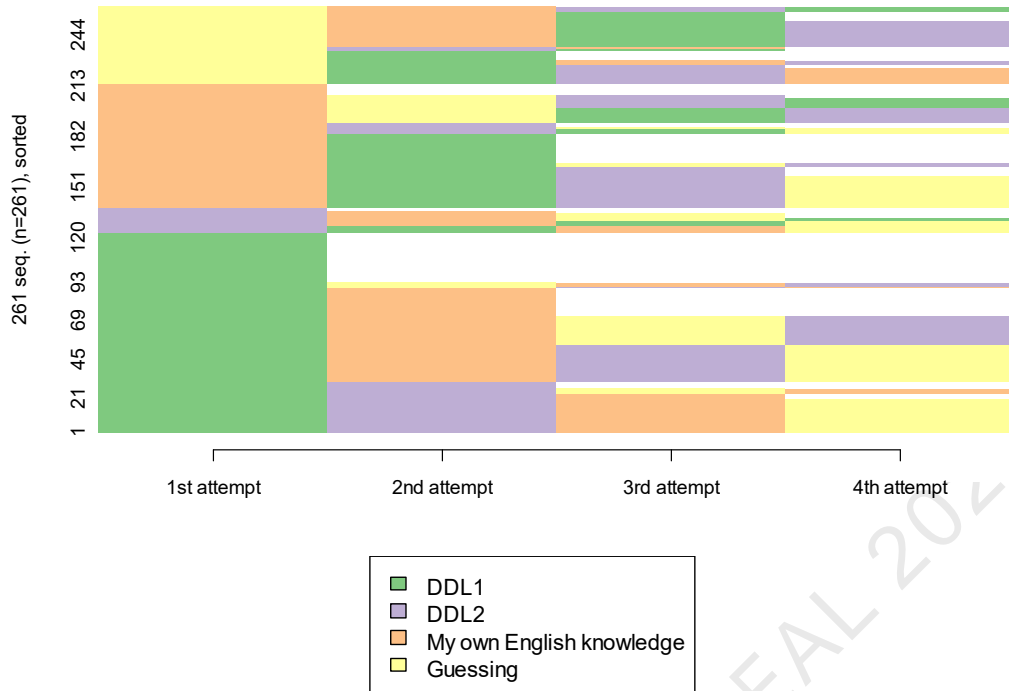
Figure 2. Sequences of consultation for target error correction by consultation attempt (DDL stage)
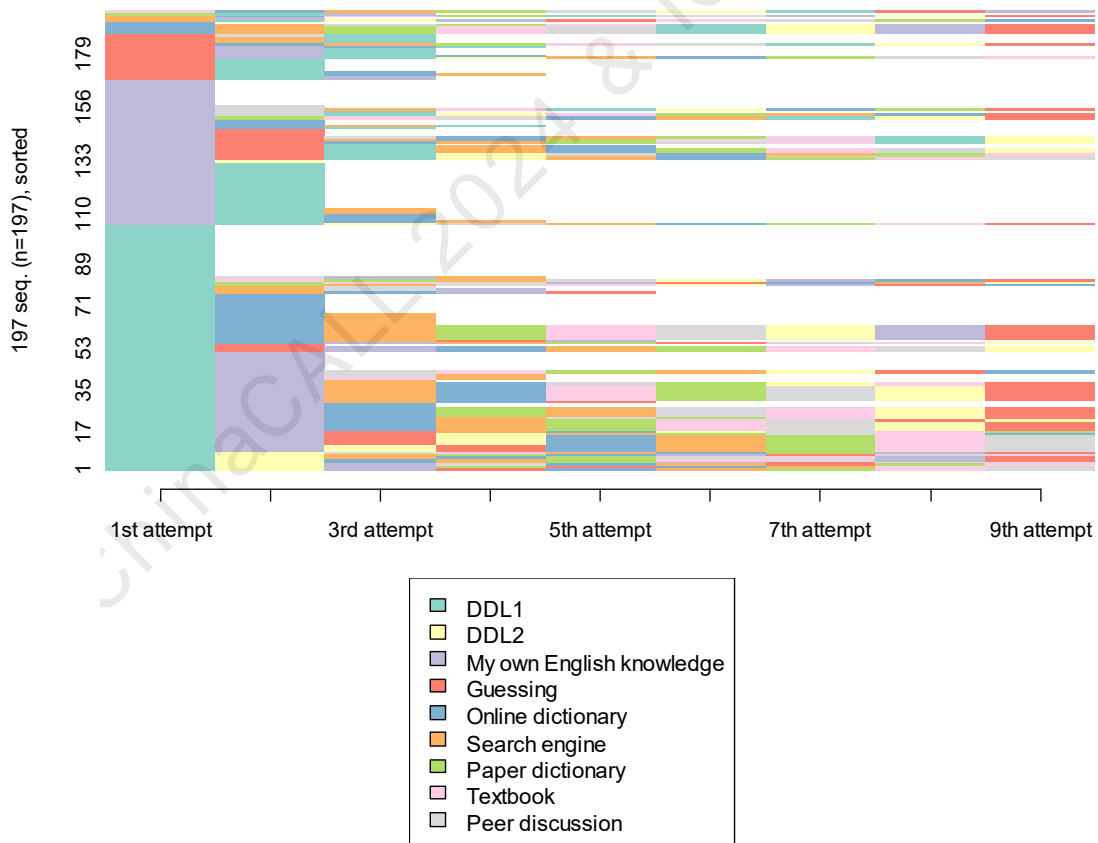
Figure 3. Sequences of consultation for target error correction by consultation attempt (DDL stage)

Table 2 illustrated the most frequent consultation sequence patterns found in three stages. Either DDL1 as the only consultation material used or its combination with students' prior knowledge occurred frequently in DDL stage and hybrid stage. Although both DDL materials were based on corpus information, DDL1 was much more favoured by students than DDL2, and DDL as consultation material profoundly challenged the use of other common materials for error correction.

Table 2. The most frequent consultation sequence patterns in three stages

	Most frequent sequence pattern	Frequency
Baseline	ME	9%
	ME-GE-OD-SE-PD-TB-PE	4%
DDL	DDL1	12%
	DDL1-ME-DDL2-GE	8%
Hybrid	DDL1	11%
	ME-DDL1	10%

Note: ME-My own English knowledge (Prior knowledge), GE-Guessing, OD-Online dictionary, SE-Search engine, PD-Paper dictionary, TB-Textbook, PE-Peer discussion

Although the general and descriptive data above highlighted the dominant position of DDL in error correction activities, we realised there might be some biases when we considered students had different knowledge levels regarding these error types. Thus, a further clustered sequence analysis based on target error types should be conducted to provide more insights about how common consultation materials and DDL materials were referred to during the study. This further approach for categorising patterns consisted of computing pairwise distances between them by means of sequence alignment algorithms and other suitable metrics. Many sequence dissimilarity measures have been proposed in the literature, of which the most popular in social sciences is the optimal matching (OM). TraMineR offered a general function that can compute the OM dissimilarity as well as a set of other dissimilarity measures. The function allowed control over the amount of information that the representative set should convey.

By default, TraMineR clustered and plotted the smallest representative set that, with a neighbourhood radius of 10%, covered at least 25% of the sequences among error type groups. In the resulting plots (Figure 4-6) the selected representative sequences were plotted bottom up according to their representativeness score with bar width proportional to the number of sequences assigned to them. We observed that throughout all three stages of study, verb errors were majorly addressed by students' prior knowledge, although sometimes students added extra resources in the following steps. For other three error types, students showed mixed engagements with available materials in different stages. We extracted some typical consultation modes in the next section for detailed exploration.
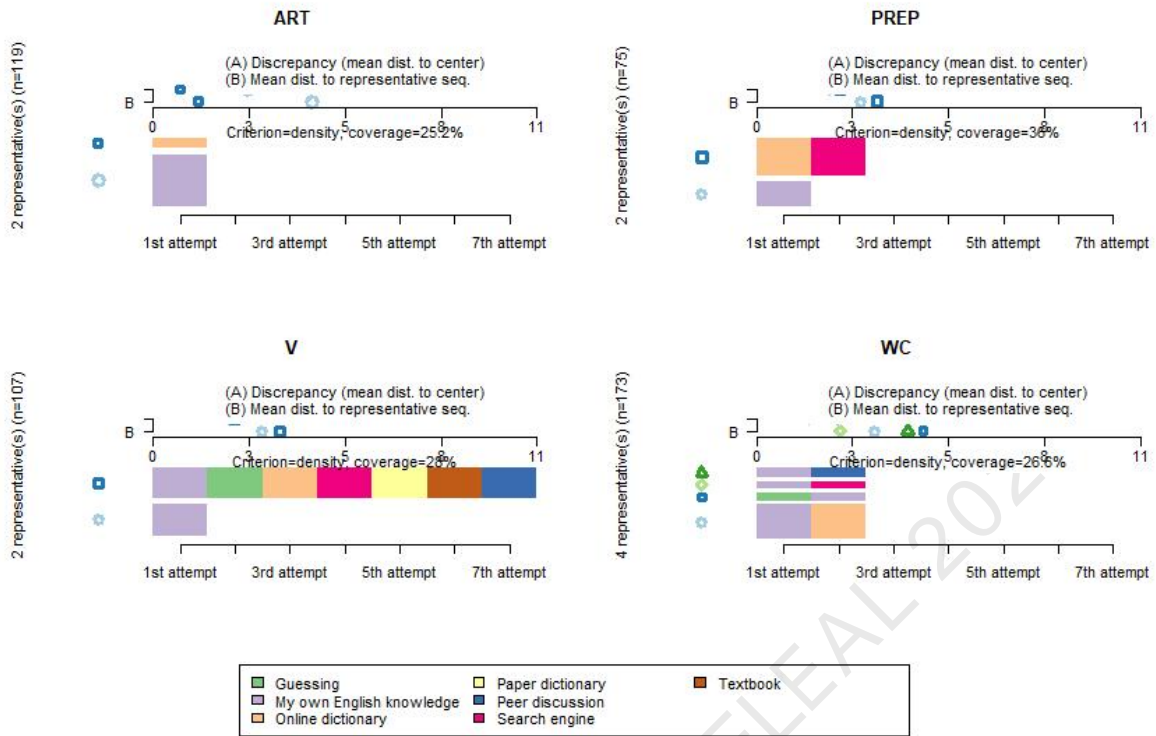
Figure 4. Representative patterns of consultation sequences based on error types in baseline stage
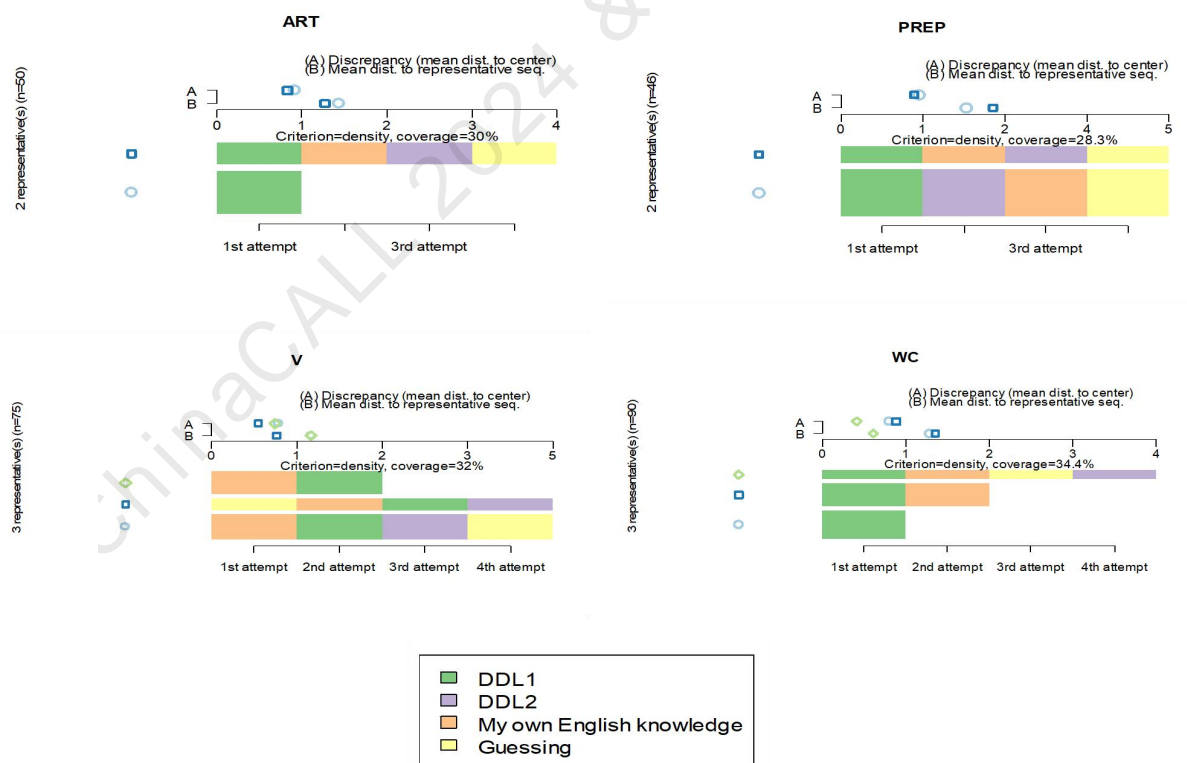
Figure 5. Representative patterns of consultation sequences based on error types in DDL stage
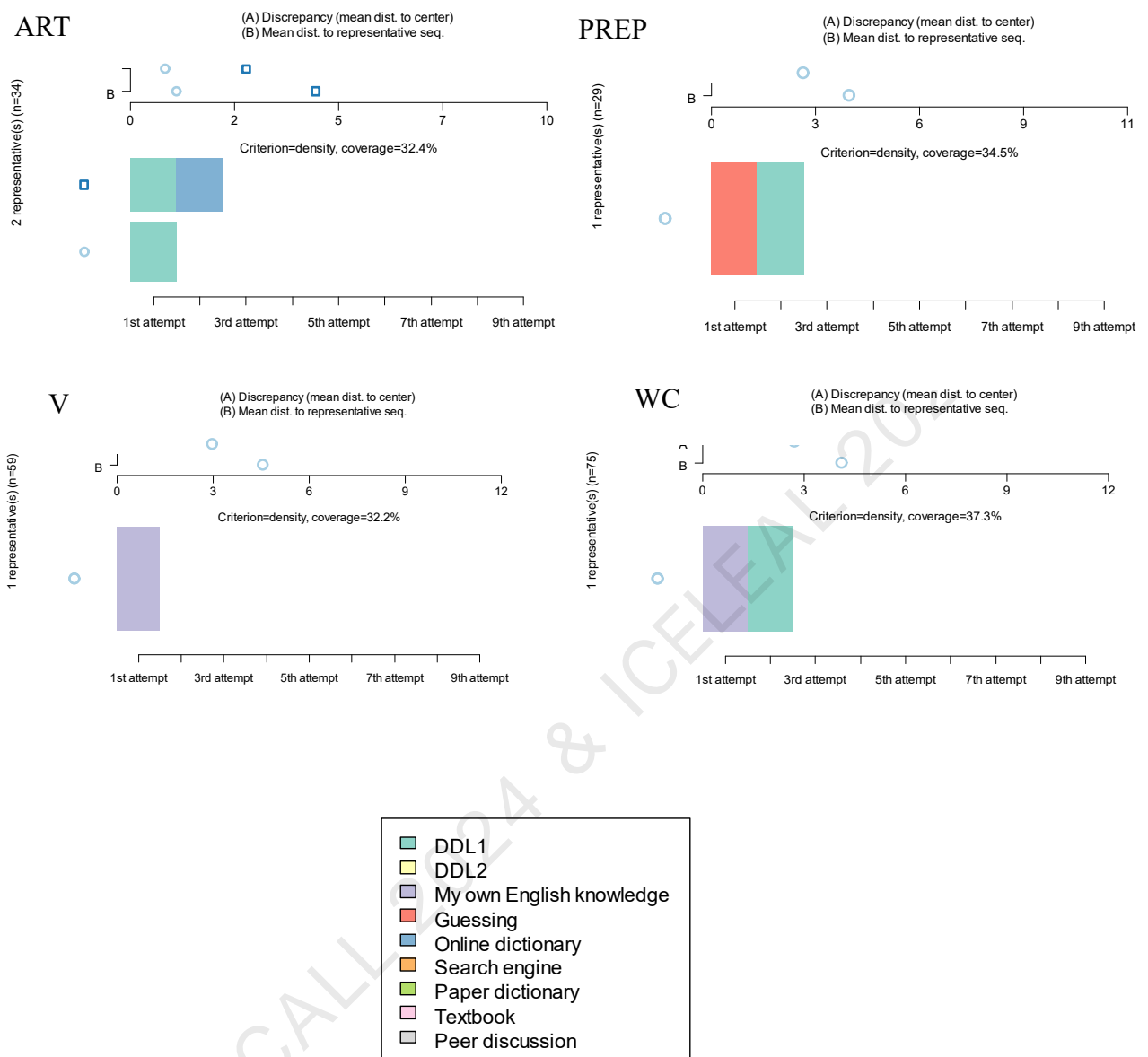
Figure 6. Representative patterns of consultation sequences based on error types in hybrid stage

RQ2 Cognitive interaction with consultation materials for error correction

The examination of qualitative data from stimulated recall interviews revealed that the processes for most error correction activities were determined by whether students had sufficient prior knowledge of target errors or not. Three main conditions were identified: sufficient knowledge without further confirmation, insufficient knowledge needing confirmation, no relevant knowledge.

In baseline stage, when students were not confident enough to use prior knowledge for error correction, they sought help from common consultation materials, especially from online resources. These behaviours were often connected with article and word choice error corrections, as participants admitted

that they were weak at mastering knowledge about them. They reported that during consultation, they repeated reading sentences with errors or translated them into Chinese to try to figure out problems. Then, they came up with some keywords to type either in search engines or online dictionaries. Finally, they read search results and tried to digest explicit contents (rules or meanings) listed for error correction solutions. However, due to the uncertainty of their interpretation, the consultation could extend to other resources. In terms of error correction results, participants often expressed the difficulties in deciding whether they truly understand what errors they generated, how they could fix them and avoid them in the future, especially for those errors they had no idea about. Emotionally, students reported the error correction process sometimes was not as satisfying as expected.

In DDL and hybrid stage, the introduction of DDL materials changed the situation. The interview data indicated that when a participant had relevant knowledge about target errors but was not confident enough, s/he would use DDL1 as a source to prove prior knowledge. When participants failed to get any useful information from their prior knowledge, most of time they directly read DDL1 for learning potential new knowledge. Interestingly, several participants reported that in some cases, after reading DDL1, they realised that they had relevant knowledge but just failed to recall immediately. No matter what scenarios the participants belonged to, their verbal reports mentioned a series of similar cognitive strategies used during DDL1 consultation. Those strategies including noticing, hypothesising, grouping, comparing, differentiating, inferring, generalising and summarising. Most interviewees agreed that consulting DDL1 was efficient for enhancing their existing knowledge or learning new knowledge, and they were confident about their follow-up corrections.

However, DDL-integrated error correction was not problem-free in this study. During interview, some students complained that DDL1 was too customised to learn universal language rules. They sometimes preferred reading DDL2 because it provided general information about target errors and concordances attached were just a piece of evidence to support. But even if students chose DDL2 for extra knowledge, a few of them expressed their worries about DDL materials limited their exploration of other alternative solutions for their target error corrections. Additionally, some participants asked about if it was possible to teach them how to search in corpus individually, as they did not want to heavily rely on curated concordances for the fear that receiving DDL material from teacher was not sustainable in life-long language learning.

Discussion and conclusion

This study aimed to investigate how Chinese EFL learners utilised common consultation materials as well as DDL materials for correcting errors of article, verb, preposition and word choice. With the aid of sequence analysis with stimulated recall, this study supported findings from previous research and showcased a picture that was often missed in terms of DDL-integrated error correction processes.

In terms of relationships between DDL materials and other common/typical consultation materials, the results extended the findings of Muller & Jacobsen (2015) and Lai & Chen (2015) to include more

consultation resources, showing that concordances, online dictionary and search engine were prioritised in students' error correction activities. The sequence visualisation also developed the settings of Liou's study (2019) for a better view about the engagement between learners and materials and confirmed the statement from Zhu (2021) that curated concordances to a large extent replaced the use of online dictionary and search engine as the time went, not to mention other common materials.

Error correction processes analysed in this study found that DDL helped participants either retrieve prior knowledge or explore new linguistic knowledge with multiple cognitive strategies. While error correction consultation with typical resources often combined with lower-level cognitive strategies such as sentence repeating and translating, participants tended to use higher-level cognitive strategies with DDL materials, such as noticing, hypothesising, grouping, comparing, differentiating, inferring, generalising and summarising. Liu and Chen (2023) believed that learners with higher language proficiency level employed more cognitive strategies in DDL-integrated activities, but their study only focused on DDL material only. This study supported the perspective from Zhu (2021) that features of DDL material, such as the involvement of students' attention and notice of erroneous areas, the exposure to authentic language input, and the implicit clues behind concordances, determined the frequent use of higher-level cognitive strategies, and this reflected a series of theories such as learning theories (constructivism and sociocultural theory) and SLA theories (noticing hypothesis).

Another contribution this study made to DDL field was the innovation of research methods. Sequence analysis helped with exploring DDL-integrated error correction process, and it generated rich data that existing literature did not disclose. The data visualisation and clustering further facilitated the stimulated recall analysis to be more systematic and purposeful when elaborating representative error correction behaviours. We believed this innovative study could inspire research in the future to focus on process investigation so that both researchers and practitioners of DDL may have a better understanding of its functions in EFL context, carefully tailor DDL according to a bunch of language learning theories and allay potential concerns in application.

References

Bridle, M. (2015). *Error correction through corpus consultation in EAP writing: an analysis of corpus use in a pre-sessional context* (Doctoral dissertation, University of Huddersfield).

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language learning*, *67*(2), 348-393.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, *25*(3), 66-89.

Chang, J. (2013). A computer-aid error analysis of Chinese college students' English compositions. *American International Journal of Social Science, 2(6)*, 103-107.

Crosthwaite, P. (2017). Retesting the limits of data-driven learning: feedback and error correction. *Computer Assisted Language Learning*, *30*(6), 447–473.

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of statistical software*, *40*, 1-37.

Jichun, P. (2015). A corpus-based study on errors in writing committed by Chinese students. *Linguistics and Literature Studies*, *3*(5), 254-258.

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria, 10,* 14-34.

Lai, S. L., & Chen, H. J. H. (2015). Dictionaries vs concordancers: actual practice of the two different tools in EFL writing. *Computer Assisted Language Learning*, *28*(4), 341-363.

Liou, H. C. (2019). Learner concordancing for EFL college writing accuracy. *English Teaching & Learning*, *43*(2), 165-188.

Liu, T., & Chen, M. (2023). An investigation into learners' cognitive processes in data-Driven learning: Case studies of six learners of Chinese. *Chinese Journal of Applied Linguistics*, *46*(4), 544-561.

Mizumoto, A. (2023). Data-driven learning meets generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics*, *3*(3), 100074.

Mueller, C. M., & Jacobsen, N. D. (2016). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL*, *28*(1), 3-21.

O'Keeffe, A. (2020). Data-driven learning–a call for a broader research gaze. *Language Teaching*, 1-14.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, *69*(2), 165–177.

Sun, J. (2010). A corpus-based study of errors in Chinese English majors' English writing. *Asian Social Science*, *6*(1), 86-94.

Tung, C. A., Chang, S. Y., & Peng, F. M. (2016). Correcting language errors in EFL writing by the use of COCA. *Malaysian Journal of ELT Research*, *11*(1), 13.

Wang, M. (2013). An empirical study on the writing ability of college students with the error analysis approach. *English Language and Literature Studies*, *3*(2), 54.

Yoon, H., & Jo, J. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology*, *18*(1), 96-117.

Zhao, L. (2015). A corpus-based investigation on Mongolian students' writing in English. *Studies in Literature and Language*, *11*(6), 35-39.

Zhu, F. (2021). Supporting EFL writing during the pandemic: the effectiveness of data-driven learning in error correction. *English as a Foreign Language International Journal*, *1*(2), 8-28.

Web-based language teaching, learning & assessment

技术赋能外语教学、学习与测评研究

Evaluating Key Indicators of Web-Based Learning Effectiveness in Teaching English Language to Filipino Students in the Post-COVID-19 Era: A Country Case Study

Jomel Ancheta Ocampo[1] and Jansen Marcos Cambia[2]

[1]*Isabela State University-Echague Main Campus, Isabela, Philippines*

[2]*Department of Public Health, International Health Program, National Yang Ming Chiao Tung University, Taipei City, Taiwan*

Corresponding Author: Jansen Marcos Cambia, PhD, International Health Program, National Yang Ming Chiao Tung University, No.155, Sec. 2, Linong St., Beitou District, Taipei City 112, Taiwan. Email: jansencambia1@gmail1.com

Abstract

The importance of web-based strategies in teaching the English language and the need for effective communication have become increasingly evident in today's globalized world, especially in the wake of the COVID-19 pandemic. This paper aims to evaluate the key indicators of web-based learning's effectiveness in teaching the English language in the Philippines during the post-COVID-19 era, as well as how the country has adapted to societal changes to improve the English language curriculum. Employing a case study in the Philippines, we assessed the key indicators of an effective teaching and learning process, as defined by the Organization for Economic Cooperation and Development, in the context of teaching the English language. Our findings revealed that the implemented English online curriculum among universities prioritizes greater flexibility, monitoring, collaboration, student-centered learning, and sensitivity to evolving individual and collective social circumstances. When classes resumed after the pandemic closed most schools in the Philippines for more than a year, they implemented three modalities: modular distance learning, online distance learning, and TV or radio-based instruction. Most parents and students chose online and modular learning. Additionally, the home learning environment posed the greatest challenge to web-based English learning, while technological literacy and competency posed the least challenge. Also, the most critical concerns revolve around comprehension of learning content, student engagement, and internet connectivity, all of which participants identified as disadvantages of flexible learning. Given the potential for continued online implementation of the English language curriculum in the Philippines for the foreseeable future, it remains crucial to persist in exploring the discourses, practices, and ideologies influenced by and shaping the individuals inhabiting this digital realm.

Keywords: Web-based; Assessment; English language; COVID-19; Case study

Terminology

ADB: Asian Development Bank

CHED: Commission on Higher Education

DepEd: Department of Education

ECQ: Enhanced Community Quarantine

FL: Flexible Learning

HEIs: Higher Education Institutions

IT: Information Technology

MELC: Most Essential Learning Competencies

OECD: Organization for Economic Cooperation and Development

1. Introduction

The COVID-19 pandemic has been one of the most challenging dilemmas the world has faced in its history. Also, the delivery of education has never been tested as it has during this time. According to UNESCO, as of 2020, around 1.6 billion learners, which is more than 90% of total enrolled learners, and 100 million educators from over 190 countries had to remain at home to prevent the spread of the virus (United Nations Educational Scientific and Cultural Organization (UNESCO) 2021). The situation was no different in the Philippines. During March to May, 2020, several schools shut down as the country was placed under the extremely strict lockdown called Enhanced Community Quarantine (ECQ). Classes at all levels were suspended, along with mass transportation and non-essential businesses (Economist 2020).

Responses like community lockdowns and quarantines in several countries have forced students and teachers to study and work from home, leading to the widespread adoption of online learning platforms. The current pandemic has compelled the several countries to adapt to online, distance learning from home (Crawford, Butler-Henderson et al. 2020). In the Philippines, this method is relatively new, with many institutions just beginning to understand it. Issues such as educators' readiness, technology resources, and geographical challenges need to be considered for the effective delivery of online education, especially in teaching English language (Ignacio 2021, Boquia 2023). This transition is not a simple shift from traditional face-to-face classroom settings to a virtual environment.

Employing a case study design, this paper aims to assess the key indicators of web-based learning effectiveness in teaching and learning the English language among students in the Philippines following the COVID-19 pandemic. The study uses the key indicators in education developed by the Organization for Economic Cooperation and Development (OECD) (The Organization for Economic Cooperation and Development (OECD)) to understand how web-based learning for teaching and learning the English language has impacted education output, education access, and the learning environment among Filipinos.

2. Key Indicators of Web-Based Learning Effectiveness
2.1. Educational Output and Learning Impact

Due to school closures caused by the COVID-19 pandemic, teachers worldwide have had to abruptly shift their traditional face-to-face classes to an online format. This rapid and unprecedented transition to what is now referred to as *"quaranteaching"* and remote learning has posed significant challenges, leaving teachers with minimal time to prepare for the intricacies of virtual instruction. As a result, educators have had to quickly adapt to new technologies, modify their teaching strategies, and find innovative ways to engage students in a completely digital environment (Pace, Pettit et al. 2020). During this time, many teachers were not equipped to handle the remote teaching environment. Moreover, no one expected the pandemic to extend beyond two years. To tackle the issues arising from the pandemic, government bodies like the Commission on Higher Education (CHED) and the Department of Education (DepEd) promoted online courses and flexible learning solutions. Additionally, people were required to follow health protocols, including maintaining physical distance, to reduce the spread of the virus (Ignacio 2021). Given that Flexible Learning (FL) imposes no restrictions on the time or pace of study, it appears to be the most practical approach during challenging periods of the global pandemic. FL offers the most convenient method of instructional delivery in such times. Although many Higher Education Institutions (HEIs) have adopted flexible learning, it is no longer a novel concept in higher education. However, numerous HEIs still struggle with a lack of pedagogical knowledge, empowerment, and readiness to develop their own content, relying instead on teaching predetermined material (Cortes 2020).

A thematic content analysis conducted in the Philippines to assess perceptions of FL design in teaching English and student coping mechanisms among college participants at Cotabato State University revealed several insights during the pre-and-post COVID-19 pandemic. Students were compelled to study independently without physical contact with their teachers, and online classes facilitated efficient studying and familiarization with various digital platforms. Additionally, it promotes student-centered learning, where students

3

acquire extra knowledge, such as English grammar and phrases, beyond what is provided by their teachers. The availability of online platforms for content searching also encouraged students to read more extensively. This aspect of FL, which requires students to read and understand instructions, further enhanced their English reading comprehension skills (Boquia 2023). Since part-time jobs are common among students in the country, the Flexible Learning (FL) teaching design allows students to choose their own pace of learning. Most students can multitask by learning English courses while simultaneously attending their part-time jobs, enabling them to support themselves and continue their academic pursuits (Payusan, Arrofo et al. 2022, Boquia 2023).

Language classes must be interactive, emphasizing peer-to-peer interaction in addition to teacher-student interaction. A lack of student interest and involvement in online learning activities has emerged as another issue in online teaching and learning. Practicing English language conversation in the country should ideally be conducted face-to-face, which might improve conversational skills, rather than relying on online communication and messaging (Boquia 2023). On the other hand, Web-based learning can promote faster and easier communication among students, enhancing their ability to collaborate on group tasks in English language classes (Lee 2024). However, since there are instances when you can't meet the person in time, online communication might not always be the best option for group tasks. You often need to wait for others to be available online to respond to group queries, which can delay progress (Alawamleh, Al-Twait et al. 2020, Boquia 2023).

2.2. Access to Education, Participation and Progression

As of January 20, 2022, schools in the Philippines have been closed for 61 weeks, causing 24.9 million pre-primary to upper secondary students to miss almost all classroom instruction time from March 2020 to September 2021 (Center for Disease Control and Prevention (CDC) , UNESCO 2022). As in other countries, enrollment rates in the Philippines decreased during the pandemic. In October 2020, about 25 million students from kindergarten to senior high school enrolled, which was only 90% of the total enrollment from the previous school year. Gross enrollment for high school dropped from 94% in the 2019–2020 school year to 86% in 2020–2021. The decline was even more pronounced in low-income regions. For instance, enrollment in the Bangsamoro Autonomous Region in Muslim Mindanao fell by 41%. Students and schools, particularly in geographically isolated and disadvantaged areas, lack the necessary tools for distance or blended learning. Consequently, more students are at risk of dropping out and finding it harder to return, especially given the demands of the new normal (Asian Development Bank (ADB) 2022).

According to DepEd, last-mile schools are characterized by having fewer than four classrooms, often makeshift and nonstandard; lacking electricity and funds for repairs or new

4

construction; and being located more than an hour away from the nearest town, accessible only through difficult terrain. These schools may also have multi-grade classes, fewer than five teachers, and fewer than 100 students, with more than 75% possibly being indigenous peoples. The Asian Development Bank (ADB) and the Japan Fund for Prosperous and Resilient Asia and the Pacific, in collaboration with the Philippines' Department of Education (DepEd), are assisting the Philippine government in ensuring that all students have equal access to quality education during and after the pandemic. They are implementing low-cost EdTech solutions in last-mile schools, targeting junior and senior high school students (grades 7–12). This project provides secondary students in selected schools nationwide with access to quality education through excellent digital content, resources, and enhanced teacher skills based on DepEd's most essential learning competencies (MELC). These digitized learning materials will cover at least 50% of the MELCs, which are challenging for junior and senior high school students, including English language. In consultation with DepEd, the project selected 21 last-mile schools in hard-to-reach communities across three regions to pilot technology-based distance education using low-cost tablets and local area network (LAN) connections powered by solar energy. The initiative will distribute 3,500 tablets to secondary students and provide 21 LAN sets to schools to help students continue learning amid the pandemic. The project will also support the capacitation of DepEd's information technology (IT) officers to install and maintain the networks, while also training teachers and administrators in using technology for seamless learning. This training package includes assistance in content development, personalized learning, student learning assessment, and transitioning to classroom sessions. Additionally, the project will develop knowledge products such as digital training tools for educators, parents/guardians, and students; guides; social marketing tools to support distance education; and lessons learned from the project. The success of this pilot program could be extended to other remote areas in the country. (Asian Development Bank (ADB) 2022).

When classes resumed across most areas of the Philippines in 2022 after a hiatus of over two years, three teaching modalities were still in use: modular distance learning, online distance learning, and TV or radio-based instruction. Many parents and students opted for online and modular learning. However, issues persisted with modular distance learning, as it necessitated the printing and distribution of modules, which heightened concerns among parents amidst the ongoing pandemic (Lu 2023). Additionally, the concept of flexible learning emerged as an option for online education, particularly in higher institutions in the Philippines. Flexible learning allows students to choose the pace, place, and mode of their learning, which can be facilitated through appropriate pedagogical practices. This approach provides a solution for students with limited access to internet connections and modern gadgets needed for distance learning (Gordon 2014). Learners are given the option to decide

5

how they will continue their studies, including where and when they can proceed and how they can meet requirements and demonstrate learning outcomes. Flexible learning and teaching encompass various approaches to address the diverse needs of learners. These approaches include independence in the time and location of learning, offering choices in the curriculum (such as content, learning strategies, and assessment), and utilizing contemporary information and communication technologies to support a range of learning strategies (Alexander 2010).

2.3. Learning Environment and Organization of Schools

Although the Philippines received some support during and after the COVID-19 pandemic (Asian Development Bank (ADB) 2022), the shift to online learning widened the gap between those with internet connectivity and those without. Continuous academic engagement has been challenging for both teachers and students due to issues with access and internet connectivity. Boquia's (2023) study reveals that one of the significant themes in students' negative experiences with flexible learning is limited access to technology. Many Filipino students lack the necessary gadgets for online classes, as some cannot afford smartphones, laptops, or other essential devices. Another major issue is the unreliable internet connection. Students reported that they struggled to understand their teachers' discussions to English courses due to poor connectivity, with some students getting disconnected in the middle of class. This disruption significantly impacts their learning process (Boquia 2023).

During the pandemic, the learning environment posed the greatest challenge for students, especially due to distractions at home and limitations in learning space and facilities (Barrot, Llenares et al. 2021). Online learning challenges during the pandemic differ somewhat from those experienced in a pre-pandemic online learning environment. One possible explanation is that mobility restrictions may have exacerbated these challenges since students could not access schools or other learning spaces outside their homes. Lockdowns restricted students' learning experiences such as internships and laboratory experiments, limited their interactions with peers and teachers, and caused depression, stress, and anxiety among students. Additionally, they depleted the financial resources of those from lower-income groups, all of which adversely impacted students' learning experiences. This aligns with previous findings on the negative effects of lockdowns on students' learning and the challenges posed by the home learning environment (Kapasia, Paul et al. 2020, Day, Chang et al. 2021). However, further studies are needed to validate the impact of mobility restrictions on students' online learning experiences. Another factor that may explain these findings is students' socioeconomic backgrounds. Studies by Adarkwah (2021) and Day et al. (2021) reveal that the pandemic exposed many inequities in educational systems within and across countries. In the context of a developing country, families from lower socioeconomic strata

6

have limited learning space at home, access to quality internet service, and online learning resources. This explains why the learning environment and resources recorded the highest level of challenges. The socioeconomic profile of the students also accounts for the frequent mention of financial problems. These students often linked the lack of financial resources to their access to the internet, educational materials, and equipment necessary for online learning (Adarkwah 2021, Day, Chang et al. 2021).

Among all the different online learning challenges, students experienced the least difficulty with technological literacy and competency. This is not surprising given the extensive research confirming that new generations such as Generation Z, iGen, iGeneration, and Generation XD—terms describing the digital generation now entering adulthood—have never known a world without the internet. This generation includes those born after 1995 (the year the internet was commercialized) or 2000. They are the first generation to have permanent access to the internet and possess high levels of technological and digital literacy (Ng 2012, Barrot, Llenares et al. 2021).

A key component in transitioning to a flexible learning modality is integrating flexibility into the essential dimensions of teaching and learning. One major consideration is incorporating flexibility into the curriculum. Adjustment on the types of assessment measures is a major factor amid the pandemic. There is a need to limit requirements and focus on the major essential projects that measure the enduring learning outcomes like case scenarios, problem-based activities, and capstone projects. Authentic assessments have to be intensified to ensure that competencies are acquired by the learners. In the process of modifying the curriculum amid the pandemic, it must be remembered that initiatives and evaluation tasks must be anchored on what the learners need including their safety and well-being (Dayagbil, Palompon et al. 2021). Curriculum recalibration involves not only the content to be learned and taught but also the methods of learning, teaching, and assessment in response to the challenges posed by the pandemic. A flexible curriculum design should be learner-centered and consider the demographic profile and circumstances of learners, including access to technology, technological literacies, different learning styles and capabilities, and diverse knowledge backgrounds and experiences. It should also ensure varied and flexible forms of assessment (Ryan and Tilbury 2013). The quantitative online survey study conducted by Dayagbil (2021) reveals that students expressed concerns about their learning experiences. Most respondents reported that teachers made adjustments to course outcomes and syllabi. However, many students felt that the learning activities were not flexible enough to be completed either offline or online (Dayagbil, Palompon et al. 2021). This result warrants further investigation into the impact of curriculum adjustments that incorporate flexible

learning experiences, as the study was conducted during the pandemic within a single school population in the country.

3. Discussion

To our knowledge, this study is the first to assess the key indicators of web-based learning effectiveness in teaching English in the Philippines during the post-COVID-19 pandemic, applying main indicators for learning effectiveness from the OECD. We focused on the following indicators: (a) educational and learning impact; (b) access to education, participation, and progression; and (c) learning environment and organization of schools. Details about these indicators are summarized in Table 1.

In this country case study, we found that educational output and learning impact commonly benefit from greater school time flexibility, more student-centered learning, and faster, more convenient group-task communications. However, distance learning was found to be less effective for group tasks and, due to reduced face-to-face interaction, may not promote a better learning experience for practicing English conversation skills among students. There was also a positive impact of post-pandemic initiatives on student access to education, participation, and progression. The Philippine government and other institutions provided low-cost EdTech solutions to last-mile schools, promoting digitalized learning materials and initial training for IT officers, school administrators, and teachers. Additionally, three main teaching modalities were used: modular distance learning, online distance learning, and TV or radio-based instruction. Flexible learning emerged as a solution to equity issues in internet access and the distribution of modern gadgets for online learning. However, modular distance learning posed challenges as it required printing and distributing modules, raising concerns among parents amidst the ongoing pandemic. On the other hand, school learning environments and organizations began redesigning school curricula to enhance students' access to learning experiences, even during pandemic times. However, many students felt that the learning activities were not flexible enough to be completed either offline or online under the new curriculum. Additionally, the home learning environment and limited access to internet connections posed the greatest challenges to web-based learning, while technological literacy and competency posed the least challenge among the new generation of youths (Gen Z).

This study has several limitations. As it is a case study by design, more prospective or even longitudinal studies are needed to clearly understand the long-term effects of web-based learning during the post-COVID-19 pandemic on the educational system in the country. We also recommend identifying factors associated with or predictive of the common negative impacts of web-based learning during the post-pandemic era, as highlighted in this study's literature review. Additionally, since the pandemic situation varies widely between developed

8

and developing countries, more studies are needed to assess the impact of web-based learning during the COVID-19 pandemic in diverse country contexts and its global implications.

4. Conclusion

The pandemic crisis has unequivocally necessitated the adoption of web-based learning, particularly for English courses, in the Philippines. This shift has impacted significant indicators of learning effectiveness, including educational impact, access to education, participation, progression, and the organizational aspects of learning environments in schools. Given the likelihood of ongoing online implementation of the English language curriculum in the Philippines, it remains imperative to continue investigating the discourses, practices, and ideologies shaped by and influencing individuals within this digital educational landscape.

9

Table 1. Summary table for the impact of the web-based learning in teaching English language among Filipino students in the post-COVID-19 Era.

Key Indicators	Positive Impact	Negative Impact
Educational Output and Learning Impact	• Greater Flexibility • Student-centered Learning • Faster and convenient group-task communication.	• Might not be best option for group-task. • Less face-to-face interaction might not promote a greater learning experience for practicing English conversation skills.
Access to Education, Participation and Progression	• Low-cost EdTech solutions in last-mile schools (e.g., digitalized learning materials, initial training among IT officers, school administrators, and teachers). • Three main teaching modalities used: modular distance learning, online distance learning, and TV or radio-based instruction. • Emerging flexible learning to address equity issues in the distribution of internet connections and modern gadgets for online learning.	• Modular distance learning necessitated the printing and distribution of modules, which raised concerns among parents amid the ongoing pandemic.
Learning Environment and Organization of Schools	• Redesigning the school curriculum to provide students with greater access to learning experiences even during pandemic times. • Gen Z possesses high levels of technological and digital literacy.	• Limited access to internet connectivity. • Many students felt that the learning activities were not flexible enough to be completed either offline or online under the new curriculum. • The learning environment posed the greatest challenge for students at home during the pandemic.

10

References

Adarkwah, M. A. (2021). ""I'm not against online teaching, but what about us?": ICT in Ghana post Covid-19." Education and information technologies **26**(2): 1665-1685.

Alawamleh, M., L. M. Al-Twait and G. R. Al-Saht (2020). "The effect of online learning on communication between instructors and students during Covid-19 pandemic." Asian Education and Development Studies **11**(2): 380-400.

Alexander, S. (2010). "Flexible learning in higher education."

Asian Development Bank (ADB). (2022). "Getting Education Back on Track." Retrieved July 01, 2024, from https://www.adb.org/multimedia/partnership-report2021/stories/getting-education-back-on-track/.

Barrot, J. S., I. I. Llenares and L. S. del Rosario (2021). "Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines." Education and Information Technologies **26**(6): 7321-7338.

Boquia, A. H. (2023). "Using Flexible Learning Design in Teaching English: Narratives of First-Year College Students of Southern Philippines Amidst Pandemic." World **13**(8).

Center for Disease Control and Prevention (CDC). "Transmission of SARS-CoV-2 in K-12 schools " Retrieved June 15, 20204, from https://www.cdc.gov/coronavirus/2019-ncov/science/sciencebriefs/transmission_k_12_schools.html.

Cortes, S. T. (2020). Flexible learning as an instructional modality in environmental science course during COVID-19. Aquademia, 4 (2), ep20024.

Crawford, J., K. Butler-Henderson, J. Rudolph, B. Malkawi, M. Glowatz, R. Burton, P. A. Magni and S. Lam (2020). "COVID-19: 20 countries' higher education intra-period digital pedagogy responses." Journal of applied learning & teaching **3**(1): 1-20.

Day, T., I. C. C. Chang, C. K. L. Chung, W. E. Doolittle, J. Housel and P. N. McDaniel (2021). "The immediate impact of COVID-19 on postsecondary teaching and learning." The Professional Geographer **73**(1): 1-13.

Dayagbil, F. T., D. R. Palompon, L. L. Garcia and M. M. J. Olvido (2021). Teaching and learning continuity amid and beyond the pandemic, Frontiers Media SA.

Economist, T. (2020). "The Philippines' fierce lockdown drags on, despite uncertain benefits." Retrieved June 11, 2021, from https://www.economist.com.

Gordon, N. (2014). "Flexible pedagogies: Technology-enhanced learning." The Higher Education Academy **1**(2): 2-14.

Ignacio, A. (2021). "Online Classes and Learning in the Philippines During the Covid-19 Pandemic." International Journal on Integrated Education **4**(3): 1-6.

Kapasia, N., P. Paul, A. Roy, J. Saha, A. Zaveri, R. Mallick, B. Barman, P. Das and P. Chouhan (2020). "Impact of lockdown on learning status of undergraduate and postgraduate students during COVID-19 pandemic in West Bengal, India." Children and youth services review **116**: 105194.

Lee, C. M. (2024). "Online Learning versus Face to Face Learning toward Students: Which can be an effective way of Learning Methodology to our current Educational System?".

Lu, B. J. (2023). "Addressing effects of Covid-19 among Filipino learners." Retrieved July 01, 2023, from https://www.pna.gov.ph/index.php/opinion/pieces/704-addressing-effects-of-covid-19-among-filipino-learners.

Ng, W. (2012). "Can we teach digital natives digital literacy?" Computers & education **59**(3): 1065-1078.

Pace, C., S. K. Pettit and K. S. Barker (2020). "Best practices in middle level quaranteaching: Strategies, tips and resources amidst COVID-19." Becoming: Journal of the Georgia Association for Middle Level Education **31**(1): 2-13.

Payusan, A., M. J. Arrofo, L. A. Codilla, S. Gaco, D. K. Dakay and V. Cortes (2022). "Priorities and Intentions of Working-Students Working amidst COVID-19 Pandemic." The Normal Lights **16**(2).

Ryan, A. and D. Tilbury (2013). "Flexible pedagogies: New pedagogical ideas." Higher Education Academy.

The Organization for Economic Cooperation and Development (OECD). "Key indicators on education." Retrieved June 11, 2024, from https://www.oecd.org/education/school/keyindicatorsoneducation.htm.

UNESCO. (2022). "COVID-19 impact on education as of Jan 20, 2022." Retrieved July 01, 2024, from htpps://en.unesco.org/covid19/educationresponse.

United Nations Educational Scientific and Cultural Organization (UNESCO). (2021). "One year into COVID-19 education disruption: Where do we stand?" Retrieved June 11, 2024, from https://www.unesco.org/en/articles/one-year-covid-19-education-disruption-where-do-we-stand.

12

Investigating Learners' Academic Emotion and Cognitive Load during Chatbot-Assisted Academic Reading

Xiaotian Guan, Yuanke Li*

School of Foreign Studies, South China Normal University, Guangzhou, China

Introduction

Academic reading, as a form of "targeted and critical reading of lengthy academic texts for the purpose of researching specific disciplines" (Sengupta, 2002) often arouses negative emotions and places a high level of cognitive load (CL) on English as a Foreign Language (EFL) university students. To start with, English academic reading requires comprehension of sophisticated domain-specific vocabulary and complex grammatical constructions (Millin et al., 2020; Snow, 2010), causing heavy CL on EFL university students. Moreover, with difficulty processing and comprehending academic texts, EFL students also reported negative emotions such as fear, anxiety, and stress towards reading in English (Eriksson, 2023). To address this problem, collaborative reading has been recommended and reported to alleviate negative emotions and lower the level of cognitive load (Jin et al., 2022; Lin et al., 2024). Additionally, from the perspective of Cognitive load theory, learners engaging in collaborative reading can share collective working memory (Fransen et al., 2011) created through communication and coordination, lowering the CL distributed to each individual (Kirschner et al., 2018). However, some other studies unraveled the challenges of collaborative academic reading among peer students. Concerning learner emotion, peer collaboration may arouse affective conflicts between members, generating feelings of frustration and incompetency and hindering group performance and member satisfaction (Jehn et al., 2008). Additionally, learners' performance may be hampered by an aroused fear of making mistakes and receiving peer judgments during collaboration with peers (L. Fryer & Carpenter, 2006). Regarding CL, learners participating in collaboration may undergo additional cognitive pressure due to the lack of coordination (Janssen et al., 2010; Kirschner et al., 2018). In this backdrop, AI-powered chatbots (e.g., ChatGPT), equipped with advanced ability to comprehend texts and interact with human users, may serve as an intelligent collaborator to assist students in academic reading. However, there is a dearth of research on advanced learners' use of chatbots in

1

academic reading. Besides emotion, the challenges of imposed CL on learners by the integration of chatbots should not be overlooked, but the level of CL imposed by chatbot assistance in academic reading is still understudied. Therefore, the present study examined the AE and CL of university students during chatbot-assisted academic reading.

Literature review

Use of Chatbot in education

Chatbot-assisted learning has received significant attention from educational experts. In the writing setting, Guo et al. (2023) argued that the chatbot is no longer merely a tool used in the activity but more a collaborative partner of the learner within a writing community. Similarly, Liu (2024) focuses on the chatbot's implementation in digital multimodal composing (DMC), and students in the chatbot group were found to outperform those in the traditional writing group in generating detailed content and critical thinking. Besides, in the speaking contexts, Guo et al. (2023) investigated students' engagement in debate activities with the support of a text-based chatbot, Argumate. As the data suggests, Argumate had a positive impact on students' cognitive, behavioral, and affective engagement, along with students' high learning performance and perceived usefulness of the chatbot.

Regarding chatbot assistance in reading, pre-programmed chatbots had been applied to company learners while reading even before the recent advancement with Gen-AI. Michaelis & Mutlu (2017) developed a learning companion robot for children, with the findings suggesting the robot's positive effects on children's reading engagement and interest. To further investigate the effectiveness of chatbot companion, Xu et al. (2021) compared the children's story comprehension and their verbal interaction between children-chatbot and children-adult collaborative reading, and equal outcomes in supporting reading comprehension of the two approaches were reported. On the other hand, after the integration of AI, the chatbot can perform better in responding to learners' real-time questions, providing much more support beyond the pre-designed programs. Among the few pieces of literature was Liu et al.'s (2022) study, in which they designed an AI-powered chatbot to communicate with learners while reading and reported that students in the AI chatbot group showed a high perceived social connection with the chatbot and maintained a steady level

2

of interest compared to the counterparts who reported a significant drop in interest level. Nonetheless, existing studies have primarily focused on children as the subject, leaving a dearth of research on the use of reading-assisting chatbots by advanced learners. Hence, this study sheds light on university students' experience of chatbot-assisted academic reading.

Academic emotion, cognitive load, their associations, and influential factors

Academic emotion refers to the emotions that are directly related to the learning process, classroom activities, and learning achievements (Pekrun et al., 2002). According to Pekrun's (2006) Achievement Emotion Theory, AE can be classified into four types, namely positive activating emotion (PAE), positive deactivating emotion (PDE), negative activating emotion (NAE), and negative deactivating emotion (NDE). Specifically, PAE contains enjoyment, hope, and pride; PDE consists of relaxation, assurance, and relief; NAE includes anger, anxiety, and shame; NDE encompasses boredom, hopelessness, and disappointment.

First proposed by John Sweller (1988), the Cognitive Load Theory (CLT) suggests that the human memory system can be segmented into working memory and long-term memory, and CL refers to the load imposed on the working memory system when solving a specific task (Paas et al., 2003; Sweller, 1988, 2011). According to Sweller's (Sweller, 1988) classification framework, CL can be divided into intrinsic cognitive load (ICL), extraneous cognitive load (ECL), and germane cognitive load (GCL). The ICL is closely related to information capacity and interactivity, as well as the learning level of learners. The ECL is mainly related to the presentation of information, teaching processes, and instructional design, while the GCL is the CL required for learning, specifically, the working memory resources used to handle the ICL. However, it was argued that GCL is redundant due to its theoretical overlap with intrinsic load (Kalyuga, 2011). Responding to the discussion, Sweller et al. (2019) revised their conceptualization of CLT and proposed a dyadic model consisting only of ICL and ECL, excluding GCL from the taxonomy.

From the perspective of the Cognitive-Affective Theory of Learning with Media (CATLM) (Moreno & Mayer, 2007), learners' affective and cognitive states are interdependent and correlated. Following this assumption, numerous studies have been conducted to probe the complex correlations between learners' cognition and emotion. For instance, in the VR-enhanced learning setting, affective motivation was identified to have a direct and negative effect on CL

3

(Huang, et al., 2022). In the context of digital game-based learning, the beneficial effects of positive emotions and GCL on cognitive performance with the detrimental effects of negative emotions, ICL, and ECL on cognitive processing were highlighted (Chang & Yang, 2023). Similarly, in the setting of e-textbook usage, learner emotion was found to be negatively intertwined with CL (Chang & Chen, 2022). The relationship between AE and CL has been examined in different technologically supported learning settings, but their complex interplay within the emerging context of chatbot-assisted academic reading is still under-studied. In this light, the current study attempts to explore the nuanced relationship between AE and CL experienced by university students while engaging in chatbot-assisted academic reading.

Besides correlations, previous studies have probed into the factors influencing AE and CL in various technology-enhanced learning contexts. Concerning factors of AE, as suggested by Pekrun's Control-Value Theory (CVT) (2006), learners' control perception shows a positive relationship with academic emotion (e.g., King & Gaerlan, 2014; Putwain et al., 2013). Besides, learners' prior knowledge of the specific subject matters or the technology used may impact their appraisals of the task and further influence their emotional response (Richter et al., 2018). Besides, the communities play a vital role in mediating AE through cognitive support (Chang & Yang, 2023), and motivation inspiration by sharing their value of learning (Frenzel et al., 2007). In addition, the aesthetic design of the learning tool may influence learners' perception of the learning material or task, thus mediating academic emotion (Loderer et al., 2020).

Similar to AE, influential factors of CL can also be categorized into learner-related factors, community-related factors, and tool-related factors. Firstly, the types of self-regulated learning (SRL) task is one of the significant individual factors, with different levels of CL demonstrated in different SRL activities (Wang et al., 2023). Besides individuals, a community consisting of teachers, peer students, and AI was reported to influence learners' CL. For instance, teacher presence in the learning video can increase ICL and improve learning outcomes (Yu, 2022); learning in social media within a group of peers contributes to higher collective working memory and the reduction of CL (Li et al., 2021); AI provides guidance of rules and access to tips and examples which significantly release learners' CL (Liu, 2024).

In summary, based on the existing studies and with the goal of filling the current research gap, the present study aims to investigate AE and CL experienced by university students in

4

chatbot-assisted academic reading.

The research questions guiding this study are as follows:

（1） What levels of AE and CL do university students experience when engaging in chatbot-assisted academic reading?

（2） What are the correlations between university students' AE and CL when engaging in chatbot-assisted academic reading?

（3） What factors affect learners' AE and CL when engaging in chatbot-assisted academic reading

Methods

Participants

Based on the principles of purposive sampling and convenience sampling, 100 questionnaires were distributed for this study in March 2024. Participants who participated in the questionnaire survey were 93 undergraduate students in various majors such as English Education and Business English. et al. Descriptions of the demographic information of participants are provided in Table 1.

Table 1 Demographic information of interviewers

Name	Major	Year of study	Duration of PDFgear assisted reading experience
Liang	English education	Year 3	2 weeks
Huang	Business English	Year 3	2 weeks
Wang	English education	Year 3	2 weeks
Rui	English education	Year 4	1 months
Chen	English education	Year 4	1 months

Experimental procedures

Before experiencing chatbot-assisted academic reading, participants received guidance and training on useful strategies for chatbot-assisted academic reading, and practiced for a week so that they were familiar with using the chatbot to assist their academic reading. Afterward, participants engaged in the chatbot-assisted academic reading experience and completed

5

self-evaluation after a specific period using the achievement emotion questionnaire and cognitive load questionnaire. Besides, 5 participants each received a semi-structured interview lasting approximately 40 minutes after their reading experience.

Instruments

Reading materials

Instead of creating experimental reading tasks for participants, this study examined learners' genuine reading experience for real demands such as lesson presentations and graduate thesis writing. Therefore, no restrictions on reading materials are given, and participants can choose the academic articles that suit their real learning demands.

The chatbot (PDFgrear)

The chatbot used in the current study is ChatGPT-3.5-Turbo integrated into PDFgear (https://www.pdfgear.com/), a PDF reader. Users can communicate with the chatbot while reading the PDF-formed academic article, and the chatbot's responses are all based on the content of the PDF document being read with the application.

Academic emotion questionnaire

The modified achievement emotion questionnaire in this study was adapted from Pekrun et al.'s (2011) Likert-scale questionnaire. For a shorter completion time of the questionnaire and increased validity, this study retained the original four-dimensional structure but included only one subtype of emotion within each dimension. The questionnaire used contains four items on hope (PAE), two on relief (PDE), four on anxiety (NAE), and four on disappointment (NDE). The selection of the remaining emotion type is determined by a pre-questionnaire (N=25) in which participants were required to choose the two most frequently experienced types of emotion in each dimension. Moreover, minor modifications were made to the items to align with the context of chatbot-assisted academic reading. The Cronbach's α value of the adapted questionnaire was 0.828, revealing that internal consistency among items was adequate.

Cognitive load questionnaire

The current study adapted the Likert-scale questionnaire developed by Leppink et al. (2013). The adapted questionnaire comprised two dimensions: ICL (4 items) and ECL (4 items), with

GCL-related items removed based on the newly proposed two-dimensional framework (e.g., Kalyuga, 2011; Leppink & van den Heuvel, 2015). Minor adjustments were made to the items to align with the context of chatbot-assisted academic reading. The Cronbach's α value of the adapted questionnaire was 0. 856, indicating adequate internal consistency among items.

Semi-structured interview

The semi-structured interview in this study mainly focuses on the influential factors of learners' AE and CL during the chatbot-assisted academic reading process. The interview questions were designed based on the AT framework, covering the six activity elements, namely subject, tools, goals, community, division of labor, and rules. Besides the theoretical underpinning, the interview outline was developed with reference to the interview questions used in Chen et al.'s (2022), Guo et al.'s (2023), and Kwong and Churchill's (2023) studies.

Data analysis

SPSS was used for quantitative data analysis. First, descriptive analyses were used to examine the levels of AE and CL experienced by university students during the chatbot-assisted academic reading. Second, Pearson's correlational analysis was employed to unravel the interrelation between AE and CL during chatbot-assisted academic reading.

For qualitative analysis, the interviews were audio-recorded and transcribed into text form. Since the interviews were conducted in Chinese, the transcribed data was then translated into English. To understand the process and unravel the factors affecting AE and CL during chatbot-assisted academic reading, a thematic analysis approach was adopted. The researcher first got familiarized with the interview transcription corpus and then generated vivo codes, that is directing using participants' words as codes (Braun & Clarke, 2006; Ravitch & Carl, 2016, pp. 249). After that, the researcher carefully checked and refined the identified themes (Braun & Clarke 2006) by rereading and re-examining the data to gain a deeper understanding of participants' interpretations.

Results

Levels of AE and CL

As Table 1 shows, the mean values of PAE (M=4.020) and PDE (M=4.025) students

generated fell within the range of 3.41-4.20, while the levels of NAE (M=2.388) and NDE (M=2.643) conceived by learners were comparatively low, with average numbers positioning between 1.81 and 2.60. This revealed the generally positive emotional states of learners during chatbot-assisted academic reading, which could be attributed to the support and companionship provided by the chatbot during the reading process, potentially enhancing learners' engagement, confidence, and reducing feelings of fear, isolation, and frustration.

In terms of CL, learners reported experiencing a moderate level of ICL (M=3.328) and a low level of ECL (M=2.438). This indicated that the chatbot responses with great accessibility and clarity did not occupy many cognitive resources of learners, and learners experienced a relatively higher level of ICL resulting from cognitive processing of intricate academic articles than ECL generated from chatbot interactions. This highlighted the positive emotional impact and effective CL management facilitated by chatbot integration in academic reading activities, underscoring the usefulness of employing chatbots to help EFL learners achieve better emotional and cognitive outcomes when engaging in academic reading.

Table 1 Levels of AE and CL experienced by students in the experiment

Items	N	Min	Max	Mean	Std. Deviation	Median
PAE	100	2.500	5.000	4.020	0.609	4.000
PDE	100	2.000	5.000	4.025	0.629	4.000
NAE	100	1.000	5.000	2.388	0.758	2.250
NDE	100	1.000	5.000	2.643	0.838	2.500
ICL	100	1.500	5.000	3.328	0.717	3.250
ECL	100	1.00	5.00	2.438	0.713	2.500

8

Table 1 Levels of AE and CL experienced by students in the experiment

Items	N	Min	Max	Mean	Std. Deviation	Median
	0	0	0			

Note. **PAE (positive activating emotion), PDE (positive deactivating emotion), NAE (negative activating emotion), NDE (negative deactivating emotion), ICL (intrinsic cognitive load), ECL (extraneous cognitive load)**

The correlations between AE and CL

To get insight into the correlations between various dimensions of AE and CL, a Pearson's correlational analysis was conducted.

As demonstrated in Table 2, there is a significant weak positive correlation between ICL and PAE, NAE, and NDE ($r_{PAE-ICL}=0.245$, $p=0.014<0.05$; $r_{NAE-ICL}=0.216$, $p=0.031<0.05$; $r_{NDE-ICL}=0.230$, $p=0.031<0.05$). This indicates that under the assistance of a chatbot, increasing difficulty in reading tasks not only generates negative emotions like anxiety, hopelessness, and disappointment, but also triggers learners' emotions such as enjoyment, hope, and pride, as they can accomplish challenging work that may be impossible for them to handle individually.

In addition, ECL shows a significant weak negative correlation with PDE ($r_{PDE-ECL}=-0.354$, $p=0.000<0.01$) and significant moderate positive correlations with both NAE and NDE ($r_{NAE-ECL}=0.557$, $p=0.000<0.01$; $r_{NDE-ECL}=0.539$, $p=0.000<0.01$). This shows that a better emotional experience can be fostered by diminishing the CL produced from chatbot interaction by improving the accuracy, clarity, and user-friendliness of the chatbot.

However, there are no significant correlations between ICL and PDE ($r_{PDE-ICL}=0.191$, $p=0.056>0.05$) as well as ECL and PAE ($r_{PAE-ECL}=-0.127$, $p=0.210>0.05$), suggesting that these two pairs of AE and CL may not be directly related, but potential relationships with mediations may be under-discovered.

Table 2 The correlations between academic emotion and cognitive load

AE		ICL	ECL
PAE	Coefficient	0.245*	-0.127

9

Table 2 The correlations between academic emotion and cognitive load

AE		ICL	ECL
PDE	p value	0.014	0.210
	Coefficient	0.191	-0.354**
	p value	0.056	0.000
NAE	Coefficient	0.216*	0.557**
	p value	0.031	0.000
NDE	Coefficient	0.230*	0.539**
	p value	0.021	0.000

* $p<0.05$ ** $p<0.01$

Note. **PAE (positive activating emotion), PDE (positive deactivating emotion), NAE (negative activating emotion), NDE (negative deactivating emotion), ICL (intrinsic cognitive load), ECL (extraneous cognitive load)**

Influential factors of AE and CL

To unravel the factors that affect learners' AE and CL during chatbot-assisted academic reading, this study conducted semi-structured interviews with five students engaged in chatbot-assisted academic reading, and a thematic analysis was performed on the interview data based on the framework of AT. Through thorough analysis, factors in relation to subject, community, division of labor, tool, and rule are found.

Learners' subjective factors

Through thematic analysis of the interview data, two significant subject-related factors are discovered: (1) learners' expectations of the chatbot support and (2) learners' prior knowledge.

To start with, there is a complex and dynamic interplay between learners' expectations of chatbot assistance and their emotional experience. As the data suggested, students' inflated expectation of the effectiveness and capabilities of chatbot support may result in their negative emotional responses when encountering unsatisfactory chatbot replies. Additionally, reading with no preset aim also leads to negative emotions such as boredom and hopelessness. On the contrary, an objective and critical perspective on what chatbot can and cannot perform contributes to

10

positive emotions.

Besides subjective expectations, learners' prior knowledge of academic reading was found to impact their AE and CL in chatbot-assisted academic reading. As the data suggest, participant Liang's prior understanding of the structure of academic articles empowers her to approach chatbot assistance in a more informed and strategic manner, enabling her to leverage the chatbot assistance more effectively and contributing to an improved emotional response possibly due to a sense of control and competence. Furthermore, Liang's reduced CL can be attributed to his active engagement with the material with a balanced reliance on both his own knowledge and the chatbot support, where the chatbot serves as a supplementary resource rather than the primary source of information.

Learner-chatbot interaction

In the chatbot-assisted academic reading process, the chatbot and learners can be seen as partners within a learning community with shared learning aims. During the learner-chatbot interaction process, both sides contribute efforts to achieve the reading goal, maintain a positive emotional state, and minimize the CL. Chatbot offers multiple powerful functions that were reported to enhance learners' positive emotions and reduce CL, including (1) real-time query assistance, (2) selective reading guidance, and (3) provision of reference. Meanwhile, human learners also make efforts to regulate the chatbot for more successful learner-chatbot communications and better comprehension of the article, such as (1) regulation of the reading process, and (2) individual proofreading.

Throughout the whole reading process, the chatbot is always available for queries and responses. First, with the awareness of the ability to enquire chatbot any time they want, learners may feel like being companied by a more knowledgeable person who is there to provide scaffolding, thus reducing learners' feelings of loneliness, stress, and fear generated by the solitary and challenging nature of academic reading. Second, chatbot's clear and detailed responses can effectively resolve learners' confusion and enhance their comprehension of the article, boosting learners' sense of achievement and self-efficacy. Concerning learners' cognitive process, the chatbot is reported to reduce learners' CL by completing some basic cognitive tasks and providing well-structured explanations that offer learners a sophisticated scheme for cognitive processing.

Apart from simple Q&A, learners also view chatbot as a guide for selective intensive reading. Some learners believe that human is the center of the learning process, determining the reading objectives and controlling the pace, and they also recognize that individual reading is irreplaceable even under the assistance of chatbot. In the selective reading process, chatbot offers powerful selection and reading recommendations based on the learners' needs, which is unraveled to release learners' fear of massive academic reading, cultivate confidence and hope for successful reading, and also save learners' cognitive resources from reading irrelevant content.

When the chatbot's textual explanation fails to satisfy the learner, the chatbot's provision of reference and learners' individual proofreading are revealed to be crucial for learners' AE and CL in chatbot-assisted academic reading. It is reported that the chatbot would automatically list the page number of the reference source at the end of its response. As participant Chen's experience demonstrates, the provision of reference offers learners a chance to seek the answer independently. These individual efforts of "verifying or cross-referencing" and "skimming" (Chen) help learners regain a sense of control over the reading process, ease their anxiety, and generate a sense of accomplishment as well as self-efficacy. Regarding cognitive experience, the effect of chatbot's reference provision shows an unstable and dual character. On one hand, when the referencing function works, learners can selectively read the pages offered according to their reading aims without having to read the whole article, which saves cognitive processing on unrelated contents, reducing their CL. On the other hand, if it malfunctions, it may cause unnecessary analysis and comprehension process, thus adding to the level of CL.

Disciplinary community

Through the process of chatbot-assisted academic reading, a disciplinary community was discovered alongside the learner-chatbot community. In this community, exports from both real-world and online contexts were found to have a positive impact on learners' emotional states and cognitive processes, such as inspiring hope and sharing learning experience.

For participant Wang, his intramural teacher's research interest in chatbot-enhanced teaching and learning raised the perceived value and affordances of chatbot technology for him, further fueling his passion and optimism about leveraging chatbot to assist his academic reading. Towards participant Huang, the professionals or more knowledgeable peers online stimulated her positive

outlook on future progress in academic reading and writing through experience-sharing videos on online platforms such as YouTube and Bilibili. Through sharing summarized approaches for reading and writing academic articles, these online experts also provided Huang with more effective and efficient cognitive strategies for processing academic articles she was reading, thereby decreasing her CL throughout the process.

Interface design of the chatbot

In terms of interface design of the chatbot, its consistency with learners' former learning tools demonstrated positive impact on learners' AE and CL through chatbot-assisted academic reading. The similar characteristics of tools were discovered to trigger learners' sense of familiarity, joy, and relief, etc., while decreasing learners' CL due to the avoidance of learners' familiarizing and adapting process with the chatbot.

Participant Chen believed that there is little difference on interface, operational approach, etc. between PDFgear and her previous PDF readers while participant Wang regards the interface of PDFgear as similar to his own paper-based note-taking modal. These similarities brought them a sense of familiarity with the tool, alleviated their pressure about using a new tool, and made her feel relieved and relaxed just like "working with an old friend" (Wang). Additionally, these similarities saved their cognitive resources from understanding the user interface, memorizing the operational approaches, and adapting their own reading procedure or habits to fit the tool, thus diminishing the CL imposed.

The prompting rules

Similar to collective reading among human peers, there are also rules that mediate the human-chatbot collaborative academic reading. One significant rule found to affect learners' AE and CL was the rule for prompting. Prompting in human-chatbot interactive context refers to the user's initiating a request, command, or input to guide the chatbot's behavior or response, including asking questions, giving instructions, or providing information, etc. As the interview demonstrated, learners who have not acquired this rule may experience conversational breakdowns (McTear et al., 2016) and consequently get into negative emotional states and suffer a high CL. Conversely, learners with an awareness of the rule may experience more positive emotions and lowered CL by utilizing the rule to avoid breakdowns and promote the quality of the

chatbot responses.

Discussion

The effect of chatbot support on improving learners' AE and CL

The descriptive analysis of the quantitative data demonstrated that students engaged in chatbot-supported academic reading mostly experienced strong positive emotions with minimal negative emotions. This highlights the chatbot's power to cultivate positive emotion and reduce negative emotion, which is consistent with previous studies in settings of story reading (Michaelis & Mutlu, 2017), academic writing (Nazari et al., 2021), and speaking activities (Han, 2020). Concerning CL, the results also showed that learners underwent a moderate level of ICL along with a low level of ECL, which is in accordance with existing literature probing learners' CL under chatbot assistance in writing tasks (Liu et al., 2023) and gamified lessons (Chen & Chang, 2024), etc.

The association between learners' AE and CL

In terms of Pearson's correlational analysis, the correlations identified in the current study (i.e., positive correlations of ICL with NAE and NDE, positive correlations of ECL with NAE and NDE, and negative correlations of ECL with PDE) are mainly in consistence with previous studies in other technology-enhanced learning settings (e.g., Chang & Chen, 2022; Chang & Yang, 2023). However, it is quite interesting to find that, in the chatbot-assisted academic reading setting, PAE also exhibited a rising trend with the increase of ICL, which contradicts previous findings indicating a negative relationship between positive emotion and CL (e.g., Chang & Chen, 2022). This contradicting situation may be a result of the integration of chatbots which offer in-time and effective support to elevate senses of confidence, self-efficacy, and hope by assisting learners to accomplish the tasks they perceived as difficult to carry out.

The influential factors of learners' AE and CL

Following the AT perspective, factors in relation to subject, community, division of labor, tool, and rule are found to affect learners' AE and CL through thematic analysis.

First, the finding of the dynamic impact of subjective expectation enriches the idea of CVT (Pekrun, 2006) that the relationship between value attribution and positive emotion is U-shaped

14

rather than linear. Second, during the process of learner-chatbot interaction, learners and the chatbot both contributed efforts to reach the reading goal. The finding partly align with Guo et al.'s (2023) activity system of chatbot-assisted essay writing while providing more insights into learners' individual regulation and critical thinking as exemplified by the proofreading activity. Third, the positive influence of experts form the disciplinary community align with former studies stressing the cognitive support and motivation inspiration from the communities (e.g., Chang & Yang, 2023; Frenzel et al., 2007). Fourth, the emphasis on the consistency between the chatbot interface and learners' former tool broadens the scope toward the learning tools in technology-enhanced learning, which previously focused on the aesthetic design (e.g., Loderer et al., 2020) and other tools besides the targeted one (e.g., Guo et al., 2023). Finally, the rule of prompting was found to mediate learners' AE and CL. Prompts were provided in previous studies as scaffolding for learners (e.g., Ayedoun et al., 2020; L. K. Fryer et al., 2020; Tai & Chen, 2024), but the present study contributes to the field by scrutinizing the significant influence of learners' acquisition of particular prompting rules on their emotional and cognitive experience.

Conclusion

This study explored learners' AE and CL during chatbot-assisted academic reading. As the results indicated, learners mainly experienced robust positive emotion with minimal negative emotion while undergoing moderate ICL and low ECL. Additionally, complex correlations were identified between AE and CL in chatbot-assisted academic reading. Specifically, ICL correlated positively with PAE and NAE, while ECL correlated positively with NAE and NDE and negatively correlated with PDE. In addition, no significant correlations between ICL and PDE as well as ECL and PAE were identified. Furthermore, based on an AT framework, learners' subjective factors, learner-chatbot interaction, the disciplinary community, the interface design of the chatbot, and the prompting rules are revealed to influence learners' emotional and cognitive experience. These findings contribute to a broader view of chatbot-assisted learning and the emotional and cognitive aspects involved throughout the process.

Reference

Ayedoun, E., Hayashi, Y., & Seta, K. (2020). Toward Personalized Scaffolding and Fading of

Motivational Support in L2 Learner–Dialogue Agent Interactions: An Exploratory Study. *IEEE Transactions on Learning Technologies*, *13*(3), 604–616. https://doi.org/10.1109/TLT.2020.2989776

Chang, C.-C., & Chen, T.-C. (2022). Emotion, cognitive load and learning achievement of students using e-textbooks with/without emotional design and paper textbooks. *Interactive Learning Environments*, 1–19. https://doi.org/10.1080/10494820.2022.2096639

Chang, C.-C., & Yang, S.-T. (2023). Learners' positive and negative emotion, various cognitive processing, and cognitive effectiveness and efficiency in situated task-centered digital game-based learning with different scaffolds. *Interactive Learning Environments*, 1–20. https://doi.org/10.1080/10494820.2023.2209600

Chen, C.-H., & Chang, C.-L. (2024). Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12553-x

Chen, Z., Chen, W., Jia, J., & Le, H. (2022). Exploring AWE-supported writing process: An activity theory perspective. *Language Learning*.

Eriksson, L. (2023). Difficulties in academic reading for EFL students: An initial investigation. *Language Teaching*, *56*(1), 149–152. https://doi.org/10.1017/S0261444822000246

Fransen, J., Kirschner, P. A., & Erkens, G. (2011). Mediating team effectiveness in the context of collaborative learning: The importance of team and task awareness. *Computers in Human Behavior*, *27*(3), 1103–1113. https://doi.org/10.1016/j.chb.2010.05.017

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Perceived learning environment and students' emotional experiences: A multilevel analysis of mathematics classrooms. *Learning and Instruction*, *17*(5), 478–493. https://doi.org/10.1016/j.learninstruc.2007.09.001

Fryer, L., & Carpenter, R. (2006). Bots as Language Learning Tools. *Language Learning*.

Fryer, L. K., Thompson, A., Nakao, K., Howarth, M., & Gallacher, A. (2020). Supporting self-efficacy beliefs and interest as educational inputs and outcomes: Framing AI and Human partnered task experiences. *Learning and Individual Differences*, *80*, 101850. https://doi.org/10.1016/j.lindif.2020.101850

Guo, K., Li, Y., Li, Y., & Chu, S. K. W. (2023). Understanding EFL students' chatbot-assisted argumentative writing: An activity theory perspective. *Education and Information*

16

Technologies. https://doi.org/10.1007/s10639-023-12230-5

Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Investigating students' engagement in chatbot-supported classroom debates. *Interactive Learning Environments*, 1–17. https://doi.org/10.1080/10494820.2023.2207181

Han, D.-E. (2020). The Effects of Voice-based AI Chatbots on Korean EFL Middle School Students' Speaking Competence and Affective Domains. *Asia-Pacific Journal of Convergent Research Interchange*, *6*(7), 71–80. https://doi.org/10.47116/apjcri.2020.07.07

Huang, W., Roscoe, R. D., Craig, S. D., & Johnson-Glenberg, M. C. (2022). Extending the Cognitive-Affective Theory of Learning with Media in Virtual Reality Learning: A Structural Equation Modeling Approach. *Journal of Educational Computing Research*, *60*(4), 807–842. https://doi.org/10.1177/07356331211053630

Janssen, J., Kirschner, F., Erkens, G., Kirschner, P. A., & Paas, F. (2010). Making the Black Box of Collaborative Learning Transparent: Combining Process-Oriented and Cognitive Load Approaches. *Educational Psychology Review*, *22*(2), 139–154. https://doi.org/10.1007/s10648-010-9131-x

Jehn, K. A., Greer, L., Levine, S., & Szulanski, G. (2008). The Effects of Conflict Types, Dimensions, and Emergent States on Group Outcomes. *Group Decision and Negotiation*, *17*(6), 465–495. https://doi.org/10.1007/s10726-008-9107-0

Jin, T., Jiang, Y., Gu, M. M., & Chen, J. (2022). "Their encouragement makes me feel more confident": Exploring peer effects on learner engagement in collaborative reading of academic texts. *Journal of English for Academic Purposes*, *60*, 101177. https://doi.org/10.1016/j.jeap.2022.101177

Kalyuga, S. (2011). Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

King, R. B., & Gaerlan, M. J. M. (2014). High self-control predicts more positive emotions, better engagement, and higher achievement in school. *European Journal of Psychology of Education*, *29*(1), 81–100. https://doi.org/10.1007/s10212-013-0188-z

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R., J. (2018). From Cognitive Load Theory to Collaborative Cognitive Load Theory. *International Journal of Computer-Supported Collaborative Learning*, *13*(2), 213–233.

17

https://doi.org/10.1007/s11412-018-9277-y

Kwong, C.-Y. C., & Churchill, D. (2023). Applying the Activity Theory framework to analyse the use of ePortfolios in an International Baccalaureate Middle Years Programme Sciences classroom: A longitudinal multiple-case study. *Computers & Education*, *200*, 104792. https://doi.org/10.1016/j.compedu.2023.104792

Leppink, J., Paas, F., Van Der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Li, X., Li, X., Song, M., & Zhang, J. (2021). Longitudinal Reading Performance Comparison and its Variation with Print, Mobile and Social Media Tools. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(13), 273. https://doi.org/10.3991/ijet.v16i13.21563

Lin, X.-F., Wong, S. Y., Zhou, W., Shen, W., Li, W., & Tsai, C.-C. (2024). Undergraduate Students' Profiles of Cognitive Load in Augmented Reality–Assisted Science Learning and Their Relation to Science Learning Self-efficacy and Behavior Patterns. *International Journal of Science and Mathematics Education*, *22*(2), 419–445. https://doi.org/10.1007/s10763-023-10376-9

Liu, C., Hou, J., Tu, Y.-F., Wang, Y., & Hwang, G.-J. (2023). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, *31*(9), 5614–5632. https://doi.org/10.1080/10494820.2021.2012812

Liu, C.-C., Liao, M.-G., Chang, C.-H., & Lin, H.-M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, *189*, 104576. https://doi.org/10.1016/j.compedu.2022.104576

Liu, M. (2024). *Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing*.

Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, *70*, 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

McTear, M., Callejas, Z., & Griol, D. (2016). *The Conversational Interface*. Springer International Publishing. https://doi.org/10.1007/978-3-319-32967-3

18

Michaelis, J. E., & Mutlu, B. (2017). Someone to Read with: Design of and Experiences with an In-Home Learning Companion Robot for Reading. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 301–312. https://doi.org/10.1145/3025453.3025499

Millin, T., Millin, M., & Pearce, J. (2020). *Unpacking the efficacy of Reading to Learn using Cognitive Load Theory*.

Moreno, R., & Mayer, R. (2007). Interactive Multimodal Learning Environments: Special Issue on Interactive Learning Environments: Contemporary Issues and Trends. *Educational Psychology Review*, *19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of Artificial Intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, *7*(5), e07014. https://doi.org/10.1016/j.heliyon.2021.e07014

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). *Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research*.

Putwain, D., Sander, P., & Larkin, D. (2013). Academic self-efficacy in study-related skills and behaviours: Relations with learning-related emotions and academic success. *British Journal of Educational Psychology*, *83*(4), 633–650. https://doi.org/10.1111/j.2044-8279.2012.02084.x

Richter, J., Scheiter, K., & Eitel, A. (2018). Signaling text–picture relations in multimedia learning:

19

The influence of prior knowledge. *Journal of Educational Psychology*, *110*(4), 544–560. https://doi.org/10.1037/edu0000220

Sengupta, S. (n.d.). *Developing Academic reading at tertiary level: A longitudinal study tracing conceptual change*.

Snow, C. E. (2010). Academic Language and the Challenge of Reading for Learning About Science. *Science*, *328*(5977), 450–452. https://doi.org/10.1126/science.1182597

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tai, T.-Y., & Chen, H. H.-J. (2024). The impact of intelligent personal assistants on adolescent EFL learners' listening comprehension. *Computer Assisted Language Learning*, *37*(3), 433–460. https://doi.org/10.1080/09588221.2022.2040536

Wang, T., Li, S., Huang, X., Pan, Z., & Lajoie, S. P. (2023). Examining students' cognitive load in the context of self-regulated learning with an intelligent tutoring system. *Education and Information Technologies*, *28*(5), 5697–5715. https://doi.org/10.1007/s10639-022-11357-1

Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner. *Computers & Education*, *161*, 104059. https://doi.org/10.1016/j.compedu.2020.104059

Yu, Z. (2022). The effect of teacher presence in videos on intrinsic cognitive loads and academic achievements. *Innovations in Education and Teaching International*, *59*(5), 574–585. https://doi.org/10.1080/14703297.2021.1889394

Research on Foreign Language Aptitude of International Talents with in the Context of 'Trinity' in the New Era

Lijie GUO

Shandong University of Science and Technology

Abstract: This article explores the role of foreign language teaching and technology-assisted language teaching and learning in predicting foreign language aptitude in the new trinity era. Taking college English as an example of foreign language teaching, this article investigated the English scores of 180 non-English majors and their teachers' teaching habits, textbook use, and technology-assisted English teaching and learning. The results of the study were analysed by SPSS with ordinal logistic analysis. It was found that textbook had a significant positive effect on students' English scores, teaching habits were less correlated than textbook, and technology-assisted language teaching and learning mainly played a facilitating role. This study contributes to accelerating the pace of foreign language textbook reform, taking students' interests as the breakthrough of teachers' teaching, and enhancing the effect of technology application in foreign language teaching and learning.

Keywords: Teacher Teaching, Textbook, Technology-assisted Teaching and Learning, Language Aptitude, International Talents

1.Introduction

International talents are those who have a global vision and excellent foreign language skills, who are familiar with their own country's history and traditions, and who are proficient in the achievements of the world's civilization[1]. A good command of foreign languages or excellent foreign language aptitude is the key for international talents to go out into the world. Foreign language education plays a fundamental role in improving the foreign language ability of international talents. A large number of countries and regions attach great importance to the foreign language education of talents in order to promote economic and social development in the era of globalization, especially the cultivation of international talents. However, the development of science and technology has brought new challenges to foreign language education. More and more countries are emphasizing the link between technology and education. What accelerates the application of technology in education and increases the technological literatures of foreign language teachers has become a common demand.

In November 2022, the report of the 20th Party Congress in China stressed that "education, science and technology, and talents are the basic and strategic support for the comprehensive construction of a socialist modernizing country", which is the first time that the Party integrates education, science and technology, and talents together for systematic planning to serve the construction of an innovative country. The new trinity strategy requires universities to integrate the basic role of education and the supporting role of science and technology, which promotes the development of international talents and provides abundant talent reserve for the construction of China's innovative country. Foreign language ability is the ability that international talents must possess, and it is also one of the goals of talent cultivation in foreign language education.

Previous researches are also abundant and diversified in foreign language teaching and the function of technology on language teaching and learning. However, there are fewer studies on the interactive effect of them on the foreign language aptitude of talents. In a new trinity era, in addition to the improvement of foreign language teaching and the development of technology, it is more important for the cultivation of international talents to have a high degree of interaction between them. Therefore, based on the above discussion, this study tries to explore the interaction between foreign language teaching and science and technology application in the cultivation of foreign language aptitude of nationalized talents.

2. Literature Review

2.1 Effects of FL teachers' teaching on foreign language aptitude

Foreign language (FL) aptitude generally refers to the specific talent for learning a foreign or second language (L2)[2][3], which is a concept mainly rooted in educational psychology. Additionally, the interpretation of FL aptitude in applied linguistics is inevitably affected by developments in the neighboring fields of education and psychology. In recent years, after decades of rapid developments within the cognitive revolution[4], human cognitive factors have increasingly become significant research directions in FL aptitude. As Dörnyei[5] pointed out, FL aptitude has increasingly become something of a kind of hybrid construct related to a number of cognitive factors that form a composite measure considered as the general ability to master an L2. To summarize, educational psychology and cognitive construction have been research points including affective variables[6][7], language learning strategies[8][9], learning styles as contributors to aptitude and aptitude as a cognitive construct affected by language variables[10][11], individual differences in language learners[5][12] and the importance of phonological/orthographic processing for foreign language learning[13][14] etc.

2.2 Effects of CALL or MALL on Foreign Language Aptitude

Compared with internal factors (psychology & cognition), it is noted that with the advent of the internet, computer-assisted language learning (CALL) and mobile-assisted language learning (MALL) also occupy a crucial position in language learning. In 1980s, various types of CALL including the use of film, radio, television, language laboratories and videotapes, computers, and interactive video and some innovative uses of software, became commonplace[15]. Subsequently, the use of multimedia [16][17][18], the Internet (especially the -World Wide Web) [19][20] and various forms of distance learning[21][22][23][24] are widely used. Generally speaking, interest, convenience and effectiveness make CALL and MALL as tools to support language learning, vocabulary[25][26] grammar[27][28][29] and acquisition[30][31] which has become a trend for language educators and language learners. In fact, the application of CALL or MALL in language learning affects the language aptitude in a certain way.

Nevertheless, it can be noticed that more attention is paid to its internal factors (psychology & cognition). Its external aspects (CALL or MALL, referred to here as technology) generally concentrate on learners or teachers' attitudes, beliefs and functions etc., ignoring the effect on language aptitude. Grace Wiebe[32]have made a comparison between students' attitudes and instructors' perceptions towards the use of educational technologies. This study revealed a discrepancy between students' awareness of instructors' goals for using new technologies and the importance instructors placed on CALL, as well as between the types of technologies that instructor thought were useful for students' success and those that students thought were useful for their own success. Some significant research has examined language learners' beliefs about particular technologies used in foreign language classrooms[33]. The study demonstrated that increased exposure to effective and varied models of CALL may result in more positive beliefs towards it. However, other research has shown that while foreign language (FL) teachers may have positive attitudes towards informal MALL, many of them hold negative sentiments towards MALL in the classroom[34]. Additionally, MALL or CALL may have a negative effect on the improvement of autonomy and motivation[35]. It can be concluded that technology, specifically CALL and MALL, has been emphasized in language learning. However, there is no agreement on whether technology contributes to foreign language learning, and few researchers have compared the effectiveness of foreign language teaching with technology applications on language aptitude. Based on the above considerations, the analysis of the different effects of foreign language teaching and technology-assisted in language teaching and learning on language aptitude is necessary.

The following are main questions:

1. What is college English teachers' teaching habit in standing position and language choice in class?

2. What's the degree of satisfaction with the textbook and topic preference in textbooks for non-English major students?

3. What's the situation of technology-assisted in college English teaching and learning?

4. How do foreign language education (teachers' teaching habits, the satisfaction of textbook) and technology-assisted teaching and learning affect non-English major students' aptitude?

3. Research Design

3.1 Participants and Procedure

Given that English is a fundamental subject in China's foreign language education, this study takes college English as a factor of foreign language and investigates the situations of college English teaching and technology application. Additionally, the participants of this study are freshmen and sophomores of non-English major college students from the author's university. Those students have already learned English for several years before university and modern technology is widely used in college English teaching and learning in a way. After at least one academic year of technology-based college English learning experience, students gain better understanding of foreign language teaching and learning and technology-assisted on foreign language learning. Moreover, this study also surveyed the English final examination score of students to reflect the impact of foreign language teaching and technology-based on language aptitude.

This study randomly selected several classes, through online and offline distribution of questionnaires at intervals before class. Those students were required to finish the questionnaire at least 5 minutes. Consequently, 200 questionnaires were given out and 180 valid questionnaires were collected. Before conducting the questionnaire survey, the study obtained the students' consent and informed them that all the research data collected were anonymized to protect participants' privacy. Students' participation was cooperative and voluntary, and thus they completed the questionnaire carefully.

3.2 Questionnaire items

The total number of questionnaire items are 23 which are composed of three main themes: personal information, foreign language teaching and the application of technology: 1) personal information contains 4 items, students' grade, gender, times of College English course in a week and their final English scores. 2) foreign language teaching generally can't devoid of foreign language teachers and relative textbooks. There are 9 items in the questionnaire, foreign language teachers' teaching including teaching habits, teaching preferences of contents; and textbooks including satisfaction and topic preference. 3) in the aspects of technological application, the questionnaire consists of 10 items with 2 dimensions application of technology in College English teaching and College English learning.

The satisfaction investigation items were ranked on a five-point Likert Scale ranging from

strongly disagree to strongly agree. The data was analyzed based on the assigned values of 1 to 5. The information was collected using Questionnaire Star and analyzed using the Questionnaire Star system and SPSS.26 software. Furthermore, the questionnaire was tested for reliability and validity, with a Cronbach's α value of 0.753 and a KMO value of 0.770, respectively. These results indicate that the questionnaire is highly reliable and valid.

4. Results and Discussion

4.1 Teachers' habits

Teachers are essential figures in influencing students' learning level and development potential[36]. Teachers' teaching habits and contents influence students' learning habits, attitudes and interests, etc., which will ultimately affect students' foreign language aptitude. Teachers' teaching habits are varied with its own personalities. The following presents statistics on the location of teachers in the classroom and their language options when teaching (see Table 1&Figure 1).

4.1.1 Teachers' Location in Class

The teachers were classified as either 'podium-centered' or 'position-random' based on their location in the classroom. The former emphasizes the teacher's authority, while the latter promotes equal participation during the lesson. Based on the statistical results presented in Table 1, 46.11% of the teachers were perceived as 'podium-centered' while 53.89% were considered 'position-random'. This indicates that over half of the college English teachers actively engage with their students, viewing them as collaborators and participants, and prioritizing interaction to foster positive teacher-student relationships. In contrast, 46.11% of the teachers prefer to stand at the podium and observe the students and classroom dynamics. They aim to grasp the overall effect, participate in classroom teaching as an observer or reviewer, and maintain a certain distance from the students in order to build up teachers' prestige. These two different habits reflect the teachers' approach to teaching. Teachers adjust their own classroom strategies based on their students' performance and strive to achieve positive outcomes. According to the Pearson correlation analysis (p>0.05), there is a negative correlation between teachers' position and students' English scores, indicating that educators' positions do not necessarily reflect students' English aptitude.

Table 1: Teacher's position in class.

	Podium-centered	Position-random	Amount
Number	83	97	180

Proportion	46.11%	53.89%	100%

4.1.2 Language Choice and Question Times in Class

Verschueren[37] argued that language use is a conscious and unconscious process influenced by both intra- and extra-linguistic factors. Language choice is generally determined by cultural, social and other objective factors. From the statistical results presented in Figure 1, it is clear that nearly 70% of college English teachers use English as the primary language in the classroom, supported by Chinese. Only a few teachers use either Chinese or English. In fact, in order to create an English learning environment for students, the majority of teachers choose English as the language of instruction, that is, in general, teachers choose the language according to the needs of the class including intra- and extra-linguistic factors. However, the number of questions in more English is lower than in the situation of more Chinese or even, which indicates that although teachers choose English as the main language to teach, they ignore the participation of students in class.
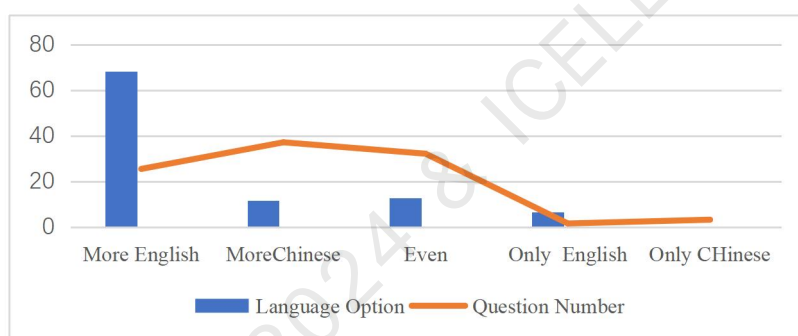
Figure 1: Language Option & Question Times.

4.2 Teachers' Teaching Instrument Textbook

The following statistics on students' satisfaction with textbook and topic preference in textbook are useful for teachers' teaching and textbook revolutionary.

4.2.1 Satisfaction & Usage Frequency

The statistical results of non-English students' satisfaction with the college English textbooks show that only more than 60% of non-English students are content with their English textbooks (Figure 2). At least one-third students keep neutral attitude. However, compared with satisfaction, the trend of textbook usage frequency shows differences. Explicitly, the proportion of agree attitude of textbook use is lower than its satisfaction but disagree attitude is higher. It indicates that although the reform of foreign language textbook in colleges has achieved certain results in recent years and these textbooks basically meet the needs of non-English major students in learning

English, they have not been fully utilized. Moreover, students' use of textbooks is primarily influenced by various external factors, completing homework, exam requirements, and teacher supervision etc., which clarifies the low initiative of students with textbooks learning English.
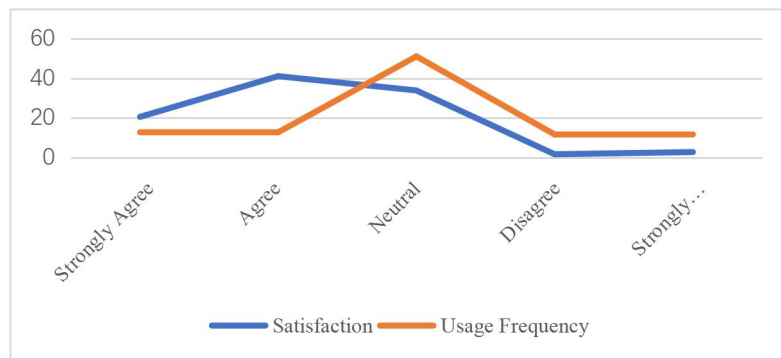
Figure 2: Textbook Satisfaction & Usage Frequency.

4.2.2Topic Preference in Textbook

Figure 3 displays the percentage of students' preferences for articles in textbooks: social life (50.56%) > language and culture (38.89%) > current affairs and politics (35%) > education and psychology (18.33%) > others (16.11%) > ethics and morals (12.78%).   The data shows that college students are interested in social, linguistic and current affairs and politics topics. Besides with a strong desire to learn English, they especially keep eyes on national affairs in textbook with a strong sense of family and country. However, at the same time, the data also demonstrate ethical and didactic articles in the textbooks are not popular among them. Living in an era of rapid technological development and diverse ideas, most of them pursue individualism and challenges traditional rules and ethics. These changes have brought new challenges to teachers and textbook compliers in designing effective teaching strategies.
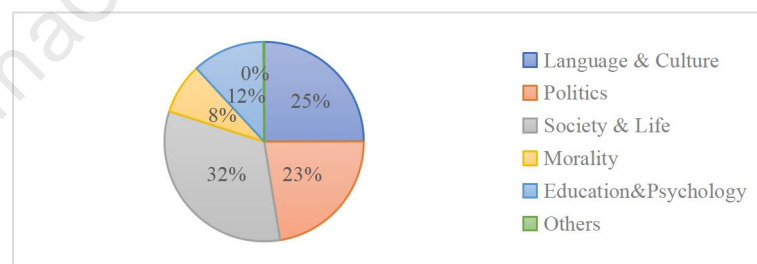
Figure 3: Topic Preference in Textbook.

4.3 Technology-assisted in College English Teaching and Learning

Computer-based teaching has become a symbol of the power of science and technology to promote the development of education in the Internet era. Especially in higher education,

technology plays an increasingly important role in teaching and learning. The following demonstrates the application of technology in college English teaching and learning and its satisfaction.

4.3.1 Technology-assisted Teaching and Learning

Technology has been widely used in English teaching and learning. Research indicates that many teachers and students have a positive attitude towards technology in education. Figure 4 illustrates the role of technology in English teaching and learning, showing that the use of technology in teaching is more prevalent than in learning. It is worth noting that teachers tend to use technology more frequently in listening exercises, while students tend to use it more for translation purposes. Teachers prioritize their students' development, while students focus on learning and understanding English words meaning. Both groups find technology useful in achieving their respective goals.
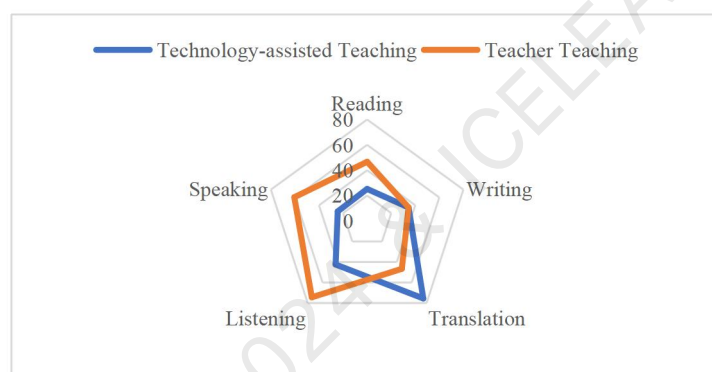
Figure 4: Technology-assisted Teaching & Learning.

4.3.2 Satisfaction on Technology-assisted Teaching and Learning

Based on the statistical results presented in Figure 4, it can be observed that satisfaction with technology-supported learning is higher than with teaching. The majority of students are satisfied with technology-supported learning, but almost half of the teachers are neutral. Nevertheless, according to Figure 3, the use of technology in teaching is more widespread than in learning, which indicates that the acceptance level of students is higher than that of teachers in technology-assisted application. Teachers have an objective approach to technology-assisted teaching and use it for students' knowledge acquisition and facilitate the achievement of teaching objectives, which shows the professionalism and rigor of college English teachers.
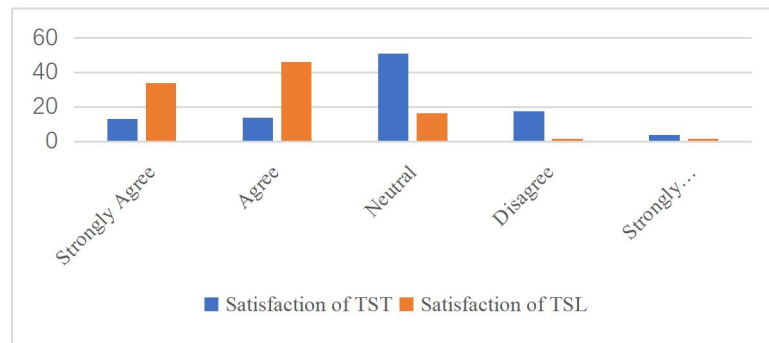
Figure 5: Satisfaction Comparison of Technology-assisted Teaching & Learning.

4.4 Effects of Both Educational Teaching and Technology-assisted Teaching and Learning on Foreign Language Achievement

The independent variables in this study are the situation of teaching, the use of textbook, and technology-assisted in English teaching and learning, and the dependent variable is students' final English grades, which are poor (below 60), average (60-69 points), good (70-79 points), very good (80-89 points), and excellent (above 90 points), and the final ordered by SPSS.26 logistics regression analysis.

4.4.1 Parallel Lines Test and Model Fitting

The ordered multi-categorical logistics regression model needs to be tested for the assumption that the coefficients of the independent variables are equal (also known as the parallel lines test). The parallel lines test is used to test whether the effect of each level of the value of the independent variable on the dependent variable is the same in each regression equation. According to the statistical results, the result of the parallel lines test of the questionnaire $p=0.168 > 0.05$ indicates that the model passes the parallel lines test and can be analyzed further. In addition, the model fit information is $p=0.001<0.05$, which means that this constructed model is meaningful.

4.4.2 Parameter Estimation Analysis

According to Table 2 below, it can be seen that: the regression coefficient of the situation of textbook satisfaction is 0.756, and it is significant ($p=0.000<0.05$), and the regression equation is: $y=0.756x+16.12$, which indicates that students' textbook satisfaction has a significant positive influence on the relationship with students' performance in English, and the use of textbooks has an explanatory degree of 75.6% of students' performance in English. However, the p-values for teacher teaching, technology-assisted teaching and learning are all greater than 0.05 and therefore not statistically significant.

Table 2: Summary of ordered logistics regression model analysis.

		Parameter estimation						
		Estimation	standard error	Wald	DOF	S.	95% CI	
							lower	Upper
VPT	[SCORE = 1]	-1.839	.904	4.144	1	.042	-3.610	-.068
	[SCORE = 2]	.012	.871	.000	1	.989	-1.695	1.720
	[SCORE = 3]	2.142	.889	5.806	1	.016	.400	3.884
	[SCORE = 4]	4.970	1.012	24.115	1	.000	2.986	6.954
Location	textbook	.756	.188	16.116	1	.000	.387	1.125
	teaching	-.077	.229	.113	1	.737	-.526	.372
	technology	-.300	.249	1.447	1	.229	-.789	.189

Correlation function: partial logarithm.

5. Conclusion and Implications

From the above statistical results, the more familiar students are with the textbook, the easier it is for them to get into the classroom learning environment quickly, which can attract their attention and contributes to following teacher's teaching progress and grasping teacher's key and difficult points. In addition, technology-assisted in teaching and learning doesn't have a significant effect on the students' English aptitude, which indicates that the application of educational software in teaching and learning is not an important reason for affecting the English aptitude of students, namely although as a kind of teaching tool and learning means, it can help improve the efficiency of English teaching and learning and enrich the way of teaching and learning, it can't drive the students to learn English independently or form the consciousness of learning English. On this basis, this paper puts forward the following suggestions:

Teachers should pay attention to the role of college English textbook in teaching, but also prevent the supremacy of textbook. The content of foreign language teaching should meet the needs of the times when students can show the world to China as well as China to the world[38]. In addition, textbooks are extremely authoritative, which requires teachers to pay particular attention to the content of college English textbooks which should satisfy needs of students to learn knowledge and cultivate their interest in the subject, as well as meet the requirements of mainstream values. At the same time, we also need to prevent the theory of textbooks and supremacy, and take textbooks as the basic teaching materials, as a basis for expanding the content of learning, cultivating students' interest in learning English independently, developing self-driven, and setting up correct values as the main focus. Finally, the quality of compliers should be improved to ensure that the pace of reform of college for the improvement of the textbook quality.

College English teachers should change the teaching concept from "textbook-supreme" to "talent-first", and from enriching teaching content to nurturing people with moral values. Since

entering the new era, in order to respond to the needs of trinity development in the national education, science and technology and talent, foreign language teachers should not only know how to teach and explain, but also know how to educate students, and strengthen the construction of civics and politics in foreign language courses[39]. Secondly, foreign language teachers can recommend films, books, forums, etc., which are in line with the teaching theme, which can not only cultivate students' interest in learning foreign languages, broaden their horizons, but also spread national streamline value. In addition, teachers organize some learning games, such as foreign language word games, memorizing English scenes and describing them in their own words, etc., so as to let students come into contact with English in entertainment, and to reduce the fear and anxiety of students who are reluctant to open their mouths to speak English and are afraid of making mistakes. Finally, foreign language teachers learn to take the initiative to walk into the students' lives as students' mentors and friends.

Colleges should give full play to the important roles of "technology-assisted in teaching and learning". Meanwhile, colleges should also promote the feeding role of "talents helping science". The purpose of scientific and technological inventions is to serve mankind. In the process of foreign language education, teachers should not only learn the basic teaching technology to assist teaching, but also keep abreast of the times, learn in-depth technology-assisted translation, and improve personal technological literature etc. for the purpose of making full use of technology to impart knowledge to students. Students should also combine information technology and network with what they have learn to maximize the use of technology to assist personal learning. Last but not least, technology serve for human beings, but people also immerging into exploring science and technology, struggling for multi-directional and multi-level technological breakthroughs in foreign language.

Time creates talents, education guides talents, and talents revolutionize the times. This paper takes college English in foreign language teaching as an example and randomly examines the influence of teachers' teaching and technology-based talent cultivation. According to the statistical results, it is found that compared with the traditional influencing factors of language aptitude (LA), teachers are still the main reason for teaching and nurturing; the satisfaction of teaching materials is an important indicator influencing LA, which develops synchronously with talents; and technology can only play a helping role. In other words, teachers' teaching habits reflect how to cultivate talents and what kind of talents to cultivate, textbook satisfaction will directly affect the effect of students' learning, while the application of technology in teaching and learning still has a lot of room for improvement. It is urgent that how to use modern technology to cultivate students' interest in learning and improve their learning efficiency. In addition, it should be noted that due to the limitations of the sample size and influencing factors used in this study, the conclusions drawn in this paper need to be further examined and tested, and it is necessary to further increase the sample size and expand the research dimensions for subsequent studies.

References

[1] *Jiang Feng. Cultivating high-level international talents with global vision and world vision[J]. China Higher Education,2020,(21):26-28.*

[2] *Carroll, J. Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), Individual differences and universals in language learning aptitude[J]. Rowley, MA: Newbury House,1981,83-117.*

[3] *Skehan, P. (1986a). Where does language aptitude come from? In P. Meara (Ed.), Spoken language[J]. London: Centre for Information on Language Teaching, 1986,95-113.*

[4] *Carroll, J. Cognitive abilities in foreign language aptitude: Then and now. In T. Parry & C. Stansfield (Eds.), Language aptitude reconsidered. Englewood Cliffs[J].NJ, Prentice-Hall.1990,1-29.*

[5] *Dörnyei, Z. The Psychology of the Language Learner: Individual Differences in Second Language Acquisition (1st ed.)[M]. Routledge,2005.*

[6] *Jee, M. J. Affective factors in Korean as a Foreign Language: anxiety and beliefs. Language, Culture and Curriculum[J], 2014,27(2), 182–195.*

[7] *Zhi, R., & Wang, Y. (2023). English as a foreign language teachers' professional success, loving pedagogy and creativity: A structural equation modeling approach[J]. Thinking Skills and Creativity, 2023,49, 101370.*

[8] *Oxford, R., & Nyikos, M. Variables affecting choice of language learning strategies by university students[J]. The modern language journal, 1989,73(3), 291-300.*

[9] *Tai, K. & Zhao, Y. Success factors for English as a second language university students' attainment in academic English language proficiency: exploring the roles of secondary school medium-of-instruction, motivation and language learning strategies[J]. Applied Linguistics Review,2023, 15(2), 611-641.*

[10] *Sparks, R., & Ganschow, L. Aptitude for learning a foreign language[J]. Annual Review of Applied Linguistics,2001, 21, 90–111.*

[11] *Sparks, R.L., Alamer, A. How does first language achievement impact second language reading anxiety? Exploration of mediator variables[J]. Read Writ, 2023,36, 2667–2692.*

[12] *Llompart, M. & Dąbrowska, E."Foreign" language aptitude predicts individual differences in native grammatical proficiency[J]. Linguistics, 2023,61(5), 1165-1193.*

[13] *Bosmans, D., & Hurd, S.. Phonological attainment and foreign language anxiety in distance language learning: A quantitative approach[J]. Distance Education,2016, 37(3), 287-301.*

[14] *Wei, X., Gunter, T. C., Adamson, H., Schwendemann, M., Friederici, A. D., Goucha, T., & Anwander, A. White matter plasticity during second language learning within and across hemispheres[J]. Proceedings of the National Academy of Sciences,2024, 121(2), e2306286121.*

[15] *Iandoli, L. J. CALL and the profession: The current state[J]. French Review,1990,64(2), 261-272.*

[16] *Akbulut, Y. Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language[J]. Instructional Science,2007, 35, 499-517.*

[17] *Sun, H., & Yin, B. Multimedia input and bilingual children's language learning[J]. Frontiers in Psychology,2020, 11, 570505.*

[18] *Teng, M. F. The effectiveness of multimedia input on vocabulary learning and retention[J]. Innovation in language learning and teaching, 2023,17(3), 738-754.*

[19] *Kessler, G. Technology and the future of language teaching[J]. Foreign language annals,2018, 51(1), 205-218.*

[20] *Xu, S., Li, X., & Yang, X. Multilateral interactive teaching model for improving foreign language learners' autonomous learning ability by using multimedia network technology[J]. The International Journal of Electrical Engineering & Education, 2021,002-77.*

[21] *White, C. J. Effects of mode of study on foreign language learning[J]. Distance Education,1997, 18(1), 178-196.*

[22] *White, C. Distance learning of foreign languages[J]. Language teaching, 2006,39(4), 247-264.*

[23] *White, C. The distance learning of foreign languages: A research agenda[J]. Language Teaching, 2014,47(4), 538-553.*

[24] *Pannatier, M., & Béntrancourt, M. Learning from academic video with subtitles: When foreign language proficiency matters[J]. Learning and Instruction,2024, 90, 101863.*

[25] *Nick C. Ellis (1995) The psychology of foreign language vocabulary acquisition: implications for call[J], Computer Assisted Language Learning,1995, 8:2-3, 103-128.*

[26] *Hanafiah, W., Aswad, M., Sahib, H., Yassi, A. H., & Mousavi, M. S. The impact of CALL on*

vocabulary learning, speaking skill, and foreign language speaking anxiety: the case study of Indonesian EFL learners. Education Research International, 2022.

[27] Metcalfe, P. CALL, the foreign-language undergraduate and the teaching of grammar: a linguistic and political battlefield[J]. ReCALL,1992, 4(7), 3–5.

[28] Pirasteh, P. The effectiveness of computer-assisted language learning (CALL) on learning grammar by Iranian EFL learners[J]. Procedia-Social and Behavioral Sciences,2014, 98, 1422-1427.

[29] Nurhayati, N., & Sedubun, S. Investigating the Role of Interactive Activities in Enhancing Pre-Intermediate English Grammar Learning: A Qualitative Inquiry at Universitas Klabat. AL-MIKRAJ Jurnal Studi Islam dan Humaniora,2024, 4(02), 869-881.

[30] Yang, C. A formalist perception on language acquisition. Ling[J]. Approach. Biling. 2018,8, 665–706.

[31] Burston, J., Athanasiou, A., & Giannakou, K. Quantitative experimental L2 acquisition MALL studies: A critical evaluation of research quality[J]. ReCALL, 2024,36(1), 22–39.

[32] Grace Wiebe & Kaori Kabata. Students' and instructors' attitudes toward the use of CALL in foreign language teaching and learning[J], Computer Assisted Language Learning, 2010,23:3, 221-234.

[33] Emily A. Hellmich. CALL beliefs in context: a study of US high school foreign language learners, Computer Assisted Language Learning, 2021,34:7, 845-867.

[34] Liu, H., Tao, W., & Cain, W. Investigating mobile assisted English foreign language learning and teaching in China: Issues, attitudes and perceptions[J]. In Handbook of research on foreign language education in the digital age,2016, 315-333.

[35] Adara, R. A. Negative Effects of MALL on the Improvement of Learners Autonomy and Motivation[J]. Prosiding ICoISSE,2020, 1(1), 613-622.

[36] Zhou Yan." Teachers are the key to improving English teaching in foreign language learning environments." [J]. Foreign Language Teaching and Research 2010,42(04),294-296.

[37] Verschueren，J． Pragmatics as a Theory of Linguistic Adaptation［C］IprA Working Document Antwerp: International Pragmatics Association, 1987.

[38] Wen Qiufang. "70 years of foreign language education in new China: achievements and challenges." [J]. Foreign Language Teaching and Research,2019,51(05),735-745+801.

[39] Huang Guowen,Xiao Qiong. Six Elements of Civic and Political Construction in Foreign Language Courses[J]. China Foreign Language,2021,(02),1+10-16

Promoting speaking through contests: A Soeasy case study

Junyan Guo, Wuxi University; The University of Auckland,

email: graceguojunyan@163.com

Xueqing Jiang, Wuxi Taihu University

Xiaoxiang Li, Wuxi Taihu University

Abstract: The growing availability of mobile technologies has contributed to an increase in mobile-assisted language learning (MALL) applications (apps) such as large-scale commercial apps Duolingo and Babbel. However, because digital cultures differ between countries, locally developed apps also need to be explored. The current study examined and compared EFL learners' speaking performance with a popular app in China (Soeasy) from a contest perspective－2024 Internet English Listening and Speaking Challenge. Data from 60 students who have entered the National Final were collected. Their scores and performances in three rounds of the contest (Provincial Preliminary, Provincial Final, and National Final) were statistically analyzed to examine their gains in speaking. Surveys were employed to gain their perceptions, attitudes, and views on the effectiveness, scoring accuracy, and potential use of the app. This study also discusses the implications of these findings for applying Soeasy in formal and informal language learning, teaching, and assessment.

Key words: Mobile-assisted Language Learning (MALL); Soeasy; Contest; effectiveness

1. Introduction

The rapid development of artificial intelligence (AI) technology has provided students with a richer and more optimized experience in language learning. In particular, the growing availability of mobile technologies has contributed to an increase in mobile-assisted language learning (MALL) applications (apps) which allow students to study anytime and anywhere. Previous research has explored learners' perceptions, beliefs, experiences, and motivations, and identified

the potential and benefits of both large commercial apps such as Busuu (Rosell-Aguilar, 2018), Duolingo and Babbel (Kessler et al., 2023) as well as self-developed apps (Zou et al., 2018). Moreover, researchers have zoomed in to examine their effectiveness in improving specific linguistic knowledge such as vocabulary and grammar, and specific skills such as reading, writing, listening, speaking, and translation (e.g., Kessler et al., 2023; Kohnke, 2020; Loewen et al., 2020; Morgana, 2015; Steel, 2012). Among them, speaking, as the more challenging aspect of language learning has drawn comparatively less attention. Under this circumstance, the current study examined and compared contestants' performance with a popular app in China (Soeasy) from a contest perspective – 2024 Internet English Listening and Speaking Challenge.

2. Literature Review

2.1 Previous research on the use of apps for developing speaking skills

Despite the multiple calls for researchers to explore the effectiveness of apps for language learning (e.g., Heift & Chapelle, 2012; Plonsky & Ziegler, 2016), empirical evidence of L2 learning in many MALL studies (Burston, 2015) remained limited. The commercial nature of the apps (e.g. cost) can partially account for this. Another possible explanation is the absence of "objective and quantifiable measures of learning outcomes" (Loewen et al., 2019, p. 294). Appendix 1 presents an inexhaustive list of previous studies investigating the effectiveness of learning apps for developing speaking skills. Major characteristics are summarized as follows. Firstly, both survey and pre-post test results indicated that learners obtain greater learning gains in receptive knowledge such as vocabulary and grammar than in speaking skills. Second, learner and (speaking) task characteristics (e.g., type, feedback) and measurement of improvement in speaking were not clear. Although most apps take speaking as their marketing strategy such as "speaking a new language with confidence" (Babbel), "speaking a language within 10 minutes a day" (Busuu), "learning to speak anytime, anywhere" (Rosetta Stone), the lessons and exercises in these apps are designed to improve or advance learners' overall language competencies. In addition, the pedagogical approach incorporates heavy input with limited opportunities to produce the language. Only two studies (Lehman et al., 2020; Zou et al., 2023) focusing on speaking apps specified speaking task type and feedback provided. In general, there are two types of tasks: reading

2

sentences/text aloud and spontaneous speaking (topic statement). Feedback includes scores, practice suggestions, sample answers, and text evaluation. The evaluation mainly targets fluency and vocabulary.

2.2 Soeasy and The Internet English Listening and Speaking Challenge (IELSC)

The National University Students' Innovation and Invention Contest: The Internet English Listening and Speaking Challenge was hosted by the China Information Industry Association (CIIA). The procedure of the contest is demonstrated in Figure 1.

Figure 1 Procedures of the IELSC
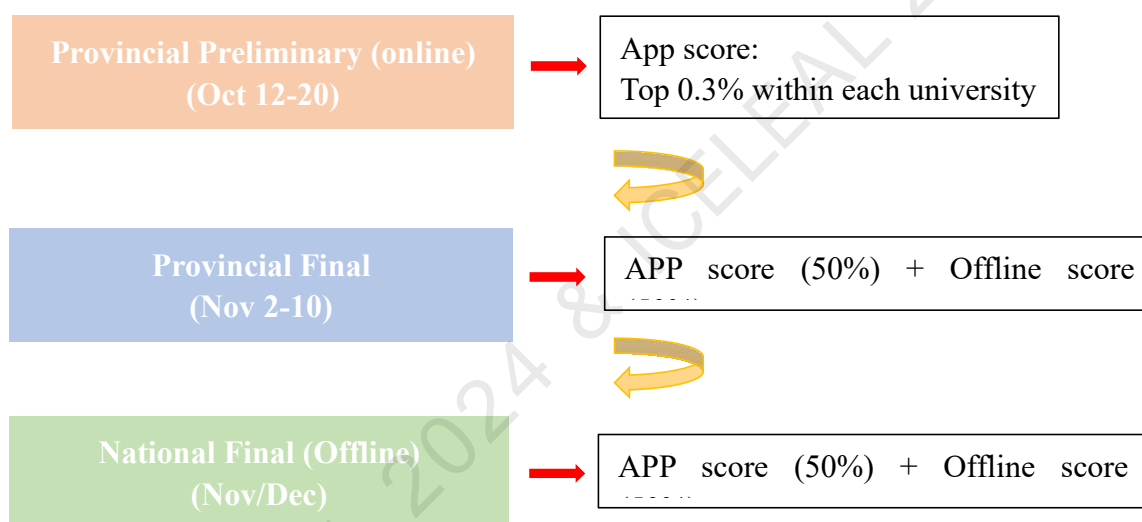
The Soeasy app was developed to provide an easily accessible resource for tertiary-level English language learners to practice listening and speaking. It was adopted in this contest. The provincial preliminary test has four challenges (see Table 1). The first three challenges each offer contestants three chances and the last challenge has unlimited chances. For the first three Challenges, the app will automatically select the highest score within the three chances. Contestants need to get 80 within the three tries within Challenge 1 to enter Challenge 2, 85 within Challenge 2 to enter Challenge 3, and 90 within Challenge 3 to enter Challenge 4. Contestants can retry multiple times to gain a higher score. *Soeasy* app will give a holistic score and subscores of pronunciation, intonation, fluency, and completeness.

Table 1 Details of the Provincial Preliminary Test

3

Challenge	Name	Description	Score Percentage	Minimum score to enter next challenge
1	Follow reading	Read after the recording (no script). The recording can be played twice. It has 10 sentences.	20%	80
2	Multiple choice	Select the relevant phrase or sentence according to the recording. The recording can be played twice. It has 10 sentences.	20%	85
3	Retell	Retell after the recording (no script). The recording can be played twice. (10 sentences)	25%	90
4	Reading	Read according to the script. (three paragraphs)	35%	/

The online test of the Provincial Final Test follows the same process as the provincial preliminary test and scores given by the Soeasy APP. In addition to the online test, a panel of judges will give scores based on the retelling section. Their criteria are accuracy (keywords and logic), pronunciation, and fluency. Similar to the Provincial Final Test, the national final test consists of two parts − APP + offline (see Table 2). The top 10% were awarded the first prize, 10%-20% second prize, 20%-30% third prize, and the rest winning prize.

Table 2.2 Details of the National Final Test

4

Name	Description	Score Percentage	Raters
Follow reading	Read the sentence after the recording (no script). The recording can be played twice. (10 sentences)	50%	APP
	Read according to the script. (3 paragraphs)		
Retelling	Retell the paragraph according to the recording. (1 minute)	50%	A panel of judges

2.3 The current study

The current study intends to investigate the effectiveness of Soeasy in developing the listening and speaking abilities of tertiary-level EFL speakers and sets out to probe into the following research questions:

RQ 1: How effective is Soeasy in developing the speaking skills of tertiary-level EFL learners in IELSE?

RQ2: What are the perceptions and experiences of these learners in this process?

3. Methodology

To investigate the learning gains of enhancing speaking skills via Soeasy in IELSE, a mixed methods design was adopted. Quantitative methods were used to examine the speaking development whereas qualitative findings served to complement and validate quantitative findings and elicit contestants' perceived gains.

3.1 Participants

The participants are 60 students who have entered the National Final, 15 from each group (1st, 2nd, 3rd, and winning award).

3.2 Analysis

Table 3.1 Details of data analysis

Research Questions	Methods	Description
1	quantitative	Statistical analysis 1. Scores of Provincial Preliminary, Provincial Final, and National Final 2. Fluency: speech rate 3. Vocabulary: diversity and accuracy
2	qualitative	Survey results of perceptions and experiences of learning gains in speaking

References:

Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL, 27*(1), 4-20. https://doi.org/10.1017/S0958344014000159

Heift, T., & Chapelle, C. A. (2012). Language learning through technology. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 555-569). Routledge. https://doi.org/10.4324/9780203808184-41

Kessler, M., Loewen, S., & Gönülal, T. (2023). Mobile-assisted language learning with Babbel and Duolingo: comparing L2 learning gains and user experience. *Computer Assisted Language Learning*, 1-25. https://doi.org/10.1080/09588221.2023.2215294

Kohnke, L. (2020). Exploring learner perception, experience and motivation of using a mobile app in L2 vocabulary acquisition. *International Journal of Computer-Assisted Language Learning and Teaching, 10*(1), 15-26. https://doi.org/10.4018/IJCALLT.2020010102

Lehman, B., Gu, L., Zhao, J., Tsuprun, E., Kurzum, C., Schiano, M., Liu, Y., & Tanner Jackson, G. (2020, 2020//). Use of adaptive feedback in an App for English language spontaneous speech. Artificial Intelligence in Education, Cham.

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL, 31*(3), 293-311. https://doi.org/10.1017/S0958344019000065

Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign*

language annals, 53(2), 209-233. https://doi.org/10.1111/flan.12454

Morgana, V. (2015). Investigating students' and teachers' perceptions of using the iPad in the EFL classroom. . *Paper presented at the EuroCALL 2015 Conference, Padova, Italy.*

Plonsky, L., & Ziegler, N. (2016). The CALL-SLA interface: Insights from a second-order synthesis. *Language Learning & Technology, 20*(2), 17-37. https://doi.org/http://dx.doi.org/10125/44459

Rosell-Aguilar, F. (2018, 2018/11/02). Autonomous language learning through a mobile application: a user evaluation of the busuu app. *Computer Assisted Language Learning, 31*(8), 854-881. https://doi.org/10.1080/09588221.2018.1456465

Steel, C. (2012). Fitting learning into life: Language students' perspectives on benefits of using mobile apps. . ascilite2012 Conference proceedings. ASCILITE 2012 - Annual conference of the Australian Society for Computers in Tertiary Education., Wellington, New Zealand.

Vesselinov, R. (2009). *Measuring the effectiveness of Rosetta Stone*.

Vesselinov, R., & Grego, J. (2016). *The busuu efficacy study*. Retrieved from http://comparelanguageapps.com/documentation/The_busuu_Study2016.pdf.

Vesselinov, R., & Grego, J. (2018). *The italki efficacy study*.

Vesselinov, R., Grego, J., Tasseva-Kurktchieva, M., & Sedaghatgoftar, N. (2021). *The Busuu efficacy study*.

Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An Investigation Into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *SAGE Open, 13*(3), 1-9. https://doi.org/10.1177/21582440231193818

Zou, B., Li, H., & Li, J. (2018). Exploring a curriculum app and a social communication app for EFL learning. *Computer Assisted Language Learning, 31*(7), 694-713. https://doi.org/10.1080/09588221.2018.1438474

7

Appendices

Appendix 1

A. Participants' characteristics of previous research on the use of apps for developing speaking skills

Author(s)	Participants Characteristics				
	APP/Participant Number	First Language	Average Age	Level of proficiency	Occupation
Vesselinov (2009)	Rosetta Stone, n=135	Varied	39	Beginner/novice	Varied
Vesselinov and Grego (2016)	Busuu, n=196	Varied	36	Beginner/novice	Varied
Vesselinov and Grego (2018)	italki, n=102	Varied	39.5	Beginner/novice	Varied
Lehman et al. (2020)	The English Language Artificial Intelligence, n=94	Chinese	/	/	University students
Loewen et al. (2019)	Duolingo, n=9	Varied	/	novice	8 graduate students, 1 professor
Loewen et al. (2020)	Babbel, n=58	/	/	Intermediate or lower	/
Vesselinov et al. (2021)	Busuu, n=114	Varied	38.2	Beginner/novice intermediate	varied
Kessler et al. (2023)	n=59 1. Duolingo (n=32) 2. Babbel (n=27)	Varied	Duolingo: 30.1 Babbel: 30	/	Varied
Zou et al. (2023)	n=40	Chinese	/	/	undergraduates

	1. Liulishuo (n=12) 2. EAP Talk (n=16) 3. IELTS Liulishuo (n=18) 4. Shanbay (n=13)				

B. Task characteristics of previous research on the use of apps for developing speaking skills

Author(s)	Task Characteristics			
	Learning Language	Task Type	Feedback	Length of Study
Vesselinov (2009)	Spanish	/	/	55 hours
Vesselinov and Grego (2016)	Spanish	/	/	2 months (average 22.5 hours)
Vesselinov and Grego (2018)	Spanish	/	/	2 months (average 19 hours)
Lehman et al. (2020)	English	Spontaneous speaking (topic statement)	1. Fluency: filler words, pauses, repeated words, speaking rate, stressed words 2. Vocabulary: vocabulary diversity	1 month
Loewen et al. (2019)	Turkish	Translation, dictation, vocabulary matching, and sentence unscrambling	Displaying a correct answer or indicating a "right/wrong" evaluation	34 hours
Loewen et al. (2020)	Spanish	Partial dictation, vocabulary matching, oral repetition, gap-filling exercises	Correct/incorrect feedback	≥3 hours ≥20 minutes weekly
Vesselinov et al. (2021)	Spanish	/	/	2 months, 2 lessons per week

9

Kessler et al. (2023)	Turkish	/	1. Fluency: filler words, pauses, repeated words, speaking rate, stressed words 2. Vocabulary: vocabulary	8 weeks
Zou et al. (2023)	English	1. Reading texts aloud; 2. Spontaneous speaking (topic statement)	Score, practice suggestion, sample answer, highlighted color, textual evaluation	1 month

C. Factors of speaking assessment of previous research on the use of apps for developing speaking skills

	Speaking Assessment		
Author(s)	**Aspect of speaking**	**Measurements of effectiveness**	**Results**
Vesselinov (2009)	Score/level	ACTFL OPIc: Elicit a sample of speech via recorded, computer-adapted voice prompts	64.4% increased their ACTFL results with at least one level
Vesselinov and Grego (2016)	Score/level	Oral Proficiency Interview (OPIc)	75% increased by at least one level
Vesselinov and Grego (2018)	Score/level	Oral Proficiency Interview (OPIc)	72% increase by at least one level
Lehman et al. (2020)	1. Fluency: filler words, pauses, repeated words, speaking rate, stressed words 2. Vocabulary: vocabulary diversity	Changes in spoken response performance over the course of app interaction	Significant improvement on aspects of filler words, repeated words, speaking rate, stressed words, and vocabulary diversity
Loewen et al. (2019)	score	Participant journals; Duolingo progress test; Turkish 151 exam	Less successful on oral language than written.
Loewen et al. (2020)	score	1. ACTFL oral proficiency interview-computer	Less gains in communicative skill than

		(OPIc): Listen and respond to 12 to 17 prompts 2. Progress questionnaire 3. Interview	receptive knowledge 59% showed improvements in speaking
Vesselinov et al. (2021)	score	TrueNorth Test (TNT) 1. elicited imitation 2. free speech	46% improved (CEFR estimate) 48% improved (ACTFL estimate)
Kessler et al. (2023)	/	1. Questionnaire: users' perceived effectiveness 2. Post-test: picture description (5 sentences)	Babbel users felt Duolingo users were relatively more successful in speaking (not significant)
Zou et al. (2023)	1. Overall score 2. Scores of pronunciation, fluency, vocabulary, and grammar	1. questionnaire; 2. semi-structured interviews; 3. Pre- and post-tests	Significant in overall score and scores of fluency and vocabulary

11

Appendix 2 Survey of learning gains in speaking through the Soeasy app

This research survey aims to find out more about your use of the Soeasy app. It should take about 5 minutes to complete. The survey is part of the research on the use of language learning apps. Your answers will be completely anonymous. You may withdraw from the survey at any time.

1. Your age: _____

2. Your gender: _____

3. Your major: _____

4. Your year at university: _____

5. How do you describe your level in English? _____ (beginner/elementary/intermediate/upper intermediate/advanced/nearly fluent/fluent)

6. Which device do you mostly use to access the Soeasy app? _____

7. How often do you use the Soeasy app during the Challenge?

8. How long in total do you spend using the Soeasy app?

9. Which task do you like most in the Soeasy app?

10. What do you think of the feedback you get about your performance on completing the tasks?

11. To what extent do you agree with the following statement: Using the Soeasy app has helped me improve my listening and speaking?

12. Please rate the overall quality of the Soeasy app. (Very good, Good, OK, Not so good, Very bad)

13. What other apps do you use to practice speaking English?

14. Please share any other thoughts you have about learning listening and speaking with the Soeasy app.

Appendix 3 Task type of 2024 Internet English Listening and Speaking Challenge on Soeasy

App

1. Following reading

1) Hello! Welcome to our online oral English contest!

2) Hi, Diana! Haven't seen you for ages! How're you doing?

2. Multiple choice (Listen and choose)

1) She didn't find out the secret code until she was told by her colleague.

A. She found out the code by herself.

B. She didn't know the code.

C She knew the code later.

3. Follow reading

1) Learning a foreign language is necessary for mental development in an ever-globalized world.

4. Retell

1) Peter's uncle lived in the country. Once Peter went to stay with him for a few weeks. Whenever they went for a walk and passed somebody in a car, his uncle waved. Peter was surprised, and said, "Uncle George, you know everyone here. Where did you meet them all?"

2) "I don't know all these people," said his uncle.
"Then why do you wave to them?" asked Peter.
"Well, Peter," answered his uncle. "When I wave to someone, he is pleased if he knows me. He continues his journey with a happy heart.

3) But when I wave to someone, and he doesn't know me, he will be surprised and say to himself, 'who is that man? Why did he wave to me?' So he has something to think about during the rest of his journey, and that makes his journey seem shorter. So I make everybody happy."

13

大学生英语写作反馈素养指标体系构建

李思莹，冀娇娇，任佩玲

北京邮电大学

摘要：反馈素养是影响学生学业成绩的关键因素。本研究在明确英语写作反馈素养的概念并分析其研究现状的基础上，探讨了大学英语写作反馈素养指标体系的组成及其具体内涵。该指标体系以学习者在不同情境中的反馈行为为基本内容，采用德尔菲法进行专家评议，旨在为科学评估大学生写作反馈素养水平提供量化标准。

关键词：反馈素养；反馈知识；英语写作；指标体系构建

一、引言

作为对学习和成就最具影响力的因素之一，反馈被概念化为"由代理人(如老师、同伴、书籍、自我、经验、父母)提供的关于个人表现或理解方面的信息" (Hattie & Timperley 2007)。反馈持续收到教育学及外语教学等领域研究者的关注。随着对反馈研究范式的深入，研究者逐渐将研究视角从学生的"被动接受" 转化为"主动寻求及利用"，更加关注学生对反馈内容的理解和应用。为此有学者提出了学生反馈素养概念，Carless & Boud (2018) 将学生反馈素养定义为"理解信息并使用信息来增强工作或者学习策略所需的知识，能力和倾向"。强调学生认识到其对学习的重要性，并在反馈过程中积极参与。

国内外学者围绕英语写作反馈素养的测量工具进行了广泛研究，如董哲和高瑛（2023）编制的《英语写作同伴互评反馈素养量表》，Yu et al.(2022) 编制的《二语学生写作反馈素养量表》，以及解冰和高瑛（2022）编制的《英语写作在线同伴反馈学习投入量表》。然而，前人研究大多聚焦于同伴互评情境，对于当前兴起的AI 工具的融入鲜有涉及。因此，本研究尝试初步探索编制一个集中于反馈知识、涵盖多种反馈情境的大学生英语写作反馈素养指标体系，以期为今后的反馈素养教学和研究提供借鉴。

二、研究设计

(一) 大学生英语写作反馈素养指标体系构建流程

本研究分为三个阶段进行。第一阶段旨在收集国内外大学生英语学习者反馈素养的指标描述语。在这一阶段，通过梳理国内外相关文献，筛选出以英语写作反馈和反馈素养为主题的文章，并提取其中与反馈或反馈素养相关的描述语。本阶段共收集了 38 篇国内外文献，提取描述语 322 条。

其次，将描述语按维度进行分类和标注，具体划分维度包括学习者个体、反馈环境、反馈行为及反馈阶段四个方面。本研究重点关注学习者个体，即学习者对反馈知识的掌握程度。反馈环境包括教师反馈、同伴反馈、自我反馈及智能环境反馈，我们重点关注学生这一主体在不同反馈环境中所受到的影响及其表现。反馈行为包括提供反馈、寻求反馈、辨别反馈和利用反馈四方面。反馈阶段涵盖反馈行为发生的前中后三个阶段，因为在不同阶段，学生的行为会有所不同。

最后提取和归纳大学生英语写作反馈素养一级和二级指标，并参考已有的反馈素养框架，进行完善和补充，最终形成大学生英语写作反馈素养指标体系。

（二）大学生英语写作反馈素养指标体系主要内容

通过文献研究和对指标描述的归纳，本文开发了一套针对大学生英语写作反馈素养的指标体系。该体系包括五个方面：写作知识、语言知识、评价知识、反馈来源及 AI 协助写作反馈。

本研究重点关注学生反馈素养中对于反馈知识的理解和应用。反馈知识作为反馈素养的重要组成部分，源自 Carless & Boud（2018）对学生反馈素养的定义，意指学生对信息的理解。在反馈情境中，反馈知识可以概括为学习者对反馈内容的理解和认知。本研究的重点是探讨学生需要掌握哪些知识，才能顺利参与反馈活动并从中受益。通过国内外相关文献的手机和整理，经过两轮筛选，共筛选出五个一级指标。

1. 写作知识

写作是反馈任务的核心，是学生在二语课堂中学习的主要内容。因此，学生必须具备相关的写作知识，才能顺利开展反馈活动的第一步。在这一一级指标下，包括了体裁知识、文章结构、写作技巧、写作规范及写作过程五个二级指标。学生需在写作任务中结合课堂上学到的二语写作课程内容，按照写作任务的要求完成写作。写作知识是这一过程的基础。

2. 语言知识

二语写作中，对于英语语言的运用也是学生反馈活动的重要组成部分。这其中不仅包括英语写作中的使用到的语言知识也包括在提供或者接受外界反馈时所需的语言知识。具体来讲，语言知识可以包括词汇、语法、句法和技术操作四个方面。

3. 评价知识

在反馈活动中，学生参与反馈以及对反馈的理解程度决定着学生的反馈素养水平。比如在同伴反馈情境下，学生就需要了解同伴反馈的评价标准，反馈活动的反馈程序设计等知识。因此，涵盖多种反馈情境的该一级指标评价知识共有评价标准、反馈程序、评价技术知识三方面。

4. 反馈来源

反馈的来源和形式多种多样，因此，为了有效进行反馈活动，了解不同来源反馈的特点至关重要。

5. AI 协助写作反馈

随着科技的发展，各种 AI 工具逐渐融入学生的日常学习生活。一些研究开始关注 AI 在二语和教育领域的应用（陈茉、吕明臣，2024；魏爽、李璐遥，2023）。在科技飞速发展的浪潮下，了解各种 AI 工具的特点以及如何将其有效融入写作反馈中，已成为大学生必备的知识之一。因此，本研究将 AI 协助写作反馈作为一级指标，并将其分为四个二级指标：学习活动支持、交互反馈功能、语言编辑功能和语料库功能。

以上这五个要素并不是简单的相加，它们之间是相互联系、相互促进的。学生反馈素养的反馈知识是一个完整的结构，其中语言知识和写作知识是基础；评价知识和反馈来源是核心；AI 协助写作反馈是现代写作中不可或缺的支持工具。各要素相辅相成，共同提升学生的写作能力和反馈素养。

（三）大学生英语写作反馈素养指标体系

经过文献筛选、划分反馈描述语维度及标注，并提取描述语的共同点以确定一级和二级指标，最终构建了大学生反馈素养中的反馈知识指标体系。在第一

轮构建完成后，研究者邀请反馈素养领域的专家教师进行了第二轮审阅，最终完成了下表所示的指标体系。

一级指标	二级指标	指标描述
写作知识	体裁知识	理解不同体裁的写作特点和要求
	写作技巧	掌握一定的写作方法和技巧
	文章结构	熟悉文章的结构布局和段落组织
	写作规范	遵循语法、标点和格式规范
	写作过程	了解写作的计划、草稿和修订过程
语言知识	词汇	有一定的词汇量并正确使用词汇
	语法	正确应用语法规则
	句法	了解构造连贯和复杂的句子
	技术操作	注意拼写检查和语法检查
评价知识	评价标准	理解评分标准和评价要求
	反馈程序	了解反馈活动的步骤和流程
	技术评价知识	熟悉评估软件和工具的使用
反馈来源	识别反馈来源	了解不同反馈来源的特点
AI 协助写作反馈	学习活动支持	用 AI 工具辅助学习和写作活动
	交互反馈功能	使用 AI 提供的即时互动反馈
	语言编辑功能	借助 AI 进行语法和文本编辑
	语料库	使用 AI 语料库资源丰富写作内容

三、结语

反馈是促进学生学习的重要教学策略，但由于个人因素的影响，学生往往难以充分理解反馈的含义 (Hill & West, 2020)。研究者将这些个体因素称为学生反馈素养 (Han & Xu, 2019)。本研究针对反馈素养中的关键环节——反馈知识，进行了深入探讨。通过结合国内外在写作反馈教学及写作反馈素养研究领域的重点文献，我们提取并筛选出体现写作反馈知识的描述语，最终构建了大学生英语写作反馈素养指标体系。然而，需要指出的是，本研究的量表尚未经过严格的验证。未来的研究将进一步完善这一写作反馈指标体系，并对其适用性进行量表验证，以提高反馈素养测量工具的科学性和实用性。通过不断的研究和改进，我们希望能为教育实践提供更有效的指导和支持，促进学生在英语写作中的全面发展。

参考文献

[1] Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

[2] Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. doi:10.3102/003465430298487.

[3] Han, Y., & Xu, Y. (2019a). Student feedback literacy and engagement with feedback: a case study of Chinese undergraduate students. Teaching in Higher Education, 26(2), 181–196. https://doi.org/10.1080/13562517.2019.1648410

[4] Hill, J., & West, H. (2019). Improving the student learning experience through dialogic feed-forward assessment. *Assessment & Evaluation in Higher Education*, 45(1), 82–97. https://doi.org/10.1080/02602938.2019.1608908

[5] Yu, S., Di Zhang, E., & Liu, C. (2022). Assessing L2 student writing feedback literacy: A scale development and validation study. *Assessing Writing*, 53, 100643. https://doi.org/10.1016/j.asw.2022.100643

[6] 陈茉,吕明臣.ChatGPT 环境下的大学英语写作教学[J].当代外语研究,2024,(01):161-168.

[7] 董哲,高瑛.英语写作同伴反馈素养量表的编制与应用[J].解放军外国语学院学报,2023,46(03):18-25.

[8] 魏爽,李璐遥.人工智能辅助二语写作反馈研究——以 ChatGPT 为例[J].中国外语,2023,20(03):33-40.DOI: 10.13564/j.cnki.issn.1672-9382.2023.03.007.

[9] 解冰,高瑛.英语写作在线同伴反馈学习投入评价量表编制与分析[J].外语电化教学,2022(05):52-58+112.

AI 反馈和 AI+教师混合反馈对中国英语学习者写作质量的影响研究

罗美娜 [1] 胡心怡 [2]

（1. 浙江师范大学外国语学院，浙江金华，321004；

2. 伦敦大学学院，英国伦敦，WC1E 6BT）

摘要：人工智能（AI）为 EFL 写作反馈提供了新的途径。虽然已有研究比较了 AI 与教师反馈的效果，但缺乏对二者合作的探索。本研究比较了 AI 单独反馈与 AI+教师混合反馈对学生记叙文和议论文质量的影响，旨在探索两种模式的优势和分工机制。浙江 12 名十年级学生参与了本研究。他们撰写并根据 DeepL Write 的 AI 反馈和教师反馈修改了文本。通过 Coh-Metrix 量化评估写作质量后，研究发现 AI 在词汇和语法方面表现出色，但在处理复杂文体和连贯性问题时存在局限，需教师介入。基于此，本文提出一种教师先处理全局问题，AI 再优化细节的反馈模式，并建议教师在使用 AI 辅助写作教学时考虑不同文体的要求。

关键词：AI 反馈；教师+AI 混合反馈；文体；AI 写作工具

1. 引言

在 EFL 写作教学中，反馈是帮助学生缩小当前与期望的写作水平差距的重要途径（Duijnhouwer, 2012）。因此，为学生写作提供高质量的反馈非常重要。但教师提供高质量反馈需要耗费大量时间和专业技能，这促使人们对 AI 写作工具产生浓厚兴趣。这些工具能够提供即时、反复且个性化的反馈，对传统教师的反馈形成有效补充（Link et al., 2022）。

尽管 AI 反馈在检测拼写和语法错误方面表现出很高的准确性（Fitria, 2021），并可能促进学习自主性（Nazari et al., 2021），但其建议的修改方法有时可能存在错误（Park, 2019）。此外，AI 反馈还被认为过于侧重语言形式，忽视了写作的交际目的（Alharbi, 2023）。研究者大多认同 AI 反馈应作为教师反馈的补充（Almusaed et al., 2023）。尽管一些研究指出 AI+教师混合反馈有助于提升学习自主性（Lu, 2019），但其对不同文体的写作任务的影响缺乏实证研究（Ajabshir & Ebadi, 2023b）。因此，本研究探究 AI 单独反馈与 AI+教师混合反馈在议论文和记叙文写作中的影响，以及教师和 AI 工具在合作中的具体角色，为优化 AI 在 EFL 写作教学中的应用提供实证依据。

2. 研究方法

采用定量方法评估 AI 单独反馈与 AI+教师混合反馈对学生议论文和记叙文写作质量的

影响。研究分析了两种文体类型的三个版本：初稿（Draft A1，Draft N1）、仅用 AI 反馈修改的二稿（Draft A2，Draft N2）和用 AI+教师混合反馈的三稿（Draft A3，Draft N3）。

2.1 研究对象

本研究在浙江省一所公立高中进行，12 名十年级学生自愿参与。已从学生、他们的英语老师及学校行政部门获得知情同意。所有参与者均通过高中入学考试，表明他们具有中等英语水平。前期测试表明，参与者的英语写作水平大致相当。

2.2 工具

本研究利用 DeepL Write 生成 AI 反馈，并通过 Coh-Metrix 3.0 量化写作质量，以解决传统评分的偏见和效率问题。DeepL Write 使用自然语言处理技术优化词汇和语法，提供改写建议。Coh-Metrix 是先进的计算文本评估工具，分析维度覆盖文体、连贯性、词汇多样性及语法范围等领域。

2.3 研究流程

共 12 名学生参与本研究，持续时间 11 周。在前五周，学生每周两次参加关于 DeepL Write 使用及写作技巧的培训。随后，学生在监督下撰写议论文和记叙文初稿，并先后使用 DeepL Write 和教师反馈进行两轮修改。共产生三个版本的 72 篇文稿，然后使用 Coh-Metrix 3.0 的 14 个指标评估写作质量，最后通过 SPSS 27 统计结果。

3. 结果

通过双因素方差分析，研究确定了反馈类型、文体及其交互作用对写作质量的主要影响。结果显示，写作类型和反馈类型的交互效应在文体要素 PCNARp (F=8.074，p =.001)、语法多样性（SYNSTRUTt (F=10.629，p<.001); SYNSTRUTa (F=8.115，p =.001)），和词汇多样性 WRDFRQa (F=6.437，p =.003)，精确度 WRDHYPnv (F=14.423，p <.001)方面显著。

文体主效应在内容选择(PCNARp (F=43.426, p<.001))、情景模型(SMCAUSvp (F=4.597, p=.036)); SMTEMP (F=5.545; p=.022))、因果、逻辑和时间连词使用(CNCCaus (F=4.225, p=.044); CNCLogic (F=39.912, p<.001); CNCTemp (F=12.508, p=.001))、局部指代连贯性 (CRFANP1 (F=12.661, p=.001)、潜在语义连贯性（LSASSP (F=21.132, p <.001); LSAPP1 (F=68.353, p <.001)）、句间句型多样性(SYNSTRUTa (F=4.428, p=.039))及词汇精确度 (WRDHYPnv (F=149.713, p<.001))方面显著。

反馈方式的主效应表现在句法连贯性（CRFANP1 (F=3.363, p =.041)）、潜在语义连贯性（LSASSP (F=3.990, p =.023); LSAPP1 (F=4.146, p =.020)）、语法多样性（SYNSTRUTa (F=9.759，p=0.000); SYNSTRUTt (F=12.644，p <.001)）、词汇多样性（WRDFRQa (F=4.857，

p =.011)）和精确度（WRDHYPnv (F=9.505, p <.001)）。

采用 LSD 事后检验比较两种文体中两种反馈方式的效果。议论文在 AI 单独反馈后叙事性略有提升（β (Draft A2-Draft A1) = 2.130, p = .723），但随后的教师反馈显著降低了该指标（β (Draft A3-Draft A2) = -17.413, p = .006）。记叙文在 AI 反馈后叙事性显著提高（β (Draft N2-Draft N1) = 9.836, p = .020），但在 AI+教师混合反馈后增长减弱（β (Draft N3-Draft N2) = 2.856, p = .485）。

议论文在 AI 单独反馈后语法多样性 SYNSTRUTa 分数降低（β (Draft A2-Draft A1) = -.044, p = .005），教师反馈后变化不大（β (Draft A3-Draft A2) = .002, p = .883）。记叙文中，AI 反馈导致 SYNSTRUTa 显著下降（β (Draft N2-Draft N1) = -.053, p = .015），但 AI+教师反馈后该指标显著上升（β (Draft N3-Draft N2) = .094, p < .001）。句间句型多样性 SYNSTRUTt 在议论文中 AI 反馈后下降（β (Draft A2-Draft A1) = -.046, p < .001），AI+教师反馈后无显著变化（β (Draft A3-Draft A2) = -.005, p = .696）。记叙文的句型多样性在两次 AI+教师反馈后持续下降（β (Draft N2-Draft N1) = -.032, p = .013; β (Draft N3-Draft N2) = -.065, p < .001）。

议论文中词汇的重复频率在 AI 单独反馈和后续教师反馈后连续下降（β (Draft A2-Draft A1) = -.101, p = .011； β (Draft A3-Draft A2) = -.119, p = .003）。记叙文中词汇的重复频率在 AI 单独反馈后减少（β (Draft N2-Draft N1) = -.368, p = .019），在教师反馈后增加（β (Draft N3-Draft N2) = .430, p = .007）。议论文词汇特异性在 AI 单独反馈后提高（WRDHYPnv (β (Draft A2-Draft A1) = .151, p =.042)），教师反馈后进一步提高（WRDHYPnv score (β (Draft A3-Draft A2) = .176, p =.019)）。记叙文中词汇在 AI 反馈后更加具象（WRDHYPnv (β (Draft N2- Draft N1) =.216, p <.001)），但教师反馈后变得笼统抽象（WRDHYPnv score (β (Draft N3-Draft N2) = -.262, p <.001)）。

4. 讨论

记叙文和议论文在风格和连续性上的区别主要源于它们的不同交际目的。记叙文按时间顺序叙述事件，展现出明显的叙事风格(McNamara et al., 2014)。悬念和省略等常用叙事技巧在特定情节中策略性地扭曲和隐瞒信息，激发读者的期待和好奇心，这样的叙事方式可能会牺牲局部句法连续性，从而增强叙事效果，故本研究中参与学生的 PCNARp、SMTEMP、CNCTemp 得分更高。相反，议论文围绕特定主题通过逻辑顺序展开论证，强调清晰的逻辑和因果联系，需要频繁使用逻辑和因果连接词以及短距指代，并展现更强的潜在语义相似性，从而本研究中参与学生在 SMCAUSvp，SMCAUSvp、CNCCaus、CNCLogic、LSASSP 和 LSAPP1 分数上表现较高。

议论文和记叙文在词汇和语法多样性上的显著差异归因于学生对文体知识的掌握差异。参与学生在英语课程中记叙文写作训练多，因此学生能积累丰富的记叙文相关文体知识。此外，记叙文话题与学生生活贴近，他们因此有更多的背景知识可以运用。面对复杂、相对陌生的议论文，参与学生难以展现同等语言多样性。根据 Skehan 的有限注意力容量模型（1998），在处理复杂议论文时，学生可能倾向于简化语言以减轻认知负担并确保表意准确。相比之下，记叙文任务因认知要求较低，使得学生有更多认知资源处理和优化语言使用。

此外，AI 写作工具 DeepL Write 在区分不同文体元素方面表现不佳，导致 AI 单独反馈的修订往往与既定文体不符，因此教师在提供文体相关反馈和修改中的角色显得尤为重要。AI 工具依赖预设的语料库或样本进行文本分析，无法涵盖所有潜在写作错误，对写作风格等非语言元素的精确识别和调整还需深入理解特定文本的语境，当前 AI 在这方面仍不足（Grimes, 2010）。DeepL Write 的句子改写功能不改变原意和风格，但教师的建议更能有效减少风格不一致。由于 DeepL Write 只能逐句处理文本，无法进行整体调整，这与 Chen & Cui (2022)的研究相符，指出 AI 对文本整体连贯性的敏感度低。尽管教师反馈在提升局部句法连贯性上更有效，但受时间和认知需求限制，教师反馈提供的词汇、句型有限，且在针对连续性提供反馈时，教师多关注单句问题，说明提高全局连贯性可能需要多次反馈和修改。

结合研究结果，本文提出优化后的合作反馈机制。教师首先识别文体偏差和明显的连贯性问题，为节省时间和减轻负担，暂时不关注词汇语法表现。接下来的修改指导由 AI 写作工具主导，主要为词汇和句法改进建议。因为基于教师反馈的修改可能在用词、句型方面存在不足，而 AI 的建议可能会影响整体风格或连贯性，所以通过多次轮换使用两种反馈能够让两者有效互补，提升整体写作质量，并通过迭代反馈过程深化学生对写作技巧的理解和应用。

参考文献

Ajabshir, Z. F., & Ebadi, S. (2023). The effects of automatic writing evaluation and teacher-focused feedback on CALF measures and overall quality of L2 writing across different genres. *Asian-Pacific Journal of Second and Foreign Language Education*, *8*(1). https://doi.org/10.1186/s40862-023-00201-9

Alharbi, W. (2023). AI in the Foreign Language Classroom: A Pedagogical Overview of Automated Writing Assistance Tools. *Education Research International*, *2023*, 1–15. https://doi.org/10.1155/2023/4253331

Almusaed, A., Almssad, A., Yitmen, I., & Homod, R. Z. (2023). Enhancing Student Engagement: Harnessing "AIED"'s Power in Hybrid Education—A Review analysis. *Education Sciences*, *13*(7), 632. https://doi.org/10.3390/educsci13070632

Chen, M., & Cui, Y. (2022). The effects of AWE and peer feedback on cohesion and coherence in continuation writing. *Journal of Second Language Writing*, *57*, 100915. https://doi.org/10.1016/j.jslw.2022.100915

Duijnhouwer, H., Prins, F. J., & Stokking, K. M. (2012). Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance. *Learning and Instruction*, *22*(3), 171–184. https://doi.org/10.1016/j.learninstruc.2011.10.003

Fitria, T. N. (2021). Grammarly as AI-powered English Writing Assistant: Students' alternative for Writing English. *Metathesis: Journal of English Language, Literature, and Teaching*, *5*(1), 65. https://doi.org/10.31002/metathesis.v5i1.3519

Grimes, D., & Warschauer, M. (2010). Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation. *The Journal of Technology, Learning and Assessment*, *8*(6). http://files.eric.ed.gov/fulltext/EJ882522.pdf

Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605–634. https://doi.org/10.1080/09588221.2020.1743323

Lu, X. (2019). An empirical study on the Artificial Intelligence Writing Evaluation System in China CET. *Big Data*, *7*(2), 121–129. https://doi.org/10.1089/big.2018.0151

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press.

Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of Artificial Intelligence powered digital writing assistant in higher education: randomized controlled trial. *Heliyon*, *7*(5), e07014. https://doi.org/10.1016/j.heliyon.2021.e07014

Park, J. (2019). An AI-based English Grammar Checker vs. Human Raters in Evaluating EFL Learners' Writing. *Multimedia-assisted Language Learning/Multimedia Assisted Language Learning*, *22*(1), 112–131.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

The Effects of Online Peer Feedback and Teacher Feedback on Student Feedback Literacy: A Perspective of Taking Actions

Peiling Ren[1*], Jiaojiao Ji[2], Siying Li[3]

[1 2] School of Humanities, Beijing University of Posts and Telecommunications

[*] panacea@bupt.edu.cn

Abstract: With the growing emphasis on student feedback literacy in educational settings, understanding how different feedback sources influence it is crucial. This study aims to compare the effects of online peer feedback and teacher feedback on student feedback literacy from the perspective of taking actions. Forty-four EFL students enrolled in an English argumentative writing course participated in the study over a 16-week semester. They were randomly divided into two groups: an online peer feedback group, where students received feedback from their peers, and a teacher feedback group, where students received feedback from the teacher. Students wrote six argumentative essays on various topics. For each writing task, both groups revised their first drafts based on their received feedback and submitted the final drafts. The collection of peer feedback, teacher feedback, first drafts and final drafts were collected for analysis. The collected peer and teacher feedback were analyzed and compared to determine any significant differences. Additionally, the initial and final drafts were compared to assess the impact of each feedback method on students' literacy to take actions effectively. The findings highlight the effectiveness of both peer and teacher feedback in enhancing students' feedback literacy.

Key words: Online Peer Feedback; Teacher Feedback; Student Feedback Literacy; Taking actions

1. Introduction

Current research in university English pedagogy highlights that teacher feedback is often focused on error correction within classroom settings (Lee, 2008). However, with a growing number of students and limited teacher availability, providing comprehensive and timely feedback becomes increasingly difficult. In traditional teacher-centered classrooms, feedback primarily comes from instructors, leading students to passively receive and implement corrections. To address these challenges, peer feedback has emerged as a practical solution. By incorporating peer feedback, students become more actively engaged in the feedback process, both as givers and receivers. This approach not only eases the teacher's workload but also promotes a more interactive and

collaborative learning environment, helping students develop their feedback literacy and take a more active role in their own learning.

2. Literature review

2.1 Student feedback literacy

The concept of feedback literacy was initially proposed by Sutton (2012), who defined it as the ability to read, interpret, and understand written feedback. Carless and Boud (2018) later redefined student feedback literacy as "the understanding, capability, and inclination needed to comprehend and use information to improve work or study strategies."

The theoretical framework of student feedback literacy has been explored through various lenses, leading to the development of models that describe how students engage with feedback. For instance, Carless and Boud (2018) proposed a model where students with feedback literacy proactively appreciate feedback, manage their emotional responses, and make consistent and informed judgments about their work. This model underscores the active role students play in the feedback process and their ability to integrate feedback into their learning practices. Molloy et al. (2020) further refined the concept by defining student feedback literacy as the capability to understand, utilize, and benefit from feedback. In a more recent development, Yu et al. (2022) introduced a L2 student writing feedback literacy scale, which encompasses five dimensions: appreciating feedback, recognizing different feedback sources, making judgments, managing emotions, and taking action. Among these dimensions, "taking action" refers to the processes by which students store, organize, and generalize feedback information, as well as modify their learning goals when necessary. This dimension reflects the practical application of feedback literacy, emphasizing how students translate feedback into concrete changes and improvements in their work.

2.2 Online peer feedback and Teacher feedback

Peer feedback, known as peer assessment, or peer review, this commonly refers to students providing oral or written feedback on their peers' written work in pairs or small groups (Yu & Lee, 2016). Online peer feedback involves students providing and receiving feedback from their peers in a digital environment. This method has been shown to enhance students' critical thinking, collaborative skills, and engagement with the writing process. Recent studies highlight the positive impact of online peer feedback on students' writing performance and feedback literacy. For instance, Zhang and Mao (2023) discovered that students enhanced their feedback literacy in multiple areas, including seeking feedback, making judgments, taking action, acknowledging feedback, and managing emotions, through activities including peer feedback and teacher feedback . Similarly, Zhang et al.

(2023) observed that student feedback literacy was improved in writing through active engagement, such as making revisions, during peer review sessions.

There is limited research specifically comparing the effects of teacher feedback versus online peer feedback on students' feedback literacy. The relative impact of these two methods on students' ability to utilize feedback effectively remains insufficiently examined. This study aims to address this gap by applying Yu (2022)'s framework, particularly focusing on the "taking actions" perspective. This framework will guide the investigation into how different feedback methods influence students' uptake of feedback and their subsequent revision behaviors. The research questions are as follows:

1) What are the changes of students' uptake to received feedback?

2) How did the two types of feedback affect students' uptake of feedback?

3. Research Design

3.1. Research Context and Participants

Forty-four EFL students enrolled in an English argumentative writing course participated in the study over a 16-week semester. They were randomly divided into online peer feedback group and teacher feedback group. Throughout the semester, students wrote six argumentative essays on various topics. For each writing task, peer feedback group submitted the first draft of their essays to Peerceptiv, an anonymous peer review system, for peer review. Students made revisions based on the peer comments they received and submitted their final drafts. However, the teacher feedback group just revised their drafts based on the teacher feedback they received. Peer feedback, teacher feedback, first drafts and final drafts were collected for analysis.

3.2. Data Collection and Analysis

Data for this study were selected from two writing tasks, Task 1 and Task 5, which included both the initial and final drafts of students' essays along with all feedback provided during these tasks. The peer feedback group comprised comments from peer reviews, while the teacher feedback group included feedback from the instructor.

To ensure a detailed analysis, the feedback received by students was first segmented to ensure that each piece of feedback addressed only one specific issue. This segmentation allowed for a precise examination of how individual pieces of feedback were addressed in the revisions. Next, the types of feedback were classified into categories such as praise, summary, and criticism. This classification helped in understanding the nature of the feedback provided and its potential impact on student revisions. Subsequently, a comparison was made between the students' initial drafts and their revised final drafts to determine whether and how the criticism was utilized in

making revisions. This involved a detailed analysis of the revisions to identify which aspects of the critical feedback were incorporated and how they influenced the changes made. Additionally, the effectiveness of peer feedback and teacher feedback was evaluated by analyzing the revisions resulting from each type of feedback. This assessment provided valuable insights into the ways in which different feedback methods impacted students' feedback literacy and their ability to effectively incorporate feedback into their writing.

4. Results and Discussion

4.1 The changes of students' uptake to received feedback

In the peer feedback group, students' average uptake of comments in Task 1 is 9.88 and 10.67 comments in Task 5 (see Table 1). The paired sample t-test results show a t-value of -0.580, indicating that the change in the number of comments taken up from Task 1 to Task 5 is not statistically significant. The stable uptake rate of comments suggests students' consistent engagement with peer feedback throughout the semester.

In the teacher feedback group, students took up an average of 7.80 comments in Task 1 and 5.75 comments in Task 5. The paired sample t-test results show a t-value of 2.877, which is statistically significant ($p < 0.01$). The result indicates a significant decrease in the number of comments taken up from Task 1 to Task 5. Possible explanations could be that students became less dependent on teacher feedback as the semester progressed, or their writing skills improved, reducing the need to take up as many comments.

These findings highlight the potential of peer feedback to sustain students' revision efforts over time, while teacher feedback may be more useful in the early stages of writing development.

Table 1 Feedback Adoption by Students in Two Writing Tasks

	Peer feedback			Teacher feedback		
	Task1	Task 5		Task1	Task 5	
	Mean	Mean	t	Mean	Mean	t
Feedback uptake	9.88	10.67	-0.580	7.80	5.75	2.877**

Note: *p<0.05; **p<0.01

4.2 The effects of online peer feedback and teacher feedback on taking actions

For Task 1, students in the peer feedback group took up an average of 9.88 comments, while students in the teacher feedback group took up an average of 7.80 comments (see Table 2). The t-value for this comparison is 1.582, indicating that the difference in the number of comments uptaken between the two groups is not statistically significant. This suggests that at the beginning of the semester, there was no significant difference in

the uptake of feedback between peer and teacher feedback groups. Therefore, students in both groups were at a relatively same level of literacy in taking actions at the beginning of the semester.

When it comes to the end of the semester, some changes had been made. The peer feedback group took up an average of 10.67 comments, whereas the teacher feedback group took up an average of 5.75 comments in Task 5. The t-value for this comparison is 3.627, which is statistically significant ($p < 0.01$). This indicates a significant difference in the number of comments uptaken between the peer and teacher feedback groups by the end of the semester. Students in the peer feedback group took up significantly more comments compared to those in the teacher feedback group.

These results suggest that while both feedback types are initially effective, peer feedback becomes more influential in fostering student revisions as the semester progresses. This is different from Ruegg (2015)'s finding, which held that teacher feedback was more often uptaken.

Table 2 Comparison of Feedback Adoption by Students in Two Writing Tasks

	Task1			Task 5		
	Peer feedback	Teacher feedback		Peer feedback	Teacher feedback	
	Mean	Mean	t	Mean	Mean	t
Feedback uptake	9.88	7.80	1.582	10.67	5.75	3.627**

Note: *p<0.05; **p<0.01

5. Conclusion

This study examined the the effects of online peer feedback and teacher feedback on their revision actions in an EFL writing course. Main findings have been drawn that peer feedback is increasingly effective in promoting substantial revisions while teacher feedback is more effective at the early stages. In summary, EFL teachers could integrate both peer and teacher feedback in writing instruction, with an emphasis on peer feedback's sustained influence on student revisions. However, there are some limitations. The lack of qualitative analysis means the reasons behind these patterns remain unexplored. Future research should investigate the qualitative aspects to provide deeper insights into students' feedback adoption and revision behaviors. Additionally, taking action is only one aspect of feedback literacy, and future research will explore more dimensions.

References

[1] Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.

[2] Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*(2), 69–85.

[3] Molloy, E., Boud, D., & Henderson, M. (2019). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education, 4*5(4), 527–540.

[4] Ruegg, R. (2015). Differences in the uptake of peer and teacher feedback. *RELC Journal, 46*(2), 131–145.

[5] Sutton, P. (2012). Conceptualizing feedback literacy: knowing, being, and acting. *Innovations in Education and Teaching International, 49*(1), 31–40.

[6] Yu, S., Di Zhang, E., & Liu, C. (2022). Assessing L2 student writing feedback literacy: A scale development and validation study. *Assessing Writing, 53*, 100643.

[7] Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching, 49*(4), 461–493.

[8] Zhang, T., & Mao, Z. (2023). Exploring the development of student feedback literacy in the second language writing classroom. *Assessing Writing, 55*, 100697. https://doi.org/10.1016/j.asw.2023.100697

[9] Zhang, F., Min, H., He, P., Chen, S., & Ren, S. (2023). Understanding EFL students' feedback literacy development in academic writing: A longitudinal case study. *Assessing Writing, 58*, 100770.

基于《中国英语能力等级量表》的 ChatGPT 二语阅读理解能力评估

杨娜 武汉大学

摘　要：以 ChatGPT 为代表的人工智能技术在二语教学应用中正广泛展开，然而其能力水平如何尚缺少深入评估。本文基于《中国英语能力等级量表》，以中国高考英语试题对 ChatGPT 的阅读理解能力水平进行评估。研究发现，ChatGPT 各方面的阅读理解能力水平表现为：(1)对叙述文和指示文的理解水平低于对说明文和论述文的理解水平；(2)话题类别的理解水平由低到高为： 日常生活 ＜ 社会生活 ＜ 人 ＜ 物；(3)在认知能力上，概括/分析能力低于识别/提取能力；(4)内容结构方面，隐性信息理解能力低于显性信息理解能力；语言结构方面，篇章结构理解最差，词义理解最佳，句子结构理解水平居中。值得注意，ChatGPT 的阅读理解能力水平总体上可达到某些高等级水平，但在某些低等级能力水平上表现却可能欠佳，这与人类阅读理解能力水平一般由高到低的表现不同。ChatGPT 的二语阅读理解能力表现与相关文本特征与机器理解机制等有关。

关键词：ChatGPT；中国英语能力等级量表；阅读理解能力

1 引言

聊天机器人 ChatGPT 的强大理解生成能力吸引了全球关注。各个领域对 ChatGPT 积极反应，二语教育领域也多对 ChatGPT 采取积极拥抱的态度，探索其在二语教学中的应用(秦颖 2023；魏爽、李璐遥 2023)。然而，ChatGPT 的二语能力究竟能达到何种水平？在二语教学应用中在何种程度上可信赖它？二语教学领域尚未见专门评估。为此，本文评估 ChatGPT 的二语能力水平，以供二语教学应用参考。

二语能力包括语言理解能力与语言表达能力；其中，语言理解能力相比更基础，又包括阅读理解能力与听力理解能力，分别是面向书面表达与口头表达的理解能力（教育部、国家语言文字工作委员会 2018）。根据 ChatGPT 的当前能力基础与二语教学应用实际，并结合英语教育在中国二语教育中的主导地位，本文评估 ChatGPT 的英语二语阅读理解能力。一般认为，阅读理解不仅是单一技能的展现，而是涉及多种能力的综合运用，包括识别能力、理解能力、推断能力等(Magliano *et al.* 2007；王淑花 2012；曾用强 2017)。考虑到评测体系的公信度，本文以《中国英语能力等级量表》(以下简称《量表》)(阅读理解分量表)（教育部、国家语言文字工作委员会 2018；曾用强 2017）为评测体系多维度评估 ChatGPT 等能力表现与水平等级。考虑到评估材料的可靠性，本文所使用评测试题为中国高考英语试题的阅读理解部分。

2 研究背景及相关工作

ChatGPT 问世以来，二语教学领域迅速响应，主要从理论或实践上进行了有关探索。理论上，研究者提出并展望了 ChatGPT 在二语教学中一系列的应用价值，包括但不限于：为学习者创造个性化学习环境和体验；为

教师提供教学材料；节约教育资源，降低教育成本等(Kohnke et al., 2023；冯志伟、张灯柯, 2023；秦颖, 2023；张震宇、洪化清, 2023；胡加圣、戚亚娟, 2023；周忠良, 2023；郭茜等, 2023；杨连瑞, 2024)。实践上，已有研究将 ChatGPT 应用于具体二语教学实践中，以二语阅读和二语写作为主。二语阅读方面，研究发现，参加大学英语课程的两个班的学生中，使用 ChatGPT 辅助的实验组在阅读英文原版小说时表现出更高的完成度和理解度(Wang & Feng, 2023)；ChatGPT 能够基于 CET-4 阅读材料和阅读子技能生成与人工编写的题目质量相当的阅读理解多选题(Lin & Chen, 2024)；ChatGPT 生成的二语阅读理解测试材料与人工材料在流畅性和表达自然度方面相似，但在选项吸引力和完成度上需改进(Shin & Lee, 2023)。二语写作方面，ChatGPT 主要被用来提供写作反馈。研究发现，ChatGPT 在二语写作反馈中能够同时提供内容与结构的整体反馈和语言层面的局部反馈(魏爽，李璐遥, 2023)；ChatGPT 能够作为一种自动化写作评估工具，在泰国英语教学中评分高效且具有一致性(Bucol & Sangkawong, 2024)。

以上关于 ChatGPT 的二语教学应用，虽有一些涉及应用评估，但总体上并不针对其二语能力水平。相比而言，其他领域有较多对 ChatGPT 能力水平的考察。如研究发现 ChatGPT 在医学领域如执业医师资格考试中可以达到通过水平(Gilson et al., 2023；Kung et al., 2023)；在物理、数学、法律、经济等领域的测试中的表现低于平均水平(Frieder et al., 2023；Kortemeyer, 2023；Newton & Xiromeriti, 2024)。

综上，对于 ChatGPT，二语教学领域已有相当的理论与实际应用探索，而很多领域也探讨了其应用能力水平，但鲜有针对其在二语能力水平的直接评估。这种情况可能与在技术原理上，ChatGPT 并不区分一般语言能力与二语能力有关，然而，我们有理由相信，一般(母语)语言能力与语言教学与二语能力与二语教学有着根本区别，而如果在二语教学中应用 ChatGPT，就有必要从其现实需求考虑 ChatGPT 的语言能力水平。有鉴于此，本文基于《量表》，使用中国高考英语试题测试评估 ChatGPT 的二语能力水平，根本上是将 ChatGPT 定位为一名英语二语学习者。我们认为，这种评估相对具体地考虑了 ChatGPT 的二语教学应用问题，有其现实意义。

3 研究设计

本文以《量表》(阅读理解分量表)为分析评测体系，以 ChatGPT 对中国高考英语阅读理解试题材料的响应结果为数据材料，具体研究以下问题： (1) ChatGPT 二语阅读理解能力在《量表》(阅读理解分量表)各维度的表现如何？(2) 其总体达到了何种能力水平等级？进一步，结合一般理解机制与机器理解机制，探讨相关原因。

3.1 《量表》

《量表》是国家语言文字规范文件(GF 0018－2018)，由国家教育部和国家语言文字工作委员会(2018)发布。其将语言能力分为 9 个等级，一般认定"达到某一能力等级的英语学习者和使用者具备该等级及以下所有等级所描述的能力，该等级不再重复描述其以下等级的能力"(第 5 页)。根据《量表》主要研制者的研究，阅读理解分量表主要从文本类型、话题类别、认知能力、内容结构等维度界定阅读能力，本文根据《量表》的阅读理解分表相关的描述语参数(曾用强 2017)确定测试题的各种参数类型。最终根据 ChatGPT 的测试正确/错误率，分

析其阅读理解能力水平。

3.2 数据采集与标注

选取 2021 年至 2023 年新高考全国 I 卷的英语阅读理解部分作为测试题(https://www.eol.cn/e_html/gk/gkst/)由 ChatGPT 作答(2023 年 11 月)。试题包括阅读理解和七选五两个题型，每套试卷包含 15 道阅读理解题，5 道七选五题，三套试卷共 60 题。一般将高考试题直接输入 ChatGPT 由其作答，但若试题涉及图表则将图表转化成文本形式。得到 ChatGPT 给出的答案之后，将其与正确答案比对，得到相关数据。

为深入分析，对试题材料按照《量表》阅读能力量表的参数类型(附表 2)进行标注。其中，文本类型分为六小类：叙述、描写、说明、论述、指示、交流。话题类别按照文本谈论的主题分为人/事/物/景、日常生活、社会生活、文化、科技等。认知能力包括识别/提取、概括/分析和批判/评价。限于试题材料皆为选择题，没有主观评价题，考察能力仅涉及识别/提取和概括/分析。内容结构包括显性信息、隐性信息和语言结构三类。根据主要特征，每个试题在各维度认定属于一种类别。

4 研究结果

(1) ChatGPT 的二语阅读理解能力表现

如表 1 所示，ChatGPT 二语阅读理解能力在各方面主要表现为：文本类型上，对说明文和论述文的理解能力高于叙述文和指示文；话题类别上，对"物"的理解能力高于对"人"、"社会生活"和"日常生活"的理解能力；在认知能力上，识别/提取能力高于概括/分析能力；在内容上，对显性信息的理解能力高于对隐性信息的理解能力；在语言结构上，对词义的理解能力高于对句子结构的理解能力，而对句子结构的理解能力高于对篇章结构的理解能力。

表 1　ChatGPT 的二语阅读理解能力

试题参数类型		错题数/涉及此参数类型的题数	错误率
文本类型	叙述	3/13	23.08%
	说明	2/24	8.33%
	论述	0/4	0.00%
	指示	5/19	26.32%
话题类别	人	2/17	11.76%
	物	0/4	0.00%
	日常生活	3/10	30.00%
	社会生活	5/29	17.24%
认知能力	识别/提取	0/6	0.00%
	概括/分析	10/54	18.52%
内容结构	内容	显性信息	0/6	0.00%
		隐性信息	10/54	18.52%
	语言结构	词义	0/3	0.00%
		句子结构	3/29	10.34%
		篇章结构	7/28	25.00%

(2) ChatGPT 的二语阅读理解能力水平等级

结合《量表》、试题难度（周碧，2020）与以上具体评估结果，大体认定，ChatGPT 基本可以达到二语阅读理解能力五级水平的要点。但也发现，其在较低等级能力水平上却不一定能完全达到，如其对相对简单的指示文的理解欠佳，无法满足三级能力水平的"能读懂简单的应用文，提取关键信息"的要求；再如，其对篇章结构的理解能力欠佳，未能很好满足四级标准中"能通过分析句子和篇章结构读懂语言较复杂的材料，理解意义之间的关系"的要求。这和《量表》所假设高等级能力水平一般可以覆盖低等级能力水平有所不同。

5 讨论

(1) ChatGPT 各维度表现结果差异原因分析

ChatGPT 在二语阅读理解各维度的表现有差异，结合一般理解机制与机器理解机制，相关原因主要在于：

第一，文本信息的直明与否和自洽度高低。ChatGPT 基于数据驱动，一般来说直接言明信息且自洽度高的文本更容易被理解，否则不易理解(Zhu et al. 2024)。表现在文本类型上，说明文旨在说明读者未解之事物及其情况、特征、本质等，论述文旨在阐述读者未解之理及其论据、论证，二者一般直接向读者言明未知之知识、事理等，文本自洽性高，理解难度小；对比之下，叙述文呈现事件发展，而事件背后蕴含的动机、意图、情感等，则多不直接言明，指示文的操作性理解往往需要读者结合相关现实场景等，由此两种文体的文本信息自洽度低，理解难度大。表现在话题类别上，关于"日常生活""社会生活"的信息因更易为人类读者所熟知而不必在文本中详细说明，由此其文本信息自洽度低，理解相较于其他话题对 ChatGPT 来说更困难。表现在认知能力

上，需要识别/提取的信息多是直接言明的，自洽度高，而概括性信息一般需从具体信息中抽象推理而来并未直接言明，由此后者更难理解。表现在内容结构(内容)上，显性信息直接呈现于文本，自洽度高，易于理解，隐性信息则相反。

第二，文本理解的全局与局部依赖差异。ChatGPT 理解的基础在语言概率模型，并且高层语言单位与结构的理解基于底层语言单位及结构的理解，如此越是高层语言单位与结构关系的理解对于 ChatGPT 越具有挑战性(Brown et al. 2020)。表现在内容结构(语言结构)上，从词汇到句子与语篇，其理解所依赖的语义范围越大，语义抽象表征的计算难度越大，理解越加困难。

(2) ChatGPT 二语阅读能力水平等级表现的原因与二语教学启示

本文基于《量表》等级分析发现，ChatGPT 具备五级的二语阅读理解能力水平，但在一些低等级能力水平上并不完全表现良好，这意味着其语言能力水平并不按《量表》语言能力水平逐级包含式提高，也意味着人工智能的语言能力发展与人的语言能力发展有所差异。一般认为，人的语言能力是逐级提高的。如生成语法(Generative Grammar)假设，语言习得过程是逐步发展的，从简单的语言结构到复杂的句法规则(Chomsky 1995)；二语习得理论的自然顺序假说(Natural Order Hypothesis)认为，语言学习者以一种可预测的顺序习得语言规则，从简单到复杂(Krashen 1981)。根本上人类语言能力的发展人类理解能力的发展受具身经验和生理神经机制等的限制，并以此为基础由低到高循序形成理解能力水平。相比较，人工智能语言能力的发展基础却不同，如 ChatGPT 语言理解能力发展主要受数据、算法、算力等计算机制的限制(冯志伟等 2023)。而这正是在二语教学中应用 ChatGPT 时所要警惕的，即对人来说难以理解的，对机器则未必；反过来，对人来说理解容易的，对机器也未必。总之，理解人类智能与 ChatGPT 等人工智能的工作机制差异，有利于更好的利用人工智能服务于包括二语教学在内的人类学习、工作、生活等；而反过来也有利于推动人工智能朝着更高级的人类智能发展。

6 结语

本文基于《量表》多维度评估了 ChatGPT 的二语(英语)阅读理解能力表现，评估显示 ChatGPT 基本达到了五级的二语阅读理解能力水平，但却可能不具备某种低层次能力。对此，在二语教学中应用 ChatGPT 等人工智能工具尤需谨慎。由于试题难度和试题类型的限制，ChatGPT 是否能够达到更高水平还有待更深入考察验证；ChatGPT 本文仅初步考察了 ChatGPT 的二语阅读理解能力，更全面的能力水平评估更有待全面考察；人工智能产品迭代迅速，更新更多的人工智能产品的评估还有待更进一步的探索。

<div align="center">

参 考 文 献

</div>

[1]   Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. & Agarwal, S. Language models are few-shot learners [R]. In *Proceedings of NeurIPS*. 2020.

[2]   Bucol, J. L. & Sangkawong, N. Exploring ChatGPT as a writing assessment tool [J]. *Innovations in Education and Teaching International*. 2024.

[3]   Chomsky, N. *The Minimalist Program* [M]. Cambridge: The MIT Press. 1995.

[4]   de Winter, J. C. Can ChatGPT pass high school exams on English language comprehension? [J]. *International Journal of Artificial*

Intelligence in Education. 2023.

[5] Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Lukasiewicz, T., Petersen, P. & Berner, J. Mathematical capabilities of chatgpt [R]. In *Proceedings of NeurIPS*. 2023.

[6] Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A. & Chartash, D. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment [J]. *JMIR Medical Education*, 9(1). 2023.

[7] Kohnke, L., Moorhouse, B. L. & Zou, D. ChatGPT for language teaching and learning [J]. *RELC Journal*, 54(2). 2023.

[8] Kortemeyer, G. Could an artificial-intelligence agent pass an introductory physics course? [J]. *Physical Review Physics Education Research*, 19(1). 2023.

[9] Krashen, S. D. *Second Language Acquisition and Second Language Learning* [M]. Oxford: Pergamon Press Inc. 1981.

[10] Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J. & Tseng, V. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models [J]. *PLOS Digit Health*, 2(2). 2023.

[11] Lin, Z. & Chen, H. Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items [J]. *System*, 123. 2024.

[12] Magliano, J. P., Millis, K., Ozuru, Y. & McNamara, D. S. A multidimensional framework to evaluate reading assessment tools [A]. In McNamara, D. S. (ed.). *Reading Comprehension Strategies: Theories, Interventions, and Technologies* [C]. New Jersey: Lawrence Erlbaum Associates. 2007.

[13] Newton, P. & Xiromeriti, M. ChatGPT performance on multiple choice question examinations in higher education. A pragmatic scoping review [J]. *Assessment & Evaluation in Higher Education*. 2024.

[14] Shin, D. & Lee, J. H. Can ChatGPT make reading comprehension testing items on par with human experts? [J]. *Language Learning & Technology*, 27(3). 2023.

[15] Wang, X. & Feng, Y. An experimental study of ChatGPT-assisted improvement of Chinese college students' English reading skills: A case study of *Dear Life* [R]. Presented at Proceedings of the 15th International Conference on Education Technology and Computers. New York: Association for Computing Machinery. 2023.

[16] Zhu, Y., Moniz, J. R., Bhargava, S., Lu, J., Piraviperumal, D., Li, S., Zhang, Y., Yu, H. & Tseng, B. Can large language models understand context? [OL] https://arxiv.org/abs/2402.00858 2024.

[17] 冯志伟,张灯柯.GPT 与语言研究[J].外语电化教学,2023,(02).

[18] 冯志伟,张灯柯,饶高琦.从图灵测试到 ChatGPT——人机对话的里程碑及启示[J].语言战略研究,2023,8(02).

[19] 郭茜,冯瑞玲,华远方.ChatGPT 在英语学术论文写作与教学中的应用及潜在问题[J].外语电化教学,2023,(02).

[20] 胡加圣, 戚亚娟. ChatGPT 时代的中国外语教育: 求变与应变[J]. 外语电化教学, 2023(01).

[21] 教育部, 国家语言文字工作委员会. 中国英语能力等级量表: GF0018—2018[S]. 北京: 高等教育出版社, 2018.

[22] 秦颖. 人机共生场景下的外语教学方法探索——以 ChatGPT 为例[J]. 外语电化教学, 2023(02).

[23] 王淑花. 中国学生英语理解能力量表的构建及验证研究[M]. 北京: 知识产权出版社, 2012.

[24] 魏爽, 李璐遥. 人工智能辅助二语写作反馈研究——以 ChatGPT 为例[J]. 中国外语, 2023, 20(03).

[25] 杨连瑞.ChatGPT 大语言模型背景下的二语习得[J/OL].现代外语,[2024-07-16].

[26] 曾用强. 中国英语能力等级量表的"阅读量表"制定原则和方法[J]. 外语界, 2017(05).

[27] 张震宇, 洪化清. ChatGPT 支持的外语教学: 赋能、问题与策略[J]. 外语界, 2023(02).

[28] 周碧.高考英语阅读（NMET）与中国英语能力等级量表（CSE）的匹配研究[D].广东外语外贸大学,2020.

[29] 周忠良. ChatGPT 在翻译教学中的应用: 变革、挑战与应对[J]. 北京第二外国语学院学报, 2023, 45(05).

人机混合反馈模式下的学生评语采纳研究

——以 Peerceptiv 和 ChatGPT 为例

朱茜 北京邮电大学

摘要：外语教育技术的创新对提升我国学生二语水平有着重要的作用。以 ChatGPT 为例的生成式人工智能模型的应用为二语写作教学提供了新兴的人机混合反馈模式。研究聚焦评语采纳环节，考察了学生在同伴反馈和 ChatGPT 反馈中的评语采纳情况。结果表明 ChatGPT 的评语采纳率高于同伴反馈的采纳率，且采纳的评语类型存在差异：在同伴反馈中学生更愿意采纳写作内容类评语，而在 ChatGPT 反馈中学生更倾向采纳语言形式类评语。研究进而探讨将同伴反馈与 ChatGPT 反馈有机结合，充分发挥人机混合反馈模式的优势，以期更好促进学生写作能力与智能化教学发展。

关键词：二语写作 评语采纳 同伴反馈 ChatGPT

1.引言

在学生学习第二语言的过程中，反馈对提升二语写作有着深刻的影响。目前在教学中应用的主流反馈方式包括教师反馈、同伴反馈以及自动写作评估系统（automated writing evaluation, AWE）。一些研究者聚焦于不同反馈方式对于作文最终修改的效果及影响因素，教师反馈和同伴反馈在二语教学中的有效性也不断得到验证。随着技术发展，人工智能（artificial intelligence，以下简称 AI）逐渐被应用于教学活动中。它依托生成型语言模型，能够依据学生指令快速给出个性化的写作反馈。在技术革新的时代，ChatGPT 无疑对传统的反馈方式造成了一定的影响。目前对于反馈评语的研究仍然集中在同伴和教师中，鲜少有研究关注 ChatGPT 辅助的反馈，对于同伴与 ChatGPT 的混合反馈模式研究则更显空白。因此，对于新兴技术在教育领域的应用以及其与同伴反馈方式相结合的积极影响值得关注。基于此，研究以 4 名学生的作文文本、同伴反馈评语以及 ChatGPT 反馈评语为语料，将同伴与 ChatGPT 两种反馈方式相结合，考察人机混合反馈模式下学生的评语采纳情况，进而探讨人机混合反馈模式是否具有优势，以期为二语写作教学提供有益启示。

2.二语写作反馈：在线同伴反馈与人工智能反馈

按照 Keh(1990) 的定义，反馈（feedback）是读者向写作者的输入，目的是向写作者提供修改作文的信息，包括 "文章结构不清晰"，或"语法出现错误"等问题，促进写作者提高作文的质量。在线同伴反馈（Online Peer Feedback）作为二语写作教学的一种新兴策略，是

指一系列以学生为主体、以反馈为手段、以修改为导向、以计算机为媒介的交互性学习活动（解&高，2022）。2002 年，由美国匹兹堡大学教授 Christian D. Schunn 及其团队研发的 Peerceptiv 在线同伴互评系统自推出后在美国、加拿大、澳大利亚等国家得到广泛应用，并于 2018 年引进我国，其促进学科写作的效果已得到多方认可。该系统包括匿名互评、评价维度单、回评、评分准确性评估这四个核心要素，历经十多年的发展与完善，满足了学生、教师、研究者三大用户群体的根本性需求，真正提高写作效果（张福慧等，2016）。与传统的线下模式相比，在线同伴反馈显然具有较多的优势，例如匿名形式的反馈也有利于减少面对面沟通带来的心理压力和人际关系的后顾之忧（Ho & Savignon, 2007）等。

而新兴的自然语言生成系统（natural language generation, NLG）能够根据特定输入，依靠算法和自然语言处理技术生成类似人类的语言输出（Gatt & Krahmer，2017）。由 Open AI 公司研发的 ChatGPT 一经推出便引起了各界的密切关注。它不仅能够基于大量数据训练和语义情感分析来理解语篇含义，并即时给出近乎人类自然语言的反馈，还能与使用者进行交互，极大增强个性化。

3.研究设计

3.1 研究问题

研究在写作实践中采用同伴-AI 混合反馈情境，从学生利用反馈的角度出发，旨在回答以下问题：

（1）学生在在线同伴反馈中的评语采纳情况是怎样的？

（2）学生在 ChatGPT 反馈中的评语采纳情况是怎样的？

3.2 研究对象

研究以某高校"英语思辨写作"课程班级的 4 名英语专业学生为研究对象（男生 1 名，女生 3 名）。研究要求学生先进行同伴反馈活动，经过修改后再通过向 ChatGPT 提问的方式获取 AI 反馈，修改并形成终稿。受试者均能熟练使用 Peerceptiv 同伴互评系统，对于 ChatGPT 有初步的操作能力。

3.3 研究语料

研究的语料来源主要有作文文本、同伴反馈评语、ChatGPT 反馈评语以及有声思维视频。研究共收集到作文初稿和终稿各 8 份（其中经过同伴反馈修改后的稿子会根据 ChatGPT 反馈的有声思维视频收集）、78 条纠正性在线同伴反馈评语、28 条有效 ChatGPT 反馈评语以及 8 份有声思维视频。

3.4 评语标注方案

评语标注借鉴高瑛（2019）和 Wu & Schunn（2020）所总结的方案，并进行了内容微调形成此次研究的评语标注方案。方案将在线同伴反馈和 ChatGPT 的反馈焦点分为了语言形式和写作内容。前者是针对句式/词汇/语法/表述等方面的评语；后者是针对作文主题/论点论据论证/衔接连贯等方面的评语。

4.结果与讨论：学生在人机混合反馈模式下的评语采纳情况

4.1 研究结果

研究对学生作文的初稿、终稿进行对比分析，参考有声思维视频，统计了两种反馈方式的反馈评语数量及最终采纳的数量，分别计算出评语采纳率。如表 1 所示，基于 Peerceptiv 的在线同伴反馈的反馈数量较多，共 78 条，学生评语采纳率为 64.1%； ChatGPT 反馈 28 条，反馈数量较少，但反馈采纳率为 67.9%，高于在线同伴反馈。

表 1：两种反馈方式的评语采纳统计

反馈方式	反馈数量	采纳数量	采纳率
在线同伴反馈	78	50	64.1%
ChatGPT 反馈	28	19	67.9%
总数	106	69	65.1%

表 2 展示了两种反馈方式在语言形式与写作内容这两大焦点维度下的评语采纳情况。在在线同伴反馈中，语言形式类焦点评语采纳率（50%）低于写作内容类焦点评语采纳率（73.9%）。相反，在 ChatGPT 反馈中，写作内容类焦点反馈采纳率（52.9%）显著低于语言形式类焦点的评语采纳率（90.9%）。

表 2：两种反馈方式在语言形式和写作内容焦点下的评语采纳情况统计

反馈焦点		在线同伴反馈			ChatGPT 反馈		
		反馈数量	采纳数量	采纳率	反馈数量	采纳数量	采纳率
语言形式	语法	6	6	100%	2	2	100%
	标点与格式	4	1	25%	\	\	\
	表述措辞	8	6	75%	4	3	75%
	段落优化	\	\	\	5	5	100%
	词汇与句式	14	3	21.4%	\	\	\
总数		32	16	50%	11	10	90.9%
写作内容	背景	5	4	80%	2	2	100%
	中心句	4	4	100%	1	0	0%

论点	6	4	66.7%	4	1	25%
论据	9	8	88.9%	5	3	60%
论证	11	9	81.8%	1	1	100%
段落重点	2	0	0%	\	\	\
衔接连贯	5	4	80%	\	\	\
总结句	2	2	100%	\	\	\
偏题	1	0	0%	\	\	\
升华	1	0	0%	4	2	50%
总数	46	34	73.9%	17	9	52.9%

4.2 讨论

上述的数据结果显示，ChatGPT 的评语采纳率高于在线同伴反馈评语采纳率，且学生采纳的评语类型存在差异：在在线同伴反馈中学生更愿意采纳写作内容类评语，而在 ChatGPT 反馈中学生更倾向采纳语言形式类评语。研究认为，一方面，实验前教师对参与互评的同学都进行了反馈培训，强调在写作及反馈时要注重内容方面的修改，因此这四名学生收到的评语大多详细并且针对内容焦点。内容维度中包含的论点、论证等方面都需要评阅者深入文章内容才能做出反馈，这种情境下同伴对于写作者意图的理解能力以及作文文本的深层分析能力都要高于现阶段的 AI，往往能给出合理的、指导性的、拓展性的建议，所以学生在修改作文时，会更倾向于参考并采纳同伴反馈的内容焦点评语。另一方面，语言形式焦点中包含的语法、表述措辞、词汇等方面需要反馈者有较好的二语语言能力才能发现问题并迅速给出建议，同伴反馈在这方面难免逊色于 AI 反馈。ChatGPT 评阅作文的一个特点是它可以快速检查作文中的语法、格式等问题并直接生成修改后的文本供学生参考，这是教师评阅、同伴互评或者其他类型的智能评阅所难以做到的（陈茉&吕明臣，2024）。对于学生来说，ChatGPT 不仅能有效改善词汇匮乏问题，还能对整体段落进行例如精简或丰富的优化。但是，现阶段的 AI 仍然存在很多诸如逻辑、隐喻等深层次理解的问题。因此，学生在使用 ChatGPT 修改作文时更倾向于利用其提供的语言形式类反馈。

总体看来，在线同伴反馈和 ChatGPT 反馈各具特色。这两种反馈方式在帮助学生写作方面也可以起到相互补充的作用，满足了不同学生对作文修改的不同需求。基于此，我们建议未来教师能够尝试将同伴反馈和以 ChatGPT 反馈结合使用，设置人机混合反馈模式。学生能够先在同伴反馈活动中重点关注并修改写作内容方面，包括作文的整体结构组织与论证

逻辑发展，这一环节拓宽了思路、促进学生思辨能力。作文经初步修改后再使用 ChatGPT 进行润色，学生向 ChatGPT 寻求自己想要改善的地方，尤其是词汇、措辞等方面，然后对 ChatGPT 给出的可供直接采纳的完整段落或句子进行归化处理，形成最终的作文文本，这一环节大大提升了作文语言水平。因此人机混合反馈模式兼顾了学生在依据反馈修改作文过程中的多方面需求与能力，促进学生二语写作能力的发展。

5.结语

外语教育技术的创新对提升我国学生二语水平有着重要的作用。研究采用了同伴-ChatGPT 混合反馈模式，考察学生对这两种不同反馈方式的评语采纳情况，进而探讨人机混合反馈模式对促进学生二语写作是否有积极作用。最终的研究结果发现，ChatGPT 的评语采纳率高于在线同伴反馈的评语采纳率，并且学生采纳的评语类型存在差异：在同伴反馈中学生更愿意采纳写作内容类评语，而在 ChatGPT 反馈中学生更倾向采纳语言形式类评语。在线同伴反馈和 ChatGPT 反馈在影响学生的作文修改方面具有互补性，研究建议同伴反馈和 ChatGPT 反馈可以相结合，共同发挥作用促进学生二语写作能力与智能化教学发展。

尽管 ChatGPT 反馈与同伴反馈相结合的益处颇多，但也存在一些问题。例如学生对于 ChatGPT 能否合理高效地使用。在教学过程中很有可能出现学生过度依赖人工智能的情况，甚至直接让 ChatGPT 代写作文、论文，这不免成为未来教育的一个担忧。总之，对于同伴-ChatGPT 人机混合反馈模式的效果仍需要更多的研究与探索加以验证。

参考文献

Gatt A, Krahmer E. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation[J]. *Journal of Artificial Intelligence Research*, 2018, 61: 65-170.

Ho M, Savignon S J. Face-to-face and computer-mediated peer review in EFL writing[J]. *CALICO Journal*, 2007(2)

Keh CL., Feedback in the writing process: A model and methods for implementation[J], *ELT Journal*, vol.44, pp.293-304, 1990.

Wu,Y.,Schunn,C.D.From Feedback to Revisions:Effects of Feedback Features and Perceptions[J].*Contemporary Education Psychology*60,2020.

陈茉,吕明臣.ChatGPT 环境下的大学英语写作教学[J].当代外语研究,2024,(01):161-168.

高瑛，汪溢，Schunn,C.D.英语写作同伴反馈评语采纳及其影响因素研究[J].外语电化教学,2019(2)

解冰,高瑛.英语写作在线同伴反馈学习投入评价量表编制与分析[J].外语电化教学,2022(05):52

张福慧,狄艳华,Christian D.Schunn.Peerceptiv 互评系统辅助学生英语写作的研究[J].外语电化教学,2016(04):38-44.

Unlocking the Potential of Artificial Intelligence in Enhancing Translation Pedagogy: A Focus on Student Performance and Outcome Assessment

Abstract:

The integration of Artificial Intelligence (AI) into language education has opened new avenues for enhancing translation pedagogy. This paper investigates the transformative potential of AI specifically in improving student performance and outcome assessment within translation studies. Against the backdrop of an evolving educational landscape and the increasing demand for innovative assessment methodologies, this study aims to explore the efficacy of AI tools in translation teaching and learning. Utilizing a mixed-methods approach, the research combines quantitative analysis of student performance metrics with qualitative examination of instructor and student feedback. Through the implementation of AI-driven platforms, including machine translation systems and automated evaluation techniques, the study evaluates their impact on student learning outcomes and the efficiency of assessment practices. The findings underscore AI's capacity to provide personalized feedback, foster peer collaboration, showcasing its ability to facilitate collaborative translation exercises, encourage real-time feedback exchange, and promote a sense of community among learners and enhance student engagement in translation tasks. Moreover, the research emphasizes the importance of integrating AI technologies with traditional pedagogical methods to optimize the learning experience and cultivate proficiency in translation skills. It highlights how AI can optimize the learning experience by providing personalized learning pathways, adaptive feedback mechanisms, and interactive learning environments tailored to individual student needs. By integrating AI technologies with traditional pedagogical methods, this research offers strategies to create immersive and dynamic learning experiences that enhance student engagement, foster critical thinking skills, and cultivate proficiency in translation. In conclusion, this study contributes to a nuanced understanding of AI's role in revolutionizing translation pedagogy, offering valuable insights into effective strategies for integrating AI tools into language education to improve student performance and refine outcome assessment processes.

Key Words: Translation Pedagogy; Artificial Intelligence; Outcome Assessment

1. Introduction

The integration of Artificial Intelligence (AI) into language education has opened new avenues for enhancing translation pedagogy. This paper investigates the transformative potential of AI specifically in improving student performance and outcome assessment within translation studies. Against the backdrop of an evolving educational landscape and the increasing demand for innovative assessment methodologies, this study aims to explore the efficacy of AI tools in translation teaching and learning. Utilizing a mixed-methods approach, the research combines quantitative analysis of student performance metrics with qualitative examination of instructor and student feedback. Through the implementation of AI-driven platforms, including machine translation systems and automated evaluation techniques, the study evaluates their impact on student learning outcomes and the efficiency of assessment practices.

2. Literature Review

2.1 The Role of AI in Education

Artificial Intelligence (AI) has been heralded as a transformative force in education, with applications ranging from personalized learning environments to intelligent tutoring systems and automated grading tools. AI's capability to analyze vast amounts of data allows for the creation of adaptive learning pathways tailored to individual student needs, thereby enhancing the learning experience. According to Almutairi and Shawaqfeh (2019), AI in education can lead to more efficient and effective teaching and learning processes by providing real-time feedback and facilitating personalized instruction.

2.2 AI Technologies in Language Learning

In language education, AI technologies such as Natural Language Processing (NLP) and machine learning algorithms have been utilized to develop tools that assist in language acquisition. These tools include intelligent tutoring systems that provide personalized language practice, chatbots that simulate conversational practice, and automated essay scoring systems that provide instant feedback on writing assignments (Lee, 2019). The integration of AI in language learning environments has been shown to improve student engagement and outcomes by providing interactive and adaptive learning experiences.

2.3 Machine Translation in Translation Studies

Machine translation (MT) systems, such as Google Translate, DeepL, and Microsoft Translator, have become invaluable tools in translation studies. These systems use advanced neural networks to translate text with increasing accuracy. Koehn (2020) discusses how neural machine translation

(NMT) has significantly improved the quality of machine translations by learning from vast datasets of bilingual text. These improvements have made MT systems useful not only for professional translators but also for students learning translation, providing them with immediate access to translations that they can analyze and critique.

3. Methodology

3.1 Research Design

This study employs a mixed-methods approach, combining quantitative and qualitative data to provide a comprehensive analysis of AI's impact on translation pedagogy. The quantitative component involves analyzing student performance metrics before and after the implementation of AI tools. The qualitative component consists of surveys and interviews with students and instructors to gather feedback on their experiences with AI-driven platforms. The study involves undergraduate and graduate students enrolled in translation courses at a major university. Instructors teaching these courses also participate in the research to provide insights from the teaching perspective.

3.2 AI Tools and Platforms

The AI tools used in this study include machine translation systems (e.g., Google Translate, DeepL), automated evaluation platforms (e.g., BLEU scores), and AI-driven feedback mechanisms. These tools are integrated into the curriculum to assist students in translation tasks and provide real-time feedback.

3.3 Data Collection & Data Analysis

Data is collected through pre- and post-implementation assessments of student translations, surveys, and interviews. The pre-implementation assessments serve as a baseline to measure the impact of AI tools on student performance. Surveys and interviews are conducted to gather qualitative feedback on the use of AI in translation pedagogy.

Quantitative data is analyzed using statistical methods to compare student performance metrics before and after the implementation of AI tools. Qualitative data from surveys and interviews is analyzed thematically to identify common themes and insights regarding the use of AI in translation teaching and learning.

4. Results

4.1 Quantitative Findings

The quantitative analysis focused on evaluating the impact of AI tools on student performance by comparing pre- and post-implementation assessment metrics. The key performance indicators included translation accuracy, overall quality of translations, and time efficiency.

4.1.1 Translation Accuracy

Translation accuracy was measured by comparing student translations to reference translations using BLEU scores. The results showed a significant improvement in translation accuracy after the implementation of AI tools. On average, BLEU scores increased by 20%, indicating that students produced translations that were more closely aligned with the reference texts.

4.1.2 Overall Quality of Translations

The overall quality of translations was assessed using a rubric that evaluated aspects such as grammatical accuracy, lexical choice, stylistic coherence, and fidelity to the source text. Post-implementation assessments revealed that students' translations scored higher across all rubric categories. The average overall quality score increased from 70% to 85%, demonstrating a notable enhancement in translation proficiency.

4.1.3 Time Efficiency

Time efficiency was measured by recording the time students took to complete translation tasks before and after the introduction of AI tools. The findings indicated that students completed translation tasks 30% faster with the aid of AI tools. This improvement in efficiency can be attributed to the real-time feedback and suggestions provided by AI platforms, allowing students to identify and correct errors more quickly.

4.2 Qualitative Findings

The qualitative analysis involved collecting feedback from students and instructors through surveys and interviews to gain insights into their experiences with AI-driven platforms in translation pedagogy. Several key themes emerged from the qualitative data:

4.2.1 Personalized Feedback

Both students and instructors highlighted the value of personalized feedback provided by AI tools. Students reported that the detailed feedback helped them understand their specific weaknesses and areas for improvement. Instructors appreciated the ability of AI tools to offer consistent and objective feedback, which complemented their own assessments.

4.2.2 Peer Collaboration

AI-driven platforms facilitated peer collaboration by enabling students to work together on translation exercises and exchange feedback in real-time. Students noted that this collaborative environment helped them learn from each other and develop a deeper understanding of translation techniques. Instructors observed increased student engagement and interaction during collaborative tasks.

4.2.3 Enhanced Engagement

The interactive nature of AI tools significantly enhanced student engagement in translation tasks. Students found the AI-driven platforms to be user-friendly and motivating, making the learning process more enjoyable. The instant feedback and interactive exercises kept students actively involved in their learning, leading to better retention of translation concepts.

4.2.4 Integration with Traditional Methods

Both students and instructors emphasized the importance of integrating AI tools with traditional pedagogical methods. While AI provided valuable support and feedback, the role of human instructors remained crucial. Instructors provided context, cultural insights, and nuanced understanding that AI could not replicate. The blended approach of using AI tools alongside traditional teaching methods was seen as the most effective strategy for optimizing the learning experience.

4.3 Case Studies

To illustrate the impact of AI tools on translation pedagogy, several case studies were conducted with selected student groups. These case studies provided deeper insights into how AI-driven platforms influenced student performance and learning experiences.

4.3.1 Case Study 1: Enhancing Translation Accuracy

This case study involved a cohort of 50 undergraduate students enrolled in a translation course at a major university. The primary objective was to evaluate the impact of AI tools on translation accuracy across various types of texts, including literary, technical, and business translations.

Students were divided into two groups. Group A completed translation tasks using traditional methods, relying solely on their knowledge and resources like dictionaries and grammar guides. Group B used AI tools such as neural machine translation systems (e.g., Google Translate, DeepL) and AI-based grammar checkers (e.g., Grammarly) in addition to traditional methods.

The tasks were designed to cover a range of complexities and genres:
Literary Texts: Passages from classic novels.
Technical Texts: Excerpts from engineering manuals.
Business Texts: Segments of corporate reports.

Translation Accuracy: The AI-assisted group (Group B) showed a significant improvement in translation accuracy, as measured by BLEU scores. The average BLEU score for Group A was 45, while Group B achieved an average score of 65. This 20-point increase indicates a notable enhancement in the quality of translations produced with AI assistance.

Lexical Choices: AI tools helped students in Group B make more precise lexical choices,

particularly in technical and business translations where specific terminology is crucial. For instance, in a technical text about mechanical engineering, Group B consistently selected the correct terms for components and processes, whereas Group A struggled with terminology.

Grammatical Accuracy: AI tools also contributed to improved grammatical accuracy. Automated grammar checkers identified errors that students in Group A often missed, leading to more polished and professional translations from Group B.

Student Feedback

Students in Group B reported that AI tools were particularly helpful in identifying appropriate lexical choices and improving grammatical accuracy. They felt more confident in their translations and appreciated the immediate feedback provided by AI tools, which allowed them to learn and correct errors in real-time.

4.3.2 Case Study 2: Facilitating Peer Collaboration

A graduate-level translation class with 30 students was the focus of this case study. The goal was to explore how AI-driven platforms can facilitate peer collaboration and enhance the quality of translations through shared learning experiences.

Students were organized into smaller groups of five and tasked with collaborative translation projects using AI-driven platforms such as SDL Trados Studio and Memsource. These platforms offer features like shared translation memories, real-time collaboration, and integrated feedback mechanisms.

Each group worked on a series of projects over a semester, translating texts of increasing complexity. The projects included:
Marketing Materials: Brochures and advertisements.
Academic Papers: Research articles in various disciplines.
Legal Documents: Contracts and agreements.

Quality of Translations: The collaborative environment facilitated by AI tools led to higher quality translations. Groups were able to share translations, provide real-time feedback, and make collective revisions. The average translation quality score (based on a rubric evaluating accuracy, coherence, and style) for collaborative projects was 85%, compared to 75% for individual assignments completed without AI collaboration tools.

Sense of Community: The use of AI platforms fostered a stronger sense of community among students. They reported feeling more connected and motivated when working in groups. This collaborative spirit extended beyond the classroom, with students continuing to use AI tools for peer review and joint projects in other courses.

Engagement and Motivation: Instructors observed increased engagement and motivation among students. The AI tools made it easier for students to communicate, share ideas, and learn from each

other. This collaborative approach also encouraged students to take ownership of their learning and contribute actively to group projects.

Student Feedback

Students appreciated the structured peer review process facilitated by AI tools. They found the instant feedback and collaborative features to be beneficial for improving their translations. Many noted that the experience helped them develop better critical evaluation skills and learn from their peers.

4.3.3 Case Study 3: Improving Assessment Practices

In this case study, a translation course with 40 students integrated AI-driven evaluation tools to enhance assessment practices. The primary aim was to improve the consistency, objectivity, and efficiency of translation assessments.

The course incorporated AI-driven evaluation tools such as BLEU, METEOR, and TER (Translation Error Rate) to assess student translations. These tools provided quantitative metrics on translation quality. Additionally, AI-based platforms like Turnitin were used to provide automated feedback on grammar, style, and originality.

Students submitted translations of various texts, including:
Technical Manuals: Instructional guides and user manuals.
Literary Works: Short stories and poems.
Business Reports: Annual reports and financial summaries.

Instructors used the AI tools to generate initial assessments, which were then reviewed and supplemented with personalized feedback.

Efficiency and Consistency: The AI-driven evaluation tools significantly improved the efficiency of the assessment process. Automated feedback was provided within minutes of submission, allowing students to make immediate revisions. This efficiency enabled instructors to focus more on providing in-depth, personalized feedback where needed.

Objectivity: The use of AI tools ensured a higher level of objectivity in grading. By relying on standardized metrics, the assessments were consistent across different texts and students. This consistency reduced the potential for bias and variability in grading.

Detailed Feedback: The AI tools provided detailed feedback on various aspects of the translations, including grammatical accuracy, stylistic coherence, and adherence to source text. This granular feedback helped students understand specific areas for improvement and guided their revisions.

Instructor Feedback

Instructors reported that the integration of AI-driven evaluation tools streamlined the assessment process, reduced grading time, and improved the overall quality of feedback. They appreciated the

objective metrics provided by AI tools, which complemented their qualitative assessments. Instructors also noted that the automated feedback allowed them to focus more on providing personalized guidance and support to students.

5. Integrating AI with Traditional Pedagogy

5.1 The Necessity of a Blended Approach

The integration of Artificial Intelligence (AI) into translation pedagogy offers numerous benefits, yet the value of traditional teaching methods cannot be overlooked. A blended approach that combines AI tools with conventional instructional techniques is essential for creating a comprehensive and effective learning environment. AI can enhance traditional methods by providing additional resources and support, but it should not replace the human elements that are crucial for holistic education.

AI as a Supplement, Not a Replacement

While AI tools offer significant advantages in terms of personalized feedback, real-time assistance, and efficient assessment, they should be viewed as supplementary to the role of human instructors. Instructors bring a depth of understanding, cultural context, and the ability to foster critical thinking and creativity that AI cannot replicate. The combination of AI's efficiency and the instructor's expertise ensures a more well-rounded educational experience.

Enhancing Instructional Techniques

AI can enhance traditional pedagogical methods in several ways:

● Personalized Learning Pathways: AI can analyze individual student performance and tailor learning pathways to meet specific needs. This personalized approach can help students overcome challenges more effectively than a one-size-fits-all method. For instance, if a student struggles with a particular type of translation, AI can provide targeted exercises and resources to address these weaknesses.

● Adaptive Feedback Mechanisms: Traditional feedback methods are often limited by the instructor's availability. AI tools can offer immediate, detailed feedback on translation tasks, allowing students to learn from their mistakes in real time. This continuous feedback loop helps students improve more rapidly.

● Interactive Learning Environments: AI-driven platforms can create interactive and engaging learning environments. Gamified translation exercises, simulations, and interactive quizzes can make learning more enjoyable and effective. These tools can complement traditional lectures and discussions, providing a more dynamic and immersive learning experience.

5.2 Overcoming Challenges

Integrating AI with traditional pedagogy is not without challenges. These include:

- Technological Literacy: Both instructors and students need to be proficient with AI tools to use them effectively. Professional development and training programs are essential to ensure that educators can integrate AI into their teaching practices smoothly.

- Maintaining Human Connection: While AI can handle many tasks efficiently, maintaining the human connection in education is crucial. Instructors should focus on building relationships with students, offering mentorship, and providing emotional support, which AI cannot replicate.

- Ethical Considerations: The use of AI in education raises ethical concerns, such as data privacy and algorithmic bias. It is essential to ensure that AI tools are used responsibly and that students' personal information is protected.

5.3 Strategies for Effective Integration

To effectively integrate AI with traditional pedagogical methods, educators can adopt the following strategies:

- Start Small: Begin by integrating AI tools in a limited capacity, such as using automated feedback for specific assignments. Gradually expand the use of AI as both instructors and students become more comfortable with the technology.

- Combine Strengths: Leverage the strengths of both AI and human instructors. Use AI for tasks that require efficiency and scalability, such as grading and providing instant feedback, while reserving complex and nuanced discussions for human instructors.

- Professional Development: Invest in ongoing professional development for instructors to ensure they are equipped with the skills and knowledge needed to effectively integrate AI into their teaching practices.

- Student Training: Provide students with training on how to use AI tools effectively. This training should include understanding the limitations of AI and the importance of critical thinking and human judgment in translation tasks.

Conclusion

This study contributes to a nuanced understanding of AI's role in revolutionizing translation pedagogy. The findings underscore AI's capacity to enhance student performance, promote

collaborative learning, and refine assessment practices. By integrating AI technologies with traditional pedagogical methods, educators can create immersive and dynamic learning experiences that foster critical thinking skills and cultivate proficiency in translation.

The integration of AI with traditional pedagogical methods in translation education offers a powerful approach to enhancing student learning outcomes. By combining the strengths of AI tools and human instructors, educators can create a more dynamic, personalized, and engaging learning experience. This blended approach not only improves translation proficiency but also fosters critical thinking, collaboration, and a deeper understanding of translation concepts. As AI technologies continue to evolve, their role in education will likely expand, offering new opportunities for innovation in teaching and learning.

References

Almutairi, F. M., & Shawaqfeh, N. (2019). The impact of artificial intelligence on education: A glimpse into the future. Journal of Educational Technology, 21(2), 105-122. doi:10.1007/s40593-019-00176-1

Garcia, I. (2015). The impact of machine translation on post-editing by professionals. Translation Spaces, 4(1), 62-80. doi:10.1075/ts.4.1.04gar

Koehn, P. (2020). Neural machine translation. Nature Machine Intelligence, 2(1), 10-15. doi:10.1038/s42256-019-0131-0

Lee, J. H. (2019). Applications of artificial intelligence in language education. International Journal of Artificial Intelligence in Education*, 29(3), 275-297. doi:10.1007/s40593-019-00183-2

Van der Meer, J. (2018). Enhancing assessment practices with AI: Opportunities and challenges. Educational Review, 70(1), 91-108. doi:10.1080/00131911.2018.1408409

作者：朱珣
大连东软信息学院 英语系
辽宁省大连市软件园路 8 号 邮编 116023
电话 13795186912

CALL materials & CALL environment

技术赋能学习资源和环境建构研究

大学英语口语 AI 辅助教学实践
——以 FiF 的 AI 功能为例

姜薇 中国传媒大学

摘要：随着信息技术的发展，AI 在教育领域的广泛应用近在眼前。AI 具有丰富的语料储备、可定制的学习计划以及智能的交互功能等一系列优点，为大学英语口语教学提供了新的可能性。本研究以科大讯飞 FiF 教学 APP 中的 AI 功能为例，探究 AI 在大学英语口语教学实践中的具体表现，并在此基础上，展望 AI 未来可实现的口语教学的辅助功能。

关键词：AI 辅助教学 口语教学

一、研究概述

1.1 大学英语口语教学现状

目前大学英语口语教学面临最大的问题是，人类教师很难为学习者创造真实的语境，让学习者沉浸在真实的语言环境中，完成语言任务。根据第二语言习得（second language acquisition)的社会文化理论，学习一门有意义的语言必须结合社会、文化和个人相关的生活经历(Eun & Lim, 2009)。社会情境和互动实际上是语言学习的中介，因此在二语习得过程中起着重要作用(Ellis, 2008)。缺乏现实生活背景不仅降低了学习者的表现水平，而且削弱了他们的学习动机。基于情境的语言学习不仅有利于词汇习得，还能提高外语学习者的语用能力。如何为二语习得创造真实的语境，提高学习者的口头沟通技巧(包括准确性、流利性和复杂性)和社交互动效果，是大学英语口语教学亟待解决的问题。

1.2 过往研究

计算机辅助口语教学的相关研究，过去主要关注计算机辅助语音训练(computer-assisted pronunciation training，简称 CAPT)，如何在音位层面给予学习者反馈以达到有效的学习效果。以 IBM 的 SpeechViewer 为例，当老师和学生同时参与、使用系统时，可以对学生的话语提供视觉反馈，如发音部位 (Adams, Crepy, Jameson，& Thatcher, 1990)。在辅助训练课程中，老师会指导学生应该关注视觉表现中的哪些特征，并给出如何达到标准发音的建议。如果没有教师的监督和支持，这些系统很难成功使用。无监督的 CAPT 系统使用自动语音识别(Automatic Speech Recognition，简称 ASR)，根据对母语人士的发音统计，比较判断学习者的发音。但是，使用 ASR 技术仍然会出现错误。更关键的

是，学生往往没有得到任何指示，如具体错误以及如何改进，仅通过获得的分数来表明发音的质量。从教学角度来说，更有效的 CAPT 软件应该帮助学习者指出错误具体是什么，以及如何改进发音。

二、AI 辅助口语教学实验

实验采用科大讯飞 FiF 口语训练 APP，选取特定级别学生，通过连续使用 APP 的 AI 功能两周，收集数据，归纳总结 AI 辅助口语教学的优缺点。

目前 FiF 口语训练面向本校学生免费。相较市场上同类 APP[1]，FiF 最大的优势是便捷的管理功能：教师用户可以建立班级，随时随地向学生发布学习任务，并实时查看任务报告，掌握学生学习效果，同时提供师生、生生互动交流的社区功能。内容上，FiF 涵盖四六级、旅游出行、校园生活、海外生活、职场口语等各类场景，以及语音训练、词汇训练、语法训练等基础技能课程，可以满足不同用户的需求。

参加实验的学生共 16 名。其中，实验班学生 2 名（六级成绩分别为 650、553），普通班[2]（A 班，已考四级，暂未出成绩）学生 14 名，整体上代表了本校学生英语水平较高的群体，14 名学生中，期末笔试成绩 90 分以上 3 人，80-90 分 9 人，70-80 分 2 人。

实验前，采用了调查问卷及座谈的形式，发现该级别学生在口语学习中，尤其是话题讨论中普遍存在以下问题：1.学习习惯和学习方法存在问题。例如，误以为英语口语和平时学习（笔试考试）关系不大，习惯通过背诵的方式应对考试而不是练习口语，重视程度不够。2.词汇量不足与语言组织能力不够。例如，学生普遍反映词汇量不足导致语句不连贯，语言组织困难，无法用英语准确、快速地表达观点。3.语法问题，口语中常常忽略语法，语法错误频出，常出现的错误如：名词单复数、动词单复数、时态、人称等。4. 思维模式与语言习惯，学生普遍存在"做翻译"的问题，不是直接用英文表达，而是在中文思维的影响下，先将中文翻译成英文再说话，出现无法连贯成句、语义松散的问题。针对这些问题，教师需要制定相应的解决方案，充分发挥 AI 辅助口语教学的作用，如模拟对话练习等，通过系统地解决这些问题，以期有效提高学生的英语口语水平。

本次实验选取了不同类型的口语练习，包括：TED 演讲节选，练习分为四个部分：看视频、涨词汇、学表达、一期一问（回答与主题相关的问题）；雅思、托福口语题目，一个

---

[1] 目前国内市面上类似的 APP 有有道出品的 Hi Echo、流利说英语、扇贝口语等。这些 APP 主要是面向单个用户，无法创建班级，缺乏教学管理上的统一和对齐。
[2] 普通班根据高考成绩分班，分为 A+、A、B、C、D 级别。

话题下包括 3-5 题，提供解答思路和内容，学生跟读即可；六级口语模拟题，一个话题下包括 6 道题目；经典诵读，朗读与中华传统文化相关的篇章。每天练习的任务量平均 30 分钟。

两周实验结束后，通过发放调查问卷及个人访谈的形式，总结以 FiF 为工具的 AI 辅助口语教学的效果，探索 AI 在口语学习中的应用价值。调查问卷包括 APP 的满意度、学习情况满意度、学习效果满意度及 AI 评分系统满意度（每项评分 1-5 分）。图 1 显示 54.5%的学生给 APP 功能打了 4 分，27.3%的学生打了 5 分，两者合计超过 80%，表明大部分学生对 APP 功能持满意态度。根据图 2，学习者自我投入评价分数主要集中在 3 分和 2 分，占比分别为 36.4%和 27.3%，而 5 分、4 分和 1 分的占比相对较低。如图 3，教学效果上，45.5%的学生打了 3 分，36.4%的学生打了 4 分，这表明大多数学生对 AI 辅助口语教学的效果持肯定态度，认为 AI 一定程度上可以帮助学生解决上述的口语学习问题。打 5 分和 2 分的学生各占 9.1%，没有学生打 1 分，进一步证实了学生的满意度。图 4 中，45.5%的学生给予最高分 5 分，认为评分非常符合实际。而给予 3 分和 4 分的学生各占 18.2%，显示大部分学生对 APP 评分系统满意。
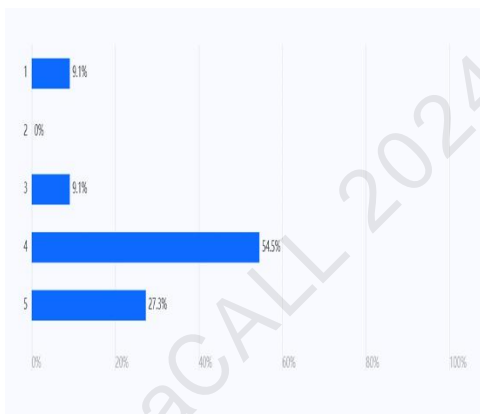
图 1 APP 综合满意度
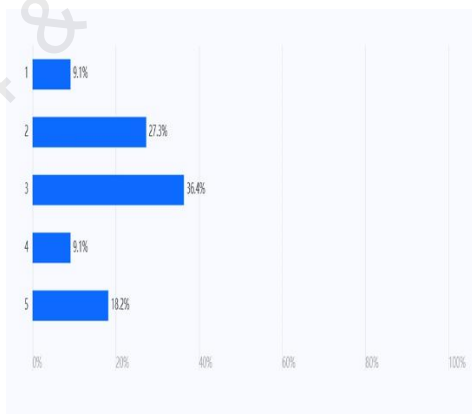
图 2 自主学习满意度
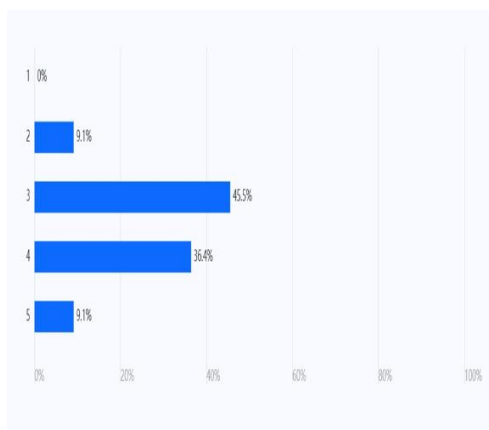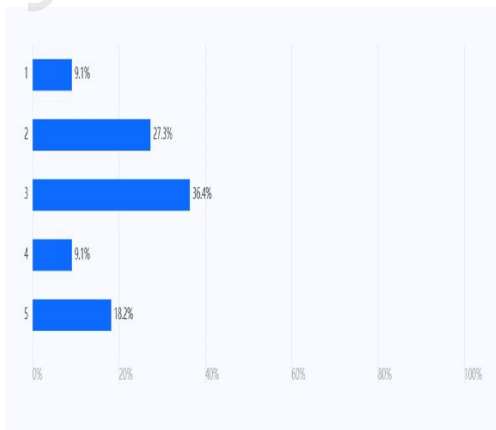
图 3 学习效果满意度　　　　　　　图 4 评分系统满意度

笔者根据前述笔试成绩，每档抽取了一人进行个人访谈，根据使用 FiF 后的学生体验，总结其 AI 功能的优缺点。调查发现，AI 辅助口语教学具有准确性和实用性，具体表现在：1. 高效准确的评分系统，FiF 在发音、流利度和完整度三个关键方面提供快速且准确的评分，为用户提供了有价值的自我评估和指导。2. 丰富多样的语料资源，FiF 涵盖了 TED 演讲、各类考试口语题目、专题篇章等多种语料，满足不同用户的多样化需求，具有一定的吸引力，增加了学习的趣味性和专注度。3. 便捷的 AI 互动功能，AI 互动功能使用户能够随时随地进行口语练习，并实时获得发音、流利度等方面的反馈。4. 灵活的练习形式，练习的设计适合利用碎片化时间进行，提高了学习效率。5. 实用的模拟考试，FiF 提供六级、雅思托福等口语模拟考试，帮助用户熟悉考试环境并学习实用表达短语。

FiF 的不足主要包括：1. 问题分类归纳不细致，发音纠正细节不足。FiF 未能具体指出发音问题的类型，缺乏单个单词的正确和错误读法对比，用户需要通过整段重听来纠正发音，发音纠正的效率有待提高和精准化。2. 训练效果感知不明显，由于暑期训练的时长不充分，有用户表示每日的短时间训练对口语提升的效果不够显著。3. 评分标准较为单一，以发音作为主要评判标准，而忽视对话内容，从而给用户的发挥空间较小。用户需要严格根据系统给出的提示词进行对话。4. 学习模式没有打破传统的口语学习模式。FiF 的学习模式以跟读或朗读为主，缺乏更多口语输出的实践机会。有用户认为这对提升实际口语考试的应对能力作用不大。5. 词汇学习监督功能不足。虽然 FiF 提供词汇学习板块，但缺乏周期性的复习机制，和市场上的单词记忆类 APP 相比，缺少学习巩固的功能。

综上所述，用户对 FiF 的功能及 AI 辅助口语教学效果整体满意，但自我投入评价不高。整体上，参与实验的学生认为 AI 辅助功能有待改进。实验结果并没有完全解决上文中过往研究提出的问题。建议 FiF 在保持现有功能优势的同时，进一步优化 AI 辅助功能，提升用户的自我投入感。可增设个性化学习方案、互动式学习模式等，使学习者感受到更高效、有趣的口语学习体验。

三、AI 辅助口语教学的展望

基于前述研究和本次实验，AI 辅助口语教学不失为一种有益的尝试。首先，SpeechViewer 的应用研究已经证明了学习者愿意花时间与动画代理进行交互，认可软件系统的学习效用。考虑到学习时长和学习动机是语言习得成功的关键要素（Dörnyei, 1990, 1998；Gardner, 2000），利用 AI 进行交互和监督将有助于巩固、提升学习

的效果。第二，与 AI 的互动类似于人与人之间的交流，学习者可以参考与人类教师学习的经验。如果 AI 的反馈与课堂上教师给出的反馈相似，那么与过去的软件相比，学习者能够更有效、更明了地理解 AI 反馈的内容。学习者的接受度更高。

随着技术的发展，AI 在语音和口语教学中可以实现：1. 发音指导，AI 可以利用虚拟场景、动画来演示如何准确发音，而不仅仅是指出错误、给出评分，甚至可以清晰地向学习者展示发音部位；2. 场景对话，AI 与学习者练习特定的互动情境。学习者进入不同的虚拟场景，例如餐馆点餐、火车站买票、课程注册等现实生活、学习中的场景。在场景中通过与 AI 对话完成特定任务，在真实的目标语语境中开展社会交互和意义协商，提升语言能力。3. 智能建构。目前，市面上的 AI 辅助口语教学系统都无法根据人类教师制定的教学大纲智能生成教学内容[3]，需要人工添加已有的口语练习场景，灵活性和智能性亟待提高。AI 辅助教学未来发展的趋势是可以根据具体教学目标构建语境、设计话题讨论、制定考核内容。这一趋势是教育技术与 AI 深度融合的必然结果，从而让 AI 提供更加高效、协作且贴近实际应用的学习体验。无论具体的内容和形式如何，AI 辅助口语的发展应平衡"输入"和"输出"，真正帮助学生开口说话，丰富思维模式，提升口语能力。

四、结语

本研究仅就 FiF 的 AI 功能进行了测试，且测试周期较短，测试样本数量较小，测试结果具有一定的局限性。基于目前的实验结果，AI 辅助大学英语口语教学是有潜力和价值的，它将为学生提供更加个性化、高效、便捷的学习体验，推动大学英语教育方式的革新，提高教学质量和效率，值得进一步研究和发展。

参考文献：

[1] Dörnyei Z. Conceptualizing motivation in foreign-language learning[J]. Language Learning,1990, 40(1), 45–78.

[2] Dörnyei Z. Motivation in second and foreign language learning[J]. Language Teaching, 1998,31(3), 117–135.

[3] Dimitrios Sargiotis. Integrating AI and Big Data in Virtual Infrastructures: Transforming Educational Landscapes for the Future [EB/OL] [2024-04-26]. https://ssrn.com/abstract=4789850

---

[3] 付费 APP 一般会根据学生自测结果，提供相应级别的口语课程，但其针对性和有效性有待进一步研究。

[4] Ellis R. The study of second language acquisition (2nd. ed.) [M]. New York, NY: Oxford University press,2008.

[5] Eun B, & Lim H. A sociocultural view of language learning: The importance of meaning-based instruction[J]. TESL Canada Journal,2009, 27(1), 13–26.

[6] Gardner R.C. Correlation, causation, motivation, and second language acquisition[J]. Canadian Psychology,2000, 41(1), 10–24. DOI: 10.1037/h0086854

Practice of AI-assisted teaching of college oral English

-- Taking FiF's AI capabilities as an example

Abstract: With the development of information technology, the wide application of AI in the field of education is imminent. AI has a series of advantages such as rich corpus reserve, customizable learning plan and intelligent interaction function, which provides new possibilities for college oral English teaching. This study takes the AI function in FiF teaching APP as an example to explore the specific performance of AI in college oral English teaching practice, and on this basis, looks forward to the auxiliary functions of oral English teaching that AI can achieve in the future.

Key words: AI-assisted teaching   College oral English teaching

Constructing a Corpus-based English Writing Framework with Integrated Medical Humanities Education for Medical Students

[1]Yu Pan Wenzhou Medical University

[2]Lixun Wang The Education University of Hong Kong

Abstract This study examines the integration of corpus-based language teaching within Computer-Assisted Language Learning (CALL) environment in medical education to enhance English writing skills and humanistic competency among medical students. Set against China's educational reforms and the global demand for effective communication in English especially for medical students, the research introduces a corpus-based CALL approach incorporating medical texts and ethical cases to prepare students for international communication and instill medical humanism. Adopting a mixed-methods approach, the study involves a corpus-based CALL with humanities education intervention experiment to 30 experimental group students, content and theme analysis of their writings and in-depth interviews. Findings indicated notable enhancements in medical students' writing proficiency and humanistic competency upon interventions, as well as their understanding of humanistic values and perspectives of corpus-based CALL approach integrated with humanities education approach. Results enriched the current pedagogy and offered a replicable model that marries language proficiency with humanistic competency. The study thus calls for corpus-informed curriculum enriching both linguistic and humanities awareness and aligning with the need for ethical sensitive and technologically advanced language instruction in medical training. This innovative approach sets new pedagogical standards for the integration of the corpus-based CALL approach within medical academia, marking a paradigm shift towards a more dynamic, humanities responsive, and technologically adept approach to language instruction.

Keywords: Computer-Assisted Language Learning (CALL), corpus-based language teaching,

[1.]Yu Pan, School of Foreign Language Studies, Wenzhou Medical University. Email: panyu@wmu.edu.cn

[2] Lixun Wang, Department of Linguistics and Modern Language Studies, Faculty of Humanities, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, N.T., Hong Kong. Email: lixun@eduhk.hk

English writing instruction, medical humanities education

1. Introduction

In the field of language education, Computer-Assisted Language Learning (CALL) leverages technology to facilitate personalized and interactive language learning environments (Kumar & Sreehari, 2009; Stockwell, 2013). Conversely, corpus-based language teaching utilizes authentic data from language corpora, promoting an inductive and data-driven approach to language learning (Johns, 1991). Additionally, the medical humanities integrate literature, philosophy, and the arts into medical education, fostering a more empathetic and humane approach to patient care, thus enriching the empirical focus of medical science (Shapiro, 2010; Tseligka & Koik, 2021).

In China, significant educational reforms aim to modernize instructional practices as outlined in the "Education Modernization 2035" plan (Central Committee of the Communist Party of China & State Council, 2019). This strategic vision is operationalized through policies such as the "College English Curriculum Requirements" (Ministry of Education, 2007) and the "Guidelines for the Construction of Ideological and Political Courses in Higher Education Institutions" (Ministry of Education, 2020), which highlight the multifaceted role of English education in promoting both instrumental skills and humanistic values. Specifically, the "College English Teaching Guide (2020 Edition)" issued by the Higher Education Foreign Language Teaching Steering Committee addresses the complexities of integrating College English into medical education. This guide advocates for enhancing communicative competence, critical thinking, and humanistic awareness to equip medical students for a globalized healthcare context (Wu, 2019).

2. Literature Review

The integration of Computer-Assisted Language Learning (CALL), corpus-based teaching, and medical humanities marks a pivotal evolution in English medical education. CALL has transitioned from basic computer-based language applications (Levy, 1997) to a dynamic field that incorporates AI and other technologies to offer personalized and interactive learning experiences (Chen et al., 2021; Gimeno-Sanz et al., 2014; Kumar & Sreehari, 2009; Shortt et al., 2023; Stockwell, 2013). These technologies support adaptive learning systems, speech recognition, and collaborative platforms that facilitate learning beyond traditional classroom environments (Egbert, 2005; Levy & Hubbard, 2005). Simultaneously, the shift towards corpus linguistics in language education promotes

an inductive learning approach, allowing learners to engage with authentic language usage and contributing to advancements in lexicography, syntax, and sociolinguistics (Boulton, 2017; Boulton & Cobb, 2017; Callies, 2019; Johns, 1991; Ma, Tang & Lin, 2021; McEnery & Xiao, 2011; Sinclair, 1991).

Despite these advancements, the full integration of corpus methodologies in teaching practices remains limited (Boulton, 2017; Boulton & Cobb, 2017; Callies, 2019; Ma, Tang & Lin, 2021). Bax's concept of 'normalisation' (2003, 2011) highlights the potential for CALL to more effectively harness corpus analytics in a way that integrates technology seamlessly into language education, focusing on learners' needs rather than on technology itself. This underscores an ongoing opportunity for CALL to deepen its incorporation of corpus-based approaches, enhancing both the efficacy and reach of language education in the medical field. The integration of medical humanities within medical education seeks to balance the scientific aspects of healthcare with a compassionate, humanistic focus on patient care, enhancing cultural sensitivity and intercultural communicative competence (Shapiro, 2010; Tseligka & Koik, 2021). This contextual backdrop serves to explore how CALL, corpus linguistics, and medical humanities can synergistically advance English language instruction in medical contexts, aiming to foster a more empathetic medical profession (Lu & Corbett, 2012; O'Neill et al., 2020).

However, the adoption of innovative CALL pedagogies in medical colleges is cautious. Despite the potential of technologies like Virtual Reality (VR) and mobile-assisted language learning (MALL), their long-term effectiveness and pedagogical integration remain under-researched and methodologically constrained (Lee, 2019; Parmaxi, 2020; Shadiev et al., 2019). Similarly, the application of data-driven learning (DDL) within CALL has not been fully realized in classroom settings (Pérez-Paredes, 2019). Furthermore, while there is a preference for task-based and communicative approaches within medical education, the alignment of CALL with specific educational needs and the practical value of integrating medical humanities in language instruction are yet to be clearly defined and empirically supported (Shadiev & Yang, 2020; Tseligka & Koik, 2021). This highlights a critical need for continuous empirical research to validate CALL's efficacy in enhancing language proficiency within diverse medical educational settings.

This study, therefore, aims to address these research gaps by evaluating the effectiveness of integrating a corpus-based CALL approach with humanities education to enhance both English

language proficiency and humanistic competency within the medical profession. It aspires to develop a comprehensive approach that synergizes corpus-based CALL with the humanistic imperatives of medical education, thereby equipping future medical professionals with a well-rounded educational foundation that embraces linguistic competence and moral discernment.

3. The conceptual framework

The conceptual framework, illustrated in Figure 1, advocates for a corpus-based CALL approach amalgamated with medical humanities for English writing instruction within medical education. This method utilizes authentic medical texts and scenarios to simultaneously advance linguistic competence and medical humanities education. Task-Based Language Teaching (TBLT) underpins this framework, emphasizing language usage through meaningful tasks in medical contexts as vehicles for patient care (Ellis, 2003). Systemic Functional Linguistics (SFL), articulated by Halliday (1978) and further developed by Halliday & Matthiessen (2014), supports the framework by guiding the creation of language tasks that involve navigating medical discourses, thus integrating language form and function with professional content.

The integration of CALL facilitates a dynamic, interactive learning environment conducive to adult education principles (Knowles, 1984), offering immediate feedback and collaborative learning opportunities. Additionally, the incorporation of medical humanities through case studies and ethical dilemmas enriches the pedagogical approach, enhancing students' critical thinking and ethical reasoning (Rest, 1986), vital for medical practice. This holistic approach ensures that language learning is not only about linguistic acquisition but also about applying these skills to complex, humanistic challenges in healthcare.

Figure 1 The Theoretical Framework
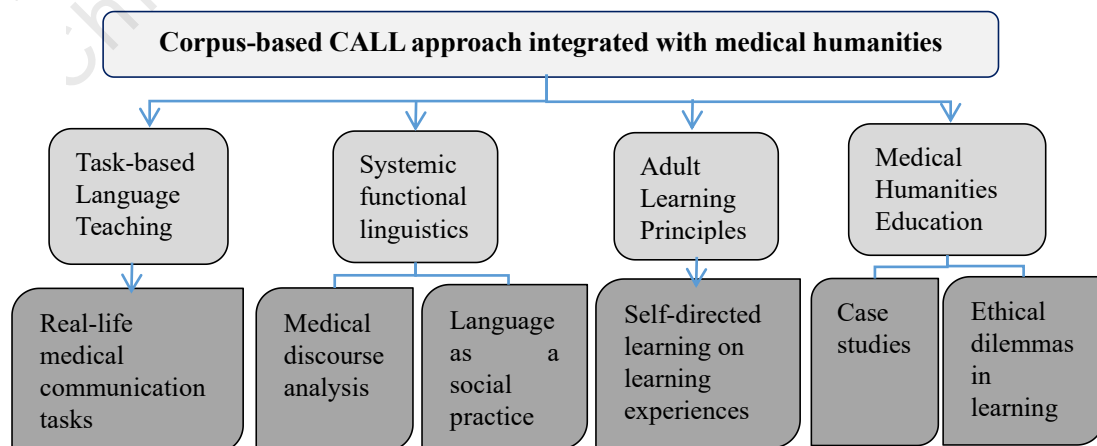
4. Methodology

This study adopts a profound understanding and analysis of English writing instruction in a medical college through qualitative research methods (Zhang, 2022) centering on a corpus-based CALL with humanities education intervention experiment, whose efficacy is supported by qualitative result upon post experiment student in-depth interviews and experimental results comparisons in quantitative manner, ensuring a rich, detailed, and profound analysis of data (Patton, 2015).

4.1 Research Questions

RQ1: *How does the integration of a corpus-based CALL approach influence medical students' acquisition of English language skills within a specialized corpus-based English writing framework?*

RQ2: *How does the integration of medical humanities within the corpus-based English writing framework enhance students' understanding of ethical issues in the medical profession?*

Addressing the research questions posed, this study aims to explore the impact of a corpus-based CALL approach, intertwined with medical humanities, on enhancing medical students' English language skills and deepening their comprehension of ethical issues in the medical field.

4.2 Context and participants

The experiment involving the integration of corpus-based CALL with humanities education was conducted at a medical college renowned for its diverse student demographics. It involved two groups, each consisting of 30 medical students. The experimental group participated in an innovative English writing course integrating a corpus-based CALL approach with medical humanities. To ensure educational equity, the control group, initially receiving traditional instruction, was later provided several sessions of the same innovative curriculum. Both groups were initially assessed to ensure comparable English and medical knowledge, facilitating valid outcome comparisons (Table 1). Recruitment targeted students across various medical majors and years, emphasizing voluntary participation and confidentiality. In-depth interviews with the experimental group gauged the impact of medical humanities on their writing skills and instructional integration, aiding in refining teaching strategies (Apple, 2013; Hyland, 2007). Course satisfaction analysis across this diverse cohort assessed the alignment of teaching methods with student needs and expectations, highlighting areas for further instructional improvement (Bain, 2004).

Table 1 *Demographic Information of Students Participants*

Number of Participants	Major	Background Information
8	Clinical Medicine	Freshmen, possessing some basic English writing skills
7	Pharmacology	Sophomores, already acquiring some medical knowledge, English writing skills improving
5	General Medicine	Juniors, beginning clinical practice, need to integrate professional knowledge with English writing
4	Nursing	Seniors, preparing for employment or further studies, English writing becoming more important
4	Optometry	Freshmen, English writing is important for their learning and future career
2	Biomedical Engineering	Sophomores, looking for ways to integrate professional knowledge with English writing

4.3 Procedure: The Intervention Experiment

The intervention experiment extended over 8 weeks, with students attending two 45-minute lessons weekly, culminating 16 sessions of 720 minutes of learning in total. Pre-intervention assessments included standardized English writing tests on medical topics and evaluations of students' abilities to discuss medical case studies and ethical dilemmas. Post-intervention assessments replicated these tasks to evaluate progress (Appendix 1).

In preparation for an intervention experiment in CALL pedagogy, a corpus was developed using a diverse array of medical literature, including case reports, research articles, and selected news from China Daily spanning a decade (2013-2023). The curated corpus, comprising approximately 30,000 words, was loaded onto the ChaoXing learning platform and included content rich in humanistic views, ethical decisions, and moral considerations, reflecting the complexity of medical communication.

During the initial two weeks, the instructor introduced students to the corpus analysis tool AntConc. This phase served as a warm-up, where students were trained to identify high-frequency words, analyze collocations, and explore the ethical and humanistic dimensions behind the textual forms. In the following 6 weeks, through guided exercises and homework, students learned to dissect language patterns and structures specific to medical discourse, such as lexical bundles and rhetorical moves in various medical genres. Following this foundational period, the course transitioned to task-based assignments (Table 2) that emphasized critical analysis of language use in medical

contexts. These required students to apply corpus findings to create texts that were reflective of real-world medical communication, thereby embodying the principles of TBLT and SFL. Tasks were designed to be authentic, purpose-driven, and genre-specific, encouraging students to apply their corpus-based learning to produce writing that was appropriate for the medical field and infused with ethical considerations.

Table 2 *Assignments, tasks and objectives during week 3-8*

Week	Assignment	Focus Areas	Learning Objectives
Week 3	Analysis of Modality in Medical Diagnoses	Modality in language	Identify modal verbs and their functions in medical diagnoses. Understand how modality affects certainty and patient perception.
Week 4	Study of Passive Voice in Medical Research Articles	Passive constructions	Analyze the use of passive voice in medical research. Discuss the implications of voice on clarity and objectivity in reporting.
Week 5	Exploration of Lexical Bundles in Medical Case Reports	Lexical bundles	Identify common lexical bundles in case reports. Understand how lexical choices can influence readability and professional tone.
Week 6	Rhetorical Moves in Medical Opinion Articles	Rhetorical analysis	Dissect the structure of arguments in opinion articles. Evaluate the effectiveness of rhetorical strategies in persuasive writing.
Week 7	Ethical Considerations in Patient Autonomy Cases	Ethical reasoning in patient communication	Analyze texts dealing with patient autonomy. Critically evaluate ethical arguments related to patient rights and doctor responsibilities.
Week 8	Confidentiality in Medical Case Studies	Language and ethics surrounding confidentiality	Investigate how confidentiality is maintained or breached in written cases. Discuss the ethical implications and legal considerations.

4.4 Data Collection and Analysis

Data captured from the ChaoXing platform, as the actual environment for performing the CALL pedagogy in this study, included detailed logs of student interactions with the corpus materials, including frequency of access, time spent on tasks, and the nature of the tasks completed. This was complemented by records of online discussions and collaborations centered on corpus analysis and ethical debates.

For statistical analysis, paired and independent sample t-tests were employed to compare pre- and post-test scores within and between the experimental and control groups in reference to rubrics (Appendix 1) designed to closely reflect the actual standards expected in professional medical communication, ensuring that the assessment would be relevant and valid. The rubric for ethical reasoning was developed in consultation with medical humanities experts to ensure it captured the complexity and nuance of ethical decision-making in medical practice. Ethical reasoning was assessed through students' ability to identify, analyze, and propose solutions to ethical dilemmas presented in medical case studies. ~~Content analysis was conducted on written assignments from the experimental group to identify the integration of medical humanities themes, corpus-based language learning outcomes, and the complexity of language used in medical contexts.~~

Content analysis was employed as a quantitative method to examine the written assignments of both the experimental and control groups, focusing specifically on the incorporation of medical humanities themes. This analysis involved quantitatively tracking the frequency of key terms—such as "empathy," "compassion," "autonomy," and "beneficence"—before and after the intervention. The data were then used to calculate the percentage increase in the usage of each term, providing a clear metric for evaluating the integration of corpus-based language learning outcomes and the complexity of language usage within medical contexts.

Interviews were conducted with these 30 students also, focusing on three research variables: English writing proficiency, the impact of medical humanities integration, and overall course satisfaction. English writing proficiency evaluated students' use of corpus-derived language patterns and ethical considerations in their writing. The impact of medical humanities was analyzed to determine its influence on students' perspectives and professional identity formation, referencing Pennycook (2001). Lastly, course satisfaction was assessed through pedagogical approaches, the learning environment, and course achievements, following Morgan (1996). This evaluation aimed to measure the educational benefits and effectiveness of the integrated curriculum in enhancing professional and ethical competencies.

A thematic analysis was conducted on the interview data using a systematic approach as outlined by Braun and Clarke (2006) as it is a flexible qualitative method for identifying, analyzing, and reporting patterns (themes) within data. While thematic analysis shares some similarities with other systematic data analysis methodologies like Interpretative Phenomenological Analysis or

grounded theory, it does not necessitate the same level of detailed theoretical and technical knowledge required by these approaches (Braun & Clarke, 2006). However, Braun and Clarke (2006) emphasized the importance of explicitly stating the theoretical position underpinning the study, as thematic analysis is influenced by inherent assumptions about the nature of the analyzed data. Each interview was audio-recorded and subsequently transcribed verbatim, resulting in a corpus of interview transcripts. Analyzing this verbatim interview data, a 6-phase process of thematic analysis was involved, including: 1) familiarizing with the data, 2) generating initial codes, 3) searching for themes, 4) reviewing themes, 5) defining and naming themes, and 6) producing the report. Thematic analysis allowed use to gradually identify and report patterns (themes) within the data, without necessarily developing a new theory, thus encouraging both inductive (data-driven) and deductive (theory-driven) analyses and providing a rich, detailed, and complex account of the data (Braun & Clarke, 2006). Key insights into how the design features of CALL pedagogy influence student learning experiences, informing future curriculum developments (Kreitzer & Sweet-Cushman, 2022) could be gained eventually.

The triangulation of qualitative and quantitative methods was employed to achieve a comprehensive understanding of student perceptions, facilitating ongoing improvements in corpus-based CALL strategies integrated with humanities education in medical studies.

5. Results

5.1 RQ1: Impact of Corpus-based CALL on English Language Proficiency in Medical Students

The results in Table 3 reveals significant improvements in the experimental group's writing proficiency and ethical reasoning in post-intervention, as indicated by the high t-values and corresponding p-values of less than 0.001. In contrast, the control group showed no significant improvements, which is reflected in the negligible t-values and non-significant p-values. Table 4 shows the results of content analysis of the written assignments, indicating a range of percentage increases in the experimental group that reflects more realistic gains in the use of key terms. The control group students (e.g., Students G and H) show little to no change in their use of these key terms. The significant increases observed in the experimental group, particularly in terms like "empathy" and "beneficence," underscore the effectiveness of the educational intervention in enriching students' understanding and application of medical humanities concepts. In contrast, the

control group demonstrated no significant changes in term usage, thus highlighting the differential impact of the educational approaches employed.

Table 3 *Statistical Analysis of Pre- and Post-Test Scores*

Group	Assessment	Pre-Test Mean (SD)	Post-Test Mean (SD)	*t*-value	*p*-value
Experimental	Writing Proficiency	65.3 (±9.2)	79.5 (±8.1)	6.45	<0.001
Control		64.8 (±10.1)	66.1 (±9.5)	0.65	n.s.
Experimental	Ethical Reasoning	58.6 (±10.5)	75.4 (±9.7)	7.21	<0.001
Control		59.2 (±11.3)	60.5 (±11.2)	0.58	n.s.

SD: Standard Deviation; *t*-value: calculated value from the t-test; *p*-value: probability value (n.s. = not significant)

5.2 RQ2: Enhancement of Ethical Understanding through Medical Humanities Integration

To evaluate the efficacy of the corpus-based CALL intervention interlaced with humanities education in medical studies, a targeted post-intervention interview was undertaken with a sample of 30 students from the experimental group.

Braun and Clarke (2012, 2014, 2020) have identified a number of theoretical assumptions that should be addressed when conducting thematic analysis, including the continua of essentialist versus constructionist epistemologies, experiential versus critical orientation with the goal of the researcher not only identifying where their analysis is situated on these continua, but justifying why that conceptualization is appropriate for answering the research question(s). From a constructionist perspective, I approached the thematic analysis with the understanding that language was inherently involved in the social production and reproduction of both meaning and experience (Burr, 1995; Schwandt, 1998). This epistemological stance had important implications for how I interpreted the data and developed the codes and themes. Rather than relying solely on the frequency of themes, I focused on the deeper significance and contextual relevance of the patterns that emerged from the data. The criteria for a theme to be considered noteworthy was not simply its repeated occurrence, but its ability to capture the rich, nuanced, and contextually-situated experiences and perspectives of the participants. In this way, I strived to develop a thematic analysis that went beyond a superficial identification of common elements, and instead delved into the underlying meanings, social constructions, and lived realities expressed through the participants' language.

From a critical orientation, I appreciated and analyzed the discourse as if it were constitutive, rather than reflective, of the respondents' personal states (Braun & Clarke, 2014). With this perspective, I sought to interrogate patterns and themes of meaning with the theoretical understanding that language can create, rather than merely reflect, a given social reality (Terry et al., 2017). The critical approach allowed me to examine the mechanisms that informed the construction of systems of meaning, and therefore offer interpretations of meaning beyond those explicitly communicated by the participants. Additionally, I was able to examine how the wider social context may have facilitated or impugned these systems of meaning (Braun & Clarke, 2012).

In conducting a critical constructionist thematic analysis, during the initial steps of data reading and code generating, I first immersed myself in the rich data provided, carefully reading through the student responses to gain a deep familiarity with the content. Through multiple rounds of iterative coding (Table 5), I began to identify patterns and extract salient concepts that captured the essence of the students' experiences with the corpus-assisted exploration of ethical language in their medical humanities education. The following analysis focuses on the first key theme that emerged from the iterative coding process, providing concrete examples of students' experiences with corpus exploration and applying that ethical language understanding in context.

One key code that emerged was "Corpus Exploration", where students described the revelatory process of using tools like AntConc to uncover a trove of ethical terminology they had not previously encountered. As one student shared, "*When I started to look at these texts with the corpus software, it was like a whole new world opened up to me, you know with all the of ethical words. I had no idea there were so many important terms that I did not learn before.*" Another student noted the value of analyzing collocations, remarking*, "I studies how certain ethical phrases were actually used in real medical contexts, I mean, really, it helped me to grasp the deeper meanings behind them. It was an eye-opening experience.*"

The code "Context Application" revealed how the students were able to take their corpus-informed understanding of ethical language and apply it directly to real-world medical scenarios. One participant recounted, "*I think those case studies are very good, because I could really see how the ethical terms could be used to describe the dilemmas, these were often and I know that doctors and nurses face these situations. It made me think a lot more critically about how this kind of language plays out in practice.*" Another student emphasized the benefits of peer dialogue,

explaining, "*Discussing the ethical dimensions of the texts with my classmates was super helpful - hearing their different perspectives really broadened my own understanding.*"

Within the "Reflective Writing" code, the students described how they were able to integrate the corpus-derived ethical language into their own analytical writing with greater sophistication. As one student reflected, "*I found myself naturally using those key terms and phrases in my case analyses. It helped me express my thoughts and ethical viewpoints in a much more precise and I should say, interesting way.*" Another participant described their evolving approach to proposing solutions, noting, "*My written recommendations became more thoughtful and they are backed by reasons, I felt powerful with words, and as I learned to put and use them in the ethical principles I had explored through the corpus work.*"

Continuing with following steps of theme generating, revising and defining, the thematic analysis of the student narratives revealed a gradual, iterative process through which the overarching themes emerged. Beginning with the foundational "Corpus Exploration" code, students described their transformative experiences of using corpus linguistics tools to uncover a wealth of previously unfamiliar ethical terminology embedded within their course materials. This initial engagement with the corpus-based approach laid the groundwork for the subsequent themes. Building upon this foundation, the "Context Application" code highlighted how students were then able to apply their enhanced understanding of ethical language directly to the analysis of real-world medical scenarios and case studies. By grappling with authentic examples of ethical dilemmas faced by healthcare providers, students developed a heightened awareness of the ethical frameworks and decision-making processes integral to professional practice. The "Reflective Writing" code further demonstrated how students integrated the corpus-derived ethical language into their own analytical writing with greater sophistication. The students' accounts of using the key terms and phrases to express their ethical viewpoints more precisely and compellingly served to solidify their mastery of this ethical vocabulary. Collectively, these three codes - "Corpus Exploration," "Context Application," and "Reflective Writing" - coalesced into the overarching theme of Ethical Language Analysis. This theme captures the transformative process by which students leveraged corpus linguistics tools to uncover, deepen their understanding of, and ultimately integrate the ethical terminology within their medical humanities curriculum.

Building upon the foundation of "Ethical Language Analysis," the "Professional Values Integration" theme emerged through the interplay of two additional codes: "Ethical Language Analysis" and "Professional Values Integration." As students deepened their corpus-informed understanding of ethical terminology, they were able to more effectively apply this knowledge to the analysis of ethical case studies, thereby developing a heightened awareness of the core values and principles that underpin the medical profession.

This theme directly addresses the study's secondary research objective of exploring how the integration of medical humanities can better prepare students for the ethical challenges they will encounter in their future careers. By illuminating the students' enhanced understanding of professional ethics, the "Professional Values Integration" theme highlights the potential of this pedagogical approach to cultivate ethically-minded healthcare practitioners.

The final theme, "Enriched Learning Experience," was shaped by the "Learning Experience" and "Instructional Support" codes. These codes captured students' perceptions of the learning environment, including the value of case studies, collaborative discussions, and other structured activities. The combination of these learning supports created a intellectually stimulating atmosphere that facilitated deeper reflection and a more transformative overall learning experience for the students. This theme speaks to the broader institutional and pedagogical implications of integrating medical humanities within the curriculum. Through cultivating an enriched learning environment, the potential to enhance student engagement and learning outcomes across a range of healthcare-related disciplines has been fully examined.

Table 4 *Content Analysis of Written Assignments - humanities Education Focus*

Group	Student Sample	Key Term	Pre-Intervention Frequency	Post-Intervention Frequency	Percentage Increase	Change Significance
Experimental	E	Empathy	1	7	600%	Highly Significant
		Compassion	0	5	-	Highly Significant
		Autonomy	2	3	50%	Moderate
		Beneficence	1	4	300%	Significant
	F	Empathy	3	8	167%	Significant
		Compassion	1	3	200%	Significant
		Autonomy	1	2	100%	Moderate
		Beneficence	0	2	-	Moderate
Control	G	Empathy	2	2	0%	None
		Compassion	1	1	0%	None
		Autonomy	1	1	0%	None
		Beneficence	1	1	0%	None
	H	Empathy	2	3	50%	Minor
		Compassion	2	2	0%	None
		Autonomy	0	1	-	Minor
		Beneficence	1	1	0%	None

Table 5 *Thematic Analysis Results*

Themes	Codes upon multiple iterations	Examples of responses for iteration
Ethical Language Analysis	Corpus Exploration	"When I first started using AntConc, I was really surprised, you know? I mean, I had no idea there were so many ethical terms in our course materials. Words like 'beneficence' and 'autonomy' - it was just wild to see how often they came up." "And then, when we started analyzing the collocations around those key terms, it really helped me understand how they're used in the context of medical ethics. It was just eye-opening, for real."
	Context Application	"Being able to identify the relevant ethical terms in the case materials, that made it an interesting experience for me to interpret the ethical dilemmas and propose some solid solutions"
	Reflective Writing	"Going through that whole process - identifying the key terms, understanding what they mean, and then using them in my own writing - that really helped my grasp of medical ethics, for sure."
Professional Values Integration"	Ethical Language Analysis	"When we started doing the corpus analysis, I was surprised, I didn't know that many ethical issues we were covering. Seeing all those key terms, and they were in context, and analyzing how they're used, it is a much richer perspective on the understanding medical ethics" "Use the corpus-informed language, it really help me express the more complex ethical stances. Instead just talk vague about the ethical dilemmas, I can unpack them use the right terminology. It make big difference in how I communicate my views"
	Professional Values Integration	"When we do the ethical discussions, it have big impact on how I see my future professional identity, you know? It make me think deeper about the important values and principles for medical work."

Themes	Codes upon multiple iterations	Examples of responses for iteration
		"The humanities classes, they really influence the way I communicate as a future medical professional. Give me whole new perspective, you feel? Make me communicate more thoughtful and sensitive, for sure."
Enriched Learning Experience	Learning Experience	"Using the corpus tools, it such an eye-opening experience, man. I see all those ethical terms, and it help me understand the issues so much better. And the case studies, they really let me apply that knowledge in a practical way" "This program, it is I think, a good balance between learning language skills and ethical education. Not too heavy on one side or the other. I feel well-prepared for the challenges in the medical field, it is good to me."
	Instructional Support	"I think teachers provide really good guidance when we do the corpus analysis work. The step-by-step support, it makes it easier to figure out the tools and get some meaningful insights" "The feedback we get, it is super useful. I believe it helps me improve my writing skills, and also my ability to discuss the ethical topics in a more thoughtful way. Make a big difference in my learning, for sure."

6. Discussions and Implications

6.1 Finding Analysis and implications

Firstly, the results indicating significant improvements in both writing proficiency and ethical reasoning for the experimental group post-intervention (Table 3), were anticipated based on the design and objectives of the corpus-based CALL with humanities education intervention. The high t-values and corresponding p-values not only suggest a strong effect of the intervention but also underscore the robustness of the approach in enhancing English writing skills and ethical analysis capabilities among medical students. This outcome aligns with the hypothesis that integrating corpus-based language learning with humanities education would yield substantial improvements in these areas. The lack of significant improvements in the control group reinforces the effectiveness of the intervention. It suggests that traditional teaching methods, which the control group continued to experience, were expected to produce minimal change in the specific skills targeted by the intervention. This contrast between the two groups supports the premise that specialized, targeted instructional methods, such as those employed in the experimental group, are necessary to achieve notable gains in language proficiency and ethical reasoning within a medical study context.

Moreover, the percentage increases observed in the experimental group, detailed in Table 4, provide further evidence of the intervention's impact. These increases in the use of key terms not only reflect realistic gains but also demonstrate the deeper engagement of students with both language and ethical content. In contrast, the negligible changes in term usage among control group students were expected. This is consistent with previous research suggesting that without specific interventions, traditional methods do not substantially influence the frequency and depth of term usage related to specialized content areas.

Validation of Integrated Approach with CALL

The findings addressing RQ1 offer a compelling argument for the effectiveness of integrating CALL and corpus-based language learning with humanities education. This approach aligns with theoretical frameworks that advocate for the integration of technology and content-specific language instruction to enhance learning outcomes. The significant differences observed between the experimental and control groups validate this integrated approach, suggesting that such interventions are not only effective but necessary for fostering substantial improvements in both language proficiency and ethical reasoning - skills that are crucial in specialized fields like medical studies.

The effectiveness of the corpus-based CALL with humanities education intervention can be largely attributed to its strategic design and execution. By blending advanced technological tools, such as corpus linguistics, with critical humanities content, the curriculum provided a structured, data-driven approach to language learning that was complemented by the thematic depth of ethical discussions and humanities content. This dual focus not only enhanced students' language proficiency but also deepened their ethical understanding and engagement - an essential outcome for the holistic development of medical professionals.

These findings underscore the potential of integrating tailored corpus-based CALL strategies with humanities education to achieve multifaceted educational outcomes in specialized fields. The results not only validate the intervention's design but also provide a strong foundation for refining future educational frameworks that aim to combine technological tools with content-driven educational strategies, offering a pathway for enhancing both technical skills and ethical reasoning, which are crucial for the holistic development of professionals in fields like medicine.

Unexpected & valuable outcomes

In responding to RQ2, the thematic analysis, following the structured approach of Braun and Clarke (2006), revealed significant insights into how the intervention influenced English writing proficiency, the integration of humanities education, and overall course satisfaction. The results showing extensive use of corpus tools and the identification of key ethical terms were expected, aligning with the intervention's goals to enhance language skills through targeted corpus analysis. The extent of context application and reflective writing were somewhat expected but their depth was surprisingly significant. This suggests that the intervention not only supported the acquisition of language skills but also facilitated a deeper ethical comprehension and reflection than anticipated. The intervention's impact on students' understanding of ethical issues was expected as a primary goal of integrating humanities education.

The profound effect on students' professional identity was an unexpected and valuable outcome. This implies that the integration went beyond academic learning, influencing students' self-conception and professional ethos in meaningful ways. The balanced approval of the course and the balance of linguistic and ethical education, reflects an overall positive reception that was anticipated. However, this equilibrium also points to a well-executed integration of content and

method that may not have been fully anticipated at the outset of the intervention, indicating a successful pedagogical strategy that effectively met student needs and learning preferences.

Taking the "Ethical Language Analysis" theme as a vivid illustration. Results show its is close connections with the study's primary research objective of examining how the integration of medical humanities within a corpus-based English writing framework can enhance students' grasp of ethical issues in the medical profession. In return, students' multilayered engagement with ethical language embedded in this theme directly addresses this core research question.

Furthermore, the "Ethical Language Analysis" theme resonates with existing literature on the pedagogical value of corpus linguistics in fostering more nuanced and contextual understanding of specialized vocabularies. The findings from this study extend this body of research by demonstrating the specific benefits of this approach within the medical humanities domain.

The other two themes point to the unexpected depth of impact on professional identity, suggesting that when students are able to connect language learning with real-world ethical issues, they engage more deeply not only intellectually but also personally. This engagement likely fosters a stronger connection between professional identity and ethical practice, a critical outcome for medical students.

6.2 In comparison with previous research findings

CALL Recent research on integrating corpus-based CALL with humanistic education in medical colleges confirms that such an approach not only enhances language proficiency but also fosters ethical reasoning, extending beyond the traditional benefits highlighted by earlier studies (Ilter, 2015; Skorczynska et al., 2016; Zhou, 2018). This nuanced contribution to CALL literature is supported by evidence showing its role in the holistic development of medical students, promoting both language skills and humanistic qualities (Bahari, 2020; Pisharady & Ravikumar, 2019). The integration also aligns with stakeholder expectations, influencing CALL adoption as suggested by Bessadok & Hersi (2023) and Al-Emran et al. (2021), and addresses educational readiness (Jeon, 2024; Park & Son, 2020). This holistic approach necessitates educators' proficiency in both linguistic and ethical pedagogy, underscoring the importance of technological pedagogical and content knowledge (TPACK) (Raygan & Moradkhani, 2022; Tseng et al., 2020). Unlike Hughes (2023), who found no difference in language outcomes between task types, this study indicates that tasks integrating humanistic elements significantly impact learning outcomes. Furthermore, the need for continuous

professional development in integrating CALL with humanistic education is emphasized (Li, 2022; Miller, 2022), supporting the potential benefits of situated learning for achieving both language and ethical competencies in medical education (Meihami, 2023).

Corpus-base Language Learning The holistic pedagogy adopted in this study that encompasses linguistic skills alongside cultural and ethical communication, aligns with the educational paradigms advocated by Baker & McEnery (2015) and Breyer (2009). Corpus-based pedagogy coupled with humanities education not only enhances language proficiency but also fosters ethical reasoning, creating a synergistic effect where authentic language resources facilitate the integration of ethical themes in student writing (Cangır, 2022; Du et al., 2022; Flowerdew, 2016). Additionally, the positive response to integrating humanity-centric corpus tools in medical education reflects favorable perceptions similar to those found among TESOL trainees (Zareva, 2016). Effective teacher-student interactions in CALL environments, crucial for the success of technology-assisted learning (Lee, Warschauer, & Lee, 2019), support the development of corpus literacy and pedagogy (Ma et al., 2021; Çalışkan & Kuru Gönen, 2018). Overall, these findings demonstrate that corpus-based CALL not only advances linguistic and professional development but also nurtures a compassionate and ethically aware medical workforce.

Humanities Education Our study addresses the lack of humanistic focus previously noted in Greek medical education by Batistatou et al. (2010) and Dimoliatis (2010). This integration responds to calls for incorporating humanistic dimensions into medical curricula, as highlighted by Kossioni et al. (2012) and aligns with O'Neill et al.'s (2020) proposals for innovative continuing professional development. Also, echoing Shapiro et al. (2009) and Tseligka and Koik (2021), our research underscores the importance of medical humanities in promoting patient-centered care and reflective practice. Despite challenges in humanities assessment in medical education noted by Ousager & Johannessen (2010), akin to those in CALL implementation, our study reveals an increasing student expectation for such interdisciplinary integration and confirms the efficacy of CALL methodologies in improving English writing skills. Moreover, the increased use of corpus-derived terms and engagement with ethical themes in our study support the necessity of English proficiency in global medical discourse (Kiosses, 2017) and suggest that our corpus-based CALL approach could significantly enhance EMP courses that combine language learning with humanistic content.

7. Significance, Limitations and Future Directions

This study's significance lies in its potential to substantiate a model of English writing pedagogy that is both technologically enriched and humanistically grounded, fostering a holistic preparatory experience for burgeoning medical professionals.

The results of the study may be hampered by its geographical limitations. Additionally, the eight-week duration may not fully capture long-term language and ethical development, and using self-generated writings to measure ethical reasoning may not encompass all competencies. Future research should extend the study duration and include preliminary digital literacy training to minimize confounding variables. Expanding interactive tasks like clinical simulations could enhance skill applicability and align the study with adult learning principles, thus improving the integration of technology, humanities education, and language learning in medical education.

References

Al-Emran, M., Mezhuyev, V. and Kamaludin, A. (2021). Is M-learning acceptance influenced by knowledge acquisition and knowledge sharing in developing countries? *Education and Information Technologies, 26*(3), 2585-2606, https://doi.org/10.1007/s10639-020-10378-y.

Apple, M. W. (2013). *Can education change society*? Routledge.

Bahari, A. (2020). Computer-mediated feedback for L2 learners: Challenges versus affordances. *Journal of Computer Assisted Learning, 37*(1), 24-38. https://doi.org/10.1111/jcal.12481

Bain, K. (2004). *What the best college teachers do*. Harvard University Press.

Baker, P., & McEnery, T. (Eds.). (2015). *Corpora and discourse studies: Integrating discourse and corpora*. Basingstoke: Palgrave Macmillan.

Batistatou, A., Doulis, E. A., Tiniakos, D., Anogiannaki, A., & Charalabopoulos, K. (2010). The introduction of medical humanities in the undergraduate curriculum of Greek medical schools: Challenge and necessity. *Hippokratia, 14*(4), 241-243.

Bax, S. (2003). CAL—past, present and future. *System, 311*, 13–28. https://doi.org/10.1016/S0346-251X(02)00071-4

Bax, S. (2011). Normalisation revisited: The effective use of technology in language education. *International Journal of Computer-assisted Language Learning and Teaching, 1*(2), 1–15.

Bessadok, A., & Hersi, M. (2023). A structural equation model analysis of English for specific purposes students' attitudes regarding computer-assisted language learning: UTAUT2 model. *Library Hi Tech.* https://doi.org/10.1108/lht-03-2023-0124

Bernardini, S. (2000). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 225–234). Frankfurt: Peter Lang.

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL, 19*(3), 307–328.

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

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). American Psychological Association.

Braun, V., & Clarke, V. (2014). Thematic analysis. In T. Teo (Ed.), *Encyclopedia of critical psychology* (pp. 1947–1952). Springer.

Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*. Advance online publication. https://doi.org/10.1080/14780887.2020.1769238

Boulton, A. & Cobb, T. (2017) Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2): 348–393. https:// doi.org/10.1111/lang.12224

Breyer, Y. (2009) Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning, 22*(2): 153–172. https://doi.org/10.1080/09588220902778328

Burr, V. (1995). *An introduction to social constructionism*. Routledge.

Çalışkan, G., & Kuru Gönen, S. İ. (2018). Training teachers on corpus-based language pedagogy: Perceptions on vocabulary instruction. *Journal of Language and Linguistic Studies, 14*(4),

190-210.

Callies, M. (2019) Integrating corpus literacy into language teacher education: The case of learner corpora. In Götz, S. & Mukherjee, J. (eds.), *Learner corpora and language teaching*. Amsterdam: John Benjamins, 245–263. https://doi.org/10. 1075/scl.92.12cal

Cangır, H. (2022). Using Corpus Tools for Academic Writing in EFL Settings: A Data-Driven Learning Approach. In G. Yangın-Ekşi et al. (Eds.), *New Directions in Technology for Writing Instruction* (pp. 171－174). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-031-13540-8_9

Chen, X., Zou, D., Xie, H. R., & Su, F. (2021). Twenty-five years of computer-assisted language learning: A topic modeling analysis. *Language Learning and Technology, 25*(3), 151-185.

Dimoliatis, I. D. (2010). The Dundee Ready Education Environment Measure (DREEM) in Greek: How it should be used, and preliminary experience in Greek medical educational environment. *Archives of Hellenic Medicine, 27*(3), 509-521.

Du, X., Afzaal, M., & Al Fadda, H. (2022). Collocation use in EFL learners' writing across multiple language proficiencies: A corpus-driven study. *Frontiers in Psychology, 13*, 752134. https://doi.org/10.3389/fpsyg.2022.752134

Egbert, J. (2005). *CALL essentials: Principles and practice in CALL classrooms. TESOL, 9* (2), xi+206.

Ellis, R. (2003). *Task-Based Language Learning and Teaching*. Oxford University Press.

Flowerdew, L. (2016). A genre-inspired and lexico-grammatical approach for helping postgraduate students craft research grant proposals. *English for Specific Purposes, 42*, 1–12. https://doi.org/10.1016/j.esp.2015.10.001

Gimeno-Sanz, A. M., Donail, C. O., & Anderson, K. (2014). Supporting content and language integrated learning through technology. In S. Jager, L. Bradley, E. J. Meima, & S. Thouësny (Eds), *CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands* (pp. 107-112). Dublin: Research-publishing.net. https://doi.org/10.14705/rpnet.2014.000203

Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. Edward Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar* (4th ed.). Routledge.

Hughes, L. S. (2023). Effects of Dynamically Computer-Mediated Communication and 'Pseudo'Communication on L2 Learning. *Computer Assisted Language Learning, 24*(1), 105-136.

Hyland, K. (2007). *Genre and second language writing*. The University of Michigan Press.

Ilter, B.G. (2015). How does technology affect language learning process at an early age? *Procedia-Social and Behavioral Sciences, 199*, 311-316. https://doi.org/10.1016/j.sbspro.2015.07.552

Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning, 37*(1-2), 1-26. https://doi.org/10.1080/09588221.2021.2021241

Johns, T. (1991) Should you be persuaded: Two examples of data-driven learning. In Johns, T. & King, P. (eds.), Classroom concordancing. *English Language Research Journal*, *4*, 1–16.

Kiosses, V. N., Tatsioni, A., Dimoliatis, I., & Hyphantis, T. (2017). "Empathize with me, doctor!":

Medical undergraduates training project: Development, application, sixmonths follow-up. *Journal of Education and Training Studies, 5*(7), 20-27. https://doi.org/10.11114/jets.v5i7.2418

Knowles, M. S. (1984). *Andragogy in Action. Applying modern principles of adult learning.* Jossey-Bass.

Kossioni, A. E., Varela, R., Ekonomu, I., Lyrakos, G., & Dimoliatis, I. D. (2012). Students' perceptions of the educational environment in a Greek dental school, as measured by DREEM. *European Journal of Dental Education, 16*(1), e73-e78. https://doi.org/10.1111/j.1600-0579.2011.00678.x

Kreitzer, R.J., & Sweet-Cushman, J. (2022). Evaluating Student Evaluations of Teaching: a Review of Measurement and Equity Bias in SETs and Recommendations for Ethical Reform. *Journal of Academic Ethics 20*, 73–84. https://doi.org/10.1007/s10805-021-09400-w

Kumar, E. S., & Sreehari, P. (2009). Computer Assisted Language Learning (CALL). In *A Handbook for English Language Laboratories* (pp. 3–14). chapter, Foundation Books. https://doi.org/10.1017/UPO9788175968677.003

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Lee, S. M. (2019). A systematic review of context-aware technology use in foreign language learning. *Computer assisted language learning, 35*(3), 294-318. https://doi.org/10.1080/09588221.2019.1688836

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization.* Oxford University Press.

Levy, M., & Hubbard, P. (2005). Why call CALL "CALL"? *Computer Assisted Language Learning, 18*(3), 143-149. https://doi.org/10.1080/09588220500208884

Li, M. (2022). *Researching and teaching second language writing in the digital age.* Palgrave Macmillan.

Lu, P.Y., & Corbett, J. (2012). *English in Medical Education: An Intercultural Approach to Teaching Language and Values.* Short Run Press Ltd.

Ma, Q., Tang, J., & Lin, S. (2021). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning.* Advance online publication. https://doi.org/10.1080/09588221.2021.1895225

McEnery, T. & Xiao, R. (2011) What corpora can offer in language teaching and learning. In Hinkel, E. (ed.), *Handbook of research in second language teaching and learning* (Vol. 2). New York: Routledge, 364–380.

Meihami, H. (2023). Situated learning in CALL teacher preparation programs: an ecological perspective to student-teachers' agency. *Computer Assisted Language Learning*, 1-27. https://doi.org/10.1080/09588221.2023.2173614

Miller, L. (2022). Computer Assisted Language Learning and Teacher Education: An Interview with Mimi Li. *International Journal of TESOL Studies, 4*(3). https://doi.org/10.46451/ijts.2022.03.12

Morgan, D. L. (1996). *Focus groups as qualitative research* (2nd ed.). SAGE Publications.

O'Neil, D., Kelly, B., O'Keeffe, S., & Moss, H. (2020). Mainstreaming medical humanities in continuing professional development and postgraduate training. *Clinical Medicine, 20*(2), 208-211. https://dx.doi.org/10.7861%2Fclinmed.2019-0332

Ousager, J., & Johannessen, H. (2010). Humanities in undergraduate medical education: A literature review. *Academic Medicine, 85*(6), 988-998. https://doi.org/10.1097/acm.0b013e3181dd226b

Park, M., & Son, J. B. (2020). Pre-service EFL teachers' readiness in computer-assisted language learning and teaching. *Asia Pacific Journal of Education, 42*(2), 320-334. https://doi.org/10.1080/02188791.2020.1815649

Parmaxi, A. (2020). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive learning environments, 31*(1), 172-184. https://doi.org/10.1080/10494820.2020.1765392

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage Publications.

Pennycook, A. (2001). *Critical applied linguistics: A critical introduction*. Lawrence Erlbaum Associates.

Pérez-Paredes, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011－2015. *Computer Assisted Language Learning, 35*(1-2), 36-61. https://doi.org/10.1080/09588221.2019.1667832

Pisharady, A. S., & Ravikumar, J. G. (2019) *Effectiveness of Mobile Assisted Language Learning In Developing the Four Skills*. IOSR Journal Of Humanities And Social Science, 24 (9), 55-58.

Raygan, A., & Moradkhani, S. (2022). Factors influencing technology integration in an EFL context: investigating EFL teachers' attitudes, TPACK level, and educational climate. *Computer Assisted Language Learning, 35*(8), 1789-1810. https://doi.org/10.1080/09588221.2020.1839106

Rest, J. R. (1986). *Moral Development: Advances in Research and Theory*. Praeger.

Schwandt, T. A. (1998). Constructivist, interpretivist approaches to human inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research: Theories and issues* (pp. 221－259). Sage Publications.

Shadiev, R., Liu, T., & Hwang, W. Y. (2019). Review of research on mobile-assisted language learning in familiar, authentic environments. *British Journal of Educational Technology, 51*(3), 709-720. https://doi.org/10.1111/bjet.12839

Shadiev, R., & Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. *Sustainability, 12*(2), 524. https://doi.org/10.3390/su12020524

Shapiro, J. (2010). Take 2 poems and call me in the morning: Reading medically themed poetry for insight and renewal. *The Journal of Clinical Outcomes Management, 17*(11), 1-6.

Shapiro, J., Coulehan, J., Wear, D., & Montello, M. (2009). Medical humanities and their discontents: Definitions, critiques, and implications. *Academic Medicine, 84*(2), 192- 198. https://doi.org/10.1097/acm.0b013e3181938bca

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2023). Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning, 36*(3), 517-554. https://doi.org/10.1080/09588221.2021.1933540

Sinclair, J. (1991) *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Skorczynska, H., del Saz Rubio, M., Carrió-Pastor, M.L. (2016). Second Language Teaching and Technology. An Overview. In: Carrió-Pastor, M. (eds) *Technology Implementation in Second Language Teaching and Translation Studies*. New Frontiers in Translation Studies. Springer, Singapore. https://doi.org/10.1007/978-981-10-0572-5_2

Stockwell, G. (Ed.). (2013). *Computer-assisted language learning: Diversity in research and*

practice. Cambridge University Press. https://doi.org/10.1111/ijal.12032

Terry, G., Hayfield, N., Braun, V., & Clarke, V. (2017). Thematic analysis. In C. Willig & W. S. Rogers (Eds.), *The SAGE handbook of qualitative research in psychology* (pp. 17–37). Sage Publications.

Tseligka, T., & Koik, K. (2021). The Becoming of a Medical Professional: Integrating Humanities-Based Resources in a Medical English Course. *E.S.P. Today, 9*(1), 50-62. https://doi.org/10.18485/esptoday.2021.9.1.3

Tseng, J. J., Chai, C. S., Tan, L., & Park, M. (2020). A critical review of research on technological pedagogical and content knowledge (TPACK) in language teaching. *Computer Assisted Language Learning, 35*(4), 948-971. https://doi.org/10.1080/09588221.2020.1868531

Wu, Y. (2019). New missions, broad perspectives, new liberal arts, and extensive foreign languages. *Frontiers of Foreign Language Education Research, 2*(2), 3-7+90.

Zareva, A. (2016). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*. Advance online publication. http://eltj.oxfordjournals.org/

Zhang, L., & Li, Y. (2022). Challenges and strategies in teaching English writing in Chinese medical schools. *Journal of Medical Education*, 36(2), 123-130.

Zhou, Z. H. (2018). Second Language Learning in the Technology-mediated Environments. *Asian Education Studies, 3*(1), 18-28, https://doi.org/10.20849/aes.v3i1.307

Appendix 1 *Assessment Tasks for Pre- and Post-Intervention*

Pre-intervention English Writing Test on Medical Topics

Instructions: You will be given a prompt related to a current medical issue. Carefully read the prompt and understand the requirements. Write an essay of 500-600 words discussing the issue. Your essay should include:

1) A clear introduction stating your understanding of the issue.

2) A detailed discussion of the implications of the issue, including potential solutions.

3) Reflection on the ethical aspects involved.

4) A concise conclusion summarizing your viewpoints and solutions.

Post-intervention Ethical Dilemma Evaluation

Instructions: Write a response that outlines how you would handle the dilemma. Your response should include:

1) A brief description of the dilemma.

2) A detailed explanation of your decision-making process.

3) Justification for your decisions based on ethical principles and professional standards.

Rubric for Assessment

Criteria	Excellent (4)	Good (3)	Satisfactory (2)	Needs Improvement (1)
Content Comprehension	Demonstrates comprehensive and accurate understanding of medical and ethical issues.	Shows good understanding with minor inaccuracies.	Basic understanding shown with some relevant details missing.	Lacks understanding of the topic; many inaccuracies.
Analytical Ability	Provides insightful analysis, identifies all key issues, and proposes well-considered solutions.	Good analysis and identification of major issues with adequate solutions.	Some analysis present but lacks depth or misses some issues.	Minimal or no analysis; solutions are superficial or not relevant.
Language Proficiency	Uses precise, professional language effectively with excellent grammar and organization.	Generally clear language with minor errors in grammar or organization.	Language use sometimes unclear or disorganized; noticeable errors.	Poor language use with frequent grammatical and organizational errors.
Ethical Reasoning	Articulates nuanced ethical considerations and aligns closely with professional standards.	Good understanding of ethics with appropriate application.	Some ethical considerations are addressed but lack depth.	Fails to address ethical considerations adequately.
Originality and Depth	Offers original insights and deep reflection, enhancing understanding of the topic.	Some original thoughts and adequate depth.	Limited originality and depth; mostly surface-level reflection.	Lacks originality; very basic or generic responses.

English Language Education

英语教育

The Development of Vocational Education and the Construction of Teaching Staff under the Background of Digital Transformation

Bi Lihua

School of General Education, Jinan Engineering Polytechnic, Jinan 250200, China

Keywords: Vocational Education, Teaching Staff, Digital Transformation

Abstract: Digital transformation is crucial across industries, including vocational education, to drive national economic and social progress. Adapting vocational education and nurturing teaching staff are critical components in enhancing vocational education amidst digital transformation. Despite opportunities and challenges in recent years, vocational education must integrate emerging technologies such as AI and big data to meet evolving job market demands. Also, fostering a competent teaching staff is essential for delivering professional guidance and personalized education, promoting comprehensive student development.

Moreover, this study aims to explore vocational college educators' digital transformation journey, guiding digital teaching practices and professional growth. Through examining educators' adaptation to digital changes, strategies for overcoming challenges and leveraging technology to enhance teaching effectiveness can be identified, ultimately elevating vocational education quality and driving teaching reform.

Furthermore, employing a mixed-methods approach, this study combines qualitative and quantitative research. Qualitative methods explore teachers' digital transformation challenges and pathways, gathering their experiences and perspectives through interviews, observations, and similar means. Quantitative methods quantify teachers' current status and effects of digital transformation through questionnaire surveys and statistical analysis, providing comprehensive data support and objective conclusions.

In conclusion, this study underscores the significant impact of teacher training and professional development plans on digital transformation. Personalized training approaches and resource support facilitate enhanced digital teaching practices, while practice and feedback mechanisms are crucial for improving teaching abilities and refining methods.

1.Introduction

Digital transformation is crucial for all industries, including vocational education. Integrating emerging technologies like AI and big data into education is necessary to meet future job market demands. Developing vocational education and building a skilled teaching staff are vital for improving vocational education under digital transformation. High-quality teachers can provide professional guidance and personalized education, promoting comprehensive student development. This article discusses strategies to address the challenges of digital transformation on the teaching staff. It emphasizes the importance of teacher training in the digital transformation process and discusses cultivating vocational education teachers with digital foundations and innovative capabilities. The aim is to promote the improvement and development of vocational education.

2.Research Status in China and Abroad

2.1. Research Status in China

Under the current background of digital transformation, research and practice on the enhancement path of vocational education have become hot topics in the academic community and educational practice. Domestic scholars have also conducted a series of research in this field, proposing some insightful views and measures from the perspectives of curriculum design, teaching methods, and student ability development that can be used for reference.

On the one hand, domestic scholars have proposed innovations in core competency development. For example, it has been pointed out that comprehensive ability assessment should be based on task and project forms, which are more in line with the requirements of modern industries for vocational education. At the same time, there is also a need to optimize the cultivation system of practical operation skills by combining course theory and practice to cultivate high-quality talents that meet the needs of modern markets (Yu, 2022)[1].

On the other hand, some scholars are researching innovative teaching models for vocational education under digitization. For instance, studying the influence of digital technology on vocational education teaching modes, attempting to establish a guidance system for digital technology education models, and achieving deep integration of digital education and interactive teaching to promote the improvement of vocational education training levels (Deng, 2023)[2].

In summary, research and practice concerning advancing vocational education should prioritize practical, theoretical, and policy innovations amidst digital transformation. There is a pressing need to explore new models, such as industry-education integration, online education,

and the application of digital technologies, in order to offer effective pathways and benchmarks for standardizing and modernizing vocational education.

2.2. Research Status at Abroad

Firstly, emphasis should be placed on promoting industry-university-research cooperation and building school-enterprise cooperation platforms. Some Western universities, such as the Warsaw University of Technology in Poland, have improved vocational education curriculum design by focusing on practical training courses through cooperation with enterprises, which enhances students' practical abilities and better adapts to the needs of modern economic development (Peres & Amparo, 2020)[3].

Secondly, innovative teaching methods should be emphasized to enhance students' core competencies. Some American universities have designed more flexible and innovative curriculum systems by adopting digital technologies, such as online and intelligent teaching methods, which improve the quality of vocational education and students' core competencies, thereby better serving social development (Means, Bakia, & Murphy, 2014)[4].

Lastly, teacher training should be prioritized, and a diverse teaching team should be established. Some universities, such as Lucerne University of Applied Sciences and Arts in Switzerland, have introduced international teaching teams to guide teachers to improve their teaching standards and international education perspectives and promote the development of vocational education (Arnot & Reay, 2019)[5].

In summary, research and practices on promoting vocational education under digital transformation require strengthening educational concepts and innovation in educational work. Continuous improvement of teaching quality and cultivation of innovative capabilities could better serve modern economic development and social progress.

3. Digital Transformation under the background of Vocational Education

3.1. The Concept and Characteristics of Digital Transformation

3.1.1. Concept of Digital Transformation in Vocational Education

Digital transformation in vocational education refers to the comprehensive reconstruction and upgrading of vocational education's goals, content, methods, and modes through the application of digital technology to promote its development towards digitization, networking, and intelligence.

3.1.2. Characteristics of Digital Transformation

(1) Diversification of Teaching Methods: Digital transformation has diversified teaching methods, such as online courses and live broadcasts, improving students' learning outcomes. Schools provide online learning platforms, allowing students to learn anytime, anywhere, and choose the best learning method.

(2) Personalized Customization: Digital transformation enables personalized education, tailoring courses to students' interests, learning levels, and career needs. Our institution focuses on

3

students' personal development and provides personalized teaching plans and course selection counseling.

(3) Increased Practical Operations: Digital transformation enhances practical education through virtual simulation experiments and laboratory courses, improving students' practical experience and skills. Our institution has advanced equipment and virtual simulation laboratories to prepare students for their careers.

(4) Sharing of Educational Resources: Digital transformation allows for worldwide sharing of educational resources, improving resource utilization. Our institution collaborates with other universities and institutions to share high-quality teaching resources through an online education resource library.

(5) Optimization of Learning Evaluation: Digital transformation enables precise student evaluation through data analysis and intelligent evaluation, improving learning outcomes and evaluation accuracy. Teachers can adjust teaching based on evaluation results, helping students master knowledge and skills (Xue, 2019)[6].

4. Impacts of Digital Transformation on Vocational Education

The impacts of digital transformation on vocational education can be divided into four aspects:

(1)Improved Teaching Efficiency and Quality

Digital transformation has led to advanced educational technologies, including intelligent and virtual platforms, which enhance teaching efficiency and quality. These technologies enable the sharing of resources, diversify teaching methods, and improve learning effectiveness through online interaction and practice.

(2)Promoting Changes in the Job Market

Digital transformation in vocational education impacts the job market through industry upgrades, expanded employment opportunities, and improved employment capabilities. It accelerates informationization and automation, expands job choices with emerging professions, and enhances students' practical skills and employability.

(3)Promoting Upgrades in Education Services

Digital transformation in vocational education improves education services through precision, efficiency, and intelligence. It enables precise delivery based on data analysis and artificial intelligence, improves efficiency through online and automated services, and advances personalized education services using intelligent technologies.

(4)Strengthening Quality Control in Talent Cultivation

Digital transformation in vocational education strengthens quality control by cultivating core competencies, enhancing evaluation indicators, expanding learning channels, and improving teaching quality. It determines core competencies based on market demand, uses digital technology for objective evaluation, expands learning resources and methods, and enhances teaching effectiveness.

5.Historical Review of the Development of Vocational Education

Vocational education, as a form of education cultivating talents with specific vocational skills and knowledge, has gone through several stages of development, summarized as follows:

(1)Ancient Vocational Education

The representative of ancient vocational education is the apprenticeship system for handicrafts in China, which has existed for over 2000 years and has had a significant impact on the development of Chinese society. The handicraft apprenticeship system has trained many artisans and technicians and protected their interests (Jia & Nan, 2023)[7].

(2)Vocational Education during the Industrial Revolution

During the Industrial Revolution, vocational education received more attention. Vocational education began to spread in Europe, driven by the Industrial Revolution. In the early 1800s, Germany established the earliest vocational-technical school, the Berlin Craftsmanship Institute. Meanwhile, countries like the United Kingdom and France also established vocational-technical schools. These schools aimed to train technical personnel for factories and enterprises, directly serving industrial production (Wang & Zhang, 2022)[8].

(3) Development Stages of Public Vocational Education

Public vocational education began to develop rapidly in the first half of the 20th century. Many countries established vocational-technical and trade schools to cultivate the labor force and improve their skills. The United States was the first country to implement public vocational education, and its vocational education system had a far-reaching impact worldwide. Japan also began implementing public vocational education in the first half of the 20th century and became a model for vocational education in Asia due to its high-quality training and good results.

(4) Modernization Stage of Vocational Education

Since the second half of the 20th century, vocational education has continued to develop and modernize. The development of industrial automation and informatization has led to the renewal and change of vocational education content. Many countries began to attach importance to the renewal and reform of vocational education, closely integrating education with industrial demand. For example, Germany's "dual system" vocational education model, which combines an apprenticeship system and practical experience, has provided Germany with a large number of highly skilled workers with practical operating skills (Wang & Zhang, 2022)[8].

6.The Impact of Digital Transformation on the Vocational Education Faculty

Digital transformation has become an inevitable trend in modern society with the rapid development of digital technology. The deepening of digital transformation has brought significant changes in the industrial sector and new opportunities and challenges to the education sector. In vocational education, the impact of digital transformation on the vocational education faculty is significant. This article will elaborate on several aspects as follows.

(1)Digital transformation enhances the informationization level of the vocational education faculty

Digital technologies can quickly obtain, process, and transmit information, providing the vocational education faculty with more abundant and comprehensive educational resources. According to Zhao et al.'s (2023)[9]research, in highly developed information technology, the vocational education faculty must master various information technology skills, such as information retrieval, content management, and teaching platforms. Through digital transformation, the vocational education faculty can access a broader range of information channels and more efficient information transmission methods, improving their information processing and utilization abilities to better adapt to teaching needs in the digital age.

(2)Digital Transformation Strengthens the Vocational Competence of the Vocational Education Faculty

Digital transformation brings more teaching resources to the vocational education faculty and promotes innovation in education methods and approaches. The advantages of digital education lie in its interactivity, flexibility, and personalized customization, all of which require teachers to have higher vocational competence. Hao (2023)[10] pointed out that digital transformation requires the vocational education faculty to possess more vital systematic thinking, creative ability, global vision, and interdisciplinary knowledge. Improving these vocational competencies will help the vocational education faculty better respond to the vocational education needs of the new era.

(3)Digital Transformation Enhances Teaching Efficiency and Quality of Vocational Education Faculty

Digital transformation provides more teaching support and resources for the vocational education faculty while improving teaching efficiency and quality. In recent years, digital technology has been widely used in vocational education, including robot teaching and virtual simulation experiments (Miao, 2023)[11]. Digital teaching can improve teaching efficiency and quality for the vocational education faculty and better meet the needs of students. For example, in the practical part of vocational education, digital technology can help simplify classroom operations, reduce artificial operation errors, and improve teaching effectiveness.

In the digital age, vocational education faculty must possess information skills, vocational competence, and teaching skills. They should master modern educational technologies, understand emerging industries, and have diverse teaching and evaluation methods. Flexibility in designing teaching is also essential. Additionally, digital transformation demands continuous improvement of professional knowledge and teaching skills, active participation in educational reforms and practices, and a deep understanding of enterprise and market demands to ensure sustainable talent cultivation.

Therefore, in response to the need for vocational education transformation and talent cultivation in the digital age, the vocational education faculty needs to continuously innovate,

develop, and improve their information skills, vocational competence, and teaching skills in order to better adapt to the development of vocational education in the digital age.

7. The Current Situation and Problems of Vocational Education in the Digital Age

7.1. Presentation of Survey Questionnaire Results

The survey used QuestionStar and was promoted through various channels to ensure the accuracy and reliability of the data. One hundred twenty-five valid questionnaires were collected and analyzed to understand vocational college teachers' perceptions and needs regarding digital transformation.

The findings revealed that most vocational college teachers have extensive teaching experience and may encounter challenges in digital transformation. They generally acknowledge the importance of digital teaching in higher education, although the frequency of its use varies among teachers. While some teachers possess a certain level of understanding and skills in digital teaching, others require more training and support. The main obstacles identified include a lack of exercise and support, inadequate equipment and resources, resistance to change, and a mismatch between the curriculum and digital teaching approaches.

Regarding their expectations and suggestions for digital transformation measures, teachers expressed a desire for professional training, access to equipment and software tools, teaching resource libraries, collaborative platforms, and support and guidance. Vocational colleges are advised to provide training opportunities, support teams, subsidies for equipment purchases, innovation laboratories, and research support as measures to address these needs.

Successful digital transformation in vocational education entails improving teacher skills, enhancing student learning experiences, fostering collaboration, promoting management reforms, and cultivating a culture of transformation. Teachers aspire to drive innovation, enhance their capabilities, receive resource support, establish evaluation mechanisms, strengthen collaboration, and implement training programs. Vocational colleges play a vital role in providing support to facilitate the application of digital teaching methods and enhance overall teaching quality.

7.2. The Current Situation of Vocational Education in the Digital Age

(1)The widescale application of digital technology requires transforming and upgrading vocational education.

As digital technology is widely applied, vocational education must transform and upgrade to meet social development demands. In the current vocational education system, problems often exist, such as an outdated curriculum, inadequate teaching staff, and a monotonous teaching method that cannot meet the needs of many students.

(2)Mismatch between vocational education and industry demand

The mismatch between vocational education and industry demand is one of the most prominent problems in the current vocational education system. Many enterprises have reported

7

that due to a lack of practical work experience, most graduates cannot undertake actual job positions, leading to a high vacancy rate for job positions (Cheng, 2023)[12].

7.3. Problems of Vocational Education in the Digital Age

(1) Insufficient Construction of Digital Education Platforms

In the current digital age, vocational education needs to rely on digital education platforms to achieve informatization construction. However, there are still problems with inadequate platform construction, weak functions, and insufficient resources, especially for some small and medium-sized schools whose digital education platform construction is even weaker.

(2) Gap between Vocational Education and Practical Application

In the existing vocational education system, many courses and teaching content are still at the theoretical level, which cannot meet the needs of practical application. It also results in many graduates spending much time adapting to the work environment when looking for a job, which burdens enterprises.

(3) Low Specialization Level of Teacher Team

In the current vocational education system, the level of specialization of the teacher team varies greatly. Although some teachers are very familiar with their professional knowledge, they may be rigid in teaching methods and textbook selection and unable to guide students to think creatively. It also constrains students' career development to some extent.

8.Recommendations and Conclusions

8.1 Recommendations

In the context of digital transformation, here are several suggestions to enhance vocational education and the teacher workforce:

(1)Strengthen the construction of digital education platforms: Schools can increase investment in building more sophisticated, powerful, and abundant digital education platforms. It will enhance the quality of digital technology for teaching and learning and promote the informationization of vocational education.

(2)Enhance the setting of practical courses: Vocational education should establish more practical courses based on practical applications and vocational demands and gradually shift the teaching focus from theoretical to practical aspects. Through rich practical teaching, students can better adapt to working environments and develop self-innovation abilities.

(3)Improve the teacher workforce's professional level: Strengthen their training and cultivation and improve their teaching level and quality. At the same time, schools should recruit more teachers with practical experience and professional skills, encourage and guide teachers to participate in industry training and seminars, and constantly improve their professional level.

(4)Promote curriculum system reform: To meet evolving vocational demands, schools should promote curriculum system reform by updating and improving curriculum settings in vocational education. They should establish a communication mechanism with employers, listen to feedback, and optimize teaching content accordingly.

8.2 Conclusions

In short, with the continuous deepening of digital transformation, vocational education faces excellent opportunities and challenges. In this context, it is necessary to strengthen the construction of digital education platforms, enhance practical course settings, improve the professional level of the teaching staff, and promote curriculum system reform. Vocational education should constantly innovate, closely integrate industry needs and market changes, and cultivate new professional talents that meet economic development needs. At the same time, in the construction of the teaching staff, more attention needs to be paid to improving the practical ability and professional skills of teachers, continuously optimizing the quality of teaching to serve the development of vocational education better, and promoting digital transformation.

References

[1]Yu, S. (2022). Challenges and Responses of Characteristic Cultivation in Digital Transformation of Vocational Education. Shanxi Science and Technology Daily, B06.

[2]Deng, X. (2023). Theoretical Logic and Practical Strategies of Digital Transformation in Vocational Education. Educational Research on Electronics, 48-53.

[3]Peres, P., & Amparo, C. (2020). The Role of University-Industry Collaboration in Fostering Innovation-Led Growth in the European Union: A Critical Review of the Literature. Sustainability, 12(2), 703.

[4]Means, B., Bakia, M., & Murphy, R. (2014). Learning online: What research tells us about whether, when and how. Routledge.

[5]Arnot, M., & Reay, D. (2019). International Teachers in Education: The Responsive Challenge. Globalisation, Societies and Education, 17(4), 604-611.

[6]Xue, X. (2019). Development and Transformation of Vocational Education in China in the DigitalEconomy Era. Information and Communications Technology and Policy, 42-44.

[7]Jia, M., & Nan, H. (2023). The Historical Logic of the Century-long Changes in Chinese Modern Vocational Education: Based on the Perspective of New Institutionalism. Journal of Hebei University (Philosophy and Social Sciences Edition), 106-115.

[8]Wang, Z., & Zhang, L. (2022). A Construction of Skill-oriented Society from the Perspective of Bioethics: A Cultural-Historical Analysis of German Vocational Education. Journal of Shanxi University (Philosophy and Social Sciences Edition), 128-138.

[9]Zhao, M., & Su, G. (2023). Knowledge Landscape and Advancement Path of Research on Digital Transformation of Vocational Education. Jiangsu Vocational and Technical Education, 1-10.

[10]Hao, T. (2023). The Digital Transformation of Vocational Education: Why and How to Transform. Vocational Education Communication, 5-10.

[11]Miao, J. (2023). Application Analysis of "Holographic + 5G" Technology in the Digital Transformation of Vocational Education. Jiangnan Forum, 81-84.

[12]Cheng, Y. (2023). Discussion on Employee Career Sustainable Development under Digital Transformation. Cooperative Economy and Science & Technology, 94-96.

A Study on the Effects of Shadowing Exercise on English Listening Teaching Efficiency and Learning Motivation for Senior High School Students

Chen Zhenhong SouthWest Petroleum University

Abstract: Shadowing exercise, as an internationally recognized effective English interpretation training method, has been proven to enhance listening and speaking skills in a short period. In the current high school English curriculum, listening comprehension is often overlooked, leading to insufficient listening abilities and lackluster English learning motivation among students. To address this issue, this study introduces the shadowing exercise method. The study employed a quasi-experimental pre- and post-test control group design. The participants included 87 high school students from Tianchang Industrial School, divided into two parallel classes: 43 students in the experimental group and 44 in the control group. The study primarily conducted teaching interventions for a duration of 12 weeks. The experimental group underwent shadowing exercise training, while the control group retained traditional listening teaching methods. This research incorporated a mixed-methods approach, combining quantitative and qualitative data collection and analysis. Prior to and following the experiment, the students' listening performance and learning motivation were assessed. SPSS26.0 was used to analyze the quantitative data, and six students were interviewed at the conclusion of the study to supplement the quantitative findings. The results indicate that compared to traditional listening teaching methods, shadowing exercise is a more effective approach to listening instruction. Students in the experimental group demonstrated a faster improvement in English listening skills and overall better performance than their counterparts in the control group. Furthermore, the listening motivation of students in the experimental group was significantly enhanced. Students reported that shadowing exercise enabled them to concentrate more intently on listening training.

Key words: Shadowing Exercise; English Listening Teaching; Learning Motivation

1. Introduction

Listening comprehension is one of the crucial question types in the English college entrance examination. It serves as a significant means to assess students' listening proficiency and capability. According to the "New Curriculum Standard for Senior High School English" (2020

Revised Edition), the purpose of listening instruction is to cultivate students' ability to acquire information and apply English to solve real-life problems in daily situations. Listening proficiency constitutes an essential component of core language competence in English. However, in actual teaching practice, teachers often do not prioritize listening classroom instruction and assessment as much as they do other language skills.

The traditional classroom teaching model of "listening to recordings, doing exercises, and checking answers" fails to improve students' basic listening skills or enhance the effectiveness of classroom instruction. It also does not effectively boost students' motivation for English listening. Implementing a listening classroom teaching approach based on shadowing training requires students to simultaneously listen to the material while engaging in synchronous shadow reading of the original script. During this process, they actively perceive fundamental phonetic knowledge such as pronunciation, intonation, tone, pauses, stress patterns, linking, and sentence breaks within the dialogue. Through repeated shadowing exercise until complete mastery is achieved, students can store acquired knowledge in their brains through extensive imitation and accurately reproduce it when needed. This method demands high levels of concentration and memory from students but effectively addresses the shortcomings of traditional listening practice models.

The utilization of shadowing exercise in high school English listening teaching is not only beneficial for enhancing students' phonetic knowledge, but also for strengthening fundamental listening skills such as concentration and memory, thereby fostering their motivation to learn listening. Therefore, it is imperative and innovative to explore the application of the shadowing exercise method in senior high school English listening instruction.

2. Theoretical Basis

According to Bandura (2003), a renowned American psychologist, the social learning theory posits that observational learning is at the core of acquiring knowledge. This process involves learners observing exemplary behaviors exhibited by others or role models and subsequently imitating those behaviors. Krashen (1985), a famous American linguist, proposed the input hypothesis theory which suggests that language acquisition occurs when learners are exposed to "comprehensible input" slightly above their current language skill level. By focusing on understanding meaning and information rather than form, learners can acquire the second language. Shadow tag-reading is a learning method based on both observation learning theory and

input hypothesis theory as it mimics the form and structure of the target language.

Learning motivation refers to individuals' tendency to initiate and sustain their own learning behavior towards specific goals. It serves as an internal drive for learners' educational pursuits. Brown (1994) views learning motivation as learners' efforts in achieving goals throughout the learning process, while Dornyei (1998) argues that it encompasses actions that persist until desired objectives are accomplished without interference. Gardner & Mac Intyre (1993) propose that motivation consists of three components: desire for goal achievement, efforts exerted towards attaining those goals, and satisfaction derived from accomplishing them.

3. Methodology

3.1 Research problems

The aim of this study is to investigate the impact of shadowing exercise method on high school students' listening proficiency and their motivation towards English learning. The main objectives are as follows:

(1) What is the effect of employing shadowing exercise method on high school students' listening performance?

(2) How does the utilization of shadowing exercise methods influence high school students' motivation in learning English?

3.2 Research object

The study involved 87 students from two Grade 1 parallel classes at a high school in Tianchang City, Anhui province. One class was designated as the experimental group while the other served as the control group. The experimental group comprised of 43 participants, including 23 boys and 20 girls. The control group consisted of 44 individuals, with 24 boys and 20 girls. Their average age is approximately 16 years old. Both parallel classes are at an equivalent level and instructed by the same English teacher. The control group followed a traditional listening teaching method involving listening to recordings, writing answers, and providing responses accordingly. In contrast, the experimental group employed shadowing exercise method for their listening instruction.

3.3 Research method

This study employs a research methodology that integrates both quantitative and qualitative approaches.

Quantitative research primarily involves empirical investigations conducted in the experimental and control groups, including pre- and post-tests on listening performance as well as pre- and post-tests on the listening learning motivation questionnaire. The listening performance tests utilize the listening sections from previous year's midterm and final exams for senior one students to ensure consistency in question types, difficulty level, and prevent question repetition. The test comprises four question types: short dialogue comprehension (five multiple-choice questions), long conversation understanding (five multiple-choice questions), essay comprehension (five multiple-choice questions), and information conversion (five fill-in questions). Each question is worth one point, resulting in a total score of twenty points. The test possesses moderate difficulty suitable for assessment purposes. The English Learning Motivation Questionnaire is adapted from Gardner's (1993) scale on English learning motivation for college students, consisting of 20 items categorized into four dimensions: learning purpose (items 1-5), learning desire (items 6-10), learning attitude (items 11-15), and degree of effort (items 16-20). The questionnaire will be administered before and after the teaching intervention.

Qualitative research mainly involves semi-structured interviews aimed at supplementing questionnaire data by exploring aspects not covered by it. These interviews aim to elucidate internal factors contributing to significant differences observed in students' English listening performance.

3.4 Research procedures

Before conducting the experiment, specifically during the first week of this semester, a random sampling method was employed to select 50 students for testing the reliability and validity of the English learning motivation questionnaire. Once the standards were met, two classes were assigned and tested accordingly. Subsequently, the listening test from last year's first-year English midterm exam was administered as a pre-test during that same week. The official commencement of the experiment took place in the second week and lasted until the ninth week of this semester. Shadowing and reading-based listening instruction were implemented on a weekly basis according to unit progression. The control group received traditional listening instruction while an experimental group underwent approximately 10 minutes of intervention per unit, which included activities such as sentence retelling, imitation and reading exercises, instant translation practice, and short passage retelling aimed at enhancing basic listening skills. Following completion of the

experiment in Week 11, post-tests were conducted to assess both listening performance and motivation levels using identical questionnaires as those used in pre-testing stages; additionally, senior high school final exam listening tests from last academic year served as post-test materials. Finally, six students from each class were randomly selected for semi-structured interviews conducted face-to-face; these sessions were audio recorded for subsequent transcription into text format for analysis purposes.

Finally, the data are sorted out, and SPSS 26.0 is used to analyze and discuss the data.

4. Results and Discussion

4.1 Effect on English listening performance

Table 4-1

	Mean	N	Std.
CC pretest	11.85	44	1.436
CC posttest	12.59	44	1.627
EC pretest	12.10	43	1.748
EC posttest	14.25	43	2.196

Table 4-2

			Levene's Test for Equality of Variances		t-test for Equality of Means		
			F	Sig.	t	df	Sig.(2-tailed)
Pretest	Listening Scores	Equal variances assumed	1.658	0.200	0.849	120	0.398
		Equal variances			0.849	115.624	0.398

		not assumed					
Posttest		Equal variances assumed	5.659	0.019	4.732	120	0.000
		Equal variances not assumed			4.732	110 . 609	0.000

Table 4-3

		Paired Differences					
		Mean	Std.Deviation	Std.Error Mean	t	df	Sig.(2-tailed)
EC	Pre&post	-2.148	1.558	0.200	-10.746	60	0.000
CC		-0.738	1.182	0.151	-4.875	60	0.000

Analysis: According to Table 4-1, the descriptive statistics reveal that the mean score of pre-test English listening in the experimental group is 12.1 points, while in the control group it is 11.85 points, resulting in a difference of 0.25 between the two groups. In terms of post-test English listening level, it was observed that the average score for the experimental group was 14.25, whereas for the control group it was 16.59; thus indicating an average score difference of 1.66 between both groups. Furthermore, it was found that there was a significant increase of 2.15 points in post-test scores compared to pre-test scores for the experimental group and a rise of 0.74 points for the control group.

Referring to Table 4-2, at a significance level of 0.05, it can be noted that Sig value for both experimental and control groups regarding pre-test scores is greater than 0.05 (Sig = .398). Consequently, this suggests no significant disparity exists in pre-test scores between both groups and their overall English listening levels are essentially similar.

On analyzing post-test scores with respect to significance level (Sig = .000), which is lower than 0.05, as indicated by Table-4-2 results; hence clearly demonstrating a substantial distinction between these two specific classes where performance by students from experimental group surpasses those from control group.

Based on findings presented in Table-4-3 using paired sample size T-tests at α = .05 significance level; statistical analysis reveals p-values less than .05 were obtained for both pre-test and post-test within experimental group suggesting noteworthy differences exist before and after teaching intervention with higher scores achieved following intervention implementation. In addition, the experimental group scored 2.15 higher than the control group. Therefore, this paper believes that the shadowing exercise teaching method is more effective in improving students' English listening performance.

4.1 Effect on English learning motivation

Table 4-4

	Mean	N	Std.
CC pretest	2.92	44	0.856
CC posttest	2.93	44	0.853
EC pretest	2.91	43	0.703
EC posttest	3.38	43	0.567

Table 4-5

			Levene's Test for Equality of Variances		t-test for Equality of Means		
			F	Sig.	t	df	Sig.(2-tailed)
Pretest	Listening Scores	Equal variances	2.074	0.153	0.064	98	0.949

		assumed					
		Equal variances not assumed			0.064	94.448	0.949
Posttest		Equal variances assumed	1.238	0.268	-3.128	98	0.002
		Equal variances not assumed			-3.128	85.27	0.002

Table 4-6

		Paired Differences					
		Mean	Std.Deviation	Std.Error Mean	t	df	Sig.(2-tailed)
EC	Pre&post	-0.465	0.382	0.054	-8.619	49	0.006
CC		-0.002	0.0749	0.011	-0.189	49	0.851

Analysis: As observed from the table 4-4, prior to commencing the experiment, the average level of learning motivation among students in both the control and experimental classes was 2.92 and 2.91 respectively, both falling below the moderate level (according to a five-level Likert scale where 4 represents moderate motivation), indicating a generally low level of English learning motivation among students in both classes. Following the conclusion of the experiment, there were changes in the motivation levels of students in both classes: The mean motivation level for students in the experimental class increased to 3.38, while it remained at 2.93 for students in the control class, suggesting an overall improvement in learning motivation among students in the experimental class. The findings of this study are discussed alongside interviews conducted with

participants: Firstly, compared to primary and middle school levels, English language learning becomes more challenging and less engaging at high school level; thus most students lack interest towards English language acquisition. Under an exam-oriented education system backdrop, traditional teaching methods tend to be monotonous with limited opportunities for active student participation during classroom activities; consequently leading to a gradual decline in their learning motivation levels. Lastly, many learners fail to grasp a correct understanding of their purpose behind studying English; relying on coping mechanisms solely driven by external motivations such as preparing for college entrance examinations – which are inherently unstable motivators – potentially contributing towards fluctuations within learners' motivational levels. The application of shadow reading intervention method has resulted in an improved average motivational level among students within the experimental class; thereby signifying its positive impact on enhancing learning motivation.

It can be observed from the table 4-5 that the significance value of the F test prior to the experiment exceeds 0.05, indicating compliance with homoscedasticity assumption. Considering the t value in the first row, a significance value greater than 0.05 for the T-test accepts the null hypothesis, suggesting no significant difference between both samples. Subsequently, after conducting the experiment, both F test and t test exhibit significance values exceeding 0.05, thereby satisfying homoscedasticity assumption as well as rejecting null hypothesis respectively. This implies a significant disparity in learning motivation levels between students in control class and experimental class post-experimentation. Prior to experimentation, there was no notable distinction between both samples; however, following experimentation, a substantial discrepancy emerged signifying alterations in data.

As shown in the table 4-6, the paired sample t-test for the control group yielded a Sig. value greater than 0.05, indicating no significant difference in learning motivation levels between pre- and post-experiment phases. Conversely, the experimental group's paired sample t-test produced a Sig. value less than 0.05, signifying a noteworthy disparity in learning motivation levels before and after intervention; furthermore, negative mean differences suggest an improvement of students' learning motivation following shadow reading intervention implementation during high school English instruction. In conclusion, integrating shadow reading interventions into high school English teaching can enhance learners' overall engagement.

Table 4-7

		Mean±Std	t	p
CC	learning purpose	0.002±0.07	0.189	> 0.05
	learning desire	0.001±0.05	0.161	> 0.05
	learning attitude	0.004±0.08	0.215	> 0.05
	degree of effort	0.001±0.05	0.161	> 0.05
EC	learning purpose	0.005±0.29	0.643	> 0.05
	learning desire	0.46±0.37	8.902	< 0.05
	learning attitude	0.38±0.46	8.651	< 0.05
	degree of effort	0.49±0.38	9.461	< 0.05

Analysis: Based on the P value in Table 4-7, there is no significant change in the learning purpose, learning desire, effort and learning attitude of students in the control class before and after the experiment as the P value is greater than 0.05. However, it can be observed that applying an outcome-oriented method to senior high school English teaching leads to significant changes in students' learning desire, learning attitude and degree of effort. The most notable change was found in their learning attitude followed by their learning desire while there was no significant difference observed regarding their learning purpose before and after the experiment.

The experimental results can be discussed from the following perspectives: Firstly, the shadowing exercise method highlights the integration of "learning" and "application", thereby transforming students' perception of English learning. During the output process, learners develop a strong desire for learning, which motivates them to actively acquire knowledge instead of passively relying on traditional classroom teaching where teachers identify important and challenging points for them. Additionally, based on achievement motivation theory, learners exert more effort in completing learning tasks to achieve success or avoid failure. According to self-efficacy theory, learners enhance their self-perception of ability after successfully accomplishing a specific task and become willing to make efforts towards more challenging tasks. This fosters students' expectations and aspirations for future studies. Considering that the college entrance examination

significantly influences high school students' academic pursuits, their purpose in studying English largely revolves around passing this exam and is not easily altered.

Based on analysis from these two aspects above, it can be concluded that implementing an output-oriented approach in shadow reading instruction positively impacts students' learning motivation. Firstly, it enhances overall levels of student motivation while simultaneously bringing about positive changes in three dimensions: attitude towards learning, desire for improvement, and level of effort invested.

5.Conclusions

By comparing the listening scores of the experimental group and the control group, this study explored the effect of shadow reading intervention in English listening teaching and its influence on English learning motivation. The results showed that the students in the experimental group who received shadowing exercise did perform better in the two post-tests than the students in the control group who received traditional listening teaching. The implementation of shadow reading training can enhance learners' overall English learning motivation. Data analysis indicates that the current motivation level among high school students is relatively low, and the adoption of an output-oriented approach has significantly elevated their motivation level. Considering the four dimensions of learning dynamics, integrating this method into high school English teaching can greatly enhance learners' aspirations for learning, foster positive changes in their attitudes towards learning, and encourage them to exert more effort during the process of English acquisition.

In a word, teachers and students work together, teachers actively use textbooks and carry out shadow reading classroom teaching practice, and students actively participate in classroom teaching activities, which can alleviate students' difficulties in English listening learning to a certain extent and enhance their English learning motivation. One of the effective approaches to incorporate shadow practice into high school listening instruction is to fully utilize existing resources and align with the current teaching schedule. Taking the teaching mode of this experiment as an illustration, it is impractical to arrange additional listening classes or purchase supplementary materials. Therefore, the author leverages the unified listening practice time for the entire grade and provides guidance and adjustments based on that foundation, directing students to engage in shadow practice during their second round of listening exercises. By doing so, both existing training time and materials are utilized effectively. Consequently, teachers do not need to

allocate extra class time for training nor worry about selecting inappropriate listening materials that do not match students' proficiency level. This undoubtedly renders the application of shadowing exercise in high school English listening more feasible and realistic. The key lies in adapting to local circumstances. Herein, the author proposes an influential instructional method as depicted in the following table 4-8:

Table 4-8

When	How	Remark
Morning reading-aloud course	Broadcasting reading material and shadowing	Teachers should be always on the side to guide students to correct and adjust.
Listening training time	Shadowing during the second-listening time	

References

[1]Brown, H. D. Teaching by principles [M]. Englewood Cliffs, NJ: Prentice Hall Regents, 1994.

[2] Dornyei, Z. & Ottó, I. Motivation in action: a process model of L2 motivation [J]. Working Papers in Applied Linguistics, 1998(4): 43-69.

[3]Gardner, R. & Macintyre, P. A student's contributions to second language learning. Part II: affective variables [J]. Language Learning, 1993, 26(1): 1-11.

[4]Krashen, S. The Input Hypothesis: Issues and Implication [M]. New York: Longman, 1985.

[5]班杜拉，(2001)，思想和行动的社会基础:社会认知论[M].林颖，王小明等，译.上海:华东师范大学出版社.

The effects of incorporating humorous memes into English classrooms on junior high school students' willingness to communicate (WTC)

英语课堂中融入幽默性模因对初中生英语交际意愿的影响效应

林勇创　华南师范大学外国语言文化学院

Abstract

The research on Willingness to communicate (WTC) has predominantly focused on determining the factors that affect second language (L2) WTC in ESL contexts. This study differs by examining whether using humorous memes as ELT materials can effectively increase students' willingness to communicate (WTC) in English. Two groups of Chinese junior high school students were selected as participants and assigned to a control group (n=44) or an experimental group (n=44). The researcher integrated humorous meme pictures into speaking lessons to facilitate students to speak English. While the experimental group received the treatment over two months, the control group experienced the regular teaching methods. Four teaching and learning sessions were recorded and the post-test was conducted to measure participants' willingness to communicate. SPSS software was used to perform a t-test to compare the WTC scores between the experimental and control groups. The results revealed a significant difference, with the experimental group showing a marked increase in WTC by the end of the intervention. This is supported by the results of classroom observation, in which the experimental group demonstrated a higher degree of volunteering and participation. Qualitative analyses of semi-structured interviews also revealed students' positive perceptions of the humorous memes, such as reducing stress and attracting attention. It is concluded that humorous memes are conducive to increasing students' WTC and could be utilized by English educators to create an innovative and motivational teaching and learning process.

Keywords: humor; memes; willingness to communicate

1. Introduction

During the last decade, technology has been advancing at a rapid speed and exerts an enormous influence on people's lives. Against this backdrop, Marc Prensky coined the term 'Digital Natives' to refer to the young students who were born in the internet age (Harshavardhan et al., 2019).

Considering our students are active users of information and communication technology (ICT), they can be viewed as digital natives. In the meanwhile, they live in an era when English is a predominant language of the world. However, students are less attentive and need more motivation in the English as Foreign Language (EFL) classroom (Harshavardhan et al., 2019). They might also face constraints and challenges in grasping and understanding the language (Pranoto & Suprayogi, 2020). Particularly, it seems that many Chinese students might feel reluctant to speak in English due to a lack of sufficient linguistic contact (Peng, 2019). As MacIntyre et al. (1998) argued, in the context of L2 education, a proper objective for language educators is to create students' willingness to communicate (WTC). In addition, according to the Ministry of Education (MOE) of the People's Republic of China (2020), cultivating students' ability to use the language to communicate helps them develop their core competencies. Thus, it is of great importance to enhance students' WTC.

An innovative approach is proposed to meet the needs and expectations of today's language learners, which is featured by integrating memes into language teaching and learning. The term 'meme' refers to "[A]n amusing or interesting item (such as a captioned picture or video) or genre of items that is spread widely online, especially through social media" (Merriam-Webster, n.d., para. 1). One of the reasons for incorporating memes into language teaching and learning is that they are authentic and related to real-life scenarios (Kalyuzhna et al., 2023; Kyrpa et al., 2022). Additionally, the humorous effect is inherent to memes (Kalyuzhna et al., 2023). Given this, humorous memes can serve as a way to raise students' interest or motivation in learning a language, which facilitates English instructors to ponder over the improvement of teaching approach to better satisfy learners' needs. Therefore, this research was conducted to investigate the effect of incorporating humorous memes into English classrooms on students' WTC.

2. Literature review

2.1 Willingness to communicate

According to MacIntyre and Legatto (2011), willingness to communicate (WTC) refers to "the intention to speak or to remain silent that can be considered the most immediate determinant of L2 use, reflecting the culmination of a variety of proximal influences...and distal influences." (p.150). This definition highlights the motivational nature and influencing factors of WTC. Similarly, MacIntyre et al. (1998) defined L2 WTC "as a readiness to enter into discourse at a particular time with specific person or persons, using a L2" (p. 547). Based on this conceptualization of WTC in L2,

MacIntyre et al. (1998) initiated a heuristic model of variables influencing L2 WTC (Figure 1). The top three layers (I, II, & III) of the pyramid represent immediate situational variables and the latter three layers (IV, V, & VI) represent more stable or enduring factors of WTC. This model helps distinguish enduring and situational influences of one's WTC.
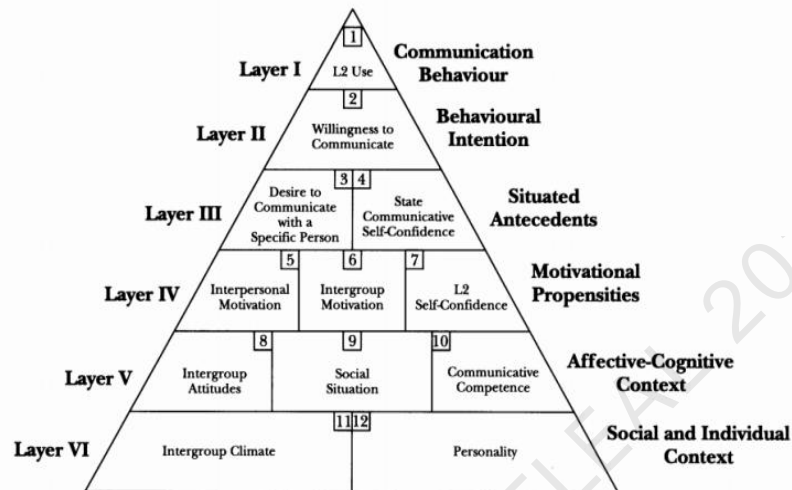
Figure 1. The Pyramid-shaped Model of L2 WTC (MacIntyre et al., 1998).

According to Peng et al. (2016), L2 WTC has trait and state characteristics. While the former level of WTC refers to the relatively stable tendency for communication, the state-level WTC highlights a temporary condition favoring communication. WTC used to be examined from the perspective of stability or trait-like predisposition. Nevertheless, attention has shifted to the more dynamic, state-like elements of WTC, and research on L2 WTC is expected to benefit from adopting a dynamic perspective by taking the state characteristics of WTC into account (Zhang et al., 2019; Yu, 2018). This tendency indicates WTC is dependent on the situation, and thus it is crucial to study the dynamic and complex features of WTC. Likewise, Peng et al. (2016) claim that L2 WTC can be regarded as a dynamic system, in which linguistic, social, cognitive, and emotional systems interact with each other and function together.

2.2 Memes

In 1976, biologist Richard Dawkins popularized the term 'memes' in his book *The Selfish Gene* to explain the manner in which cultural information spread. Memes are defined as small cultural units of transmission like genes spreading from person to person by means of copying or imitation

methods (Harshavardhan et al., 2019; Purnama et al., 2017). With the advent of the internet, memes have been widely spread via social media in the form of pictures, videos, songs, and many others (Purnama et al., 2017; Baysac, 2017). In other words, the interaction between the internet and memes contributes to the advent of internet memes.

Nowadays, internet memes are viewed as an optimal addition to enhance students' learning experience since they are massively incorporated into students' lives (Altukruni, 2022). According to Pranto and Suprayogi (2020), contextual teaching-learning conforms with the philosophy of constructivism. It emphasizes the link between the learning material and real-world situations and learners' agency of constructing knowledge in their own minds. The interaction between language and image within internet memes renders learners the opportunity to see the use of English in real-life scenarios. Besides this, the humorous elements embedded in the internet memes also account for their popularity in the field of education. In light Clark (2013, as cited in de OLIVEIRA, 2022), humor can yield the most cognitive effects in spite of the extra cognitive effort to process the information. Given this, internet memes are rendered effective for learners to decode the meaning of learning materials.

There are several studies concerning the relationship between the use of internet memes and students' WTC. The study by Farahani and Adbollahi (2018) indicates that incorporating elements of humor in the classroom can enhance students' WTC. This is collaborated by Pranoto and Suprayogi's (2020) study to examine the effect of internet memes on students' WTC. Thirty adult EFL learners were employed as participants and then grouped into the experimental and control group. While the control group received their routine material related to English for daily conversation, the experimental group received additional lessons that contained the use of humorous memes taken from 9GAG. The results of pre-tests and post-tests show that the incorporation of internet memes into language learning is effectively applicable in increasing learners' WTC. Although this study demonstrates that the incorporation of internet memes exerts a positive influence on students' WTC by means of t-test, it may fail to provide a comprehensive and in-depth analysis of learners' perceptions of using internet memes via interviews. That is to say, a qualitative approach can be supplementary to this quantitative paradigm.

Another common theme of studies is the positive reception of memes by students. In Kayali and Altuntas's study (2021), students considered the use of memes to be interesting and beneficial, highlighting the overall enjoyment and engagement it brought to the learning process. This sentiment

is echoed by Purnama (2017) and Garg (2021), where the majority of participants agreed the integration of memes made language learning enjoyable. Purnama et al. (2017) further elaborate on the intrinsic value of memes, noting that students found them attractive (30.2% of participants) and enjoyable (30.2%). In addition, the activity of creating memes was associated with students' willingness to engage, which demonstrates the integrative value of using memes. In the meanwhile, the results of Reddy et al.'s (2020) study showed that most of the students agree excessive use of memes might divert attention from the core topic, potentially leading to a less serious classroom environment.

In conclusion, there are relatively few studies examining the effectiveness of incorporating humorous memes and students' perceptions. These studies are quantitative in nature, lacking in-depth analysis of students' perceptions of using internet memes. Additionally, current research primarily focuses on the impact of internet memes on the linguistic aspects of language learning, such as vocabulary, grammar, and writing. Hence, further research on the emotional dimension of language learning is needed. Finally, learners' WTC at different educational levels should be considered. In previous studies, the research subjects mainly consisted of university students instead of junior high school students. Given these research gaps, three questions are raised in the present study:

RQ1: Does the incorporation of humorous memes exert an influence on students' WTC?

RQ2: Does the experience of incorporating humorous memes influence students' WTC in terms of classroom performance?

RQ3: How do students perceive the incorporation of humorous memes?

3. Methodology

3.1 Research design

The methods of classroom observation and quasi-experimental design were used to examine the effect of humorous memes on students' WTC. Participants included 88 junior high school students in Guangzhou (44 in the experimental group and 44 in the control group) and they are 14-15 years old. The selection of classes and students conformed with the strategy of criterion sampling. Prior to the experiment, a pretest of WTC was administered in the form of a questionnaire. Then the experimental group received the treatment over a 4-week period, whereas the control group received

the regular condition. At the end of the treatment, participants in both groups received the post-test to measure their WTC.

Experimental group ($N_{student}$=44) Pretest (beginning of the experiment) ● Questionnaire about students' WTC Meme-based treatment (during class time) ● Teachers use humorous memes during class time. ● Students are asked to create internet memes and complete the oral task. Post-test (end of the experiment) ● Questionnaire about students' WTC ● Self-report questionnaire about students' WTC in relation to the oral task	**Control group** ($N_{student}$=44) Pretest (beginning of the experiment) ● Questionnaire about students' WTC Without Meme-based treatment (during class time) ● Teachers don't use humorous memes during class time. ● Students complete the oral task. Post-test (end of the experiment) ● Questionnaire about students' WTC ● Self-report questionnaire about students' WTC in relation to the oral task

Figure 2. The experimental and control groups in the intervention of internet memes.

3.2 Data collection methods

A WTC scale was utilized as the pretest and post-test. The WTC scale was derived from the willingness to communicate (WTC) questionnaire developed by Khatib and Nourzadeh (2015). Some items have been slightly modified to suit the context of Chinese education. The survey was translated into Chinese in order to ensure all participants understood the survey items. It consisted of 23 items and used a five-point Likert scale. Before the experiment, the scale was piloted with 20-30 students in other classes to test its reliability and validity. In addition, the Motivational Orientation of Language Teaching (MOLT) Observation Scheme developed by Guilloteaux and Dörnyei (2009) was used to examine students' WTC. Considering the focus of the current research is on students, two observation categories including learners' participation and volunteering are made reference. At the end of the study, 10 interviews were conducted to investigate students' perceptions of using humorous memes in the classroom.

Figure 3. Motivational Orientation of Language Teaching (MOLT) Observation Scheme: Part A

3.3 Data analysis techniques

SPSS software version 26.0 was used to analyze the data. First, the reliability of the WTC scale was obtained. Second, an independent samples t-test was used to check if there were any significant preexisting differences between the treatment group and the control group at the beginning of the research. Third, a paired t-test was conducted to compare the pretest and post-test scores of WTC within the treatment group and the control group respectively. As mentioned earlier, the content of classroom observation included volunteering and participation. The dimension of volunteering could be measured by the number of students who raise their hands in class and the degree of participation can be measured by the number of students who took part in the oral task and interacted with others. Finally, interviews were transcribed verbatim in Chinese and then translated from Chinese to English. Each transcript was scrutinized and coded.

3.4 Teaching materials

This teaching experiment was based on the content of units 2~4 and 6 of the eighth-grade English textbook from the Shanghai Education Edition for junior high school. This could minimize disruption to students' English learning progress and align with the school's teaching schedule. Furthermore, the themes of these four units were closely related to students' real-life experiences and were considered suitable for incorporating humorous memes. Since this study examined the effect of humorous memes on students' WTC, the lesson type focused on listening and speaking. The images of humorous memes were sourced from the internet and slightly modified. Before the formal

experiment, five public school English teachers were asked to evaluate the humor of the materials using a Likert scale, and the results indicated that the materials possess a basic level of humor. With respect to the teaching procedure, the following table demonstrates three stages of organizing students' work with memes.

Stage	Activities	Purpose
Engage	1. Guess the theme of the meme pictures 2. Introduce the origin of memes	● to notice the general idea of meme pictures ● to stimulate students' interest in active work
Study	1. Describe the content of meme pictures 2. Analyze the elements of internet memes 3. Express opinions on internet memes	● to understand the meaning of internet memes ● to summarize the structure of internet memes ● to evaluate the text of internet memes
Practice	1. Create personal internet memes about festivals and celebrations by adding text 2. Perform oral tasks based on the communicative theme of the meme	● to transfer the skill of creating internet memes for self-expression ● to raise students' interest in oral practice and reduce their anxiety

Table 1. Description of the three-staged teaching procedure

4. Results

4.1 Students' WTC in English classes under the effect of humorous memes

The WTC variable was tested and found internally consistent (α=0.96) before conducting the intervention. Additionally, the KMO value was 0.633, indicating that the validity of the scale was moderate. The results of the independent samples t-test indicated that the experimental group (M=3.62, SD=0.76) and the control group (M=3.37, SD=0.50) did not differ significantly on levels of WTC prior to the experiment, $t(74.452)$=1.83, p=0.07 (p>0.05), suggesting that both groups were comparable in terms of WTC. After the experiment, the statistical results of the post-test are shown in the following Table 2. According to Table 3, it can be seen that the two groups differ significantly on levels of WTC, $t(86)$ = 3.19, p=0.002 (p<0.05), indicating that the incorporation of humorous memes can exert a positive influence on students' WTC.

	N	Mean	SD	Std. Error Mean
Experimental group	44	4.01	0.80	0.12
Control group	44	3.48	0.77	0.12

Table 2. Independent Samples Statistics: the post-test of Students' WTC

	Mean Difference	Std. Error Mean Difference	t	df	sig.
EG—CG	0.53	0.17	3.19	86	0.002

Table 3. Independent Samples: the post-test of Students' WTC

In addition, the paired samples t-test was also used to compare students' WTC of two groups. As Table 4 and Table 5 show, there was a significant difference between the post-test and pretest of students' WTC in the experimental group, while there was not in the control group.

	Mean	Std. Deviation	Std. Error Mean	t	df	sig.
Pre-test—Post-test	-0.11	0.54	0.08	-1.35	43	0.18

Table 4. Paired Samples: Students' WTC of the control group

	Mean	Std. Deviation	Std. Error Mean	t	df	sig.
Pre-test—Post-test	-0.39	0.62	0.09	-4.20	43	0.000

Table 5. Paired Samples: Students' WTC of the experimental group

4.2 Students' WTC in terms of classroom behavior

Session	1		2		3		4	
Group	EG	CG	EG	CG	EG	CG	EG	CG
Number of students	45	43	43	45	43	42	44	44
Number of students participating in the activity	44	37	41	41	41	39	42	38
Rate of participation	98%	86%	95%	91%	95%	93%	95%	86%

Table 6. Students' participation in four sessions

Session	1		2		3		4	
Group	EG	CG	EG	CG	EG	CG	EG	CG
Number of students participating in the activity	45	43	43	45	43	42	44	44
Number of students raising hands	36	19	38	35	30	25	36	20
Rate of hand raising	80%	44%	88%	78%	70%	60%	82%	45%

Table 7. Students' performance of volunteering in four sessions

Firstly, the statistics on participation are based on the period between the teacher issuing instructions for oral activity exercises and inviting students to present. As shown in Table 6, the participation rate of the experimental group students remained stable at around 95% across the four experiments, with a peak value of 98%. In contrast, the participation rate of the control group students fluctuated between 85% and 95%, consistently lower than that of the experimental group in all four experiments. Secondly, the statistics on hand-raising are based on the period from the teacher's invitation for oral presentations to the point when the first two groups of students are invited to

answer questions. As illustrated in Table 7, the hand-raising rate in both groups exceeded one-third, but the average hand-raising rate of the experimental group was approximately 80% over the four experiments, with a maximum of 88%. Meanwhile, the control group's hand-raising rate fluctuated between 40% and 85% across the four experiments, also consistently lower than that of the experimental group. From the above data, it can be concluded that, from the perspective of classroom behavior, the experimental group students exhibited a higher level of willingness to participate in classroom oral activities.

4.3 Students' perception about the incorporation of humorous memes

To begin with, students were surprised and found it novel when teachers presented and used humorous memes in class, perceiving it as different from traditional teaching methods. As S3 and S5 mentioned in interviews,

> It was quite new and interesting because I had never seen other teachers use these memes in class before. I think it's quite fun. (S5)

> I had never seen this form before, and I felt very surprised and intrigued. Traditional classes are not so interesting. (S3)

Secondly, all interviewed students expressed that using humorous memes in class made them more willing to communicate in English. The main reason was that they felt these images were relatable to their real lives and matched their interests, easily resonating with their past experiences. As S6 and S8 mentioned,

> I feel that these meme images are very fitting and align with the way I usually chat with friends. (S8)

> Humorous memes can combine elements of traditional Chinese culture with current popular internet memes, which can arouse students' interest. (S6)

Based on this interest, the interviewed students believed that humorous memes would broaden their imagination, making them more willing to engage in English oral expression.

Furthermore, all interviewees indicated that using humorous memes in class could reduce their negative emotions regarding oral expression. As S2 and S3 said,

> We are in a high-pressure learning environment, and these memes can alleviate our tension, making us more comfortable in participating in class activities. (S2)
>
> Because there are some interesting thoughts in my mind, they can replace my feelings of anxiety. When I start speaking, I find it quite amusing, which eliminates those anxious feelings and encourages me to speak boldly with my classmates. (S3)

Finally, most interviewees stated that using humorous memes helps to enhance their concentration because the content was highly engaging. As S7 mentioned,

> The content is quite funny, and especially the atmosphere, everyone is discussing. Everyone is interested, and each person's ideas are different. Listening to others' thoughts, I can also express my own ideas, so I am more focused. (S7)

5. Discussion

From the perspective of willingness, it reveals that integrating humorous memes into the English classroom helps enhance students' WTC by using a questionnaire and conducting independent sample and paired sample t-tests. This finding is supported at the behavioral level by means of classroom observations. The results of classroom observation indicate that students in the experimental group show higher participation and more frequent hand-raising compared to those in the control group. Therefore, it can be concluded that humorous memes exert a positive impact on students' willingness to communicate. Semi-structured interviews further reveal that students have positive perceptions of the use of humorous memes in the classroom. The underlying mechanisms behind the positive effect of humorous memes can be attributed to three main factors: their relevance to students' daily lives, reduction of negative emotions, and enhancement of attention.

The results of this study are comparable with the results of the previous studies reviewed in this study. This comparison is essential to understand how incorporating a specific genre of universal humor can enhance ELF learners' WTC. Previous research has typically focused on the impact of humor on learning a second or foreign language, students' perceptions of humor, and the types of humor used by teachers in EFL classrooms. There was only one study that examined the effect of meme pictures on students' WTC. Compared to Pranoto and Suprayogi (2020), this research

confirmed the findings. Their research affirmed that the utilization of humorous memes in language classes is conducive to increasing students' willingness to communicate. This study supplemented Pranoto and Suprayogi's research in that students' WTC was also investigated from the perspective of behavior, allowing for a more comprehensive exploration of the relationship between humorous memes and students' WTC. Additionally, the consistency of findings is attributed to the positive effects of humor on learners' WTC. This is corroborated by Farahani and Abdollahi's (2018) research, in which the role of humor in enhancing EFL learners' WTC was examined. Based on this, Farahani and Abdollahi's study is comparable with this present research.

The findings of this research also reveal the feature of humorous memes as a specific genre of learning material. It is stated by Harshavardhan et al. (2019) that memes and humour are gradually associated synonymously. That is, memes can be regarded as the carrier of humorous messages. This combination potentially endows English teaching with particular advantages, including reduction of stress and enhancement of attention. According to Harshavardhan et al. (2019) and Altukruni (2022), memes are conducive to reducing stress in the classroom and increasing learner's attention. In this regard, the current research provides evidence to support the previous statement. Furthermore, in Hibatullah and Ardlillah's (2019) study, it was found that the majority of learners affirmed that humour delivered by the teacher could foster a relaxing atmosphere and reduce stress during the English learning process. Regarding this, it can be noted that the findings of the current research add new insights by taking humorous memes into consideration to improve L2 students' WTC.

6. Conclusion

This research revealed the effectiveness of incorporating humorous memes into EFL classrooms to develop students' WTC. The findings discussed in this study can provide some implications for English teaching, for instance, instructors ought to be open-minded to using memes in their classroom rather than relying on conventional teaching strategies. With this respect, further research can be conducted to examine the English teachers' perceptions of using humorous memes. Besides, it is reasonable to assume that the stress level has a relationship with WTC. Therefore, the level of stress can be used as a mediator variable to examine the effect of humorous memes on students' WTC in future studies.

References

[1] Altukruni, R. (2022). A Systematic Literature Review on the Integration of Internet Memes in EFL/ESL Classrooms. *Arab World English Journal*, 13(4), 237-250.

[2] Baysac, P. (2017). Laughter in class: Humorous memes in 21st century learning. *Journal of Social Sciences*, 6(2), 267-281.

[3] de OLIVEIRA, V. (2022). Technology and humor: sample lessons to keep English learning functional during pandemic times. *Entrepalavras*, 11(3), 1-23.

[4] Farahani, A. A. K., & Abdollahi, Z. (2018). Incorporating humor to develop EFL learner's speaking ability and willingness to communicate. *Journal of Language Teaching and Research*, 9(1), 205–211.

[5] Guilloteaux, M. J. & Dörnyei, Z. (2009). Motivational Orientation of Language Teaching (MOLT) Observation Scheme: Part A. Retrieved from https://docs.wixstatic.com/ugd/ba734f_81d8062a6c984e4bb9c252a4943ce147.pdf?index=true

[6] Garg, R. (2021). Memes: Creative Technological Strategy for ESL Learners. *Language in India*, 21(12), 122-130.

[7] Harshavardhan, V., Wilson, D., & Kumar, M. V. (2019). Humour discourse in internet memes: An aid in ESL classrooms. *Asia Pacific Media Educator*, 29(1), 41-53.

[8] Hibatullah, O. F., & Ardlillah, Q. F. (2019). "Give me a joke, please!": Creating a fun learning by teachers' jokes. *PEOPLE: International Journal of Social Sciences*, 5(2), 948–958.

[9] Kalyuzhna, A., Lavrinenko, I., & Radchenko, O. (2023). Presenting grammar through Internet memes in teaching English for specific purposes. *Cognition, communication, discourse*, 26, 68-82.

[10] Kyrpa, A., Stepanenko, O., Zinchenko, V., Udovichenko, H., & Dmytruk, L. (2022). Integration of internet memes when teaching philological disciplines in higher education institutions. *Advanced Education*, 20, 45-52.

[11] Kayali, N. K. & Altuntaş, A. (2021). Using memes in the language classroom. *International Journal of Education*, 9(3), 155-160.

[12] Khatib, M. & Nourzadeh, S. (2015). Development and validation of an instructional willingness to communicate questionnaire. *Journal of Multilingual and Multicultural Development*, 36(3), 266-283.

[13] MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and

affiliation. *The Modern Language Journal*, 82(4), 545-562.

[14] MacIntyre, P. D., & Legatto, J. J. (2011). A dynamic system approach to willingness to communicate: Developing an idiodynamic method to capture rapidly changing affect. *Applied Linguistics*, 32(2), 149-171.

[15] Meriam-Webster. (n.d.). Meme. In Meriam-Webster Dictionary. Retrieved from https://www.merriam-webster.com/dictionary/meme

[16] Ministry of Education of the People's Republic of China (MOE). (2020). *General Senior High School Curriculum Standards (English)*. People's Education Press.

[17] Purnama, A. (2017). Incorporating memes and Instagram to enhance student's participation. *Language and Language Teaching Journal*, 20(1), 1-14.

[18] Purnama, A., Desiarti, E., Aflahah, N., & Ekaningrum, V. (2017). Utilizing memes to promote students' motivation in language classroom. *Linguistics, Literature and Language Teaching Journal*, 7(2), 134-153.

[19] Pranoto B. E. & Suprayogi. (2020). Incorporating 9GAG Memes To Develop EFL Learners' Speaking Ability And Willingness To Communicate. *IJEE (Indonesian Journal of English Education)*, 7(2), 130-144.

[20] Peng, J., Zhang, L., & Chen, Y. (2016). The mediation of multimodal affordances on willingness to communicate in the English as a foreign language classroom. *TESOL Quarterly*, 51(2), 302-331.

[21] Peng, J. (2019). The roles of multimodal pedagogic effects and classroom environment in willingness to communicate in English. *System*, 82, 161-173.

[22] Reddy, R., Singh, R., Kapoor, V., & Churi, P. P. (2020). Joy of Learning Through Internet Memes. *International Journal of Engineering Pedagogy (iJEP)*, 10(5), pp. 116-133.

[23] Zhang, J., Beckmann, N., & Beckmann, J. F. (2019). One situation doesn't fit all: Variability and stability of state willingness to communicate in a Chinese college English classroom. *Language Teaching Research*, 1-26.

[24] 余淼. (2018). 论外语交流意愿研究的两次转变. *现代外语*, 41(4), 12.

Supplementary material

Appendix 1. Questionnaire of WTC

初中生英语交际意愿度问卷

班级: 初二（　）班　　学号:＿＿＿＿＿＿＿＿＿＿

说明：本项问卷是关于英语交际意愿度的调查，共 23 题。请根据自己英语交际意愿度在对应选项旁打勾。

请认真阅读和完成问卷，问卷所选内容将会严格保密。

1. 开口用英语交流表达，即使其他同学笑你有一些语言错误。

A. 非常不愿意　　　　B. 不愿意　　　　C. 一般　　　　D. 愿意　　E. 非常愿意

2. 开口用英语交流表达，即使知道其他同学的口语能力可能比你好。

A. 非常不愿意　　　　B. 不愿意　　　　C. 一般　　　　D. 愿意　　E. 非常愿意

3. 在同学面前用英语做展示或报告。

A. 非常不愿意　　　　B. 不愿意　　　　C. 一般　　D. 愿意　　　E. 非常愿意

4. 开口用英语交流表达，即使老师会纠正你的一些语言错误。

A. 非常不愿意　　　　B. 不愿意　　　　C. 一般　　　　D. 愿意　　E. 非常愿意

5. 在英语课堂的小组活动中用英语和同学进行讨论。

A. 非常不愿意　　　　B. 不愿意　　　　C. 一般　　　　D. 愿意　　E. 非常愿意

6. 在英语课堂中谈论英国国家人民的生活方式。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

7. 与同学用英语谈论英语国家的历史。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

8. 在小组中用英语讨论英/美国人和中国人的文化差异。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

9. 和英语老师用英语讨论英国或美国文学。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

10. 用英语谈论你的个人生活经历。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

11. 不管教室多拥挤，会找机会开口进行英语表达。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

12. 就算坐在教室后排，会找机会开口进行英语表达。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

13. 与同学用英语谈论电影、电视剧或短视频等媒体内容。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

14. 在小组讨论中用英语谈论你所了解的伟大艺术家或其他名人。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

15. 与同学用英语讨论手机或电脑游戏。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

16. 在全班讨论中用英语谈论你最喜欢的运动。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

17. 向同学询问英语单词的正确发音。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

18. 向其他同学请教英语的语法问题。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

19. 如果听不懂英语老师上课说的内容，举手请老师重新讲一遍。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

20. 在英语课堂中举手提问或回答问题。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

21. 在英语课堂之外与同学用英语交流。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

22. 在课后与同学用英语讨论你的周末生活。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

23. 与其他班级的同学用英语交流。

A. 非常不愿意　　　　B. 不愿意　　　C. 一般　　　　D. 愿意　　E. 非常愿意

Appendix 2. Teaching plan of the first session

Teaching plan in E1		
Teaching content	Unit 2 Body language	
Teaching model	PWP	
Teaching duration	40 minutes	
Students level	Junior high school, Grade Eight	
Lesson type	Listenin&Speaking	
Teaching objectives	By the end of this lesson, students will be able to: 1. Understand the main idea and details of the listening material about the body language of ballet dancers 2. Discuss the meaning of specifc gestures 3. Express their interpretation of specifc gestures by using gerunds	
Group	Experimental group	Control group
Step 1: **Pre-listening** **(5 minutes)**	**Activity 1: notice and perceive-5'** -T presents 1 meme pictures about ballet -Ss are invited to express their feelings about the ballet by reference to the memes	**Activity 1: notice and perceive-5'** -T presents 1 picture about ballet -Ss are invited to express their feelings about the ballet
Step 2: **While-listening** **(10 minutes)**	**Activity 2: analyze and practice-10'** -Ss predict the content and analyze the part of speech of words in blanks and then listen to the audio about ballet twice -T check the answer with students	
Step 3: Speaking **(25 minutes)**	**Activity 3: review and internalize-5'** -T guides Ss to review the grammatical knowledge about gerunds	
	Activity 4: analyze and judge-10' -T shows 1 meme picture about body language -Ss discuss the meaning of each picture and share their ideas (T-Ss) -T presents the general structure of dialogue -Ss practice the dialogue with their partner **Activity 5: transfer and create-10'** -T provides 2 meme pictures for students and asks them to add sentences to the provided pictures by using gerunds. (be good at/be interested in/enjoy/finish doing sth.) -Ss demonstrate their meme pictures and express their interpretation of meme pictures through dialogue (A: Look at this picture. ____ is _____. What does —_____ mean? B: I think _____ means _____. A: I agree with you. / Well, I think _____ means _____)	**Activity 4: analyze and judge-10'** -T shows 1 picture in the textbook about body language -Ss discuss the meaning of each picture and share their ideas (T-Ss) -T presents the general structure of dialogue -Ss practice the dialogue with their partner **Activity 5: transfer and create-10'** -T presents 2 pictures in the textbook to students -Ss express their interpretation of meme pictures through dialogue (pair work)

Appendix 3. Humorous memes used in the classroom (Unit 3&6)

Create **Task:** choose a meme picture and add sentences by using the passive voice

I AM FORCED TO DO MY WORK.

我的内心是拒绝的

命中带苦

骑上我心爱的小摩托

Picture 1

Picture 2

Word bank

require, ask, force (强迫), invite, reject, accept, give, allow, hurt, remind about...

Pairwork

fight with each other

make a mess

damage the house

bark all day long

对话模版

A: I want to make a complaint about my _____.
B: What's the matter?
A: _____.
B: I am sorry to hear that.

步骤

第一步: 两人一组，选择一幅图并给其配文字（需用上副词adverbs，比如loudly, angrily, seriously...）
第二步: 根据所选图片内容补充左边的对话。
第三步: 两两进行口语练习。

Exploring teachers' perceptions of changes to teaching practice and the influence of professional development

SHI Weihao

Xi'an International Studies University

Abstract: This study explores the perceptions of English-as-a-foreign-language teachers regarding changes in teaching practices following professional development in China. Informed by a qualitative case study approach, this study collected data from 3 teachers across a semester of an English-as-a-foreign language course at a university in China. The data included 9 hours of audio-recorded classroom talk (3 for each teacher), teacher's reflections over the courses and interviews, coupled with field notes. The qualitative analysis shows that various aspects of the teachers' context and the complex interrelationship among these factors led these teachers to develop a passive attitude toward professional development learning and implementation. Despite participating in various professional development experiences throughout their careers, they seldom considered implementing changes in their teaching practices. It is recommended that both pedagogy and teacher education/development adopt coordinated efforts to bring substantial changes to the teaching practices of EFL teachers and enable them to share their expertise with other educators.

Key words: teachers' perceptions; teaching practice; professional development

1. Introduction

Globally, the teaching of English as a foreign language (EFL) is gaining increasing attention and its importance cannot be overstated. As an international language, English not only plays an important role in global communication, but also plays a key role in the education systems of various countries (Rao, 2019). Especially in China, as a foreign language, the teaching of English is not only the teaching of language skills, but also an important bridge for cultural exchange and global competitiveness.

However, Chinese EFL teachers face many challenges and demands in today's rapidly evolving educational environment. Not only do they need to adapt to changes in technology and educational policies, but they also need to continually improve their teaching competence and

professionalism in order to better respond to students' learning needs and social trends.

Although a large number of studies have explored the impact of teachers' participation in professional development (PD) on their own change (Coldwell , 2017; Borg, 2018; Mohammadi and Moradi, 2017), however, these studies have focused more on the effectiveness of the PD program than on an in-depth understanding of the process of change for individual teachers. Criteria for judging the effectiveness of PD are often used in assessing the effectiveness of PD rather than truly exploring in depth the internal changes of teachers during their participation in PD (Cordingley et al., 2015).

Understanding how teachers change through their involvement in PD is particularly important in China, especially in the context of university EFL teachers. EFL teachers in China face unique teaching and learning environments and challenges, such as the rapid development of technological applications and the restructuring of educational policies. Therefore, an in-depth exploration of teachers' changes in PD requires more listening to teachers' voices-what they think and feel.

Current research rarely represents teachers' perspectives, which leads to some one-sidedness in our understanding of how PD affects teacher change. In order to fully understand teachers' real experiences and changes in their participation in PD, it is necessary to adopt a qualitative research methodology to analyze in depth from teachers' own perspectives their thinking processes, changes in their teaching practices, and the actual impact of these changes on their teaching efficacy. This research method can not only provide more targeted suggestions for the design and implementation of PD, but also provide important theoretical support and practical experience for improving the overall quality of EFL education.

Therefore, this study aims to delve into the perceptions and impact of EFL teachers on changes in teaching practices resulting from their professional development. Through the teachers' perspectives, it is hoped to learn about the specific issues they face during their professional development, the changes they perceive, and how they translate these changes into effective strategies in their teaching practices. This will not only help to enhance the professionalism of individual teachers, but also provide useful references and suggestions for the improvement of the quality of EFL education as a whole.

The purpose of this study was to explore three Chinese university EFL teachers' perceptions

of changes in their teaching practices over the course of their careers and to gain insight into the relationship between these changes and participation in professional development. The specific research questions and sub-questions are as follows:

1. How do teachers articulate the evolution of their teaching practices throughout their professional trajectories?

2. To what extent do teachers attribute specific changes in their practice to their participation in professional development?

3. Which dimensions of PD do teachers identify as influential in shaping changes in their instructional methods?

4. How do teachers perceive the sustainability of these practice changes over an extended period?

By exploring these questions in depth, this study aims to reveal the key factors that influence changes in EFL teachers' teaching practices in Chinese universities, and to provide empirical support and guiding recommendations for teachers' professional development and PD design.

2. Literature Review

Professional development plays a crucial role in education, not only as a means for educators to improve the quality of teaching and learning, but also as a key way to keep abreast of changes in education, promote personal growth and enhance career satisfaction. Through continuous learning and growth, educators are able to improve their teaching skills, keep abreast of the latest developments in the field of education, and enhance teaching effectiveness and student learning outcomes, thereby contributing to the advancement of the education system and the spread of quality education.

EFL teachers in Chinese universities face multiple challenges, including differences in language proficiency and cultural background, limitations in teaching resources and technology, classroom management and student motivation, and pressure for personal development and career advancement. To effectively address these challenges, teachers can continue to improve the quality of their teaching and professional development by enhancing their intercultural teaching competence, actively exploring modern teaching resources and technologies, adopting effective classroom management strategies, engaging in educational training and academic research, and

promoting effective teaching evaluation and feedback mechanisms.

Prior research has focused on teachers' attitudes toward and implementation of professional development (Maskit, 2011; Wasserman and Migdal, 2019), but there is a research gap in terms of the specific contexts and influencing factors of EFL teachers. In particular, detailed analyses addressing the challenges faced by EFL teachers in Chinese universities and how they can be addressed through professional development are insufficient. This includes in-depth research on the specific impact of language proficiency and cultural differences on teaching, the role of teaching resources and technological equipment in the innovation of teaching methods, and the relationship between teachers' personal development and career advancement and teaching effectiveness has not yet been fully developed. Therefore, there is a need to further explore and understand these specific contexts and influences in order to guide and support EFL teachers' practical coping strategies in their professional development process.

3. Methodology

3.1 A Qualitative Methodology

The qualitative case study methodology was chosen for this study primarily because it allowed for an in-depth exploration of individual teachers' perceptions of and responses to changes in their teaching practices during the course of their professional development and the actual impact of professional development activities on their teaching practices. The following are a few rationalizations for the choice of this methodology. Qualitative case studies allow for an in-depth exploration of the complexity of individual experiences and perceptions (Simons, 2014), particularly in the field of education, which is critical to understanding how teachers make sense of and apply what they learn in their professional development; the case studies also allow for a detailed analysis of each teacher's changes and development in their particular teaching contexts (Levin, 2003), as each teacher faces unique challenges and opportunities in different teaching and learning environments; the study combines a variety of data sources, including classroom recordings, teacher reflections, interviews, and field observation transcripts, and this variety and depth contributed to a comprehensive understanding of teachers' responses and behaviors; a qualitative analysis approach was used to extract rich themes and patterns from the data that were not limited to qualitative descriptions, but also allowed for theoretical linkages and interpretations.

3.2 Participants

In this study, three EFL teachers from Chinese universities were selected as participants, each of whom had extensive teaching backgrounds and professional experiences, as well as interests and motivations for participating in this study.

A has more than 20 years of teaching experience as an EFL teacher in one of the leading universities in China. A has been deeply involved in the field of language education for many years and has deep insights and experience in teaching methodology and curriculum design. A would like to participate in this study to further understand and explore the potential impacts of professional development on teaching practice, especially in terms of the changes in teaching methodology and student learning outcomes.

Teacher B is a young faculty member teaching at an emerging comprehensive university. B has strong interest and practical experience in instructional technology and innovative teaching methods, with a particular expertise in utilizing technological tools to enhance language teaching and learning. B participates in this study in order to explore the impact of professional development on how teachers incorporate new technologies and pedagogical methods, and how to respond to pedagogical challenges and student needs more effectively.

Teacher C is a master's degree holder who has a long history of teaching EFL in higher education. C is concerned with education policy and curriculum reform, and is actively involved in teaching research and teacher development programs. C hopes that through this study, she will explore how professional development can help teachers adapt to the demands of education reform, enhance their own pedagogical efficacy, and positively impact student learning.

These three teachers have not only accumulated rich experience in teaching practice, but also possess profound educational concepts and the pursuit of professional development, so their participation will provide diversified perspectives and in-depth insights for the study.

3.3 Data Collection

Data collection for this study included a variety of sources designed to provide a comprehensive understanding of teachers' perceptions of changes in instructional practices and the impact of professional development. Data sources included: nine hours of classroom recordings of each teacher (three hours each) to capture their teaching practices and student interactions; teachers' reflective accounts of their professional development processes, including written and

audio recordings, to understand their subjective perceptions and understandings of changes in teaching and learning; and in-depth interviews with each teacher to explore their experiences of engaging in professional development activities and their impact on their teaching practices; Observation transcripts and notes from the researcher in the classroom and at the site of the professional development activities to supplement and validate other sources of data.

3.4 Data Analysis

The data analysis phase utilized a qualitative analysis approach designed to extract themes and patterns that provide insight into teachers' perceptions of changes in instructional practices and the impact on professional development. The main analytical steps included:

Data organization and categorization: the large amount of data collected was first organized and categorized, including transcribing classroom recordings, organizing reflections and interview transcripts.

Coding and Theme Extraction: the researcher identified key concepts and patterns in the data through step-by-step coding and comparison to form initial themes.

Pattern Matching and Interpretation: Based on the organized themes, the emergence and relationships of patterns are further analyzed in an attempt to link teachers' responses and behaviors to the specific impact of professional development activities.

Theoretical Connections: finally, the analyst connects the patterns and themes identified to existing theory and research in order to provide insights and explanations about teacher development and professional learning.

4. Results

The professional development activities in which teachers participated were viewed differently by various teachers. Some teachers viewed these activities positively, believing that they helped to enhance teaching competence and professionalism, while others had reservations, questioning their actual effectiveness and application value. Differences in teachers' attitudes towards professional development were affected by a number of factors, including the level of support from the school, attitudes of colleagues, and individual educational background and work experience. The adequacy of school resources, the level of administrative support, and the cooperative atmosphere among colleagues also had a significant impact on teachers' attitudes.

In terms of teaching practices, teachers adopted a variety of practical changes, mainly including the introduction of more interactive teaching methods, the use of more technological tools to assist teaching, and increased opportunities for student participation in the classroom. These changes are intended to enhance teaching effectiveness and students' learning experience. However, teachers face many challenges in implementing new teaching methods, such as lack of sufficient time and resources to prepare for new lessons, students' acceptance and adaptability, and the hindrance of some traditional teaching concepts. In addition, teachers' own familiarity with and confidence in the new approach is a major hindrance.

Experience sharing and mutual influence among teachers played an important role in professional development. By sharing their successes and failures in teaching practice, teachers were able to learn from each other and to improve their own teaching methods. Most teachers considered peer support and communication as an important part of professional growth. Through communication with colleagues, teachers can obtain emotional support, professional advice, and practical teaching strategies, which are important for meeting challenges in teaching and achieving personal professional goals.

5. Discussion

Teachers' attitudes towards professional development often show a tendency towards passivity. The development of such attitudes can be attributed to a number of factors, including lack of confidence in new teaching methods, reliance on traditional teaching styles, and time and resource constraints. In addition, inadequate administrative support in schools is one of the major reasons for the passivization of teachers' attitudes.

Professional development activities are closely related to teaching practice, and effective professional development not only enhances teachers' teaching ability, but also promotes the renewal and improvement of teaching methods. However, its effectiveness is often constrained by the level of teacher engagement and the supportive environment. Teachers' backgrounds and environments have a profound impact on the implementation of new teaching methods. For example, the richness of educational background and teaching experience, as well as the level of support from school resources and colleagues, directly affects the application and dissemination of new methods in practice.

Teachers' assessment of the effectiveness of professional development was largely based on actual teaching effectiveness and student feedback. Some teachers felt that the professional development activities had significantly enhanced their teaching abilities, while others had reservations and did not consider the effects significant.

To improve the situation, education institutions can optimize their professional development programmes by providing more targeted and hands-on training, as well as strengthening support and guidance to teachers as they implement new approaches. In addition, establishing more practice opportunities and communication platforms, and emphasizing the importance of sharing and collaboration among teachers in practice are also effective ways to enhance teaching quality and student learning outcomes. Through the combined promotion of the above measures, teachers' attitudes towards professional development can be gradually improved, and the innovation and improvement of teaching methods can be promoted, thereby enhancing the overall quality of education.

6. Conclusion

This study revealed that teachers generally had negative attitudes towards professional development and subsequent changes in teaching practice. The reasons behind such negative attitudes are manifold and include lack of confidence in new approaches, reliance on traditional teaching methods, time and resource constraints, and inadequate school administrative support.

Understanding these barriers is crucial in guiding the professional development of EFL teachers. Insights gained from this study emphasize the need for practical training programs tailored to the specific challenges faced by teachers. By strengthening support structures and fostering a culture that values continuous improvement, educational institutions can empower teachers to effectively incorporate innovative teaching methods into their practice.

Future research should delve into the experiences of teachers from diverse backgrounds. Investigating how factors such as educational background, teaching and learning environments, and institutional resources influence teachers' perceptions and implementation of professional development programs will provide additional insights. In addition, exploring longitudinal influences and conducting comparative studies across educational settings could enrich our understanding of effective strategies for promoting sustainable professional growth among EFL

educators.

References

Borg, S. (2018). Evaluating the impact of professional development. *RELC Journal*, *49*(2), 195-216.

Coldwell, M. (2017). Exploring the influence of professional development on teacher careers: A path model approach. *Teaching and teacher education*, *61*, 189-198.

Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., ... & Coe, R. (2015). Developing great teaching: Lessons from the international reviews into effective professional development.

Levin, B. B. (2003). *Case studies of teacher development: An in-depth look at how thinking about pedagogy develops over time*. Routledge.

Maskit, D. (2011). Teachers' attitudes toward pedagogical changes during various stages of professional development. *Teaching and teacher education*, *27*(5), 851-860.

Mohammadi, M., & Moradi, K. (2017). Exploring change in EFL teachers' perceptions of professional development. *Journal of Teacher Education for Sustainability*, *19*(1), 22-42.

Rao, P. S. (2019). The role of English as a global language. *Research journal of English*, *4*(1), 65-79.

Simons, H. (2014). Case study research: In-depth understanding in context. *The Oxford handbook of qualitative research*, 455-470.

Wasserman, E., & Migdal, R. (2019). Professional Development: Teachers' Attitudes in Online and Traditional Training Courses. *Online Learning*, *23*(1), 132-143.

KWL 教学模式对初中英语学习者阅读焦虑程度与阅读理解水平的影响研究

（孙欣 首都师范大学）

【摘要】 《义务教育英语课程标准（2022 年版）》指出，教学应"注重发挥学生的主观能动性，引导学生成为各类评价活动的设计者、参与者和合作者"。由此，高度契合当今教学理念的 KWL 教学模式开始流行。本研究基于 Carrel (1983) 的图示理论、Piaget (1936) 和 Vygotsky (1978) 的建构主义学习理论，以 19 名初三学生为研究对象，以 Ogle (1986)提出的 KWL 阅读教学模式和教学过程为依据，进行为期 10 周的行动研究教学设计，进一步探讨 KWL 教学模式对中国初中学生阅读焦虑程度与阅读理解水平的影响。研究结果表明：KWL 教学模式能够有效帮助初中生改善阅读习惯，激发阅读兴趣，对阅读教学具有积极作用。KWL 教学模式能够降低学习者的阅读焦虑程度，进一步提升学习者的阅读理解水平。

【关键词】初中英语学习者；KWL 教学模式；外语阅读焦虑；阅读理解

一、前言

阅读是人们获取信息、认识世界、发展思维的重要途径，也是提高国民文化知识水平，提升国家竞争力的重要基础 (OECD，2019; 王蔷、敖娜仁图雅，2015)。在教育教学中，阅读是一项十分复杂的社会、心理、认知活动，涉及识别、理解、推断、评价等多种能力等，受到阅读材料、阅读目的、读者语言能力和情境等多种因素影响(Grabe & Stoller，2019;曾用强，2017)。在北京市中考英语试卷中，阅读部分占据了笔试部分的近半比例。所以，阅读教学在初中英语教学中占据着重要的地位，一线教师需要不断更新阅读教学手段、教学模式，帮助学生提升阅读兴趣、提升阅读能力。

二、KWL 教学模式的发展

KWL 教学模式由 Ogle 于 1986 年提出。Ogel 指出 KWL 教学模式"有助于教师在阅读说明性材料时对学生的知识和兴趣做出更积极的反应，并为学生在阅读信息时的主动思维提供了模型。"在教学过程中，不应该由教师提示阅读的要点，应该让学生结合、引出自身的先验知识，再进行交互性的阅读。由此，Ogel 开发了一种适用于任何年级、任何阅读题材的教学模型——KWL 教学模式。Ogel（1986）通过文本分析法得出 KWL 更能引发学生对先验知识的思考、帮助学生在阅读过程中学习更多的知识。此外 KWL 图表可以有效、清晰地记录学生的学习成长变化，也能使提升学生与教师阅读过程中的互动性。Szabo（2007）表明 KWL 图表可以直观地观察学生的成长过程，这种教学方式可以帮助学生建立自信，提高学习效率。Raines (2018) 指出 KWL 图表是一种以学生为中心的、积极的学习策略，能够促进

引导学生成为独立的批判性思考者和决策者。

但是，由于研究者的广泛关注与不断研究，在 KWL 的基础上演变出了许多的衍生模式。Carr 与 Ogle(1987)共同提出 KWL+教学模式，该模式为在原本的 KWL 教学模式最后再附上 mapping 或 summarization，让学生在学习任务完成后注重概括总结。Reid（1990）等随后又发展出 KWHLS，期望提高学生的问题解决能力，其中，"H"指的是"How will I learn it and work with others"；而"S"指的是"How will I share the information I have learned"。2000 年 Miller（2002）又对 KWLS 图表进行研究，扩展的"S"指的是"What do I still need to know"，即学生还想要继续学习什么，增加的"S"栏强调学习者需要进一步的好奇心来提高对某一主题的元认知。其后，Szabo 于 2006 年在 KWL 模式的基础上发展了 KWHHL，第一个"H"栏是给学生记录阅读中感到困惑的生词，第二个"H"栏是用来记录学生的学习心得。随后，在 2012 年，Emaliana 开发了类似于 KWLS 的 KWLM 变体，其中，"M（More）"意味着要学习得更多。

KWL 教学模式不断地被研究者修改、调整。但总体来说，KWL 原始模式的研究者和研究次数是最多的，且仔细研究其模式变体的内涵意义，我们不难发现，这些演变模式都只是围绕 KWL 进行细节上的修改，并未完全颠覆 KWL 对原有教学产生的影响（董艳、聂静雨 & 蔡翔英，2020）。由此，本研究依旧选用最原始的 KWL 教学模式，并在行动研究过程中根据学生的具体反馈情况对 KWL 教学模式做出进一步调整。本研究采用的 KWL 教学模式的具体课堂实施过程修改于 Ogle（1986）与庄海滨和李琼（2019)的研究。

Figure 2.1 KWL Strategy Sheet (Ogel,1986)

三、研究方法

1. 研究问题

1.运用 KWL 教学模式前后，初中英语学习者的外语阅读焦虑程度有何改进？

2.运用 KWL 教学模式前后，初中英语学习者的外语阅读理解水平是否提高？

2. 研究对象

研究对象为北京市某中学的 19 名初三学生（14 名女生、5 名男生）。在研究进行前，研究者已经教授 19 名学生两学期的英语课程。一方面，研究者了解学生的英语学习特征、学习情况，了解学生的课堂表现及相应的应对方案。另一方面，学生也适应研究者的课堂风格和教学方式，双方较为熟悉，为课程的顺利进行提供了有力支持。19 名学生均有一定的英语水平，英语学习年限为 7-9 年不等，学生课堂状态较为积极，迫切需要能顾将课堂效益提升到最大的阅读教学模式，来帮助学生提升阅读能力水平。

3. 研究工具

本研究采用的研究工具主要包括外语阅读焦虑量表、阅读理解测试试卷、半结构化访谈三种方式。外语阅读焦虑量表测得 19 名学生在研究前后的焦虑程度变化，阅读理解测试试卷测得学生在行动研究前后的阅读成绩变化，以上两项研究工具可测得量化数据。此外，研究者将通过课堂录音、教师反思、半结构化访谈、学生课堂笔记等质性数据进行进一步分析。

3.1 外语阅读焦虑量表

为了测得 KWL 教学模式对初中英语学习者阅读焦虑程度的影响差异，本研究将分别在研究开始前后两次分发调查外语阅读焦虑测量问卷。问卷采用取自 Saito (1999) 的外语阅读焦虑量表（见附录 A），包括焦虑量表 20 个测试项目，采用李克特 5 点量表形式计分，1 代表非常不同意，2 代表不同意，3 代表中立，4 代表同意，5 代表非常同意。该问卷由研究者翻译为中文（见附录 B），最终问卷得分越高，表明被试者的焦虑程度越高，焦虑总分数应分布在 20 分至 100 分之间。

3.2 阅读理解测试

阅读理解试题全面考查学生的阅读能力，涉及获取要点和关键信息，推断隐含信息（包括事实观点类推断、语境中词义推断、作者写作目的推断等），概括段落或语篇主旨等阅读技能。本研究的前测、后测题目来源于 2022 年与 2023 年北京市中考英语真题的阅读部分，该部分共分为 A、B、C、D 四篇，共 13 小题，每题 2 分，共计 26 分。阅读理解是英语中考所占分值最大的题型，是笔试总分 60 分的 43%。阅读理解根据考察目标不同，文章类型主要分为应用文、记叙文、说明文、议论文四种，题目类型主要为细节理解题、主旨大意

题、推理判断题和词义猜测题等。

3.3 访谈

实验后的访谈能够进一步地反应 KWL 教学模式对英语学习者产生的影响，从而对行动研究计划做出适当的调整。本研究采取非结构化访谈，访谈对象将根据阅读测试分数分层抽样 6 名学生，研究者将从一系列初步问题开始指导对话，并且可以灵活根据访谈过程的具体趋势来增加更富有成果的话题。同样，如果参与者对某个特定话题没有反应，研究者也可以跳过一些问题。

四、行动计划的实施与调整

首先，根据教师反思日志、课堂教学观察、学生课堂笔记、学生访谈记录等，在第一轮行动研究中，主要发现了以下几个问题并提出了相应的解决方案：首先，KWL 教学模式的预先阶段应先将文章的生词讲解给学生，这样学生在使用的时候就会更顺利；其次，针对不同的学生应当做出不同的要求，基础稍弱就要求写出关键词、关键短语、基础稍强就要求写出完整句子、完整段落，这在一定程度上可以规避学生抄书的行为，设置符合水平的要求，更利于学生完成阅读任务；再者，在 W 栏的提问过程中发现，经常出现学生提出的问题无法在文中找到答案，而且是一些老师也无法解答的问题，所以教师可以提供一些定向的文中存在答案的问题，展示给学生，让有需要的学生自由选择使用，这就确保了 L 阶段可以顺利进行，对于学生自发提出的无法解决的问题，留作课后的拓展任务；最后，对于 KWL 栏的反馈，也得到了更好的处理方式，教师意识到需给学生一定的反馈，而不是收上来就结束任务，所以教师可以在每节课的开头简短 3 分钟，随机抽取两名同学的 KWL 表格进行点评，并且就其中容易犯的语法错误、拼写错误、词汇错误做更正。

五、数据分析

5.1 外语阅读焦虑程度

在外语学习焦虑程度前测中，19 名研究对象的焦虑分值范围为 36-88，Cronbach's α系数（Cronbach's Alpha）为 0.869。在后测中，19 名研究对象的焦虑分值范围为 20-75，Cronbach's α系数（Cronbach's Alpha）为 0.971。前后测中，α >0.8, 说明两次量表测试信度很好（见表 5.1.1）。

Table 5.1.1 Reliability Statistics of Foreign Language Reading Anxiety

	Range (20-100)	Cronbach's Alpha	Cronbach's Alpha Based on standardized Items

| FLRA（pre-test） | 36-88 | 0.869 | 0.857 |
| FLRA（post-test） | 20-75 | 0.971 | 0.972 |

数据表明，在行动研究前后，学生在后测中的外语阅读焦虑程度（M=48.16，SD=15.331）低于前测中的外语阅读焦虑程度（M = 60.42, SD = 12.646），且此差异是统计学显著的，t (18) = 3.280, p < 0.05, d = 0.753, 95% CI [4.409; 20.117] (见表 5.1.2)。由此可见，KWL 教学模式能够有效降低学生的外语阅读焦虑程度，帮助学生克服在阅读过程中产生的焦虑情绪。同时，在半结构化访谈中，6 名研究对象都对 KWL 阅读教学模式的作用持有肯定态度，研究对象表示愿意在高中英语阅读中继续使用这种教学模式，并且认为该种教学环境能够有效的帮助其自身缓解阅读过程中的紧张感，让其以更快的融入文本内容语境之中去。

Table 5.1.2 T-test of Foreign Language Reading Anxiety Before and After the Treatment

	Paired Differences				
	M	SD	t	df	Sig.(2-tailed)
Foreign Language Reading Anxiety Before the treatment	60.42	12.646	3.280	18	.004
Foreign Language Reading Anxiety Before the treatment	48.16	15.331			

Note. M = mean; SD = standard deviation

5.2 阅读理解水平

在阅读理解程度前测中，19 名研究对象的前测成绩范围为 8-24，后测成绩范围为 10-24，前后测试中，Cronbach's α系数均为（Cronbach's Alpha）为 0.996，α >0.9，说明前测量表信度很好（见表 5.2.1）。

Table 5.2.1 Reliability Statistics of Reading Comprehension Scores

	Range (0-26)	Cronbach's Alpha	Cronbach's Alpha Based on standardized Items
Scores (pre-test)	8-24	0.996	0.998
Scores (post-test)	10-24	0.996	0.999

数据表明，在行动研究前后，学生的后测中的外语阅读理解成绩（M=15.37，SD=4.669）高于前测中的外语阅读理解成绩（M = 17.79, SD = 3.326），且此差异是统计学显著的，t (18) = -2.426 p < 0.05, d = 0.557, 95% CI [-4.518 ; -.324] (见表 5.2.2)。研究对象在半结构化访谈中表明，KWL 教学模式能够帮助学生更积极的进入英语阅读氛围中，学生们无论是主动的还是被动的，他们都需要在 KWL 表格中写下问题与答案，所以这就要求了每位学生都需要一

定的参与度。一些学生还表明，KWL 表格能够帮助积累写作素材，因为在 K 栏和 L 栏学生时常会大篇幅进行写作来表达自己的观点，即便观点中很多内容是来自于文章中的生词短语，但是这能帮助他们积累这些陌生的词汇短语，用于今后的写作练习中。根据半结构化访谈与量化数据表明，KWL 教学模式能够有效帮助初中生改善阅读习惯，激发阅读兴趣，进一步提升学习者的阅读理解水平，对阅读教学具有积极作用。

Table 5.2.2 T-test of Reading Comprehension Score Before and After the Treatment

	Paired Differences				
	M	**SD**	**t**	**df**	**Sig.(2-tailed)**
Reading Comprehension scores Before the treatment	15.37	4.669	-2.426	18	.026
Reading Comprehension scores After the treatment	17.79	3.326			

Note. M = mean; SD = standard deviation

六、结论

根据量化数据与质性数据可得出，KWL 教学模式能够有效帮助初中生改善阅读习惯，激发阅读兴趣，对阅读教学具有积极作用。KWL 教学模式能够降低学习者的阅读焦虑程度，进一步提升学习者的阅读理解水平。但是，本研究也存在一定的限制还未得到妥善解决，首先，研究对象采用教学自然班属于方便抽样，在总体代表性方面较弱。其次，对于学生英语阅读理解水平与阅读焦虑程度的变化并不能排除是否受到其他因素的影响。

参考文献:

Carr, E., & Ogle, D. M. (1987). K-W-L plus: a strategy for comprehension and summarization. *Journal of Reading*，*30*（7）, 628-631.

Carrell, P. L. (1983). Some issues in studying the role of schemat or background knowledge in second language comprehension. *Reading in a Foreign Language*, 2, 81-92. https://bit.ly/40gnBiu

Emaliana, I. (2012).The effectiveness of the KWLM technique in the teaching of reading comprehension in the non-English departments（Thesis）.DISERTASI dan TESIS program Pascasarjana UM.

Grabe, W., & F.L. Stoller. (2011). Teaching and Researching Reading (2nd Edition). London: Routledge.

Miller D. (2002). Reading with meaning: teaching comprehension in the primary grades. Portland: Stenhouse Publishers.

OCED. (2019). PISA 2018 assessment and analytical framework. https://www.oecd.org/education/pisa-2018-assessment-and-analytical-framework-b25 efab8-en.htm (accessed 12/03/2022)

Ogle, D. M. (1986). KWL: a teaching model that develops active reading of expository text. *Reading teacher*, *39*（6）, 564-570.

Piaget, J. (1936). Origins of Intelligence in the Child. London: Routledge & Kegan Paul.

Raines, D. A. (2018). Using a KWL chart to bridge the theory-practice gap. *Nursing education perspectives*，*39*（3）, 182-183.

Reid, J. A., Forrestal, P.， & Cook J. (1990). Small group learning in the classroom. Western Australia: Chalkface Press.

Szabo S. (2006). KWHHL: a student-driven evolution of the KWL. *American secondary education*，*34*（3）, 57-59，61-67.

Szabo S. (2007).The K-W-L strategy: helping struggling readers build evidence of their learning.*Thinking classroom*，*8*（2）,32-37.

Vygotsky, L.S. (1978). Mind in Society. Cambridge: Harvard University Press.

曾用强. 2017. 中国英语能力等级量表的：阅读量表：制定原则和方法.《外语界》,（5）, 2-11.

董艳,聂静雨,蔡翔英.2020. 关联新旧知识的 KWL 教学模式发展研究. 电化教育研究,41(08):33-42.

王蔷、敖娜仁图雅. 2015. 中小学生外语阅读素养的构成及教学启示,《中国外语教育》,（1）,16-24.

庄海滨,李琼.2019. KWLS 模式在英语阅读教学中的实践探索. 中小学外语教学(中学篇),42(01),49-54.

A Self-Built Corpus-based Study of Directives between Non-native and Native English Teachers from University EFL Classrooms in China
(谭芳、张策 大连理工大学)

Abstract: English teachers' use of directive is crucial for effective teaching and communication in the EFL classroom. While previous studies have mainly examined the directives in classroom discourse, the comparison and contrast between the non-native and native English teachers' directives in EFL university classrooms has not been too much talked about due to the limited availability of spoken discourse corpora.

Inspired by the previous empirical studies of teachers' directives as well as the theoretical literature of politeness, based on the two self-built university English classroom corpora at the tertiary level—Chinese English teachers' (non-native English teachers') EGP classroom discourse corpus and the foreign English teachers' (native English teachers') ESL classroom discourse corpus, the present study further adjusts and specifies the classification of teachers' directives in EFL classroom; explores, compares and contrasts the structure and function of the teachers' directives in the two corpora, in order to provide implications for the English teachers' professional development in China.

Results show that non-native English teachers tend to employ imperative directives with incomplete structures, accounting for 52.81% of all sentence patterns. While native English teachers prefer declarative directives with complete structures, making up 50.76% of all sentence patterns. When using imperative directives, native English teachers favour inclusive forms such as "Let's" and avoid directly commanding bare directives. Non-native English teachers frequently employ politeness markers like "sorry" and "please" in their imperative directives. In the use of declarative directives, native English teachers adopt sentences indicating permission more frequently than non-native English teachers. Native English teachers use indirect directives to imply what students should do; on the contrary, non-native English teachers do not do like this at all.

Based on these findings, it can be inferred that non-native English teachers use classroom directives more straightforward, often relying on explicit instructions to motivate students to complete tasks, while native English teachers emphasize equal communication with students and the bidirectional flow of knowledge in EFL classrooms in China. These differing role orientations and distinguishing directive styles reflect the influence of different cultural backgrounds and educational philosophies.

Key Words: Self-built Corpus; Directives; Non-native English Teachers; Native English Teachers; University EFL Classrooms

Introduction

Directives, as one type of speech act, are "attempts by the speaker to get someone to do something (Searle, 1976: 3). Directives perform various functions like asking, ordering, requesting and advising (Searle, 1976). In the context of classroom, directives are the utterances produced by the instructor to ask the students to perform an action or to instruct the students on something (Searle, 1976). Though different types of speech acts are employed by teachers, directive speech acts are the most frequently used one in EFL classrooms (Yanti et al., 2021). Directives are embedded in the institutional setting of EFL instruction, the teachers' use of directives manifest the teachers' authority in the classroom. The frequency of production of directives is

closely related with the type of learning. The more directives from the teacher, the more teacher-centered the class is; the less directives from the teacher, the less teacher-centered the class is (Haryanto & Mubarok, 2018).

Against a backdrop of a long-lasting discussion and exploration about the aforementioned shifts in English classrooms, the present research explores directives as both a dimension of interpersonal relationship and a device of pedagogic practice in classrooms in which English is both the medium and the target of instruction. The article analyses the use of directives between non-native and native English teachers from university EFL classrooms in China for the purpose of informing ESL pedagogy.

Methodology

Research Questions

In view of the above, the purpose of the present study is to compare and contrast the use of the directives between non-native and native English teachers from university EFL classrooms in China, with the intent to inform pedagogy and teacher's profession in classroom contexts. The study tries to answer the following three questions.

1. What directives were employed by the non-native and native English speaking teachers in university ESL classrooms in Chinese mainland?

2. What are the similarities and differences between the non-native and native English speaking teachers when they are using directives in university ESL classrooms in Chinese mainland?

3. What are the teacher-student relationships between the non-native and native English speaking teachers in university ESL classrooms in Chinese mainland?

Framework and Design of Directives Classification System

In order to answer the above three questions, the research first designed the directives classification system based on the politeness strategy theoretical framework of Brown & Levinson (1987).

Brown and Levinson's (1987) politeness strategy theory encompasses four tiers:

(1) Bald-on-record Strategy: This strategy involves the direct and unadorned expression of one's intentions or opinions, devoid of any cushioning or euphemistic language. It disregards the recipient's feelings or reactions, potentially causing embarrassment or dissatisfaction, thereby instigating conflict or offense.

Examples:

<89><Female Teacher> Speak louder. *(LD_1_19_N2_FCYF)*

(2) Positive Politeness: The positive politeness strategy focuses on fulfilling the recipient's positive face needs, which encompass desires to be liked, approved of, and included. Employing this strategy, speakers typically express admiration, gratitude, empathy, or friendship towards the recipient, fostering or maintaining a close relationship.

Examples:

<173><Male Teacher>Okay. Good. Talk to your partner. *(TT_2_21_E1_MJW)*

(3) Negative Politeness: Negative politeness strategy prioritizes respecting the recipient's negative face needs, which entail a desire for autonomy, independence, and avoiding impositions or feelings of coercion. When using this strategy, speakers

refrain from directly referring to themselves and instead utilize more indirect, roundabout means to convey their wishes or requests.

Examples:

<87><Female Teacher> Sorry, sorry to interrupt, speak louder, okay? *(LD_1_19_N2_FCYF)*

(4) Off-record Strategy: This strategy involves communicating information through implication, metaphor, or other indirect means. It allows the listener to interpret the message based on their own understanding, thereby avoiding the potential awkwardness or conflict that may arise from direct expression.

Example:

<156><Male Teacher> Dead Silence.*(LD_1_19_E1_MAS)* (Hinting students to actively participate in the discussion）

So based on Brown and Levinson's (1987) politeness strategy continuum (continuum from least politeness to most politeness), this study designed a directives classification system that is composed of three levels of syntactic types, communication strategy/purpose and politeness degree. (Detailed are shown in Table 1)

Table 1 Directives Classification Continuum

Form of Strategy Implementation		Sample Utterances	Formulae
I **Imperatives**	1. Bold	<89><Female Teacher> Speak louder.*(LD_1_19_N2_FCYF)* <29><Female Teacher>Try to find out some details, some words, some sentences in this part.*(LD_1_19_N1_FHQ)* <42><Female Teacher><((Student's Name))> , (read) paragraph fifteen.*(LD_1_19_N1_FLH)*	Do X. Be + adj.
	2. Politeness Markers	<173><Male Teacher>Okay. Good. Talk to your partner. *(TT_2_21_E1_MJW)* <173><Male Teacher>Okay. Good. Talk to your partner. *(TT_2_21_E1_MJW)* <4><Female Teacher>Yes, good. Another one. *(LD_1_19_N1_FHQ)*	Okay/well/good/yeah/yes, do X.
	3. Mitigated	<13><Female Teacher>Please read this part and answer the following questions. *(LD_1_19_N1_FHQ)* <17><Female Teacher>Okay, please read this part quickly. *(LD_1_19_N1_FHQ)*	Please do X.

	4. Inclusive	<63><Female Teacher>So let's think about the question. *(LD_1_19_N2_FCYF)* <69><Female Teacher>Okay now let's come to the second part. *(LD_1_19_N1_FHQ)*	Let's do X.
	5. Apologize to the person	<87><Female Teacher> Sorry, sorry to interrupt, speak louder, okay? *(LD_1_19_N2_FCYF)* <141><Female Teacher> Uh sorry, uh just try to speak again.*(LD_1_19_N2_FCYF)*	Sorry +Noun. Sorry + do X.
II **Declaratives**	1. Performatives / Hedged Performatives	<70><Female Teacher>I will give you one minute to think about the sentence. *(LD_1_19_N2_FCYF)* <17><Female Teacher>I'd like to, uh, invite one student, to read the first paragraph. *(LD_1_19_N1_FJLZ)*	I ask you to do X. I'd like to ask you to do X.
	2. Obligation	<14><Female Teacher>You have to talk for a minute and thirty seconds. *(LD_1_19_E1_FJ)* <131><Female Teacher>You should take it with a grain of salt. *(LD_1_19_E1_FJ)*	You have to/must /should do X.
	3. Want/Need	<6><Female Teacher>I want to ask you a question. *(LD_1_19_N2_FCYF)* <7><Female Teacher>You need to put the character or the story in (.) setting. *(LD_1_19_N1_FHQ)* <72><Female Teacher>But you need to summarize the reasons by yourself, okay. *(LD_1_19_N1_FHQ)*	I want/need you to do X. You need to do X.
	4. Permissive	<79><Female Teacher>You can say a sentence you write or you can also say a sentence from the QQ group. *(LD_1_19_N1_FJLZ)* <295><Female Teacher>You will read the	You can/will/are going to do X.

		first paragraph and last paragraph, then tell me the main idea of this passage or predict what's this passage about. *(LD_1_19_N1_FLYY)*	
	5. Use of in-group identity marker	<193><Female Teacher>Okay, we're gonna go on break. *(LD_1_19_E1_FJ)* <91><Female Teacher>Okay, we can speak together. *(LD_1_19_N2_FCYF)*	We can/will/are going to do X.
	6. Hint	<16><Female Teacher> You're playing a little game on your phone? *(HQ_1_20_E2_FBJ)* <75><Male Teacher> Oh, guys. Someone calls me.(TT_2_21_E1_MJW) <156><Male Teacher> Dead Silence.*(LD_1_19_E1_MAS)*	A strategy that conveys information through implication, metaphor, or other indirect ways, allowing the listener to interpret the information based on their own understanding.
III **Interrogatives**	1. Feasibility/Ability	<43><Male Teacher>Can you please open the result on your phone? *(LD_1_19_E1_FJ)* <51><Male Teacher>Can you paraphrase those?*(LD_1_19_E1_MJ)*	Can you do X?
	2. Consultative Devices	<2><Female Teacher>Could you tell me the title of the story?*(LD_1_19_N1_FHQ)* <20><Female Teacher>Could you, could you find the name in this part? *(LD_1_19_N1_FHQ)*	Could you do X?
	3. Suggestory Formulae	<14><Female Teacher>How about, let's take a very quick one to two minutes break?*(LD_1_19_E1_FJ)*	How about X?
	4. Willingness	<22><Female Teacher>Would you please read the first paragraph? *(LD_1_19_N1_FJLZ)*	Would you like to do X?

Corpus Description

The present study is based on two self-built classroom discourse corpora. One is non-native English teachers' classroom discourse corpus named HJL-EGPCDC (Heilongjiang, Jilin, Liaoning-English for General Purposes Classroom Discourse Corpus). The other is native English teachers' classroom discourse corpus named

FT-EFL-CDC (Foreign Teacher-English as a Foreign Language-Classroom Discourse Corpus). The data of present study is chosen from the two self-built classroom discourse corpora. The data used in the present study involves a total of 69 English lectures (each class lasting approximately 45 minutes) from six Chinese and six native teachers. The non-native English teachers' classroom discourse recordings were collected from Jilin Province, Liaoning Province, and those of Native English speaking teachers were from Liaoning Province and Tianjin City in China, all of which were English courses in universities, the former primarily consisting of non-English majors' college English courses and the latter mainly consisting of oral English courses for the first-year and the second-year non-English majors. Only the native English teachers' oral classes from Tianjin was recorded in the English major's classroom.

Corpus Description of HJL-EGPCDC and the Non-native English Teachers' EGP Classroom Discourse Corpus in the Present Study

(1) Corpus Description of HJL-EGPCDC

HJL-EGPCDC (Heilongjiang, Jilin, Liaoning-English for General Purposes Classroom Discourse Corpus) is a self-built corpus from 2011 till now. These data were collected the three provinces of Northeast China: Heilongjiang, Jilin and Liaoning Province, focusing purely on English classes within the participating universities: Intensive Classes from English Majors in the lower grade, Advanced English from English Majors in the higher grade, and General English Classes from those in non-English majors. The corpus of 52 teachers' EGP classroom transcripts created from the total 12, 408 minutes' (206.8 hours') recordings of lessons from the first, second year and third year of the universities from September, 2011 to April, 2024.

The recordings were part of my research projects and the extended empirical study from the dialogic classroom theory construction in my doctoral research (Tan, 2012). All the recordings were taken in the natural EGP classes, i.e. offline classrooms. Only one teacher's partly classroom recordings were taken online during the epidemic period in 2022.

The teachers' classroom recordings were collected purposefully to vary levels of universities, two teachers are from two 985 project universities, three are from a 211 project university, one is from non-government university and all the others from the common provincial universities. The average age of the teachers is 35.8 and all of them got MA degree. The average teaching experience of the 52 teachers is 11.2 years. After excluding the Chinese, English in HJL-EGPCDC is roughly about 14, 770 types, 945, 809 tokens.

(2) *The Non-native English Teachers' EGP Classroom Discourse Corpus in the Present Study*

In the present study, six Chinese teachers' transcripts in university English classes were taken from HJL-EGPCDC. In the present study, after excluding the Chinese, the size of the corpus in the present study is roughly about 7, 198 types and 61, 484 tokens. All six teachers gave classes in non-English majors. All the courses taught by non-native teachers were offline, and all of them were college English

courses. The teachers came from 985 universities. The total duration of the recordings of non-native teachers' classes was 722 minutes.

The average age of the six Chinese English speaking teachers in universities was 51 years old, with an average teaching experience of 25 years, and 18 lectures were recorded in total. Table 2 introduces the basic information of the course and the teachers.

Table 2 Basic Information of Non-native English Teachers and Their Classroom Discourse

No.	File Name	Duration of Recording	Teachers' Age	Length of Teaching	University Level	Number of Lectures
1	LD_1_19_N1_FLH	183 Minutes 22 Seconds	62	40	985	4
2	LD_1_19_N1_FJLZ	78 Minutes 12 Seconds	59	33	985	1
3	LD_1_19_N2_FCYF	78 Minutes 12 Seconds	29	3	985	2
4	LD_1_19_N1_FHQ	109 Minutes 5 Seconds	49	27	985	3
5	LD_1_19_N1_FLYY	97 Minutes 34 Seconds	51	27	985	3
6	JC_1_04_N1_MXKW	175 Minutes 40 Seconds	53	22	985	5
		Total Duration of Recording: 12 hours 2 Minutes 5 Seconds	Average Age: 50.5	Average Length of Teaching: 25.3		Total Number of Lectures: 18

Corpus Description of FT-EFL-CDC and the Native English Teachers' ESL Classroom Discourse Corpus in Present Study

Six teachers' transcripts in university English classes were taken from FT-EFL-CDC (Foreign Teachers EFL Classroom Discourse Corpus).Among the recorded materials of the native teachers, only one teacher (with 10 recorded classes) conducted online classes (live sessions, featuring real-time interactions between teachers and students), while the rest were offline classes, all of which were English speaking courses. One of the native teachers was from a university in Tianjin participating in the "Double First-Class" Initiative (neither a 985 nor a 211 university), and the other five native teachers taught at 985 universities. The total duration of the recorded classroom sessions by native teachers amounted to 3, 002 minutes. Utilizing AntConc 3.5.8, the corpus analysis revealed 6, 353 types and 252,140 tokens in the classroom discourse of the native teachers. This study recorded the classroom discourse of six native English teachers in universities, with an average age of 34 years and an average teaching experience of 10 years, spanning a total of 51 classes. Detailed corpus statistics are showed in Table 3.

Table 3 Basic Information of Native English Teachers and Their Classroom Discourse

No.	File Name	Nationality	Duration of Recording	Teachers' Age	Length of Teaching	University Level	Offline (Y/N)	Number of lectures
1	LD_1_19_E1_MAS	U. K.	488 Minutes 1 Second	35	10	985	Y	10
2	HQ_1_20_E2_FBJ	U. S. A	535 Minutes 35 Seconds	26	2	985	Y	10
3	LD_1_19_N1_FPD	U. K.	525 Minutes 35 Seconds	Unknown	37	985	N	10
4	TT_2_21_E1_MJW	Australia	547 Minutes 55 Seconds	40	11	Provincial	Y	12
5	LD_1_19_E1_FJ	U. S. A	466 Minutes 58 Seconds	26	1	985	Y	4
6	LD_1_19_E1_MJ	U. S. A	438 Minutes 13 Seconds	42	0	985	Y	5
			Total Duration of Recording：50 hours 2 Minutes 17 Seconds	Average Age：33.8 Average Length of Teaching：10.2	Average Age：33.8 Average Length of Teaching：10.2			**Total Number of lectures:** 51

Method and Research Procedure

1. The Analytic Scheme of Directives

Taking the research of House & Kasper (1981) as well as that of Brown & Levinson (1987) into consideration, we revised the scheme of classification systems of directives proposed by Liu & Hong (2009). The revised scheme is used for tagging teachers' directives at the three levels of syntactic forms, strategy realisation and pedagogic purposes (Liu & Hong, 2009:5), in the descending order according to the degree of politeness (Brown & Levinson, 1987). The revised coding scheme provides a detailed outline of structures, types and functions teachers' directives and highlights teachers' strategies and politeness. In short, the revised scheme help us analyse teachers' directives according to linguistic, pragmatic criteria as well as pedagogic strategies. This scheme allows us (1) to distinguish the clause types of teachers' directives, i.e. whether the teachers' directives are imperative, declarative or interrogative, (2) to identify the types and functions of the teachers' teaching strategies from the use of teachers' directives, (3) to explore the degree of politeness of the teachers' directives in the EFL classroom. Consequently, it was possible for us to analyse the efficacy of teachers' directives comprehensively. On the other hand, it will help teachers to gain a sense of efficacy by giving them a roadmap for what they can do before and during instruction in classrooms.

2. Bottom-up Corpus-driven Analysis of Directives

Since the self-bulit corpus includes the monologic segments and the dialogic question-and-answer sessions of classroom discourse, both the monologic segments of the teacher's talk and the dialogic question-and-answer session between teachers and students are taken into the the scope of the the identification of teachers' directives. Most previous work on directives has focused on dialogic contexts (e.g. Bax 1986; Curl & Drew 2008; Ervin-Tripp 1976; Weigel & Weigel 1985).

3. Identification of Directives.

(1) We adopted the bottom-up corpus-driven analysis in the present study. Two researchers read line by line and identified directives manually respectively; (2) Inter-rater reliability of Directives. Directives of 90 percent agreement between the two researchers were picked out as the target directives. The result of the consistency was 91 percent. (3) Annotation and Retrieval of Directives. We annotated the teachers' directives in classroom transcripts by Notepad++. After the annotation, we employed AntConc to retrieve the concordances of teachers' directives and imported the output into Excel; (4) Analysis of Teachers' Directives. We classified and coded for the teachers' directives by the revised analytic scheme in Excel, and finally analysed the types and functions of teachers' directives between non-native and native English teachers in EFL university classrooms.

Results and Analysis

Non-native and Native English Teachers' Directives in Chinese University ESL Classrooms

For non-native English speaking teachers:

(1) Bold imperative directives including the characters of direct, low-politeness, and highly forceful tone are the preferred choice among non-native teachers .

(2) non-native teachers tend to employ limited sentence types of directives and rigid, uniform methods of directives expression.

(3) The primary communicative purpose of non-native teachers' use of directives is focused on organizing classroom teaching, and when used for purposes other than classroom teaching, the imperative force tone is particularly strong.

Table 4 Proportion of Non-native Teachers' Directives in All Patterns of Classroom Directives

Non-native Teachers' Classroom Directives Sentence Pattern		Frequencies (606 in total)	Proportion	Coverage
I		**Imperatives**	320	52.81%	
	1	Bold	143	23.60%	all 6
	2	Politeness Markers	66	10.89%	all 6
	3	Mitigated	36	5.94%	all 6
	4	Inclusive	70	11.55%	all 6
	5	Apologize to the Person	5	0.83%	2
II		**Declaratives**	227	37.45%	
	6	Performatives/Hedged Performatives	43	7.09%	5
	7	Obligation	21	3.46%	5
	8	Want/Need	30	4.95%	3
	9	Permissive	44	7.26%	all 6
	10	Use of in-group Identity Marker	89	14.69%	all 6
	11	Hint	0	0	0
III		**Interrogatives**	**59**	**9.74%**	
	12	Feasibility/Ability	34	5.61%	4
	13	Consultative Devices	10	1.65%	4
	14	Suggestory Formulae	1	0.17%	1
	15	Willingness	14	2.31%	4

(1) Bold imperative directives including the characters of direct, low-politeness, and highly forceful tone are the preferred choice among non-native teachers.

First, it is necessary to clarify that the preferred choice here refers to the highest

proportion of bold imperative directives appearing in the overall classroom directives. Such sentence patterns covered the classroom recordings of all six teachers. A total of 52.81% of the directives used by non-native teachers are imperative directives, exceeding half of the total number of directives.

Drawing on Brown & Levinson's (1987) literature on politeness strategies, we can categorize imperative sentences as relatively low in politeness, while declarative and interrogative sentences as relatively high in politeness. In the context of non-native teachers' use of classroom directives, the data indicate that imperative directives are the most frequently employed, with 320 instances accounting for 52.81% of all directives. Conversely, interrogative directives are the least used, with only 59 instances, constituting 9.74% of the total. Notably, non-native teachers favor the use of "bold" imperatives, accounting for 23.60% of all directives. (Excerpt 1, 2)

Excerpt 1

<89><Female Teacher>Speak louder.*(LD_1_19_N2_FCYF)*

Excerpt 2

<6><Female Teacher>Find something that you like, find something that you really like. *(LD_1_19_N1_FJLZ)*

Variations of imperatives and directives accompanied by various politeness markers (Politeness Markers, Mitigated, Apologize to the person) make up a combined proportion of 17.66%, occurring 107 times. Non-native teachers tend to rely heavily on the most direct and commanding form of imperatives, with "bold" imperatives being their preferred choice. In organizing classroom teaching and facilitating classroom interactions, non-native teachers employ a significant number of "bold" directives. While they also frequently utilize "mitigated" directives, directives with "politeness markers," and directives that "apologize to the person" (Politeness Markers, Excerpt 3; Mitigated, Excerpt 4; Apologize to the person, Excerpt 5), these strategies are comparatively more direct, albeit with a softer tone than "bold" imperatives. However, the imperative nature of these directives remains evident.

Excerpt 3

<121><Female Teacher>Okay, sit down. *(LD_1_19_N1_FLYY)*

Excerpt 4

<25><Male Teacher> Sit down, please. *(JC_1_04_N1_MXKW)*

Excerpt 5

<87><Female Teacher> Sorry, sorry to interrupt, speak louder. *(LD_1_19_N2_FCYF)*

(2) Non-native teachers tend to employ limited sentence types of directives and rigid, uniform methods of directives expression.

Compared to imperative sentences and their strategic forms, non-native teachers employ fewer declarative sentences and their strategic forms that are considered more polite. Among declarative sentences, the strategies of "Use of in-group identity marker", "Permissive", and "Performatives/Hedged Performatives", primarily used to organize classroom activities, are the most prevalent, occurring 176 times in total, accounting for 26.9% of all directives. Other strategic forms, such as "Want/Need"

and "Obligation", are less frequently used, with a total of 51 instances, representing 8.41% of the total. These data suggest that within different declarative sentence patterns, non-native teachers tend to rely on a limited set of three strategies: "Performatives/Hedged Performatives", "Permissive", and "Use of in-group identity marker", with the first two strategies being second only to imperatives in terms of their imperative force.

Interrogative sentences and their related forms are the most indirect, polite, and offer the greatest scope for meaning negotiation. However, non-native teachers use them with the lowest overall frequency. Among them, "Feasibility/Ability" is the most frequent, occurring 34 times, accounting for 5.61% of the total. The more euphemistic "Consultative Devices", such as "Could you do X?", appear less frequently, with only 10 instances, representing 1.65% of all directives. "Willingness", which carries the most negotiation potential, ranks second among interrogative directives, occurring 14 times, or 2.31% of the total.

Furthermore, it is evident from the following Table 4 that not all non-native teachers in the recorded English classrooms used all types of classroom directive realizations. Only three teachers employed "Want/Need" sentence patterns to guide classroom instruction, two teachers used "Apologize to the person" directives, and only one teacher utilized "Suggestory Formulae." Notably, none of the non-native teachers used "Hint" directives, indicating that they did not employ strategies that require students to interpret and engage in discussions. This also suggests that non-native teachers lack a rich repertoire of English expression and may struggle to articulate their requests or demands in a polite and flexible manner.

(3) The primary communicative purpose of non-native teachers' use of directives is focused on organizing classroom teaching, and when used for purposes other than classroom teaching, the imperative force tone is particularly strong.

The directives employed by non-native teachers are primarily utilized for classroom organization purposes, with minimal application to discipline enforcement. Among the 606 classroom directive utterances identified in the transcribed discourses of six non-native teachers , only 10 utterances, accounting for 1.65%, served the function of discipline enforcement, while the remaining 596 utterances were directives aimed at organizing classroom activities or teaching steps.

When not used for the purpose of organizing classroom activities (such as enforcing classroom discipline, arranging pre-class preparations, encouraging student interaction, or providing feedback), the selected directives tend to be more direct and imposing. Notably, when non-native teachers use directives to provide feedback on students' questions, they have even resorted to directives with a punitive undertone (Excerpt 6).

Excerpt 6

<29><Female Teacher>[C]-我告诉你是十三，回去写两百遍，下节课给我交上来。-[C] *(LD_1_19_N1_FLYY)*

The infrequent use of directives for discipline enforcement by non-native teachers suggests that university students possess strong self-discipline and adhere to classroom rules. On the other hand, it also reflects that the directives used by teachers

in responding to students' answers fail to consider students' face, fostering a hierarchical classroom environment that deprives students of autonomy and fosters passive learning.

For native English speaking teachers：

(1) Permissive declarative directives including the characters of indirect, high-politeness, and moderately forceful tone are the preferred choice among native teachers.

(2) Native teachers tend to employ a rich variety of directive sentence types and flexible expression methods, including the use of implicit hint strategies.

(3) The communicative purposes of directives used by native teachers encompass not only organizing classroom instruction but also arranging pre-class preparations for students, encouraging student interaction, and rarely enforcing classroom discipline, all with a relatively mild tone.

Table 5 Proportion of Native Teachers' Directives in All Patterns of Classroom Directives

Native Teachers' Classroom Directives Sentence Pattern			Frequencies (1, 509 in total)	Proportion	Coverage
I		**Imperatives**	**650**	**43.08%**	
	1	Bold	248	16.44%	all 6
	2	Politeness Markers	153	10.14%	all 6
	3	Mitigated	30	1.99%	3
	4	Inclusive	215	14.25%	all 6
	5	Apologize to the Person	4	0.26%	4
II		**Declaratives**	**766**	**50.76%**	
	6	Performatives/Hedged Performatives	34	2.25%	5
	7	Obligation	125	8.28%	all 6
	8	Want/Need	105	6.96%	all 6
	9	Permissive	311	20.61%	all 6
	10	Use of in-group Identity Marker	184	12.19%	all 6
	11	Hint	7	0.46%	4
III		**Interrogatives**	**93**	**6.16%**	
	12	Feasibility/Ability	55	3.64%	all 6
	13	Consultative Devices	10	0.66%	4
	14	Suggestory Formulae	13	0.86%	2
	15	Willingness	15	0.99%	5

(1) Permissive declarative directives including the characters of indirect, high-politeness, and moderately forceful tone are the preferred choice among native teachers.

Among the directives used by native teachers, declarative directives are the most prevalent, with a total of 766 instances, accounting for 50.76% of all directives. In contrast, imperative directives are used less frequently, appearing 650 times, constituting only 43.08% of the total directives. Interrogative directives are the least utilized, with a mere 93 instances, accounting for 6.16% of the total.

Native teachers predominantly employ permissive declarative directives, which occur 311 times, making up 20.61% of all directives. Notably, the "permissive" is the most frequently used among all directives employed by native teachers, representing their preferred choice (Excerpt 7). Regardless of whether the directives are used for organizational or disciplinary pragmatic purposes, native teachers extensively utilize "Permissive" directives. These data indicate that native teachers have a habit of using declarative directives with a more euphemistic tone, moderate imperativeness, and a high level of politeness.

Excerpt 7

<325><Female Teacher> You can change it.*(HQ_1_20_E2_FBJ)*

Furthermore, the use of relatively more imperative "Performatives/Hedged Performatives" within declarative sentences is minimal, occurring only 34 times, accounting for 2.25% of the total directives. In contrast, the "Use of in-group identity marker" is frequently employed, appearing 184 times, comprising 12.19% of the total directives. Directives expressing "Obligation" and "Want/Need" within declarative sentences each constitute a significant proportion, totaling 230 times and accounting for 15.24% of the total. These directives are more indirect in nature, with a noticeably softer tone and less overt imperative meaning. It can be observed that in the use of classroom directives, native teachers intentionally diminish the directness and imperativeness of their directives to enhance equal communication with students.

(2) Native teachers tend to employ a rich variety of directive sentence types and flexible expression methods, including the use of implicit hint strategies.

All the directive strategies listed in Table 1 have been utilized by native teachers. Furthermore, it is noteworthy that when using declarative directives, native teachers employed the strategy of "hint", which refers to a tactic that conveys information through implication, metaphor, or other indirect means, enabling the listener to interpret the information based on their own comprehension. The utilization of this hinting strategy did not merely manifest in the classroom data of a single native teacher but cover the classes of all four native teachers (Detailed are shown in Table 5). These data indicate that native teachers, quite commonly, employ more indirect directive strategies in English language classrooms, leaving room for students to engage in discussions.

Compared to declarative sentences and their strategic forms, native teachers use imperative sentences with lower levels of politeness and their strategic forms less frequently. Among imperative sentences, the "Bold," "Inclusive," and "Politeness Markers" strategies, primarily used for organizational pragmatic purposes, are the

most prevalent, occurring a total of 616 times, accounting for 40.82% of the total frequency of directives. Other strategic forms, such as "Mitigated" and "Apologize to the Person", are used infrequently, appearing only 34 times, which is 2.25% of the total directives. This suggests that native teachers relatively use fewer discourse markers like "please" and "sorry" , thereby fostering a closer relationship between teachers and students.

Interrogative sentences and their related strategic forms, when employed as classroom directives, are the most indirect, exhibit the highest level of politeness, and offer the greatest room for meaning negotiation. However, their overall usage frequency by native teachers is the lowest. Among them, "Feasibility/Ability" and "Willingness" have relatively higher frequencies, but they only appear 70 times, accounting for 4.63% of the total. "Suggestory Formulae" follow, occurring 13 times, or 0.86% of all directives. Apart from the modal forms, these classroom strategies in interrogative sentences belong to the relatively more polite categories.

The use of "Suggestory Formulae" by native teachers not only encompasses questions like "How about X?" (Excerpt 8) but also includes forms like "Why don't you do X?" (Excerpt 9), demonstrating a flexible application.

Excerpt 8

<14><Female Teacher>How about, let's take a very quick one to two minutes break?*(LD_1_19_E1_FJ)*

Excerpt 9

<3><Male Teacher>Why don't you just study computers now?*(TT_2_21_E1_MJW)*

These data indicate that, native teachers possess relatively rich English expression abilities, enabling them to express directives in a flexible and polite manner.

(3) The communicative purposes of directives used by native teachers encompass not only organizing classroom instruction but also arranging pre-class preparations for students, encouraging student interaction, and rarely enforcing classroom discipline, all with a relatively mild tone.

In terms of pragmatic purposes, there are 1, 507 organizational directives, accounting for 99.87%, while only 2 directives used to organize classroom discipline, accounting for 0.13%, suggesting that discipline-related directives can be negligible. It can be observed that the use of directives is primarily for classroom organization purposes rather than for disciplinary constraints; directives for disciplinary purposes have little statistical significance.

It is noteworthy that, among organizational directives, native teachers' directives are not solely employed for the traditional purpose of organizing classroom teaching but also for arranging pre-class preparations for students (Excerpt 10) and encouraging student interaction (Excerpt 11, Excerpt 12).

Excerpt 10

<1><Female Teacher> Can someone close the back door? *(LD_1_19_E1_FJ)*

Excerpt 11

<156><Male Teacher> Dead Silence.*(LD_1_19_E1_MAS)* (Hinting students to

actively participate in the discussion）

Excerpt 12

<325><Female Teacher> Someone who hasn't spoken yet.*(LD_1_19_E1_FJ)*(Hinting at students who haven't answered questions to speak up.)

Among them, the "Feasibility/Ability" interrogative "Can you do X?" used in Excerpt 10, and the "hint" declarative directives with a softer tone and a high level of politeness employed in Excerpt 11 and Excerpt 12, demonstrate that native teachers emphasize conveying respect and understanding through indirect language. They view students as active and proactive participants, encouraging them to express themselves and guiding the learning process, thereby stimulating students' learning potential and creativity.

Similarities and Differences in Directive Use by Non-native vs Native English Teachers in Chinese University ESL Classrooms
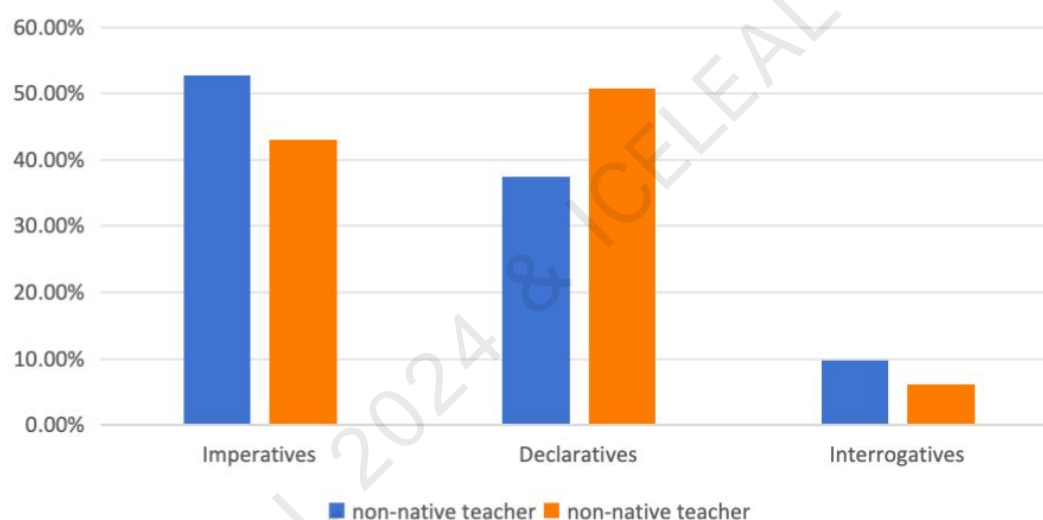
Graph 1 Proportion of Sentence Types between Non-native and Native English Teachers' Classroom Directives

Similarities between Non-native and Native English Teachers' EFL Classroom Directives

There are similarities in the use of classroom directives between Chinese and native teachers. Firstly, as can be seen from the above chart, both Chinese and native teachers use interrogative directives with the lowest frequency. Secondly, among imperative directives, non-native teachers use bold imperatives accounting for approximately 37%, and inclusive directives constitute around 22% of imperative directives. In terms of imperatives, native teachers' usage of classroom directives is quite similar to that of non-native teachers . Native teachers also prefer to use bold imperatives and inclusive directives, but the proportion of bold imperatives used by native teachers is also about 37%, while the use of inclusive directives is even higher, accounting for about 33%. Lastly, under imperative directives, both Chinese and native teachers use "apologize to the person" with the lowest frequency. Non-native teachers only account for 0.83% of the total directives, while native teachers only

account for 0.26%. Additionally, the purpose of using directives for both Chinese and native teachers primarily focuses on organizing classroom teaching.

Differences between Non-native and Native English Teachers' EFL Classroom Directives

(1) Non-native teachers tend to prefer using imperative directives with a relatively low level of politeness in English language teaching classrooms, whereas native teachers are inclined to employ declarative directives with a comparatively higher degree of politeness.

(2) Non-native teachers exhibit a limited range of sentence patterns and relatively rigid expressions when using directives in English language teaching classrooms, whereas native teachers demonstrate a diverse array of sentence structures and flexible modes of expression in their directives.

(3) Compared to non-native teachers, native teachers employ directives for a broader range of purposes, extending to pre-class preparations, encouraging interactions, and other aspects, and their tone is considerably more tactful and euphemistic.
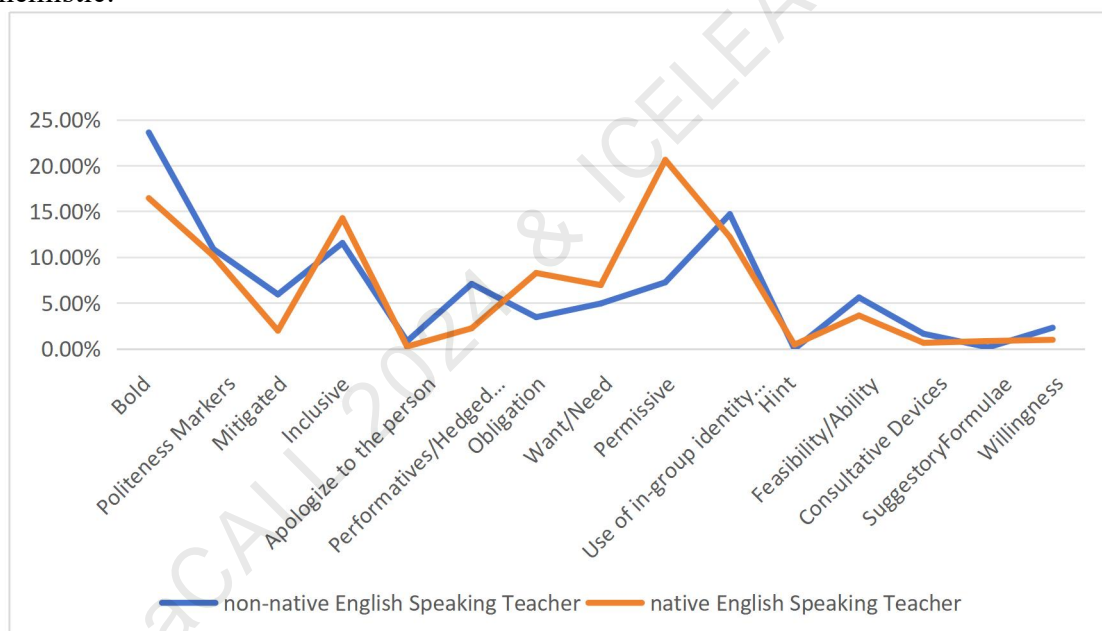
Graph 2 Proportion of Classroom Subcategories Directives between Non-native and Native English Speaking Teachers

(1) Non-native teachers tend to prefer using imperative directives with a relatively low level of politeness in English language teaching classrooms, whereas native teachers are inclined to employ declarative directives with a comparatively higher degree of politeness.

Table 6 Differences between Non-native Teachers ' and Native Teachers' Usage of Directives

Categories	Frequencies in Non-native English speaking teachers	Frequencies in native English speaking teachers	Log-likelihood	Overuse/ Underuse
Imperatives	320	650	+8.68	overuse
Declaratives	227	766	-17.05	underuse
Interrogatives	59	93	+7.24	overuse

Through the calculation of Log-likelihood, it is evident that non-native teachers significantly use imperative directives with direct, low politeness, and strong imperative meanings more frequently than native teachers, whereas the use of indirect, high politeness, and moderately imperative declarative directives is significantly lower among non-native teachers compared to native teachers.

Table 7 Proportion of Non-native teachers ' and Native Teachers' Usage of Directives

	Non-native Teacher	Native Teacher
Bold	23.60%	16.44%
Politeness Markers	10.89%	10.14%
Mitigated	5.94%	1.99%
Inclusive	11.55%	14.25%
Apologize to the Person	0.83%	0.26%
Performatives/Hedged Performatives	7.09%	2.25%
Obligation	3.46%	8.28%
Want/Need	4.95%	6.96%
Permissive	7.26%	20.61%
Use of in-group Identity Marker	14.69%	12.19%
Hint	0	0.46%
Feasibility/Ability	5.61%	3.64%
Consultative Devices	1.65%	0.66%
Suggestory Formulae	0.17%	0.86%
Willingness	2.31%	0.99%

Although direct comparisons of frequencies (Proportion) can yield some conclusions, the calculation of Log-likelihood allows for a more nuanced comparison of the significance in the use of various directive types between Chinese and native teachers (Detailed are shown in Table 8).

Table 8 Differences between the Non-native and Native English Speaking Teachers' Directives (from Log-likelihood and Overuse/Underuse)

Categories and Subcategories		Frequencies in Non-native English Speaking Teachers	Proportion of Non-native English Speaking Teachers	Frequencies in Native English Speaking Teachers	Proportion of Native English Speaking Teachers	Loglikeli-hood	Overuse/ Underuse
Imperatives	Bold	143	23.60%	248	16.44%	+11.44	overuse
	Politeness Markers	66	10.89%	153	10.14%	+0.23	overuse
	Mitigated	36	5.94%	30	1.99%	+19.30	overuse
	Inclusive	70	11.55%	215	14.25%	-2.41	underuse
	Apologize to the person	5	0.83%	4	0.26%	+2.83	overuse
Declaratives	Performatives/Hedged Performatives	43	7.09%	34	2.25%	+24.76	overuse
	Obligation	21	3.46%	125	8.28%	-16.63	underuse
	Want/Need	30	4.95%	105	6.96%	-2.87	underuse
	Permissive	44	7.26%	311	20.61%	-53.94	underuse
	Use of in-group Identity Marker	89	14.69%	184	12.19%	+2.03	overuse
	Hint	0	0	7	0.46%	-4.73	underuse
Interrogatives	Feasibility/Ability	34	5.61%	55	3.64%	+3.75	overuse
	Consultative Devices	10	1.65%	10	0.66%	+4.02	overuse
	Suggestory Formulae	1	0.17%	13	0.86%	-4.07	underuse
	Willingness	14	2.31%	15	0.99%	+4.96	overuse
* p<0.05，critical value=3.84							

From this, it can be inferred that non-native speakers significantly use "bold" imperatives more frequently than native speaker teachers when employing imperatives. In comparison to non-native teachers , native teachers, even when utilizing imperative directives, tend to place themselves on an equal status with students (e.g., by frequently using "Let's," an inclusive type), and less frequently employ "bold" imperatives with overtly commanding tones. However, when non-native teachers use imperative directives, they more frequently incorporate discourse markers such as "sorry" (Apologize to the person) and "please" (Politeness Markers), which can potentially create a sense of distance between teachers and students, in contrast to native teachers. Additionally, for interrogative directives, native teachers notably employ suggestory formulae.

(1) Non-native teachers exhibit a limited range of sentence patterns and relatively rigid expressions when using directives in English language teaching

classrooms, whereas native teachers demonstrate a diverse array of sentence structures and flexible modes of expression in their directives.

Native teachers demonstrate greater flexibility in using directives, which is evident in their tendency to provide hints to students (not just in one teacher's classroom but consistently across the classroom corpus of all four native teachers). Furthermore, native teachers more frequently utilize Suggestory Formulae-style directives compared to non-native teachers. (Details are listed in Table 8)

Hints:

Excerpt 13

<16><Female Teacher> You're playing a little game on your phone?*(HQ_1_20_E2_FBJ)* (Hint to students not to play with phones during class)

Excerpt 14

<75><Male Teacher> Oh, guys. Someone calls me.*(TT_2_21_E1_MJW)*(Hint that the class will be interrupted for a while)

Additionally, the transcribed corpus of native teachers contains directive sentences that are not commonly used

by non-native teachers , further showcasing the diversity and flexibility in their use of directives. When issuing directives, native teachers sometimes do not use "you" or the students as the subject, but instead use the first-person pronoun "I" to indicate their own actions, thereby guiding and motivating students to follow suit in their classroom instruction.

Excerpt 15

<41><Male Teacher> I'm gonna give you ten minutes, talk to your partner Task Two.*(TT_2_21_E1_MJW)*

Excerpt 16

<103><Female Teacher> So I want every single one of you to pick one person and raise your hand, and say their name, raise your hand. （*LD_1_19_E1_FJ)*

Table 9 shows the number of teachers (6 non-native teachers and 6 native teachers in total) who used various

directive phrases under different directive sentence types. Overall, it can be observed that native teachers have a higher coverage in terms of the number of individuals for each type of directive, indicating that native teachers generally possess greater flexibility in their use of directives.

Table 9 Coverage of the Non-native and Native English Speaking Teachers' Directives

Coverage	Non-native English Speaking Teacher	Native English Speaking Teacher
Bold	all 6	all 6
Politeness Markers	all 6	all 6
Mitigated	all 6	3
Inclusive	all 6	all 6
Apologize to the Person	2	4
Performatives/Hedged Performatives	5	5
Obligation	5	all 6
Want/Need	3	all 6
Permissive	all 6	all 6
Use of in-group Identity Marker	all 6	all 6
Hint	0	4
Feasibility/Ability	4	all 6
Consultative Devices	4	4
Suggestory Formulae	1	2
Willingness	4	5

(3) Compared to non-native teachers, native teachers employ directives for a broader range of purposes, extending to pre-class preparations, encouraging interactions, and other aspects, and their tone is considerably more tactful and euphemistic.

Among the non-native teacher corpus, only 10 sentences (1.65% of the total) served the function of discipline, while the remaining 596 sentences were directives used to organize classroom activities or teaching steps.

When not being used for the purpose of organizing classroom instruction (such as maintaining classroom discipline, arranging pre-class preparations, encouraging student interaction, or providing teaching feedback), the selected directives tended to be more direct and imperative in nature. Non-native teachers , when using classroom directives to provide feedback on students' issues, even employed directives with a punitive undertone (Excerpt 6).

Excerpt 6

<29><Female Teacher>[C]-我告诉你是十三，回去写两百遍，下节课给我交上来。-[C] *(LD_1_19_N1_FLYY)*

This indicates that in the process of providing feedback on students' answers, the teachers did not take into account the students' dignity at all, resulting in a rigid classroom hierarchy that deprives students of autonomy and fosters passive learning.

Among the organizational directives, native teachers' directives were not only used in the traditional sense of organizing classroom instruction but also for arranging pre-class preparations for students (Excerpt 10) and encouraging student interaction (Excerpt 11, Excerpt 12).

Excerpt 10

<1><Female Teacher> Can someone close the back door? *(LD_1_19_E1_FJ)*
Excerpt 11
<156><Male Teacher> Dead Silence. *(LD_1_19_E1_MAS)* (Hinting students to actively participate in the discussion)
Excerpt 12
<325><Female Teacher> Someone who hasn't spoken yet. （*LD_1_19_E1_FJ)*(Hinting at students who haven't answered questions to speak up.)

The use of "Can you do X?" in Excerpt 10, which falls under the category of Feasibility/Ability Interrogatives, and the hint declarative directives in Excerpt 11 and Excerpt 12, both employ a relatively mild tone and exhibit a high level of politeness. These data suggest that native teachers prioritize conveying respect and understanding through indirect language, viewing students as active participants, encouraging them to express themselves, and guiding the learning process, thereby stimulating students' learning potential and creativity.

Teacher-student Relationships in Chinese University ESL Classrooms with Non-native & Native English Teachers

In non-native teachers' classrooms, there is a more teacher-centered approach, with a highly asymmetrical power relationship between teachers and students. The authoritative status of teachers is evident, and students passively follow instructions, leading to an unequal status between teachers and students. In contrast, native teachers' classrooms are more student-centered, with a relatively symmetrical power relationship between teachers and students. The authoritative status of teachers is not prominent, and students are active participants in the classroom, resulting in a more equal status between teachers and students.

When exploring the usage patterns of directives in interpersonal interactions, a notable characteristic lies in the hierarchical variations in politeness and strategic choices, which profoundly reflect the social status and power dynamics between the communicating parties (Brown & Levinson, 1987). Specifically, the directness of directives is often positively correlated with the speaker's power status: requesters with higher power status tend to adopt more direct forms of verbal expression, whereas those with lower status lean towards more indirect or polite communication styles. As a typical authority figure, a teacher's choice of directives directly influences the nature and effectiveness of classroom interactions.

In other words, when teachers opt for more indirect forms of directives in the classroom, it essentially embodies a delegation of power, aimed at diminishing the traditional sense of authority associated with the teacher's role and thereby promoting students as the center of the learning process. The shift from the traditional teacher-centered, direct instructional strategy of "teacher teaches, student listens" to teachers' preference for more indirect directives signifies a notable increase in opportunities for student participation in classroom activities.

Non-native teachers tend to view themselves as the primary subject in the classroom, and due to their responsibility for completing teaching tasks, they frequently use imperative sentences with incomplete sentence structures, exhibiting a direct and explicit teaching style. In contrast, native teachers perceive both teachers

and students as the transmitters and receivers of knowledge, as well as the designers and executors of classroom content. They utilize more declarative directives with complete sentence structures, emphasizing equality and communication with students.

Compared to native teachers, non-native teachers more frequently employ discourse markers such as "sorry" (Apologize to the Person) and "please" (Politeness Markers) when using imperative directives, which relatively distances teachers from students. Additionally, native teachers significantly use suggestive inquiry methods for interrogative sentences. It can be inferred that non-native teachers maintain a more traditional teaching and learning relationship with students, while native teachers have a closer rapport with their students.

However, non-native teachers do not use "implicative" directives that convey negotiation, indicating they do not employ strategies that require students to comprehend instructions on their own or leave room for discussion. This suggests that non-native teachers , to some extent, overlook the importance of students' understanding in the classroom and that the English language classroom is still teacher-centered, with students in a subordinate position. Teachers have yet to create a learning environment where "teachers are the guides and students are the main participants," failing to provide equal interactive communication spaces for students. Furthermore, non-native teachers lack diverse English expression abilities and cannot express their requests or demands politely and flexibly. On the other hand, teachers have not developed a strong consciousness to research how to issue directives more appropriately. When instructing or disciplining students, they unconsciously position themselves as authorities in the classroom, rarely negotiating with students or genuinely seeking their opinions, leaving little room for choice. In other words, this patterned power hierarchy leaves students with few opportunities for negotiation and forces them to passively obey commands.

This expression may be related to the emphasis on authority in Chinese culture and the authoritative role of teachers. Direct and explicit instructions help maintain classroom order and improve teaching efficiency. Non-native teachers tend to use limited directive sentence patterns and rigid, uniform instruction expression methods. This may be associated with the emphasis on uniformity and standardization in the domestic education system. The primary purpose of using directives is to organize classroom teaching, and outside of this context, they exhibit a strong sense of compulsion. This reflects non-native teachers' strict attitude towards maintaining classroom discipline and improving teaching efficiency.

Native teachers, on the other hand, prefer to use indirect, polite, and moderately imperative permissive declarative directives. This expression reflects an emphasis on student autonomy and respect, contributing to a more open and interactive learning environment. Native teachers employ higher levels of politeness in their language, providing students with more space for thought and response. Their more indirect directives are significantly less direct and compulsory than those of non-native teachers, intentionally reducing the directness and compulsion of instructions to enhance equal communication with students.

Moreover, native English teachers are adept at using diverse directive sentence

patterns and flexible expression methods. This diversity helps attract students' attention, stimulate their interest in learning, and promote active participation in the classroom. Beyond organizing classroom instruction, native teachers also use directives to arrange pre-class preparations, encourage student interaction, and more. For instance, they employ implicative directive strategies that fully consider students' dignity, minimizing the power imbalance between teachers and students, fostering a more equal and student-centered classroom environment. In native teachers' English classrooms, directives embody an equal relationship between teachers and students. Teachers are no longer authoritative figures but rather friends and partners who learn and discuss together with students. This relationship promotes a democratic classroom atmosphere, enhancing students' learning motivation and creativity. When issuing directives, native teachers genuinely seek students' opinions, leaving them with more choices than non-native teachers . This relatively equal teacher-student role allows students to negotiate rather than merely obey commands.

Simultaneously, their relatively milder tone contributes to a more relaxed and enjoyable learning atmosphere. Furthermore, native teachers seldom use directives to maintain classroom discipline, preferring instead to cultivate students' self-management skills through guidance and encouragement. Their mild directives are not limited to organizing classroom instruction but also encompass arranging pre-class preparations, encouraging student interaction, providing feedback and encouragement, and rarely disciplining students. These data indicate that native teachers view students as active participants, encouraging them to express themselves and guide the learning process, rather than merely being passive recipients and obedient subjects. In the classroom, the roles of teachers and students have shifted, establishing a partnership based on equal communication and breaking the traditional power hierarchy.

In summary, directives in native teachers' English classrooms exhibit diversity, politeness, and negotiation, collectively creating an equal, democratic, and positive classroom atmosphere. Compared to non-native teachers' classrooms, native teachers place greater emphasis on conveying respect, understanding, and encouragement through language, thereby stimulating students' learning potential and creativity.

Conclusion

Non-native English Teachers' ESL Classroom in Chinese University has characters of teacher-centered and teachers' knowledge transmission, also with a hierarchical relationship with students. While native English Teachers' ESL Classroom in Chinese University has characters of student-centered and teacher-student' knowledge co-construction, with an equal relationship with students. What's more, both groups prioritize teaching organization but differ in syntactic types, communication strategy/purpose and politeness degree.

Regarding the directives employed, it is evident that there are distinct preferences and patterns between the two groups. Non-native English speaking teachers, predominantly non-native teachers, favor bold imperative directives characterized by directness, low politeness, and a highly forceful tone. In contrast, native English speaking teachers, or native teachers, prefer permissive declarative

directives that exhibit indirectness, high politeness, and a moderately forceful tone. This disparity underscores the cultural and linguistic nuances that influence instructional language choices.

The analysis highlights both similarities and differences in directives usage between the two teacher groups. Both groups use interrogative directives infrequently, and both prioritize organizing classroom teaching as the primary communicative purpose of directives. However, notable differences emerge in terms of politeness levels, sentence pattern diversity, and the range of communicative purposes. Non-native teachers exhibit a more limited range of sentence types and rigid expression methods, often employing directives with lower politeness levels. In contrast, native teachers demonstrate a greater variety of sentence structures and flexible expression methods, using directives for a broader range of purposes, including pre-class preparations and encouraging student interactions, with a more tactful and euphemistic tone.

These differences in directive usage reflect the distinct teacher-student relationships prevalent in the two types of classrooms. In non-native teachers ' classrooms, there is a more teacher-centered approach, with a highly asymmetrical power relationship between teachers and students. The authoritative status of teachers is evident, and students passively follow instructions, leading to an unequal status between teachers and students. When instructing or disciplining students, non-native teachers unconsciously position themselves as authorities in the classroom, rarely negotiating with students or genuinely seeking their opinions, leaving little room for choice. In native teachers' ESL classrooms, students' critical thinking over knowledge and practical application of English can be respected, student are not passively obeying commands. It can be inferred that non-native teachers maintain a more traditional teaching and learning relationship with students, while native teachers have a closer rapport with their students.

Implication

This study examines the directives employed by non-native and native English-speaking teachers in university ESL classrooms in Chinese mainland, explores the similarities and differences in their use of directives, and analyses the teacher-student relationships formed by these two groups, aiming to provide insights into effective instructional practices and intercultural communication dynamics in ESL contexts.

The distinct use of directives underscores the importance of cultural and social backgrounds in shaping instructional styles. Direct and low-politeness expression may be related to the emphasis on authority in Chinese culture and the authoritative role of teachers. Non-native teachers ' preference for bold imperative directives reflects a traditional authoritarian teaching culture, emphasizing direct control and organization. Non-native teachers ' more limited directives, though effective in maintaining discipline, may limit students' opportunities for independent problem-solving and self-expression. In the context of Chinese culture and society, respecting teachers has been of paramount importance since ancient times. Throughout history, the hierarchical difference between teachers and students has been pronounced, with a

prevalent norm of teachers issuing commands and students adhering to them obediently. This dynamic reflects a deep-rooted cultural tradition that emphasizes the authority and guidance of educators, as well as the respectful deference shown by students.

However, in Western countries, greater emphasis is placed on individual equality and independence, which naturally leads to a more egalitarian relationship between teachers and students. Native English-speaking teachers' use of permissive declarative directives aligns with a more democratic and interactive approach, fostering a student-centered learning environment.The diversity and flexibility of directives among native English-speaking teachers highlight the benefits of a rich linguistic repertoire in the classroom. Their ability to employ various sentence types and implicit hint strategies encourages creativity and critical thinking among students, enhancing learning outcomes. Native teachers' broader communicative purposes reflect a holistic teaching approach, balancing authority and collaboration in teacher-student relationships. This suggests that cultural awareness and adaptation in teaching methodologies are crucial for effective cross-cultural communication and student engagement.

References

[1] Bax, I.P. 1986. How to Assign Work in an Office: A Comparison of Spoken and Written Directives in American English [J]. *Journal of Pragmatics*, 10 (6), 673-692.

[2] Brown, P. & Levinson, S.C. 1987. *Politeness: Some Universals in Language Usage* [M]. Cambridge: Cambridge University Press.

[3] Curl, T.S. & Drew, P. 2008. Contingency and Action: A Comparison of Two Forms of Requesting [J]. *Research on Language and Social Interaction*, 41(2), 129–153.

[4] Ervin-Tripp, S. 1976. "Is Sybil There?" The Structure of Some American English Directives [J]. *Language in Society*, 5, 25-66.

[5] Haryanto, H. & Mubarok, H. 2018. Teacher's Directive Expressions Analysis in English Teaching Classes [J]. Lensa: Kajian Kebahasaan, Kesusastraan, dan Budaya 8(1), 22-42.

[6] House, J., & Kasper, G. 1981. Politeness markers in English and German [A]. In: F. Coulmas(ed). Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech [C]. New York: Mouton de Gruyter.

[7] Liu, Y.B. & Hong, H.Q. 2009. Regulative Discourse in Singapore Primary English Classrooms: Teachers' Choices of Directives [J]. *Language and Education*, 23:1, 1-13.

[8] Searle, J.R. 1976. A Classification of Illocutionary Acts. *Language in Society*, 5, 1-24.

[9] Tan, F. 2012. *An Exploratory Study of Bakhtin's Dialogic Theory in Foreign Language Classrooms* [D]. PhD. Dissertation. Jilin: Changchun, P.R.C. Northeast Normal University.

[10] Weigel, M.M. & Weigel, R.M. 1985. Directive Use in a Migrant Agricultural

Community: A Test of Ervin-Tripp's Hypotheses. Language in Society 14(1), 63-79.

[11] Yanti, D. R., Amin, M., & Amrullah, A. (2021). The Analysis of Speech Acts Used by EFL Teachers' in Classroom Interaction at SMAN 2 Mataram in Academic Year 2021/2022 [J]. *International Journal of Multicultural and Multireligious Understanding*, 8(5), 90-98.

[12] Yongbing Liu & Huaqing Hong. 2009. Regulative Discourse in Singapore Primary English Classrooms: Teachers' Choices of Directives [J]. *Language and Education* (1), 1-13.

Bio-data

TAN Fang
School of Foreign Languages, Dalian University of Technology
E-mail: tanfang0330@126.com
TAN Fang, PhD. Associate Professor, MA Supervisor in School of Foreign Languages, Dalian University of Technology. Research Interests: Applied Linguistics, Corpus Linguistics.

ZHANG Ce
School of General Education, Dalian University of Technology
Email: 2749216014@qq.com
ZHANG Ce, 3rd Year Undergraduate Student. Major: Business English.

基于教材课文的数字化多模态语篇教学活动设计

谭倩怡 华南师范大学

【摘要】新媒体时代的信息呈现方式已从单一模态转变为多模态形式。在这一背景下，《普通高中英语课程标准（2017 年版 2020 年修订）》认为培养学生"看"的能力变得尤为重要。本文介绍了基于教材课文开发的数字化多模态视频语篇的开发理念，并在此基础上设计了多种教学活动，旨在综合提升学生的语言能力。本文主要介绍六种活动：听写填空、纠错视听、关键词理解、视听完形填空、合作听写和关键词复述。

【关键词】多模态语篇；看；课文；视频

随着信息技术的发展，人们获取信息的方式和渠道更加多元，信息的呈现也渐渐融合了文字、音频、图像多种形式。在这一背景之下，《普通高中英语课程标准（2017 年版 2020 年修订）》（以下简称"新课标"）随即在四大传统语言技能的基础上增加了"看"（viewing）的能力，即利用多模态语篇中的图形、表格、动画、符号以及视频等理解意义的技能（教育部，2020，35）。为了充分利用课文资源，更好地提升学生"看"的能力，笔者的团队已为人教版和北师大版高中英语课文开发了配套多模态视频资源。这些资源不仅仅是作为课内材料的简单补充，教师也可充分利用这一系列资源设计多种英语学习活动，使得学生能在"做"中发展"看"的能力。故本文将详细介绍一些基于多模态英语课文视频资源设计的教学活动案例，以期为广大一线教师的教学提供参考。

一、开发理念

尽管"看"已经成为一项基本的语言技能，但目前各版本高中教材配套的视频资源仅包含拓展性资源，如人教版的 Video Time，北师大版的 Viewing Workshop 等，而主语篇尚未配套视频资源。为了让一线教师在教授主语篇时能够更好地融入"听、说、读、写、看"的活动，本研究团队为常用版本的高中教材开发了配套的主语篇视频资源。

在介绍多模态视频资源的应用之前，本文先对"多模态"及"看"这两个概念进行进一步的解析。首先，"多模态"是相对于"单模态"而言的。单一的书面文字、口语或不含文字、语音的图像均属于"单模态"，而如果在语言的基础上融合图像等其他信息呈现形式，则属于"多模态"（陈力，2022）。在多模态语篇中，各种模态的组合不是简单的堆砌，而要能够相互

补充、增强不同模态信息的含义（Kress & Van Leeuwen, 2001，20）。

"看"指的是理解多模态语篇的技能。有学者提出"看"的微技能体系包括预测/预览、关联、对比、推断、想象、具象、导航、把握重点、浏览、扫描、自我提问、总结/改述、综合、自我监控等内容（刘晓斌，2023）。在目前在英语科的考查内容中，"看"并未被单列为一项关键能力，而是被融合在"听"与"读"之中（陈康等，2019）。

在理论层面上看，新课标对"多模态"和"看"这两个概念的关注建立在 Paivio 的双重编码理论以及 Mayer 的多模态学习认知理论的基础之上。Paivio 认为，人脑分别通过非言语通道和言语通道对信息进行加工（Paivio，1990，54）。前者主要处理图片、视频等信息，而后者负责语言信息的处理。基于 Paivio 的理论，Mayer 认为当人们使用多模态学习材料时，图片、图表、录像等画面进入非言语通道，文字、语音等文字信息进入言语通道（Mayer，1997）。由此，多模态学习资源为学生提供视听多渠道的话语意义输入，有利于学生对学习内容获得更深的理解，并将其转化为长期记忆，进行有意义的学习（郭颖，2023）。但值得注意的是，若模态的组合不得当，可能会增加学生的认知负荷，干扰学生的注意。有研究发现，相比转瞬即逝的动画视频，静态关键帧视频能够降低学生的认知负荷，显著提升学生对视频内容的理解（Chi et al., 2023）。

基于此，笔者的团队将能够表达课文意义的静图组合成视频，并配上课文音频、字幕及音乐，制作了一套教材配套的多模态课文视频资源。本系列资源分为有字幕版和无字幕版，教师可以根据教学需要对字幕文件做挖空、注释或高亮等。研究表明，不同的字幕呈现方式对二语学习有不同效果。有完整字幕的视频有利于促进学生对视频内容的整体理解（Teng，2019），而无字幕视频能够使学生更关注音频输入，避免被文字干扰，从而能提升其"听"的能力（Hsieh，2020）。此外，带有关键词注释、增显的视频能够增强学习者对重点词汇的关注，促进学生的词汇学习（Teng，2023）。

总得来说，这套资源通过多种模态的组合增强了课文文本的意义，有利于促进学生对深层语篇意义的理解和对语言之美的感知。教师可以通过修改字幕文件，灵活组织教学活动综合锻炼学生的语言技能。接下来，本文以北师大版选择性必修一 Unit 1 Lesson 1 Teachers 中的 My Teacher, My Student 为例，具体说明利用该系列资源开展教学活动的六种方法。

二、多模态课文视频资源的应用方法

本文要介绍的具体应用方法有六种，词汇层面的有听写填空（Spot Dictation）和纠错视听（Error Correction），语篇理解层面的有关键词理解（Keyword Comprehension）、视听完形填空（Audio-visual Cloze）、合作听写（Dictogloss）以及关键词复述（Retelling with Keywords）。接下来，本文将根据 Richard 和 Rodgers 提出的教学法三维度来进行介绍，即理论(approach)，教学设计(design)和实施步骤(procedure)（Richard & Rodger，2000）。

1. 词汇层面

（1）听写填空（Spot Dictation）

Mayer 的多模态效应指出，在读文本或听录音的同时观看相应画面的学生能够比单纯接受文字输入的学生产出更多解决问题的想法（Mayer & Moreno,1998）。这与他提出的多模态学习认知理论是相一致的。两个理论共同说明，多种模态信息的组合使得学生能够更高效地解读信息并产出。

基于此，本活动一方面旨在考察学生对单元重点词汇的语音知识和拼写知识的掌握情况，另一方面旨在培养学生"看"的能力，即引导他们"理解多模态语篇中的文字信息与非文字信息在建构意义过程中的作用"（教育部，2020，37）。在实施本活动前，教师可根据单元的教学重难点及学生平时的表现情况灵活确定目标词，教师应尽量选择有同步出现能解释其意义画面的词汇进行挖空，以更好地利用多模态效应。

课前准备阶段，教师需使用电脑自带的"记事本"或"写字板"打开资源中的字幕文件，对其进行修改并保存（图1）。随后，教师使用 QQ 影音打开无字幕的视频，将修改后的字幕文件拖放到视频处，视频中便会显示相应字幕。但由于 QQ 影音的默认字幕字体较小，教师可在界面右下角找到"字幕样式"的功能，修改字体大小。在读后环节，教师可让学生边看视频边完成听写填空。在多模态听写活动中，学生在掌握语篇语言点的基础上，需要理解图片是如何解释同步的文字信息的，在完成填空后，学生又可以根据画面检查自己答案是否正确。在本案例中，教师对本单元的重点词之一"rocket"进行了挖空处理（图2），视频同步出现呈现了火箭的图片，能够帮助学生理解听到的单词，加深学生对该词汇的记忆。

图 1 字幕文件修改
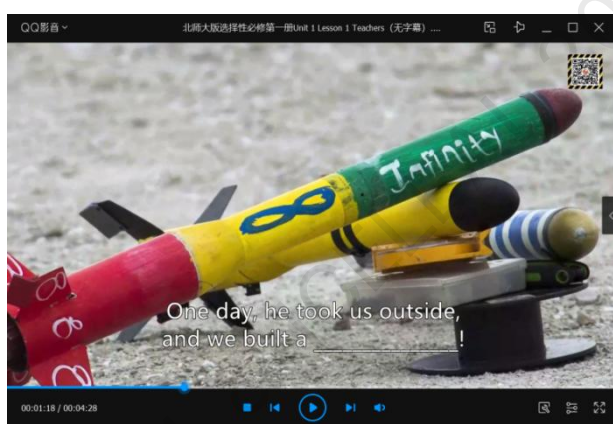
图 2 教师对"rocket"进行挖空处理

（2）纠错视听（Error Correction）

错误分析理论指出,错误是二语学习中必然出现的,是二语学习者习得语言的一个窗口。不断分析和改正错误能够有效减少错误的产生（Mitchell et al.，2019，35）。除了能有效促进学生对语言规则的掌握，语篇纠错活动也能锻炼他们的逻辑推理和综合辨析意义的能力（曾绛，2016）。

总得来说，该活动设计要求学生在理解多模态材料的过程中出现干扰时，要学会修复，解决这一问题，重新建立理解。学生不仅要运用其"看"和"听"的理解性技能去识别多模态语篇中的错误，还要运用其表达性技能，即"根据语篇的意义选择正确的词汇和语法结构"（教育部，2020，37）。

在开展这一活动前，教师结合本节课的语法点修改字幕，使其带有语法或语义错误，让学生边看视频边记录下字幕中的错误并修改。本单元的主语篇 my teacher 中的"I"原本是一个不爱学习的孩子，在遇到 Mr Jenkins 后他发现了求知的乐趣，最终成为了一名科学作家。在纠错视听活动中，教师将原文句子"I suppose I was a bit lazy"中的"lazy"改为了"diligent"，这一方面能够考察学生对"I"这个人物形象的理解，另一方面，由于视频中与这一句子同步

出现的画面是一个懒于学习的男孩，与改后的文字信息涵义相悖，故也可考察学生看图片理解言语信息的能力。此外，为考察学生对过去完成时的掌握情况，教师将视频原句"I knew I had chosen a job with a lot of stress but I love what I do"改为现在完成时形式"I knew I have chosen a job with a lot of stress but I love what I do"。

2. 语篇层面

（1）关键词理解（Keyword Comprehension）

关键词指的是能够帮助学生预测、理解文章内容的线索词，对理解语篇至关重要（Filatova, 2016）。根据双重编码理论，线索词和对应图像的输入能够分别激活言语系统中的言语码和非言语系统中的意象码。人脑中语义表征和意象表征这两种知识的表征方式相互独立又相互联系，互为补充。关键词视频提供了帮助学生理解语篇的线索词的画面，给学习者以语义和表象的双重表征，当两个编码系统同时被激活时，学习者能更好地激活已有的图式，建立已有知识和新信息之间的联系，形成对材料内容的初步认识和理解（杜华、孙艳超，2012）。

基于以上理论，该活动一方面旨在锻炼学生"预览/预览"的看的微技能（刘晓斌，2023），另一方面也可使学生在语境中加深对重点词汇意义的理解，这和新课标对词汇知识学习的要求相一致（教育部，2020，21）。

因此，教师在读前让学生观看带有关键词字幕的视频，引导学生根据画面和关键词预测课文内容，并提出以下问题："Who wrote the texts?"," What's the relationship between Mr Jenkins and Graham?", "What are the texts mainly about? "。学生回答完以上关于主旨大意的问题后，正式进入主语篇的阅读。

（2）视听完形填空（Audio-visual Cloze）

建构主义学习理论指出，当学生主动思考并解决问题，成为主动的学习者，有意义的学习就发生了（Mitchell et al., 2019：224）。同时，学习过程中，学生的已有知识会和新知识相互作用，形成新的理解。"视听完形填空"要求学生在视听的过程中，主动梳理归纳文章大意，与改述后的新文本互动，构建自己的理解。同时，改述文本提供了与原文不同的地道语言表达，可供学生欣赏、比较和学习。

总得来说，鉴于完形填空文本形式上的破碎性，学生在理解音频文本后，还需要进一步根据语境线索推理文意和填补细节信息，在此过程中，学生的信息搜索能力和信息整理能力都能够得到提高（徐剑等，2022）。视听完形填空在纯文本完形填空的基础上，给学生提供

了非文字的语境线索，是一种较新颖而有趣的语言练习活动。

活动前，教师先利用 Quillbot 网站将课文文本进行改述（paraphrase）（图 3）。随后，教师可参考 Quillbot 提供的同义词对自动生成的文本进行难度改编。最后，教师可对关键词或短语挖空。课堂上，学生先观看并充分理解无字幕视频，随后完成完形填空练习。

I've read a couple of Graham's books and seen him on TV. I always say to my wife,"Ohlook, I used to teach him!"I remember Graham was very difficult before he came intomy class. I had heard stories about his bad behaviour. Once I caught him and his friendsseeing who could jump the farthest off the school stage! But when he got interested, hechanged. The first day he walked into my class, he was dragging his schoolbag behindhim and looking bored, but as soon as i set up an experiment to show how the humanstomach works using acid and an onion, he gave me his full attention, He loved science!He was very bright and he had done very well in science subjects.

121 Words

Rephrase

Graham is a writer whose works I've read and watched on television. My wife and I often remark, "Ohlook, I used to teach him!"Before he entered my class, Graham was, in my memory, an extremely challenging student. Stories concerning his unruly demeanour had been told to me. I once saw him and his friends competing to leap from the school stage the farthest. But he changed as he became interested. He appeared bored the first day he entered my class and was dragging his suitcase behind him, but as soon as I put up an experiment to demonstrate how the human stomach functions using acid and an onion, he completely focused. He liked science!He was quite intelligent and excelled in scientific classes.

1/8 Sentences • 122 Words

图 3 Quillbot 改述语篇

（3）合作听写（Dictogloss）

互动假说指出，意义协商是二语学习中重要的一环。研究表明，"合作听写"通过给学生提供协商意义的情境，能够有效丰富学生的二语词汇，提升语法能力，提高二语写作中的体裁意识（Chun & Aubrey, 2021; Nabei,1996）。此外，合作语言学习理论指出这种学习模式可以利用小组成员之间的差异性，高效促进语言学习及批判性思维能力的发展。

基于以上理论，教师应尽可能建立异质性小组，让组员之间能够相互帮助，利用各自优势，通过协商共同构建出意义完整的新语篇。过程中，教师应引导学生将注意力放在语篇内容和语言形式上，提醒他们复述的内容应包含原文的主要信息，但其语言形式应避免与原文重合。

这一活动包含四个步骤：准备-视听-重构-分析纠正。在准备阶段，教师可引导学生就视听的话题进行小组讨论，以确保学生掌握听写材料中出现的重点词汇，也有利于激活学生已有的知识。视听阶段，教师可播放三次录音：第一次播放时，学生不记笔记，仅掌握语篇的大意；第二次播放时，学生可选择性地记录可辅助复述的关键词；第三次播放时，学生检查笔记内容并做适当补充。在重构阶段，小组成员分别用自己的话复述自己负责的部分，并由一位成员负责记录重构的语篇（图 4）。在最后分析纠正的阶段，小组成员应仔细检查复述语篇，以确保意义完整和语言准确。教师也可以让学生对照原文修改自己的语篇，并做总结反思。
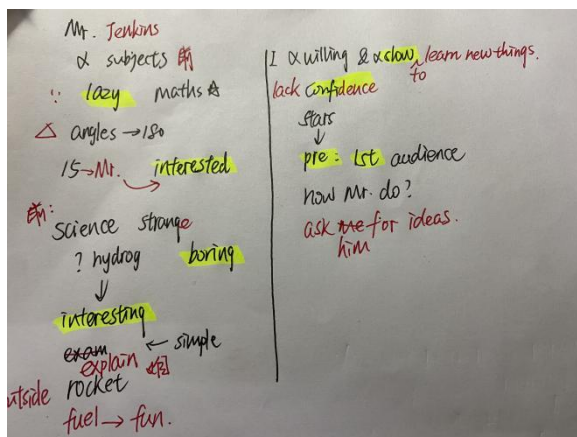
图 4 小组重构语篇的笔记

（4）关键词复述（Retelling with Keywords）

深度学习加工理论指出，被深度加工的知识更容易转化为长期记忆，在脑海中形成深刻理解。复述，是学生在深入阅读和完全理解文本后，将文本中的语言吸收、储存，并内化、整理，然后以自己的语言表达文本内容，对语篇再创作的过程（李婉莹，2021）。

新课标要求学生"能够利用所学语言知识、文化知识等，通过口头和书面等形式创造新语篇"（教育部，2020，36）。根据这一要求，教师在读后环节可利用多模态课文资源开展"关键词复述"活动。复述活动中，学生是积极的学习者，主动对文本的内容及语言进行深度学习。教师借助关键词视频，引导学生把握语篇主脉络，给学生提供脚手架，帮助学生回忆所学的语言和文化知识（雷霞辉，2020）。同时，视频画面借助视觉效果帮助学生进一步构建语篇内涵，增强了任务的趣味和真实性。

本案例中，教师在课后开展了"Watch and dub"的教学活动，视频主要选取了本单元的重点词"hydrogen"，"rocket"，"pour into"等作为关键词。为了给学生提供更多的时间进行复述，降低任务难度，教师将视频播放速率调整至 0.75 倍速。在活动开始前，教师先让学生浏览课文，并预测视频中可能会出现哪些辅助复述的关键词。活动开始时，教师先让学生观看一次视频，并给同学适当的准备时间，再邀请同学上台给视频以配音的形式复述。

三、结语

在新媒体时代下，多模态成为信息的主要呈现方式。在这一背景下成长起来的学生从小就以基于新媒体的"读图"方式来理解信息，"看"已然成为了他们获取信息的主要方式之一。因此，在教授语篇时，教师可以为学生提供丰富且多样化的多模态资源，让他们能够自然地体验到语言的意义和美感。本文基于本研究团队开发的多模态课文视听资源，详细介绍了如

何在日常语篇教学中融入"看"的活动，综合培养学生的语言技能，希望能为一线教师提供教学参考。

参考文献

Chi, C., Chen, H. J. H., Tseng, W. T., & Liu, Y. T. Efficacy of different presentation modes for L2 video comprehension: Full versus partial display of verbal and nonverbal input [J]. ReCALL, 2023, 35(1): 105-121.

Chun, L. T.&Aubrey, S. Using a modified dictogloss to improve English as a second language Llearners' use of genre-appropriate conventions and style[J]. RELC Journal, 2021, 0(0)

Filatova O. More than a word cloud[J]. TESOL Journal, 2016, 7(2): 438-448.

Hsieh, Y. Effects of video captioning on EFL vocabulary learning and listening comprehension[J]. Computer Assisted Language Learning, 2020, 33(5-6):567-589.

Kress, G.&Van Leeuwen, T. . Multimodal Discourse: The Modes and Media of Contemporary Communication[M]. London: Edward Arnold.2001

Mayer, R. E. Multimedia learning: Are we asking the right questions[J]. Educational Psychologist, 1997(32): 1-19

Mayer, R. E., & Moreno, R. A cognitive theory of multimedia learning: Implications for design principles[J]. Journal of educational psychology, 1998, 91(2), 358-368.

Mitchell, R., Myles, F., & Marsden, E. Second Language Learning Theories[M]. Routledge.2019

Nabei, T. Dictogloss: Is it an effective language learning task?[J]. Working Papers in Educational Linguistics, 1996, 12(1), 59-74.

Paivio, A. Mental Representations: A Dual Coding Approach[M]. Oxford University Press. 1990

Richard,J.C.&Rodgers,T.S.Approaches and Methods in Language Teaching [M].Foreign Language Teaching and Research Press, 2000.

Teng, F. Maximizing the potential of captions for primary school ESL students' comprehension of English-language videos[J]. Computer Assisted Language Learning, 2019, 32(7): 665–691.

Teng, M. F. Effectiveness of captioned videos for incidental vocabulary learning and retention: The role of working memory[J]. Computer Assisted Language Learning, 2023: 1-29.

陈力.多模态语篇在英语课程教学中的意义[J].小学教学设计,2022(33):1.

陈康、吴泓霖、李新煜等.基于高考评价体系的英语科考试内容改革实施路径[J].中国考

试,2019(12):33-37.

杜华、孙艳超.文字云图在英语阅读教学中的应用研究[J].中国电化教育,2012(08):115-119.

郭颖.多媒体和网络资源在中学英语单元教学中的应用[J].中小学外语教学(中学篇),2023,46(04):15-20.

刘晓斌.英语多模态语篇教学中"看"的微技能培养[J].中小学数字化教学,2023(09):5-10.

李婉莹.基于主题语境的高中英语单元词汇处理[J].中小学外语教学(中学篇),2021,44(12):1-6.

雷霞辉. 2020. 英语阅读教学中如何运用复述促进学生语言和思维能力的提升——以 Memories of Christmas 为例［J］. 英语学习，（12）：35-38.

普通高中英语课程标准[M]. 中华人民共和国教育部.人民教育出版社.2020

徐剑、唐书哲、袁辉.高考评价体系下英语完形填空的测试价值研究[J].天津师范大学学报(基础教育版),2022,23(04):46-51.

曾绎.完形、改错的语言测试与专门用途英语语能考测[J].外语电化教学,2016(03):86-90.

语料库驱动的英语近义中度副词语义韵对比研究

北京邮电大学　　王芮鑫

摘要：程度副词近义词辨析一直是词汇教学与习得中的重难点。英语母语者大规模通用语料库能够展现近义词在母语中的真实使用情况，促进二语学习者理解，提高辨析能力。本研究采用数据驱动的语料库研究方法展开近义词辨析，为英语学习者提供除查阅词典外的新的词汇习得途径。本研究基于美国当代英语语料库（COCA）中的大量真实语料，对比研究英语程度副词中近义的中度副词 *fairly，rather，relatively*。通过对比其在不同语域的词频分布，观察搭配差异，揭示语义韵特征，促进词汇习得。研究发现，在本族语者实际使用中，目标近义词项在语域分布、搭配、语义韵上存在显著差异。结果表明，在母语者实际语言运用中，*fairly* 具有显著的中性语义韵，*rather* 具有消极语义韵，而 *relatively* 具有积极语义韵，三者不可混用。本研究对比分析了三个近义中度副词的语义韵，为英语学习者提供了新的近义词学习方法，同时证明了这一方法较高的教学应用价值。

关键词：语料库；近义词辨析；程度副词；语义韵

Exploring Semantic Prosody of Synonyms in English:

A Corpus-driven Analysis of Moderate Adverbs of Degree

Abstract: Discrimination of synonyms among degree adverbs has always been challenging in vocabulary teaching and acquisition. A reliable large-scale common corpus of native English speakers can show the real usage of synonyms in the native language, promote the understanding of second language learners, and improve their discrimination ability. This study aims to facilitate synonym identification and vocabulary acquisition based on the approach to the corpus. Utilizing a substantial amount of authentic data from the Corpus of Contemporary American English （COCA）, this study is conducted on the synonyms of moderate degree adverbs in English，namely **fairly，rather and relatively**. By contrasting their frequency distributions in different linguistic domains，observing collocational differences，and revealing semantic nuances，I found that the target synonymous terms exhibit significant differences in domain distribution，collocation，and semantic nuances. Results tell that in the actual use of native speakers，the target near-synonymous items have more obvious differences in domain distribution，collocation，and semantic prosody，with **fairly having a significant neutral prosody, rather having a negative prosody, and relatively having a positive prosody**, and the three cannot be used interchangeably. This study compares and analyzes the semantic prosody of three near-meaning intermediate adverbs，provides a new method for English learners to learn synonyms，and proves that this method is of high value in teaching applications.

Key words: corpus studies; synonym discrimination; adverbs of degree; semantic prosody

1. 引言

近义词辨析是外语教学和词汇习得中的重难点之一。在现代近义词教学或词典编纂中，解释近义词语义常采用近义词"循环释义法"（Tognini-Bonelli 2001）。这一释义法基于词汇的语义共性，其不足是弱化和模糊了个体词汇独有的搭配特性和语义韵潜势（余江陵，马武林 2022）。

语料库驱动的词汇搭配和语义韵研究可以帮助学习者深入理解近义词的使用情况，在语境中理解近义词语义特征，提高近义词辨析能力。弗斯（J. R. Firth）于 20 世纪 50 年代发表了著名论断："You shall know a word by the company it keeps！"（Firth 1957/1968）。搭配是词语的结伴或共现，反映了词与词之间的相互吸引、相互预见和相互期待（卫乃兴 2003）。词项的搭配行为显示了一定的

语义趋向，整个语境随节点词（node word）的显著搭配而弥漫某种语义氛围，即语义韵（Semantic Prosody）（Sinclair 1991）。语义韵可以表示共现词汇产生的情感色彩和评价态度。Stubbs（1996）把词语搭配形成的语义韵分为消极语义（negative prosody）、中性语义韵（neutral prosody）和积极语义韵（positive prosody）三类。卫乃兴（2002）提出了语义韵研究一般采用的三种方法：建立类联接，基于数据概括和描述关键词的语义韵（data-based approach）；计算搭配词，用数据驱动方法研究语义韵（data-driven approach）；数据与数据驱动相结合，建立语义韵结构。

近年来语料库语言学的发展为通过词汇搭配行为描写近义词的语义韵差异提供了有效的量化研究途径，帮助非母语人士准确理解近义词的细微差异，避免搭配误用，提高学习效果。在语言习得中，英语中度副词数量多、语义差异小，为学习者带来诸多困扰。鉴于此，本研究通过语料库中大量的真实语料和语言事实，以英文中度副词 fairly，rather，relatively 为例，对比其在不同语域中的词频分布，观察近义词名词和形容词搭配差异，揭示语义韵特征，帮助学习者提高程度副词近义词的辨析能力，有助于优化词典的近义词释义方式。

2. 文献综述

近年来，随着语料库语言学的兴起，基于数据驱动的近义词对比研究近年来研究方兴未艾。笔者以"近义词"和"语料库"为主题词，在中国知网上检索到 20 篇北大核心和 CSSCI 刊源文献，对其进行初步分析得出以下发现：

研究方法和研究目的角度上，基于语料库的近义词对比实证研究主要可以分为两类：一类（吴欣达 2003；侯丽 2014；李倩 2018）通过对比一组近义词在本族语语料库中的语域分布、节点词搭配和语义韵等方面的差异，帮助学习者理解和使用近义词；另一类（任培红 2008；陆军 2010；余江陵、马武林 2022）通过对比一组近义词在本族语语料库和学习者语料库中的扩展意义单位等差异，分析语言学习者的近义词少用、多用和误用等现象和产生原因，帮助教师进行近义词教学。

研究对象上，近年来，国内研究聚焦于英汉两种语言，也不乏对于日语的研究（李倩 2018）。从研究工具来看，英国当代语料库（BNC）、中国英语学习者语料库（CLEC）、国家语委现代汉语语料库（CMLC）为主流语料库。研究案例聚焦于近义形容词的对比研究，且跨语言视角下的程度副词语义韵对比研究几乎未被北大核心和 CSSCI 期刊收录，基于语料库的近义副词语义韵对比研究大有可为。

3. 研究设计

3.1 研究问题

在 *Longman Dictionary of Contemporary English for Advanced Learners* （*the 6th edition*）词典中，

fairly 的释义为"more than a little，but much less than very"，rather 的释义为"fairly or to some degree"，

relatively 的释义为"something that is relatively small，easy etc. is fairly small，easy etc. compared to other

things"，是典型的近义词循环释义，同时，作为程度副词，它们也为英语学习者带来了很多困扰。

因此，本研究将 fairly，rather，relatively 这一组表示中等程度的近义副词作为本文的研究对象，主

要研究以下三组问题：

（1）fairly，rather，relatively 在不同语域的词频分布上有何特征？对词义辨析有何启示？

（2）fairly，rather，relatively 在高频搭配词上有何特征？对词义辨析有何启示？

（3）fairly，rather，relatively 在语义韵上有何特征？对词义辨析有何启示？

3.2 研究方法和步骤

本文采用 Stubbs（1996）提出的语义韵三分法，即积极语义韵、消极语义韵和中性语义韵。根

据节点此吸引的搭配词的语义色彩判断该节点词所处的搭配和语义韵所属的语义韵类别。对语义韵

的测量参照卫乃兴（2011）提出的语义韵强度计算方法，即计算该节点词表达积极、中性和消极的

搭配词占搭配词总频数的百分比，并通过内部比较来确定语义韵极性。

本研究英文语料来源于美国杨伯翰大学 Mark Davies 教授开发的美国当代英语语料库（Corpus of

Contemporary American English，COCA）。该语料库库容量超过十亿，涵盖美国 1990 年至 2019 年间

博客、网络、影视、口语、小说、杂志、报纸、学术期刊共八大类型语料，是当今世界上最大的英

语平衡语料库。通过对 COCA 中的英文语料进行在线检索，提取索引行和搭配信息，以定量分析和

定性解释相结合的方式，对比分析该组中度副词近义词的语义韵特征。

具体研究步骤如下：

首先，确定节点词。根据 Klein（1999）对英语程度副词的分类法，本研究选择了同一层级相对

应的英汉中度副词 fairly，rather，relatively 为研究对象，通过对比其在本族语者语料库中的显著搭

配词，分析英汉中度副词呈现的语义韵特征。

其次，提取目标词的语义韵特征。具体步骤为：（1） 统计近义词在不同语域中的频数和频率分

布。COCA 中的语料可以按照博客、网络、影视、口语、小说、流行杂志、报纸和学术期刊八大类

型语域分类。本研究使用语料库在线检索和人工分类得出不同领域的每百万词频分布。（2） 提取目

标词高频搭配。在英语语料库中，搭配强度指标 MI 值与共现频率为确定搭配词的指标。MI 值即相

互信息值（Mutual Information Value），是计算词语搭配强度的一种常见方法，MI≥3 具有明显搭配强

度。Jones & Sinclair（1974）发现，在±4 的跨距（span）内，搭配词的分布与语法结构紧密相关，

超出±4 的距离后，节点词已不再有太大的吸引力。因此，本研究通过 COCA 在线检索工具，以 FAIRLY，

RATHER，RELATIVELY 为节点词，跨距为±4，相互信息值 MI≥3，最小共现频率为 50 的标准，分别找出每组词词左四和右四跨度内搭配的动词和形容词及其 MI 值。（3）提取索引行，标注语义韵。由于建立语义韵至少需要 100 行索引（卫乃兴 2002），本研究从 COCA 中随机提取 100 条与节点词高频搭配的动词与形容词的 KWIC 索引行（每隔两行），确保删除重复行或不合乎义项要求的索引行。根据扩展语义单位测量标注近义词在 KWIC 索引行扩展语义单位中的语义韵极性特征。

最后，根据统计数据，对比分析 fairly，rather，relatively 在语域分布、搭配和语义韵的特征，并探讨语义韵对近义词辨析、程度副词习得等方面的启示。

4. 数据统计与讨论

4.1 语域分布情况

语域是人们在实际的语言活动中，出于交际的需要，或因其所从事的职业和兴趣相异，亦或因其话语发生的情景、说话的对象、地点和话题的不同而产生的一种言语变体，体现为语言中的不同语体风格、用语格调等，研究同义词在不同语域中的分布频率对同义词的辨析具有宏观指导作用（张继东 2005）。"频数"和"频率"是语料库语言学的重要概念，"频数"指在语料库中某一词汇或短语出现的绝对次数，"频率"指语料库中某一词汇或短语在每百万词中出现的频数，是被标准化的值，可用于比较。通过检索 COCA 语料库，得出 fairly，rather，relatively 在八个语域中的总出现频数和每百万词频率分布表，如表 1，和每百万频率对比柱形图，如图 1。

表 1   COCA 中目标节点词不同语域词频分布表

节点词	FAIRLY		RATHER		RELATIVELY	
分布语域	频数（词）	频率（每百万词）	频数（词）	频率（每百万词）	频数（词）	频率（每百万词）
博客	7259	56.44	37674	292.92	6474	50.34
网络	6002	48.3	35454	285.34	7014	56.45
影视	1344	10.49	10907	85.16	532	4.15
口语	6202	49.17	17704	140.36	3473	27.53
小说	2718	22.97	17360	146.72	1394	11.78
杂志	5103	40.47	26790	212.47	9365	74.27
报纸	3562	29.26	18228	149.73	6269	51.49
学术期刊	4366	36.45	45937	383.48	17544	146.46
总库	36556	36.81	210054	211.53	52065	52.43
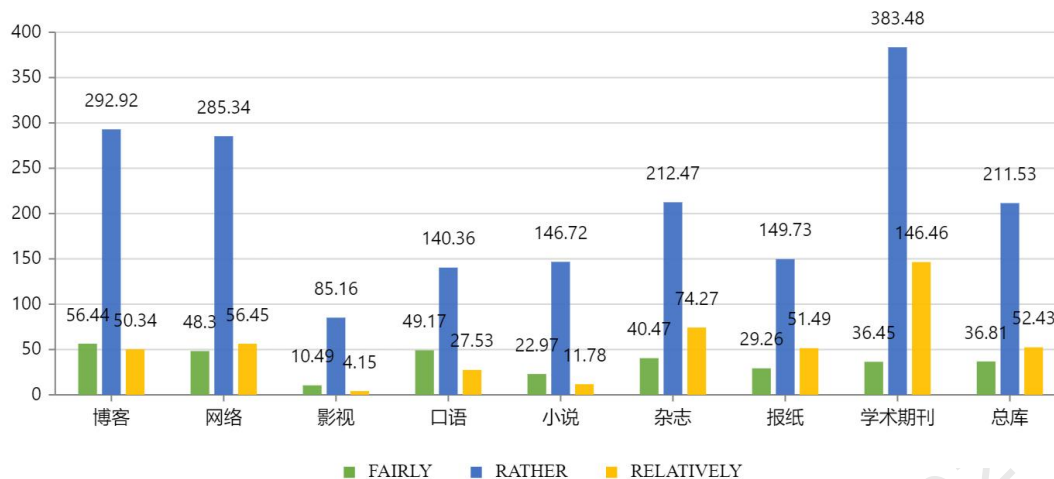
图 1 FAIRLY，RATHER，RELATIVELY 每百万词频语域分布柱形图

在 COCA 总库中，RATHER 的使用频率明显高于 FAIRLY 和 RELATIVELY，几乎为 FAIRLY 的 6 倍、RELATIVELY 的 4 倍，RELATIVELY 的使用频率介于 RATHER 和 FAIRLY 之间。这说明英语母语者在实际使用中，更常用 RATHER，较少用 FAIRLY 表示中量程度。

观察三词的不同语域分布图可知，它们的分布因语义相似而有一定的共性，但也呈现出个性特征。共性如，三词均在博客和网络语域中频繁出现，而在影视作品中出现频次均最低。个性上，FAIRLY 分布前三的语域是博客（56.44 words per mil）、口语（49.17 words per mil）和网络（48.30 words per mil），RELATIVELY 分布前三的语域则是学术期刊（146.46 words per mil）、杂志（74.27 words per mil）和网络（56.45 words per mil），而 RATHER 分布前三的语域是学术期刊（383.48 words per mil）、博客（292.92 words per mil）和网络（285.34 words per mil）。

由此可得，在书面语及正式语体中（除小说），母语者更常用 RELATIVELY 和 RATHER 表示中量程度，在小说中，FAIRLY 和 RELATIVELY 几乎全被 RATHER 取代，而在口语和非正式语体中，更常用 FAIRLY。

4.2 与动词和形容词的搭配特征

计算搭配强度的主要算法有互信息值（MI 值）、Z 值、T 值、Log-log 值、卡方值等。这些方法各有优劣，本文采用 COCA 语料库自带的 MI 值来测量搭配强度。MI 值越大，说明节点词对其词汇环境的影响越大。笔者以 FAIRLY，RATHER，RELATIVELY 为节点词，跨距为（-4，+4），相互信息值 MI≥3，最小共现频率为 50 的标准，分别对它们的左四右四的搭配动词和形容词进行检索，分别找出左四和右四跨度内，两个词各自的搭配名词和形容词及其共现频次和 MI 值，分别制表如下。

表2 COCA 中 FAIRLY 搭配的 MI≥3 的所有动词和部分形容词频数及 MI 值

序号	+ Verb	频数	MI 值	序号	+ ADJ	频数	MI 值
1	compensate	97	5	1	straightforward	337	7.14
2	treat	921	4.95	2	routine	61	6.41
3	price	78	4.59	3	predictable	124	5.56
4	distribute	130	4.2	4	inexpensive	69	5.08
5	value	55	3.5	5	uniform	72	4.98
				6	decent	201	4.91
				7	confident	215	4.8
				8	stable	199	4.78
				9	obvious	418	4.65
				10	accurate	226	4.63

需要说明的是，表 2 呈现的 FAIRLY 的搭配形容词是 MI≥3 的前十条数据。在检索中，与 FAIRLY 搭配的形容词（MI≥3）还有：typical，consistent，constant，sophisticated，modest，robust，simple，common，steady，certain，standard，reliable，well-known，mild，extensive，substantial，easy，representative，narrow，widespread，rare，conservative，rapid，reasonable，intelligent，comprehensive，regular，flat，detailed，minor （根据 MI 值大小排序）。

表3 COCA 中 RATHER 搭配的 MI≥3 的所有形容词频数及 MI 值

序号	+ Verb	频数	MI 值	序号	+ ADJ	频数	MI 值
1	bland	66	4.49	12	pointless	54	3.54
2	simplistic	58	4.24	13	fond	85	3.51
3	amusing	89	4.2	14	bizarre	122	3.47
4	then	68	4.14	15	vague	114	3.45
5	mundane	61	3.98	16	obscure	58	3.42
6	startling	58	3.88	17	peculiar	67	3.23
7	unpleasant	88	3.75	18	odd	256	3.22
8	dull	121	3.73	19	disappointing	63	3.19
9	lengthy	92	3.63	20	ironic	80	3.18
10	dubious	58	3.62	21	straightforward	68	3.18
11	bleak	53	3.58	22	unusual	258	3.05

表 3 穷尽了 RATHER 的搭配形容词，而没有搭配的动词数据，这是因为 COCA 语料库中，与 RATHER 搭配的动词其 MI 值均<3，未能达到搭配词的判断标准，研究价值较小，固本研究尊重语言数据，不考虑 RATHER 的动词搭配。

表4 COCA 中 RATHER 搭配的 MI≥3 的所有形容词频数及 MI 值

序号	+ Verb	频数	MI 值	序号	+ ADJ	频数	MI 值
1	obscure	62	4.35	1	inexpensive	668	7.85
2	remain	1054	3.7	2	unscathed	139	7.85

3	compare	526	3.37		3	painless	129	7.54
4	calm	91	3.29		4	unimportant	123	7.06
5	isolate	60	3.21		5	homogeneous	107	6.83
					6	straightforward	328	6.59
					7	undisturbed	58	6.49
					8	benign	181	6.44
					9	stable	840	6.35
					10	inexperienced	90	6.22

同样，表 4 呈现的 RELATIVELY 的搭配形容词是 MI≥3 的前十条数据。在检索中，与 RELATIVELY 搭配的形容词（MI≥3）还有：modest，unchanged，unaffected，tame，affluent，minor，harmless，low-cost，sparse，insignificant，short-lived，mild，prosperous，obscure，untouched，powerless，trivial，uncommon，scarce，inefficient，unknown，rare，autonomous，small，flat，low，shallow，short，peaceful，cheap，simple，weak，abundant，intact，uniform，easy，constant，narrow，brief，robust，affordable，moderate，quiet，static，minimal，recent，smooth，calm，limited，slow，safe，wealthy，expensive，rapid，steady，high，sophisticated，light，consistent，fixed，healthy（根据 MI 值大小排序）。

综合表 2 至表 4 的数据，根据搭配词语义，本研究将前十个搭配名词和形容词粗略分为语义趋同和语义差异两类（见表 5）。

表 5　FAIRLY，RATHER，RELATIVELY 的高频搭配词

节点词	FAIRLY		RATHER		RELATIVELY	
	+ Verb	+ADJ	+ Verb	+ADJ	+ Verb	+ADJ
语义趋同的搭配词	distribute	straightforward routine stable inexpensive		dull bland simplistic mundane	isolate	straightforward stable inexpensive
语义差异的搭配词	compensate treat price value	predictable uniform decent confident obvious accurate		amusing then startling unpleasant lengthy dubious	obscure remain compare calm	unscathed painless unimportant homogeneous undisturbed benign inexperienced

如表 5 所示，fairly，rather，relatively 的语义共核是"稍微，比较，有点"的中度程度之义，但其也各自吸引了语义特征差异较大的共现词。就动词而言，三者吸引的近义动词较少，差异较大：FAIRLY 的显著搭配动词如"定价、估值"，语义上较为客观，而 RELATIVELY 的搭配动词则有如 isolate，calm 等情感态度词，包含一定的主观色彩。就形容词来说，三者均常与 straightforward，stable，

dull，inexpensive 搭配，而不同的是，FAIRLY 的显著搭配形容词多表示事物的属性和特征，如 obvious，accurate，uniform，情感倾向弱；RATHER 的显著搭配形容词中积极情感词和消极情感词比例相当，还需要放入 KWIC 索引行中进一步甄别；RELATIVELY 的显著搭配形容词情感倾向较明显，积极语义较强。

搭配名词的典型语义类型差异可以表明近义词可能的语义倾向，从而更好地区分近义词（余江陵、马武林 2022）。因此，本文以搭配形容词为例，根据细分语义将其概括为以下三类：（1）表示性质和状态的形容词，如 uniform，obvious，stable，unscathed，homogeneous 等。（2）表示方式的形容词，如 straightforward，painless，simplistic 等。（3）表示情感的形容词，如 decent，amusing，startling，modest，unimportant，benign。

4.3 语义韵特征

根据 Stubbs（1996）对语义韵的分类和卫乃兴（2011）提出的语义韵测量方法，笔者结合上文归纳高频搭配词，提取词语在 COCA 中的 KWIC 索引行对节点词 FAIRLY，RATHER，RELATIVELY 的搭配词进行描述和分析。因显著搭配动词数量少，研究意义不大，因此这里聚焦于三个近义词的显著搭配形容词语义韵。COCA 中，节点词 FAIRLY，RATHER，RELATIVELY 的各类含义词在搭配词中所占百分比例如表 6 所示。

表 6　FAIRLY，RATHER，RELATIVELY 三类含义搭配形容词占比

	积极（%）	中性（%）	消极（%）
FAIRLY	30.00	62.50	7.50
RATHER	22.73	13.64	63.64
RELATIVELY	52.11	21.13	26.76

测量结果表明，中度程度副词 fairly，rather，relatively 中，FAIRLY 具有显著的中性语义韵，RATHER 具有消极语义韵，而 RELATIVELY 具有积极语义韵。结合 KWIC 索引行分析如下：

就节点词 FAIRLY 而言，排名前五的高频搭配形容词是 straightforward，routine，predictable，inexpensive，uniform，前四个词的 MI 值均在 5 以上，出现频数超过每百万词 60 次，其中 straightforward 的 MI 值更是高达 7.14，出现频数达每百万词 337 次。统计可知，FAIRLY 所在的索引行通常是中性语义氛围，呈现错综语义韵或是中性语义韵。如下索引行：

picking season . Within the next hour they will have a | fairly | accurate idea of how good -- or bad -- the yield really
. These revenue sources , of course , often serve as | fairly | accurate predictors of the decisions that a candidate might
sooner the better . # " I guess that is | fairly | accurate to the situation with this play , " says Murray .
will ensure complaints against the department will be fully and | fairly | acted upon if necessary . " He goes on to say :
shows an analogue of himself doing anything . His work is | fairly | adamantly not confessional (which separates it from the
of the citizenry in gory detail , whereas they have been | fairly | adequate at reporting mounting deficits and the great
or Wilkins elementary buildings , both of which are located in | fairly | affluent , predominantly white communities and have been
after Ray Allen in the offseason (and they did , | fairly | aggressively recruiting him over the phone) but they switched
and just killed him . # " I thought that was | fairly | amazing , that we all made it through the war we all
long-term , we have to distribute our remaining food equally and | fairly | among all survivors . Surely you agree with that , Abby ?
IN many places in the world , minorities are not treated | fairly | and can be destroyed by the callousness of authorities . # It
, written policies , employers must apply their policies | fairly | and consistently to all employees . Therefore , it is crucial
Second , they will want you to deal with her | fairly | and forthrightly . Do n't play games by telling her she 's
environment , in which all parties trust they are being treated | fairly | and that their opinions matter . # " I 'm a strong
leaves me cold , it 's the conclusion , which is | fairly | anticlimactic , as Dee crushes the ruby and inadvertently
time , I will work to ensure that the law is | fairly | applied and that prosecutorial discretion is not abused . Given
to explain to their children that we may be involved in | fairly | arcane legal discussions of what is sex and what is n't .
begins with discomfort about present convention , but it may be | fairly | argued that stripping away the delusions about the

就节点词 RATHER 而言，排名前五的高频搭配形容词是 bland，simplistic，amusing，then，mundane，其 MI 值均高于 3.9。除 amusing 外，其他四个词都具有消极语义，"单调、无聊"的含义更是出现了三次，而 RATHER 修饰 amusing 时，会削弱 amusing 的情感浓度，相对而言 amusing 本身也沾染上消极语义。如下索引行：

by all Apollo children , " I said matter-of-factly in a | rather | cold tone . I only replied to Jason to the hopes
they had drained much of Havrad 's personality , leaving him | rather | cold , haggard , and blank . Archer gave him credit for
. But then , they were not brother and sister , | rather | companions who had been raised in the same home , and that
felt the problem was not getting out of Vietnam , but | rather | conducting the war forcefully enough to win . # Last April we
imminent bankruptcy . I can could imagine it makes libertarians | rather | conflicted if I were at all thoughtful , or even had two
I could somehow explained him what BI is , he got | rather | confused as to what he could do to become a successful BI
shots as evidence , many of them showing DC Parker in | rather | conservative heroic poses . Neither Shad nor Parker told
on weekends. 63 The bicycle race was not discontinued , but | rather | converted into an event that used less of a precious resource .
recently , I picked Cheap and safe , living in a | rather | crap-hole'ish post-industrial steel mill area that was on the
expect good things from his life , and his devoted if | rather | critical parents have inflated his sense of self-worth . That ,
282) . Just because the statement is meant as a | rather | cruel boast does not mean it does not contain some truth .
anything in the first place . These days , I 'd | rather | deal with a torrent of Thingy comments than for a comment thread
; even the stupendously gifted Lizst refused to attempt it , | rather | defensively dismissing the piece as " a waste of time , an
that the first few equalities are n't obvious truths , but | rather | definitions of what we mean by " work " , " force
if you believed me then . Yeah . WALLACE : I | rather | did . The technicians , the camera people ... Mr-IRVING :
I WILL DECIDE WHAT I WANT . AMIGO , I WOULD | RATHER | DIE THAN TO GIVE YOU THAT BAG . (gunshot) ANY
this one is in a similar vein . Though it is | rather | different from Sharon 's book in which the main focus was on
make the critical attitude of the association to other Journals | rather | difficult to maintain ' . Another BMA witness , the
are plunged in grief , it will give no joy but | rather | displeasure . # Objection 5 . Further , as sense is to

就节点词 RELATIVELY 而言，排名前五的高频搭配形容词是 inexpensive，unscathed，painless，unimportant，homogeneous，其 MI 值均大于 6.8，出现频数均高于每百万词 100 词，inexpensive 更是高达每百万词出现 668 次，具有很强的显著性。这五个词中，前三个都有积极语义，一个消极语义，一个表示状态的中性语义。从如下索引行中可以看出 RELATIVELY 总体呈现出积极语义韵：

, if imperfect , introduction to virtual reality that 's | relatively | attainable at only $100 . But there 's an even better deal
will often mask their feelings enough for the departure to be | relatively | benign . # However , when an employee is " asked to
programs have been engaged in service delivery for a | relatively | brief period of time (11.8 years) , their volume of
humans . The Intelligence Principle has proven true over the | relatively | brief trajectory of our own species , despite notable past and
especially in 1997 , 1998 , and 1999 , over a | relatively | broad survey area in lower Cook Inlet (Fig. 3) .
come increasingly to resemble commercials , and not just in the | relatively | brutish form of " product placement " (a practice that has
. # As far as schedules go , I have a | relatively | busy one . And it 's been especially difficult because I have
. Yet when it works , it 's a noninvasive , | relatively | cheap clinical treatment that 's fairly safe , beyond potential

5. 总结与展望

本研究采用语料库驱动和对比研究方法，分析了中度程度副词 fairly，rather，relatively 这组近义词在语域分布、搭配关系和语义韵特征三个方面的差异，得出以下结论：

（1）通过观察近义词在 COCA 语料库不同语域中的分布频数和频率差异，证明了近义词确实在实际使用中遵循"语域优先"原则（杨春霞，2014）。从总频数和频率分布来看，在本族语者语料库中，rather 最为常用，relatively 次之，fairly 最不常用；其次，在书面语及正式语体中（除小说），母语者更常用 relatively 和 rather 表示中量程度，在小说中，rather 的使用频次远远高于 fairly 和 relatively，而在口语和非正式语体中，更常用 fairly。

（2）通过计算近义词与其搭配名词和形容词的 MI 值，可知近义词在实际运用中的搭配关系和搭配行为。通过语料库检索并计算搭配强度发现，COCA 中，fairly 常与表示性质、状态的形容词连用，与客观动词连用，rather 常与表示"单调、无聊"的形容词搭配使用，无强搭配动词，relatively 常与表示积极情感的形容词和动词连用，其本身也沾染较强的情感意义。

（3）通过对比近义词的语义倾向，可分析近义词在语境中呈现的积极、中性和消极语义韵特征。在计算语义韵强度和分析搭配行为的基础上，可以看出 relatively 明显表现为积极语义韵，fairly 的中性语义韵特征更突出，rather 则呈现出消极语义韵特征。

本研究的不足之处和改进之处如下。首先，在本研究用 MI 值作为提取搭配词的标准。但 MI 值的可信度在学界受到质疑（z 值的可信度更高）。其次，在语义韵测量方面，本研究抽取了 100 条索引行，数量较少，若抽取 500 条进行分析统计结论更有说服力。

总之，本研究采用语料库驱动的方法对中度副词 fairly，rather，relatively 的搭配行为和语义韵特征进行了对比研究，发现英语母语者在使用中度副词时的语言习惯和情感倾向。基于语料库的近义词辨析方法为词汇学习提供了新的视角，弥补了传统的词典释义只从概念意义或句法功能方面辨析近义词的局限。在语言学习过程中，语料库能为学习者提供丰富翔实的语料，帮助学习者在真实使用中理解和辨析近义词语义的微妙差异。鉴于此，使用语料库展开自主学习是未来语言学习，尤其是近义词辨析的大势所趋。

参考文献

[1]   Firth, J. R. 1957. *Papers in Linguistics* [M]. London: Oxford University Press.

[2]   Firth, J. R. 1957/1968. A synopsis of linguistic theory, 1950-55 [A]. In F. Palmer (ed.). *Selected Papers of J. R. Firth 1952-59* [C]. Bloomington: Indiana University Press: 168-205.

[3]   Jones, S. & J. Sinclair. 1974. *English lexical collocations: a study in computational linguistics* [J]. Cahiers de Lexicologie (2).

[4]   Sinclair, J. 1991. *Corpus, Concordance, Collocation* [M]. Oxford: Oxford University Press.

[5]   Stubbs, M. 1995. *Collocations and Semantic Profiles: On the Cause of the Trouble with Quantitative Methods* [J]. Function of Language (1): 1-33.

[6]   Stubbs, M. 1996. *Text and Corpus Linguistics* [M]. Oxford: Blackwell Publishers.

[7]   Henny K. 1999. *Adverbs of Degree in Dutch and Related Languages* [J]. The Journal of Comparative Germanic Linguistics 3: 37–44.

[8]   Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work* [M]. Amsterdam: John Benjamins.

[9]   Hunston, S. 2002. *Corpora in Applied Linguistics* [M]. Cambridge: Cambridge University Press.

[10]   Whitsitt, S. 2005. *A Critique of the Concept of Semantic Prosody* [J]. International Journal of Corpus Linguistics (3): 283-305.

[11]   Xiao & McEnery. 2006. *Collocation, Semantic Prosody, and Near Synonymy: A Cross-linguistic Perspective* [J]. Applied Linguistics (1): 103-129.

[12]   Hunston, S. 2007. *Semantic Prosody Revisited* [J]. International Journal of Corpus Linguistics (2): 249-268.

[13]   *Longman Dictionary of Contemporary English for Advanced Learners the 6th edition* [Z], 2014. London: Pearson Education ESL.

[14]   卫乃兴，2002，语义韵研究的一般方法[J]。外语教学与研究（4）：300-307。

[15]   杨惠中，2002，语科库语言学导论[M]。上海：上海外语教育出版社。

[16]   卫乃兴，2003，搭配研究 50 年：概念的演变与方法的发展[J]。解放军外国语学院学报 2003（02）：11-15。

[17]   蔺璜、郭姝慧，2003，程度副词的特点范围与分类[J]。山西大学学报（哲学社会科学版）2003（02）：71-74。

[18]   王海华、王同顺，2005，CAUSE 语义韵的对比研究[J]。现代外语 2005（03）：297-307+330。

[19]   张继东，刘萍，2005，基于语料库同义词辨析的一般方法[J]。解放军外国语学院学报（6）：49-52。

[20] 黄瑞红，2007，英语微量程度副词的极性敏感性[J]。外语教学与研究（04）：268-275+321。

[21] 任培红，2008，近义词的教与学：以 Common 和 Ordinary 为例[J]。解放军外国语学院学报（04）：57-60。

[22] 汪兴富、M. Davies，刘国辉，2008，美国当代英语语料库（COCA）——英语教学与研究的良好平台[J]。外语电化教学（05）：27-33。

[23] 王春艳，2009，基于语料库的中国学习者英语近义词区分探讨[J]。外语与外语教学 （06）：27-31。

[24] 马拯，2010，基于语料库的 HAPPEN 语义韵的对比研究[J]。外语教学理论与实践（4）：20-27。

[25] 陆军，2010，基于语料库的学习者英语近义词搭配行为与语义韵研究[J]。现代外语（03）：276-286+329-330。

[26] 卫乃兴，2011，基于语料库的对比短语学研究[J]。外国语（4）：32-42。

[27] 李晓红、卫乃兴，2012，双语视角下词语内涵义与语义韵探究[J]。现代外语（1）：30-38。

[28] 杨梅，2013，双语视角下近义词语义韵对比研究[J]。广西民族大学学报（哲学社会科学版）（03）：181-184。

[29] 侯丽，2014，基于 BNC 语料库的近义词用法对比研究——以"rather"和"quite"为例[J]。山西师大学报（社会科学版）（S4）：149-150。

[30] 杨春霞，2014，基于语料库的近义词辨析研究——以 suspect 和 doubt 为例[J]。西南科技大学学报（哲学社会科学版）31（05）：45-49+86。

[31] 余渭深、李中正，2017，基于语料库的中英文最高程度副词的扩展意义研究[J]。外语教学（05）：32-37。

[32] 濮建忠，2020，扩展意义单位模型再解读[J]。外语研究（2）：1-8。

[33] 余江陵、马武林，2022，汉英同情义近义名词语义韵对比研究[J]。外国语文（1）：91-99。

教育信息化视域下人工智能与高校英语教学的融合发展

向阳红 青岛恒星科技学院

【摘要】教育信息化逐渐成为了高等教育的重要改革方向。而人工智能作为一种前沿的"通用目的技术"，引发了教育等社会各个领域的变革。对于高校英语教学来说，将人工智能引入到教学中来，既符合教育信息化背景下的改革要求，也符合新时代高校大学生学习的现实需求。但是，在教育信息化视域下人工智能与高校英语教学的融合还存在一定的难点问题，如教学理念滞后、教学模式单一、教学评价有待革新、师资素养急需提升等等，这些问题的存在也影响了高校英语教学的智能化改革。对此，本论文立足教育信息化的大背景，对教育信息化视域下人工智能与高校英语教学的融合价值、现状问题和有效对策等展开了全面性、实践性研究，以期探索有效的改革路径，推进人工智能与高校英语教学的深度融合，让人工智能赋能高校英语教学新发展，全面提升英语教学质量。

【关键词】教育信息化，人工智能赋能，高校英语教学，融合发展

前言：

教育信息化与人工智能的结合是新时代教育发展的趋势，但如何在理论层面阐释这种结合的合理性与必要性，仍需深入探讨。通过分析人工智能在高校英语教学中的应用，能够进一步完善和丰富现有的教育教学理论体系，为教育信息化时代的英语教学提供理论支撑。其次，有助于推动跨学科理论的融合。人工智能属于计算机科学领域，而英语教学则属于语言学和教育学的范畴。通过探讨二者的融合发展，能够促进计算机科学、语言学和教育学等多个学科的交叉融合，产生新的理论生长点，为相关学科的发展提供新的思路和方法。最后，还有助于提升教育技术的理论研究水平。人工智能作为一种新兴的教育技术，其在高校英语教学中的应用尚处于探索阶段。通过深入剖析人工智能在高校英语教学中的实际应用效果，能够为教育技术的理论研究提供实证支持，推动教育技术理论的创新与发展。

教育信息化视域下人工智能与高校英语教学的融合发展，对于提高高校英语教学质量具有重要的实践价值。通过引入人工智能技术，英语教学可以更加精准地满足学生的个性化需求，提高学生的学习兴趣和学习效率。论文中提出的融合

策略和实践路径，能够为高校英语教师提供有益的参考，帮助他们更好地运用人工智能技术改进教学方法，从而提升教学质量。其次，有助于推动高校英语教学的创新改革。在传统的教学模式下，英语教学往往侧重于知识的传授而忽视了学生的实际应用能力。而人工智能技术的应用为英语教学提供了更多的创新空间。通过探索人工智能与英语教学的融合模式，能够激发教师的教学创新意识，推动高校英语教学向更加多元化、个性化的方向发展。最后，本研究的实践价值还体现在对人工智能技术教育应用的推广上。通过展示人工智能在高校英语教学中的成功应用案例，能够增强教育界对人工智能技术的认可和接受度，推动人工智能技术在教育领域的更广泛应用。

一、概念论述

（一）教育信息化

教育信息化是指在教育领域全面深入地运用现代信息技术来促进教育改革与发展的过程。它不仅仅是技术的引入，更是一种教育理念和教学模式的革新。教育信息化旨在通过信息技术手段，提高教学效果，扩大教育资源的共享，促进教育公平，培养适应信息化社会需求的新型人才。

（二）人工智能

人工智能是一种模拟、延伸和扩展人类智能的理论、方法和技术，它结合数学、计算机科学、心理学等多学科的理论，通过计算机程序来模拟人类的思维过程和智能行为。在本文中，人工智能特指应用于教育领域，尤其是英语教学中的各种智能技术和系统，如智能语音识别、自然语言处理、机器学习等。

（三）高校英语教学

高校英语教学是指在高等院校中进行的英语语言教学活动，旨在培养学生的英语听说读写能力，提高其跨文化交际能力。在本研究中，高校英语教学特指在信息化和人工智能技术支持下的新型英语教学模式和实践，强调技术与教学的融合与创新，探索如何将现代信息技术和人工智能有机地融入到高校英语教学中，以改变传统的教学模式，提升英语教学的质量和效率，培养出适应未来社会需求的高素质英语人才。

二、国内外研究现状

首先，在国外方面，美国学者 Marina Dodigovic 针对英语的非母语学习者，

开发了"学术英语智能教学系统"（Intelligent Tutor of Academic English），借助人工智能技术帮助非母语学习者纠正语言学习中的偏误，并且 Marina Dodigovic 教授实践教学中验证了该系统的有效性。Attali、Li 等人的研究中智能写作系统 Criterion 显著改善了学生的写作质量，并且 Attali 的研究发现学生作文修改的次数与提高的分数成正比例关系，蒋艳指出智能作文批改系统可以帮助学生养成反复修改的习惯，但是不能过分依靠智能写作评分系统，仍然需要有教师指导。其次，在国内方面，谢维兴（2022）分析了"智慧教育"提升学生学习适应性的途径，包括正确诊断分析，"把脉"学生的学习适应性；基于"智慧课堂"的精准教学，增强学生学习适应能力；"线上+线下"相结合的学习策略指导课程，提升学生学习适应性等。张秋菊（2024）指出，人工智能技术可以运用于高校大学生英语学习辅导、教学测评等不同环节，能够提高大学英语教育和学习的效率，为学生带来更具情境的学习体验，为学生实现个性化学习奠定坚实基础。万杰（2024）分析了人工智能时代大学英语多模态教学路径，包括熟悉多模态融合教学平台工具，加强跨文化学习、借助现代技术，提高学生学习参与度等。

总的来说，国内外的相关研究成果也给本论文的研究带来了一定的理论和经验支持，但同时也发现，当前的研究多以理论分析为主，主观性问题比较明显，而且关于教育信息化视域下人工智能与高校英语教学的融合价值、现状问题和有效对策方面的实践性研究较少，这也是本论文研究的一个出发点和创新点。

三、研究方法

1. 文献法

通过系统地搜集、整理和分析国内外关于教育信息化、人工智能在英语教学中的应用等相关文献资料，旨在明确当前的研究现状和发展趋势，找出研究的空白点和不足之处，为本研究的开展提供理论基础和依据。同时，还可以了解到国内外在人工智能与英语教学融合方面的成功案例和先进经验，为本研究提供有益的参考。

2. 问卷法

为了解当前教育信息化视域下人工智能与高校英语教学的融合现状，设计了一份详细的研究问卷，针对青岛恒星科技学院的大学生进行大规模的调查，以期

全面了解人工智能在高校英语教学中的应用情况，以及存在的问题和挑战。

3.访谈法

在问卷调查的基础上，还将选取部分高校英语教师进行深度访谈。通过面对面的交流，深入了解他们对人工智能在英语教学中应用的看法和体验，以及他们在实际教学过程中遇到的问题和困惑。访谈法有助于获取更真实、更深入的信息，获取第一手的实践经验和反馈。

4.行动研究法

结合实际情况，提出教育信息化视域下人工智能与高校英语教学的融合对策，并在实际教学中进行验证和实践。通过行动研究法，可以不断地总结与反思实践经验，及时调整和完善相关对策，以期探索出一条科学有效的教育教学改革之路。同时，行动研究法还有助于将理论与实践相结合，提升实用性和可操作性。

四、研究成果

（一）实践成果

在教学研究中，我们取得了显著的实践成果，这些成果涉及教学、学生和教师三个层面，共同构成了我们研究所取得的宝贵经验。

首先，在教学层面，通过不断的教学实践和探索，形成了一套富有创新性的教学方法。我们引入了互动式教学，通过小组讨论、案例分析、角色扮演等多种形式，激发学生的学习兴趣，提高他们的学习参与度。同时，结合多媒体技术，使教学内容更加生动有趣，增强了学生的学习体验。这些教学方法的改进，不仅提高了教学效果，也为我们进一步的教学改革提供了有力的支撑。

其次，在学生层面，教学实践成果主要体现在学生能力的提升上。通过新的教学方法，学生们的自主学习能力、批判性思维和创新能力都得到了显著提升。他们在课堂上的表现更加活跃，愿意主动提出问题、分析问题并寻求解决方案。此外，学生在团队协作和公开表达方面也展现出了明显的进步。这些变化不仅体现在学生的学业成绩上，更体现在他们综合素质的提升上。

再者，在教师层面，实践成果则表现为教师教学水平的提升和专业成长。通过参与教学研究，教师们不仅掌握了更多的教学技巧和方法，还学会了如何根据学生的实际情况灵活调整教学策略。同时，教师在与学生互动的过程中，也更深入地了解了学生的学习需求和特点，从而能够更有针对性地进行教学。这种教学

相长的过程，极大地促进了教师的专业成长，使他们成为了更加优秀的教育工作者。

（二）策略成果

1.完善教学理念，注重技术应用

在传统的高校英语教学中，教师往往侧重于语法和词汇的传授，而忽视了技术的运用和学生个性的发展。为了改变这一现状，提出了"以学生为中心，以技术为辅助"的教学理念。在这一理念的指导下，积极引入人工智能技术，如智能语音识别、自然语言处理等，来辅助英语教学。这些技术的应用不仅提高了学生的学习兴趣，还使教学更加生动有趣。

2.优化教学设计，加强教学效果

针对高校英语教学的特点，利用人工智能技术优化了教学设计。通过大数据分析学生的学习习惯和水平，为每个学生量身定制了合适的学习计划。同时，还利用 AI 技术设计了丰富多样的互动教学活动，如虚拟角色扮演、智能对话练习等，这些活动极大地提高了学生的参与度和学习效果。

3.构建教学平台，提高教学质量

为了充分利用人工智能技术，构建了一个集教学、练习、评估于一体的智能英语教学平台。这个平台不仅提供了丰富的教学资源，还能根据学生的学习进度和反馈进行智能调整。通过平台的智能推荐系统，学生能够找到最适合自己的学习资料和练习题目。同时，教师也可以通过平台实时监控学生的学习情况，及时调整教学策略。

4.完善教学评价，促进教学提升

在传统的教学评价体系中，往往只注重学生的笔试成绩。而利用人工智能技术，建立了一个多维度的教学评价体系。这个体系不仅考虑学生的笔试成绩，还包括学生的口语表达、团队合作能力、创新能力等多个方面。通过 AI 技术的辅助，能够更加客观、全面地评价学生的学习成果，从而促进教学的全面提升。

5.强化师资素养，保障教学效果

为了让人工智能技术更好地服务于英语教学，还加强了师资素养的提升。定期组织教师进行 AI 技术的培训和学习，使教师能够熟练掌握这些技术并运用到教学中。同时，还鼓励教师之间进行教学经验的交流和分享，共同提高教学效果。

五、存在的问题、建议

（一）存在的问题

首先，是技术应用不够深入。目前，虽然已经将人工智能技术引入英语教学，但其应用主要停留在表面层次，如简单的语音识别、智能推荐等。对于更深层次的语言理解、智能对话等高级功能的应用相对较少。其次，是教师技术接受度不一。部分教师对于新技术的接受和使用存在困难，导致人工智能技术在英语教学中的普及和应用受到限制。再者，是学生自主学习能力有待提高。虽然人工智能技术为学生提供了更多的自主学习资源，但部分学生仍然习惯于传统的被动学习方式，缺乏主动利用这些资源进行自我提升的意识。最后，是数据安全与隐私问题。随着人工智能技术的广泛应用，学生的个人信息和学习数据的安全性也面临着挑战。如何保障学生的数据安全和隐私，成为了一个亟待解决的问题。

（二）建议

首先，是深化技术应用。应该进一步探索人工智能技术在英语教学中的深层次应用，如利用自然语言处理技术进行更深度的语言分析和学习，或者开发更具智能性的对话机器人，以提供更贴近实际的教学辅助。其次，是加强教师培训。针对部分教师对新技术接受度不高的问题，定期组织相关培训。再者，是引导学生自主学习。可以通过设置一些激励机制，如学习积分、奖励等，来鼓励学生更多地利用人工智能技术进行自我学习和提升。同时，教师也应该在教学中引导学生培养自主学习的好习惯。最后，是加强数据安全保护。需要建立完善的数据保护机制，确保学生的个人信息和学习数据不被泄露。这包括加强数据加密、设置严格的访问权限等措施。

总的来说，虽然在人工智能与高校英语教学的融合中取得了一些成果，但仍然面临诸多挑战。对此，在未来我们将正视这些问题，并积极寻求解决方案，以确保人工智能技术能够更好地服务于高校英语教学，全面提升大学生的英语听说读写能力和综合素质。

参考文献

[1]Dodigovic M.Artificial Intelligence in Second Language Learning:Raising Error Awareness.[M].

[2]Attali, Y. Exploring feedback and revision features of Criterion [P].Presented at the annual meeting of theNational Council on Measurement in Education.San

Diego,2014,CA.April 12-16.

[3]Li,Z.,S.Link&V.Hegelheimer.Rethinking the role of automated writing evaluation feedback in ESL writinginstruction[J].Journal of Second Language Writing 2015,(27):1-18.

[4]谢维兴."智慧教育"提升高一学生学习适应性的途径探索[J].福建教育学院学报,2022,23(08):55-57.

[5]万杰.人工智能时代大学英语多模态教学研究[J].海外英语,2024,(04):141-143.

[6]王振环.数智时代背景下大学英语教师信息化教学能力及提升路径[J].公关世界,2024,(02):73-75.

[7]张晓宁,许晓艳.基于人工智能的大学英语混合教学实践与研究[J].海外英语,2024,(02):154-156.

[8]张秋菊.人工智能背景下大学英语教学变革探究[J].新课程研究,2024,(03):72-74.

数字化环境下的高中英语大单元项目式作业的设计与实施

轩雅莉 重庆市南坪中学校

摘要： 作业设计是高中英语教学评一体化设计中的重要一环，是落实新高考考查要求的重要途径。本文以新人教版高中英语选择性必修三第一单元为例，通过单元语篇整合和学情分析，确定单元整体教学目标和分课时系列作业内容。在此基础上，借助信息技术平台进行项目式作业设计、实施和评价，以促进学生英语学科核心素养的融合发展。

关键词： 高中英语；单元整体教学；项目式作业设计

引言

作业是教育教学的一个必不可少的环节，它是课堂教学的延伸和拓展，是落实立德树人、推进素质教育的重要载体（王月芬、周坤亮，2022）。但是，当前高中英语作业设计普遍存在内容难、烦、碎、多和训练方式机械重复等问题。这种缺乏大单元整体思维的作业设计，难以引导学生探究单元主题，也难以实现作业巩固知识、拓展能力、增进交流和建构意义的育人本质（刘辉、李德显，2021）。

《国务院办公厅关于新时代推进普通高中育人方式改革的指导意见（2019）》中指出要"提高作业设计质量，精心设计基础性作业，适当增加探究性、实践性、综合性作业"。在大单元教学背景下，项目式作业顺应课程改革与发展的需求，通过设置针对性和表现性较强的项目活动，整合和串联单元内各个课时的学习内容，帮助学生面对与现实生活情境相关的项目任务，促使学生在一定周期的探究和合作中构建相对完整的知识体系，完成项目成果，获得知识、能力、素养的协调发展 (桑国元等，2023)。

本文依托高中英语教材设计项目式作业，以小组合作的方式进行开放性探究活动，帮助学生逐步建构与单元主题相关的结构化知识，落实单元主题观念和语言观念，回归作业本质和满足学生个性化发展需求。

一、项目式作业与单元整体教学

1. 项目式作业

夏雪梅 (2021) 将学习素养视角下的项目化学习定义为"学生在一段时间内对学科或跨学科有关的驱动性问题进行深入、持续的探索，在调动所有知识、能力、品质等创造性地解决新问题、形成公开成果的过程中，形成对核心知识和学习历程的深刻理解，能够在新情境中进行迁移"。 项目化学习虽不能替代以系统讲解为主的传统教学方式，但它在传统学科知识和真实情境之间架构桥梁，指向知识的深度和广度(桑国元等，2023)。项目式作业是基于项目化学习衍生的一个概念，它指的是学生利用单元所学内容，通过开展复杂的、连续的和合作的成果展示活动来表达自己对主题意义的理解和评价，完成单元整合性输出活动的过程。

2.单元整体教学

单元是承载主题意义的基本单位，它承载着学科核心素养培养的具体而有所侧重的目标任务，是通过

英语课程学习促进学生成长的比较完整的课程周期（梅德明、王蔷，2018）。 新人教版高中英语教材以单元形式来呈现教学内容。虽然按单元语篇顺序展开也可以完成教学任务，但存在"课时割裂明显、教学碎片化、学用分离等常见问题"（李宝荣等，2023），缺乏对单元教学内容的有机整合。学生难以通过探究型任务建构深层次概念，教师也难以从单元角度开展持续性评价。而单元整体教学的设计与实施以核心素养为目标导向，以单元核心语篇为依托，从具体的知识点走向相对完整的学习单元。

3. 项目式作业与单元整体教学的联系

项目式作业不仅是单元整体教学中的一环，更是重要补充和推进。项目式作业可以作为评价学生单元学习效果的有效方式，贯穿整个单元教学始终（李宝荣等，2023）。首先，将评价前置能帮助教师转变教学设计的思路；其次，分课时系列作业使教师能开展持续性评价，切实从学生的角度出发，设计基于语篇又超越语篇的语言实践活动，引导学生分析问题、解决问题，为完成单元项目式作业奠定基础；最后，在完成最终的作业时，检测学生的学习效果和知识建构的程度。

二、高中英语大单元项目式作业实践
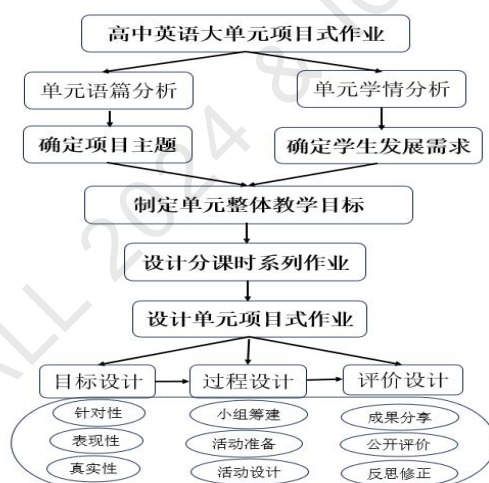
图 1 大单元项目式作业设计与实施路径图

教师要实施大单元项目式作业，就必须明确大单元项目式作业设计与实施的路径（见图1）。 首先，分析单元语篇，对教学内容进行系统整合，提炼单元主题大观念和语言大观念，初步确定单元项目式作业主题。然后，调查了解学生的主题知识和语言知识的已有基础和发展需求，以制定单元整体教学目标，确定分课时系列作业内容。最后，依据单元整体教学目标设计贯穿于单元的，由目标设计、过程设计和评价设计三个关键环节构成的项目式作业，开展教、学、评一体化的单元整体教学实践。

1.出发点：根据单元内容确定项目式作业主题和形式

（1）整体分析单元多模态语篇内容

确定项目式作业主题的前提是提炼单元主题意义。这就需要教师对单元所有语篇进行浏览性分析，整体了解单元主题语境以及各个语篇的主题意义，在此基础上梳理语篇内容之间的关联。以 Art 单元为例，这一单元围绕艺术这一主题展开，由八个主要语篇组成（见表 1）。该单元通过介绍西方绘画艺术发展简史，展示中国水墨动画的代表作品，提供艺术展览信息，逐步引导学生了解艺术的价值和意义，理解艺术家的创作动机，学会赏析艺术作品，体会艺术的魅力。

表 1 Unit1 Art 单元多模态语篇内容整体分析

语篇	语篇类型（课型）	语篇内容	主题意义
Reading and Thinking	说明文（阅读课）	西方绘画艺术发展史	感知西方绘画艺术的不同风格
Learning About Language	说明文（词汇课）	印象派代表画家莫奈的创作主题	认识印象派艺术的特点
Using Language 1:	演　讲（听说课）	以我国水墨动画片为主题的演讲	谈论艺术作品
Using Language 2:	应用文（读写课）	中国古代艺术藏品展览信息	运用艺术知识进行交流
Video Time	说明文（视听课）	《清明上河图》概况及其艺术价值	感受艺术魅力，理解作品价值
Workbook: Listening and Speaking	采　访（听说课）	了解不同的艺术创作目的	探究艺术的价值
Workbook: Expanding Your World	说明文（阅读课）	介绍米开朗基罗的雕塑作品《大卫》	了解艺术创作背后的故事
Workbook: Reading and Writing	说明文（读写课）	介绍梵高的作品《星空》	思考如何精彩地介绍艺术作品

（2）提炼主题大观念和语言大观念。

项目式作业不能仅仅依据某一课时的话题进行拓展，脱离了教材单元的主题语境，无从培养核心素养。因此，教师要对单元语篇进行整合利用，使学生循序渐进地进行主题意义探究（李宝荣等，2023）。通过单元内容分析和重组，梳理出本单元的主题大观念是：艺术有不同的形式和特点，需要尊重和欣赏多元文化中的艺术形态，探索艺术价值，树立文化自信。找到学生在单元框架内多个语篇共有的、能学习和循环使用的、完成各种学习任务的语言学习重点，有利于促进学生在新情境中完成表达活动（李宝荣等，2023）。在细致分析单个语篇的基础上发现：谈论艺术目的、作品特色以及发展变化和趋势的语言是多个多模态语篇共有的语言现象。围绕单元主题大观念和语言大观念，Art 单元项目式作业的驱动性问题可设计为：艺术是什么？它有什么价值？项目式作业的主题为介绍自己喜欢的艺术家，形式为幻灯片展示。

2. 着力点：设计系列作业，助推项目式作业

依据学情分析和单元语篇内容整体分析，确定了单元整体教学目标：梳理中西方艺术形式，探究艺术创作背后的故事；基于不同的创作元素和创作技巧，多维度评价中外艺术作品；运用艺术知识介绍艺术展览，就艺术价值表达己见。然后，教师对单元整体教学目标进行分解，按照学生的已有基础进行环环相扣

的课时规划和系列作业规划。Art 单元以艺术为主题，单元内多模态语篇在体裁、主题内容的表达视角以及语言表达特征等方面各有侧重。根据学情和教学重难点，该单元课时规划如下表所示。

表 2 Unit1 Art 单元分课时系列作业规划

课时	教材内容教学	项目式学习	分课时系列作业内容
1-3	话题：A Short History of Western Painting 语篇内容：西方绘画艺术发展史 核心语言：character, religious, humanistic, hand-painted, subject, dimension, 等表达作品特点和影响的词汇；development, breakthrough, innovation, begin, shift 等表达变化的词汇；动词不定式作表语 时间：约 2.5 课时	约 1/2 课时	绘制思维导图，总结不同阶段西方绘画作品的风格以及影响其发展变化的因素
4	话题： Qingming Shang He Tu 语篇内容：清明上河图概况及其艺术价值 核心语言：Song Dynasty, painting scroll, appeal, ordinary 等词汇 时间：约 3/4 课时	约 1/4 课时	录制一分钟视频介绍一幅中国古代名画
5	话题：Monet and the Impressionism; Michelangelo's David 语篇内容：印象派画家莫奈的创作主题;介绍雕塑作品"大卫" 核心语言： fine art, sculpture, painting, be fond of, influential, in particular, purchase, realistic, subjective, theme 等词汇 时间：约 3/4 课时	约 1/4 课时	创作自己的绘画并附有英文文本介绍
6	话题：The Starry Night 语篇内容：介绍梵高的"星空" 核心语言：1. The painting makes me feel...because...;I think Van Gogh must have felt/been...when he made the painting because...; I like/dislike the painting because... 时间：约 3/4 课时	约 1/4 课时	写一篇演讲稿介绍一幅名画
7	话题：Talk about Works of Art 语篇内容：我国水墨动画片为主题的演讲 核心语言：话语标记词和演讲的语篇结构 时间：约 3/4 课时	约 1/4 课时	用课时所学词汇和语篇结构改写演讲稿

8	话题：The Purpose of Art 语篇内容：了解不同的艺术创作目的 核心语言：art, take a second look, purpose 等词汇；Through art, we can learn about...等句型 时间：约 3/4 课时	约 1/4 课时	用课时所学词汇和句型改写演讲稿
9-10	话题：Ancient Chinese Art on Show 语篇内容：中国古代艺术品展览信息 核心语言：exhibition, vases, entry, civil, recognition 等词汇 时间：约 2 课时	约 1/4 课时	将国内某博物馆的展览公告改写成英文，并陈述艺术作品的价值和意义

3. 落脚点：开展持续性评价和支架式教学，确保项目式作业实施效果

指向核心素养的高中英语大单元项目式作业设计由目标设计、过程设计和评价设计三个关键步骤构成。目标设计主要体现出表现性、针对性和真实性三个方面。过程设计一般包括小组筹建（确定成员及前期分工）、活动准备（主题讨论、资料搜集）、活动设计（设计初稿、讨论定稿、合作完成）三个步骤。评价设计主要包括成果分享、公开评价和反思修正三个部分。

Art 单元的项目式作业是要求学生能够根据表达目的选择词汇、句型和语篇结构框架介绍艺术家生平，描述其代表作品特征，抒发艺术见解。要完成此活动，学生需要从单元不同文本中综合提取信息、语言以及表达的视角。教师要通过观察、交流等方式了解学生在作业完成过程中的难点，为学生提供项目流程、话题语言、信息技术、演讲技巧等教学支架。在作业完成之后，教师要引导学生基于教材语篇提炼用于展示和评价的重点语言和语篇特征，并进一步探讨如何借助希沃电子白板完成高质量的汇报展示。此外，师生根据教材内容改编和制定评价标准（见表3）。在完成小组自评和互评之后，教师通过提出引导性问题来帮助学生继续完善作业，引导学生回顾主题语境中所迁移使用的单元内容和语言、小组合作过程以及作业的完成度，引发学生反思并进一步完善作业，达到以评促学的目的。

<div align="center">表 3 Unit 1 Art 项目式作业评价表</div>

评价内容	评价维度	第一组	第二组	...
演讲稿	1.从汇报展示中，我清晰得了解到艺术家以及代表性作品。			
	2.从汇报展示中，我清晰地了解到汇报人对该作品的评价和对艺术的理解。			
	3.演讲稿中，使用了目标词汇、句型和语篇框架结构描述艺术作品特征。			
	4.演讲稿层次分明、思路清晰，使用了话语标记词。			
汇报展示	1.幻灯片排版制作是否清晰明了？			

	2.语法、词汇是否正确？			
	3.语音、语调是否准确？			
	4.与观众的互动是否充分？			
	5.汇报内容的逻辑是否清楚？			
请根据实际情况作答，每个方面都按照 1-5 打分，1 代表做得非常不好，5 代表做得非常好。				

四、结语

鉴于课程改革的要求和项目式作业的育人本质，本文详细阐述了高中英语大单元项目式作业的实施路径。通过围绕单元项目式作业组织和安排单元教学内容，让学生在完成分课时系列作业的过程中逐渐积累演讲所需的语言和主题知识，小组成员利用信息技术合作完成幻灯片制作，在班级和校园英文广播站进行汇报展示，将教材知识迁移到真实情境的过程中，实现单元语言知识与能力的内化，完成艺术之旅。

参考文献

[1] 国务院. 2019. 国务院办公厅关于新时代推进普通高中育人方式改革的指导意见 [EB/OL]. https://www.gov.cn/gongbao/content/2019/content_5404151.htm. ,2019-06-11.

[2] 李宝荣等. 2023.中学英语单元整体教学 [M]. 上海：华东师范大学出版社.

[3] 刘辉,李德显. 2021.中小学作业的异化及回归 [J]. 天津师范大学学报(基础教育版), 22 (04): 1-8.

[4] 梅德明，王蔷. 2018. 普通高中英语课程标准（2017 年版）解读[M].北京：高等教育出版社.

[5] 桑国元等. 2023.项目式学习教师手册 [M]. 北京：北京师范大学出版社.

[6] 王月芬,周坤亮. 2022. "提升中小学作业设计质量的实践研究"教学成果推广应用的实践路径 [J]. 中国教育学刊， (S1): 9-11+15.

[7] 夏雪梅.2021.项目化学习设计：学习素养视角下的国际与本土实践 [M]. 北京：教育科学出版社.

作者信息：

姓名：轩雅莉

职称：中学一级

单位：重庆市南坪中学校

通讯地址：重庆市南岸区白鹤路 138 号

电话：18639429487　　邮箱：1104606103@qq.com

Chinese students' perceptions of affordances and challenges, and their associated coping strategies within different contexts of intercultural communication

Chenjie Yang, Wageningen University & Research, Social Sciences School, Education and Learning Science group

Biemans H.J.A., Wageningen University & Research, Social Sciences School, Education and Learning Science group

Vitaliy Popov, University of Michigan, Ann Arbor, Department of Learning Health Sciences

Abstract

Although the challenges that Chinese students face in intercultural communication encounters have been well documented, scholars have yet to address the context-dependent features of these challenges. This study provides a qualitative examination of the perceptions of Chinese students concerning affordances, challenges, and related coping strategies within the contexts of personal interaction, group work, and class interaction, based on their individual experiences of intercultural communication. Interpretative phenomenological analysis was employed to explore the personal intercultural communication experiences of 22 Chinese international students within multicultural classroom settings at a Dutch university. The qualitative data suggest that the perceptions of these students were closely related to their intercultural interactions within the three contexts. In addition, the participants used coping strategies primarily to address language-related challenges in all three contexts, as well as to address other challenges in different contexts. The challenges and affordances perceived by participants were summarized within their respective contexts, along with their coping strategies. The analysis addresses the relationship between challenges and affordances, and the results suggest that group work is more appropriate than the other two contexts are for communication as a learning environment.

Key words: intercultural communication competence, ecological perspective, Chinese student, higher education, multicultural context

Introduction

As globalization accelerates, the demand for intercultural communication competence (ICC) among college graduates entering the international workforce is growing (Griffith et al., 2016). ICC, defined as "the ability to communicate effectively and appropriately in intercultural situations based on one's intercultural knowledge, skills, and attitudes" (Deardorff, 2006), is essential for building diverse and harmonious communities (Luo & Chan, 2022). Higher education institutions are key in providing multicultural learning opportunities, preparing students for global citizenship through ICC.

Despite efforts to internationalize campuses and support diverse student groups, intercultural interactions remain suboptimal (Popov et al., 2022; Volet & Ang, 1998). Chinese students, the largest group of international students in developed countries (UNESCO project Atlas, 2023), often face greater challenges in intercultural communication than their Western counterparts due to language barriers, cultural differences, and misunderstandings (Cai et al., 2022; Joyce & Hopkins, 2014).

The existing body of literature on ICC can be summarized into roughly three domains: theory, assessment, and practice (Koester & Lustig, 2015). Most studies focus on structured student experiences (Luo & Chan, 2022), but this study adopts a phenomenological approach to examine personal, lived experiences from the perspective of Chinese students. To explore the diverse, subjective intercultural experiences of Chinese students from their own perspectives, ecological perspectives (e.g., Gibson, 1986) were introduced, thereby stressing the importance of the relationship between the environment and the individual.

This study aims to identify factors that promote and hinder Chinese students' engagement in intercultural communication in multicultural classrooms. It explores their personal experiences, revealing the challenges they face and the strategies they use to cope. The findings contribute to ICC theory and could inform educational activities designed to foster ICC, offering insights into the influence of context on intercultural communication.

Theoretical framework

Ecological perspectives

Ecological perspectives, which highlight the socially situated nature of cognition and action (Brown et al., 1989), have been integrated into intercultural studies to better understand the context-sensitive nature of communication (Kimmel & Volet, 2012). This study examines

intercultural communication within a multicultural classroom, focusing on personal interaction, group work, and class interaction during class and break times. It aims to uncover affordances, challenges, and coping strategies in these specific contexts to enhance the comprehensive understanding of intercultural dynamics.

Affordances and Challenges

In this study, we explore the intercultural communication experiences of students through the lens of affordances (Gibson, 1986) and challenges. Affordances are opportunities perceived and utilized by students for engaging in intercultural interactions, not limited to any single context. Challenges, on the other hand, are factors that impede effective participation in these interactions. Borrowing from ecological psychology, these concepts provide a framework for analyzing the interactive process between individuals and their environment, applicable across various fields including second-language acquisition (Van Lier, 2004).

While the perceptions of Chinese students in intercultural settings have been documented (e.g., Ye & Edwards, 2015), the relationship between affordances, challenges, and specific contexts remains underexplored. Studies such as Kim et al. (2022) and Lin & Zhang (2021) have detailed student experiences but lack clarity on the contextual dimensions of these perceptions. Similarly, research on group work contexts does not compare across different settings, making it hard to generalize the findings. This study aims to bridge this gap, offering a nuanced understanding of how affordances and challenges are perceived across various intercultural communication contexts.

Coping strategies

This study examines coping strategies as individual reactions to intercultural challenges and affordances. Students' responses vary based on personal and contextual factors. For instance, British and Chinese students in a multicultural setting often self-segregate as a coping mechanism (Hou & McDowell, 2014), while some Australians view international peers as individuals, fostering more open communication (Willoughby-Knox & Yates, 2021). However, existing research lacks a clear link between coping strategies and context, obscuring the precise influence of environment on student actions.

Our study investigates the following research questions:

RQ1. What affordances and challenges do Chinese students perceive within different multicultural contexts, and why do they perceive them as such?

RQ2. How do Chinese students cope with affordances and challenges within different multicultural contexts, and why do they choose these coping strategies to address affordances and challenges?

Method

Research setting and sample

The study was conducted within the domain of life sciences at a Dutch university with a 22% international student body, primarily targeted Chinese students through WeChat and the Academic Consultancy Training (ACT) program. A snowballing strategy was used for recruitment, and interviews continued until no new themes emerged. Two participants were excluded due to limited intercultural experience (P3) or not being enrolled (P11). Participants were required to have at least three months of study in the Netherlands, and only Master's students were included to focus on classroom intercultural communication, excluding PhD candidates with more workplace experience.

Data collection

Semi-structured interviews, 40-60 minutes each, were conducted with 22 Chinese Master's students (Figure 1) to explore their intercultural communication experiences. Participants discussed challenges, affordances, and coping strategies across the contexts in terms of personal interaction, group work, and class interaction. Questions were designed to prompt detailed responses and examples. Interviews, primarily in Mandarin with occasional English, were recorded with consent and transcribed, accommodating participant preferences for in-person or online formats over five months from June to October 2023.

Figure 1 Demographic information on participants

Variables	Values	Frequency (cases)	Percentage (%)
Gender	Male	5	20.8
	Female	19	79.2
Interview form	In person	11	45.8
	Online	13	54.2
Length of stay in the Netherlands (months)	n/a	1	4.2
	3-12	5	20.8
	13-24	16	66.7
	Over 24	2	8.3
Overseas experience other	None	7	29.2

than the Netherlands	1-12	12	50.0
(months)	Over 12	5	20.8

Valid cases: 22/24

n/a=not applicable

Data analysis

This study adhered to the COREQ checklist for qualitative research rigor and transparency (Tong et al., 2007), employing interpretative phenomenological analysis (IPA) to explore the personal experiences of Chinese students in intercultural communication (Larkin et al., 2009). The focus was on identifying perceived challenges, affordances, and coping strategies within three contexts. The IPA approach facilitated an examination of lived experiences without predefined categories, suitable for the study's small sample size based on in-depth interviews.

The analysis involved six steps: carefully reading, initial noting, theme development, theme connection, case progression, and pattern identification across cases. Data were transcribed, imported into NVivo (version 12) for coding, and analyzed inductively to develop themes and identify patterns. Trustworthiness was ensured through member-checking and peer-debriefing, with codes verified by all authors and themes reviewed by external peers, culminating in a comprehensive discussion among all parties involved.

Findings

Challenges and affordances

Expression and comprehension

In this study, expression and comprehension refer to self-presentation and understanding of others using the English language. The majority of Chinese participants (16/22) found English expression and comprehension challenging across the three contexts. Issues included difficulty understanding various accents, avoiding complex topics, and inefficient communication due to language barriers. As recounted by P6, "Sometimes, I cannot understand teachers and classmates who have heavy accents, especially when they speak fast so that I can no longer join them." In contrast, P9 experienced more difficulty with expression: "I can understand others in English, but for organizing my words, I had to start by translating my words from Chinese into English. I was

therefore always a bit slower in reacting than others were." Atmosphere, passive communication, and power imbalance also affected intercultural engagement.

Atmosphere

Atmosphere is a feeling that participants perceive from others and the environments surrounding them. The atmosphere's impact on intercultural communication was significant, with 19 out of 22 participants noting its influence. A positive atmosphere encouraged communication, as seen in relaxed and friendly conversations (e.g., P5), while a negative one, such as unresponsive feedback or a closed group dynamic, hindered participation (e.g., P19). The atmosphere's openness was particularly crucial in group work, fostering an environment where all felt comfortable to contribute. Conversely, a passive communication style could be both beneficial and detrimental, depending on the context, affecting the willingness of students to engage in dialogue.

Passive communication

In this study, passive communication refers to passive involvement in intercultural communication. Students prefer responding rather than initiating, was experienced by 9 out of 22 participants. It was seen as an affordance in personal interaction and group work, with participants like P23 valuing topic sharing and P20 noting Chinese peers' willingness to contribute despite passivity. However, in class interaction, it became a challenge, as P7 found being called upon in a Dutch course extremely stressful, highlighting the theme of power imbalance.

Power imbalance

In this study, power imbalance, perceived as unequal status among interlocutors, was a challenge for half of the participants. In personal interaction, participants like P7 felt their English proficiency was insufficient for in-depth communication. A deeply ingrained teacher-student status difference was also noted, as P19 mentioned the student's lower status. In group work, power imbalance was less discussed, except by P1 who felt empowered enough to lead. Class interaction saw more reflections, with P9 feeling disadvantaged as the only Chinese person with weaker English. Overall, a lower status was consistently challenging, while a higher status could be advantageous.

Each of the four themes mentioned above were described within all three contexts. The following four themes were mentioned in only two contexts.

Talking size

In this study, talking size refers to the number of interlocutors involved in a conversation. "Talking size" influenced engagement for some participants, particularly in group work and class interaction. Smaller group sizes (4-7 people) were more conducive to self-expression, as P10 preferred speaking in smaller groups for freedom, while P19 and P9 had varying comfort levels. Three participants (P1, P4, and P13) found speaking in front of the entire class challenging and stressful due to concerns about the value of their contributions, as P8's concern over using "public time" indicated, reflecting a close tie to self-assessment.

Self-assessment

In this study, self-assessment refers to a psychological activity in which one evaluates oneself before engaging in any actual expression. Self-assessment was common among participants, particularly in group and class interactions. It served as an affordance for those willing to contribute, like P24, who was eager to improve project reports with her ideas. However, for P23, uncertainty about the value of her ideas posed a challenge. In class interaction, self-assessment often acted as a barrier, with participants needing time to be convinced of their contributions' worth, as P1's struggle to speak until confident in the quality and expression of her answer illustrates. Additionally, subjective preferences influenced the decision to speak.

Subject preference

Subjective preference refers to a judgement concerning whether certain topics or individuals are attractive, based on individual interests. Many participants (17/22) found their motivation for personal interactions hinged on private interests, such as P12 joining a discussion about Germany's education system. However, this interest-based approach could hinder conversation initiation, as P15 avoided contact unless intrigued by someone. In class interactions, attractive content like P17's example could be an affordance, but uninteresting topics, especially those lacking diverse perspectives as P16 experienced, often posed a challenge. Group work and emergent needs were not associated with subjective preferences.

Emergent need

In this study, emergent need refers to personal requests in terms of asking for academic and life assistance. Eight participants reported initiating interactions due to emergent needs, which facilitated personal conversations, such as P14 discussing local life with native students. In class interaction, this need-based motivation was common, with P19 and P22 seeking course materials from lecturers.

Talking size, self-assessment, subjective preference, and emergent need were mentioned within two of the three contexts (i.e., personal interaction, group work, and class interaction). Another two themes were mentioned in only one context: classroom culture shock and duty.

Classroom culture shock

Classroom culture shock refers to a temporary inability to adjust to a new educational paradigm, due to differences in previous educational experiences. Experienced by some participants, like P18, were unaccustomed to the freedom of interrupting teachers, a practice contrasting with their previous experiences in China where students were taught not to interrupt. Others, including P4, P5, and P8, viewed lectures as communal time and reserved personal questions unless broadly beneficial. Conversely, participants had previously associated speaking in class with negative experiences, such as being examined or subtly punished, causing shame if unable to answer correctly. A sense of duty, rather than negative perceptions, encouraged interaction.

Duty

In this study, duty refers to a sense of responsibility for accomplishing something. Nine participants felt a sense of duty in group work, which generally encouraged intercultural interaction. P7 saw active communication as a group work necessity due to shared goals, while P22 was motivated by the expectation to contribute responsibly. However, for some like P16, the pressure for efficiency in group tasks could hinder sufficient interaction, as some discussions might be deemed inefficient.

The figure below (Figure 2) was created to answer the first research question and to provide a clear overview of challenges and affordances. The reasons behind the perceptions are addressed in the Discussion section.

Figure 2 Challenges and affordances, depending on context

Theme	Context					
	Personal interaction		*Group work*		*Class interaction*	
①Expression and comprehension	O	C	O	C	O	C
②Atmosphere	O	A&C	O	A&C	O	A&C

③Passive communication	O	A	O	A	O	C
④Power imbalance	O	C	O	A	O	C
⑤Talking size	X	n/a	O	A&C	O	C
⑥Self-assessment	X	n/a	O	A&C	O	A&C
⑦Subjective preference	O	A&C	X	n/a	O	A&C
⑧Emergent need	O	A	X	n/a	O	A
⑨Classroom culture shock	X	n/a	X	n/a	O	C
⑩Duty	X	n/a	O	A&C	X	n/a

C = challenge; A = affordance; O = mentioned; X = not mentioned; n/a = not applicable

This summary of ten themes across the three contexts highlights distinct participant perceptions. Class interaction posed the most challenges, with eight themes viewed as such, though some also had affordances. Personal interaction saw the fewest themes, with a mix of challenges and affordances. Group work had a more balanced view, with themes often seen as affordances but also carrying challenges. Expression and comprehension consistently emerged as challenges, while emergent need was solely an affordance. Understanding coping strategies is key to addressing the second research question.

Coping strategies

To introduce this section, the figure below (Figure 3) was created to provide an overview of coping strategies, based on Figure 2. The challenges perceived within each of the three contexts were selected from Figure 2, and coping strategies were added to correspond to these challenges.

Figure 3 Correspondence between challenges and coping strategies

Challenge	*Context*	*Coping strategy*
①Expression and comprehension	Personal interaction	i. Learning by doing (5/22) ii. Further explanation (8/22) iii. Change topic (1/22)
	Group work	i. Seeking help (1/22) ii. Reading as an auxiliary means (2/22)
	Class interaction	i. Seeking further explanation (8/22)
②Atmosphere	Personal interaction	i. Body language (2/22)
	Group work	n/a
	Class interaction	n/a
③Passive communication	Class interaction	n/a
④Power imbalance	Personal interaction	n/a
	Class interaction	n/a
⑤Talking size	Group work	n/a

	Class interaction	n/a
⑥Self-assessment	Group work	n/a
	Class interaction	n/a
⑦Subjective preference	Personal interaction	i. Seeking common ground (4/22)
	Class interaction	n/a
⑧Emergent need	Personal interaction	n/a
	Class interaction	n/a
⑨Classroom culture shock	Class interaction	i. Observation (4/22) ii. Auto-suggestion (1/22)
⑩Duty	Group work	n/a

(xx/22) = number of participants mentioning this strategy (of 22 total participants); n/a = not applicable

Figure 3 illustrates coping strategies aimed at overcoming expression and comprehension issues in personal and group interaction, as well as addressing atmosphere, subjective preference, and classroom culture shock. Participants, however, addressed less than half of the challenges in their experiences.

Coping strategies for addressing challenges relating to expression and comprehension

To tackle expression and comprehension challenges in personal interactions, participants like P17 and P20 favored active learning to enhance language skills, while P23 found interacting with culturally close friends beneficial for initial confidence building. A common strategy was seeking further explanation, with 8 participants using it to request clarification from teachers. Changing topics, as P5 did, was another strategy to prevent conversation breakdowns. In group work, participants sought help, like P9 aiding a Chinese peer, or used reading as an auxiliary, as P7 did with meeting records, and P22 by preparing written notes for fluent expression. Class interaction saw participants, again, primarily using the strategy of seeking further explanation to address comprehension issues.

Coping strategies for addressing other challenges

To counteract atmospheric challenges in personal interaction, P20 utilized body language to project openness and friendliness, initiating communication. Finding common ground was another strategy, as P2 engaged with those she perceived shared her aesthetic preferences, and P19 used mutual interests like makeup as conversation starters. P24 suggested introducing oneself to offer commonality cues. For classroom culture shock, observation was key, with P1 overcoming her fear by noting classmates' comfort in speaking up, while P10 employed auto-suggestion to set personal interaction goals.

Discussion

This study identified various challenges and affordances Chinese students face in the three contexts, with class interaction posing the most challenges, likely due to the need to adapt to new educational paradigms. (Khoo, 2004) and (Li & Pitkänen, 2018) suggest that Chinese students are adaptable but require an "initial settlement period" to adjust, after which they engage more in class interactions. (Wang & Byram, 2011) describe this as bringing two learning cultures into relation, though not all students adapt successfully.

In group work context, a sense of duty was a strong motivator, with all participants feeling responsible, even if unsure about their exact contributions. (Popov et al., 2022) note that contribution is a challenge in multicultural groups, which aligns with our findings that students struggle to negotiate roles effectively. Providing a collaboration script or course facilitation could help students find their roles and participate more effectively.

Personal interaction context saw the least motivation for initiating communication, with participants typically engaging only when needs emerged or they were passively involved. Small talk, though not seen as a challenge or affordance, may play a vital role in building rapport and solidifying working relationships, as argued by (Pullin, 2010). Encouraging small talk could be beneficial for initiating interactions.

Participants reported fewer coping strategies than challenges, with limited approaches to addressing intercultural encounters. Learning to recognize and utilize affordances may be more beneficial than focusing solely on challenges. The study suggests that group work is a more suitable context for communication as a learning environment, recommending the use of an affordance "reminder" and balanced assessment to enhance student participation and learning outcomes.

In conclusion, the intercultural experiences of Chinese students are influenced by their perceptions of challenges and affordances within different interaction contexts. Class interaction presents the most challenges, group work offers the most affordances, and personal interaction may undervalue the role of small talk. Strategies to enhance intercultural communication should consider these context-specific perceptions and motivations.

Limitations and future research agenda

This study's reliance on interview data from Chinese Master's students at a single university may limit the findings' transferability. Bias from participants' expression and memory accuracy could affect data reliability. Future research should use diverse data-collection methods and include a broader demographic for a comprehensive understanding of intercultural communication. Longitudinal studies and quantitative assessments could further evaluate and measure students' intercultural competence development.

Reference

Brown, S. J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.

Cai, Y., Fang, F., Sun, H., & Jiang, L. (2022). Unpacking identity construction and negotiation: a case study of Chinese undergraduate students' social and academic experiences while studying abroad. *System*, *110*, 1–12. https://doi.org/10.1016/j.system.2022.102896

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, *10*(3), 241–266. https://doi.org/10.1177/1028315306287002

Gibson, J. J. (1986). *The ecological approach to visual perception*. Erlbaum.

Griffith, R. L., Wolfeld, L., Armon, B. K., Rios, J., & Liu, O. L. (2016). Assessing intercultural competence in higher education: existing research and future directions. *ETS Research Report Series*, *2016*(2), 1–44. https://doi.org/10.1002/ets2.12112

Hou, J., & McDowell, L. (2014). Learning together? experiences on a China–U.K. articulation program in engineering. *Journal of Studies in International Education*, *18*(3), 223–240. https://doi.org/10.1177/1028315313497591

IIE-The Power of International Education. (2023). Project Atlas. Retrieved from https://www.iie.org/research-initiatives/project-atlas/. Accessed February 5, 2024.

Joyce, T., & Hopkins, C. (2014). "Part of the community?" first year international students and their engineering teams. *Engineering Education*, *9*(1), 18–32. https://doi.org/10.11120/ened.2014.00019

Khoo, H. E. (2004). Implementation of problem-based learning in Asian medical schools and students' perceptions of their experience. *Medical Education*, *37*(5), 401–409.

Kim, A., Lee, S., & Choe, H. (2022). Socioenvironmental affordances of an English-medium instruction international university specializing in English as a lingua franca communication: an exploratory study of L2 student voices. *Lingua*, *273*, 1–13. https://doi.org/10.1016/j.lingua.2022.103327

Kimmel, K., & Volet, S. (2012). Understanding motivation, engagement and experiences of intercultural

Interactions at university: a person-in-multiple contexts perspective. *European Journal of Psychology of Education*, *27*(2), 227–245. https://doi.org/10.1007/s10212-011-0084-3

Koester, J., & Lustig, M. W. (2015). Intercultural communication competence: theory, measurement, and application. *International Journal of Intercultural Relations*, *48*, 20–21. https://doi.org/10.1016/j.ijintrel.2015.03.006

Larkin, M., Flowers, P., & Smith, J. A. (2009). *Interpretative phenomenological analysis: theory, method and research*. SAGE Publications.

Li, H., & Pitkänen, P. (2018). Understanding the integration of mainland Chinese students: the case of finland. *Nordic Journal of Migration Research*, *8*(2), 107–115. https://doi.org/10.1515/njmr-2018-0008

Lin, Y., & Zhang, H. (2021). "I finally marginalized myself from the mainstream": an autoethnography study of Chinese international student's development of Intercultural communicative competence. *Qualitative Report*, *26*(11), 3665–3682. https://doi.org/10.46743/2160-3715/2021.4827

Luo, J., & Chan, C. K. Y. (2022). Qualitative methods to assess intercultural competence in higher education research: a systematic review with practical implications. *Educational Research Review*, *37*, 1–18. https://doi.org/10.1016/j.edurev.2022.100476

Popov, V., Brinkman, D., Fortuin, K. P. J., Lie, R., & Li, Y. (2022). Challenges home and international students face in group work at a Dutch university. *European Journal of Engineering Education*, *47*(4), 664–678. https://doi.org/10.1080/03043797.2022.2044762

Pullin, P. (2010). Small talk, rapport, and international communicative competence: lessons to learn from BELF. *Journal of Business Communication*, *47*(4), 455–476. https://doi.org/10.1177/0021943610377307

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, *19*(6), 349–357. https://doi.org/10.1093/intqhc/mzm042

Van Lier, L. (2004). *The ecology and semiotics of language learning: a sociocultural perspective*. Kluwer Academic Publishers.

Volet, S. E., & Ang, G. (1998). Culturally mixed groups on international campuses: an opportunity for inter-cultural learning. *Higher Education Research & Development*, *17*(1), 5–23.

Wang, L., & Byram, M. (2011). "But when you are doing your exams It Is the same as in China" - Chinese students adjusting to Western approaches to teaching and learning. *Cambridge Journal of Education*, *41*(4), 407–424. https://doi.org/10.1080/0305764X.2011.625001

Willoughby-Knox, B., & Yates, L. (2021). Working toward connectedness: local and international

students' perspectives on intercultural communication and friendship-forming. *Higher Education Research and Development*, *41*(7), 2506–2522. https://doi.org/10.1080/07294360.2021.1985087

Ye, L., & Edwards, V. (2015). Chinese overseas doctoral student narratives of intercultural adaptation. *Journal of Research in International Education*, *14*(3), 228–241. https://doi.org/10.1177/1475240915614934

信息科技特色高校全球胜任力外语课目标重构与实践[1]

张钫炜 闫鑫凯 赵欣
北京邮电大学

摘要：

在全面建设社会主义现代化强国的进程中，教育、科技、人才一体化发展有着深刻的内在逻辑。强国建设，外语何为？信息科技特色院校的外语课应积极回应科技强国和教育强国建设对高素质科技人才培养的需求。本研究基于北京邮电大学的"网信报国，网络强国"的人才培养目标，针对该校拔尖人才的全球胜任力培养要求，从知识、态度、技能和行动四个维度重新设计了大学英语教学目标，以团队搭建的 LEGO 教学理念为指引，开展了长达三学期的教学实践。研究采用混合研究方法，依托问卷调查和电子档案袋的文本挖掘，结合定量和定性数据，全面评估教学实践的效果。数据分析结果表明，经过系统教学，学生在全球胜任力的各个维度上均表现出显著提升。

关键词：信息科技特色；全球胜任力；大学英语教学

一、引言

在全球治理体系变革的背景下，新一轮科技革命和产业变革深入发展，高技术领域成为国际竞争最前沿和主战场，深刻重塑全球秩序和发展格局。为进一步抢占科技竞争和未来发展制高点，我国科技人才培养范式需要各学科参与的范式变革。在这一过程中，外语学科的作用愈加凸显。外语教育不仅是语言能力的培养，更是文化理解与跨文化沟通能力的培育。具备全球视野和跨文化交流能力的科技人才，能够更好地理解国际科技前沿动态、参与国际科技合作与竞争，以及在全球化背景下更有效地传播中国的科技成果和声音。本文结合北京邮电大学特色，面向未来科技和产业发展对人才综合素质的新要求，重构以全球胜任力为导向的拔尖人才英语课程体系，主动融入"大通信观、大网络观、大数据观、大安全观"的育人体系[1]，培养具有国际竞争力的青年科技人才后备军和创新团队[2]。

二、文献回顾

最早开展全球胜任力理论研究的哈佛大学就已经关注到科技人才应具备跨文化沟通能

力、创新思维和对新兴技术的快速适应能力[3]，以确保在全球化的技术竞争中脱颖而出。亚洲协会与经合组织（简称为 OECD）联合发布的《PISA 全球胜任力框架》，将全球胜任力划分为知识、态度、技能以及价值观四大维度[4]。多项研究使用美国工程与技术教育认证委员会[5]文件中对工程师教育的定位，如跨文化沟通能力，多学科团队合作、全球视野和文化理解、适应性和灵活性等全球胜任工程师的核心胜任力，探索全球胜任工程师需具备的技能[6, 7]。我国正面临全球产业格局调整及经济发展环境变化对高等教育工程人才培养体系提出的挑战[8]。以实现"面向产业、面向世界、面向未来"，培养可持续胜任力为人才培养目标核心的"新工科"建设背景要求高等教育在现有全球胜任力框架的基础之上，既要体现办学特色和办学优势，同时要培养具有前瞻性、能够引领未来发展的技术创新人才[9]。

国内学者在构建全球胜任力本土理论的过程中都强调外语能力的重要性,清华大学的全球胜任力框架甚至将语言作为六大核心素养之一。对于外语课全球胜任力的教学设计，学者从教学理念、教学内容、教学方式、教师素养[10]、语种开设、支撑体系等多个维度进行了较为全面的讨论[11]。实践研究不多，有学者以全球胜任力和语言能力发展为目标开展 OBE 教学模式实证研究[12]，但在数据收集和分析上有一定的局限性。

在此背景下，探讨如何结合高校类别、层次和所在区域，主动对接相关学科和相关领域，提升学生的专业能力与语言能力，同时拓宽国际视野，是培养高水平国际化拔尖创新人才的关键。这不仅满足国家对紧缺人才的需求，也为高校外语课程的教学设计和实践提供了新的方向。

三、北邮全球胜任力英语课程教学目标重构与实践

在现有全球胜任力框架中，OECD 所提出的全球胜任力模型相对完备。新时代对卓越工程科技人才培养在专业理论、工程基础、管理科学、实践能力、职业道德、敬业精神、社会责任感、人文素养、创新意识、终身学习、团队协作和国际交流等方面提出了要求[13]，《工程教育认证通用标准解读和使用指南》也对学生的社会责任感、人文社科素养和跨文化沟通能力有明确的规定[14]。

课程基于 PISA 全球胜任力知识、态度、技能与价值观框架，结合工程教育认证通用标准解读和使用指南[14]（2022）和北邮办学特色，将全球胜任力重新将知识、态度、技能、以及行动四个维度进行划分（图 1），下设 8 个二级指标，见表 1。

图1　全球胜任力英语课程教学目标重构思路
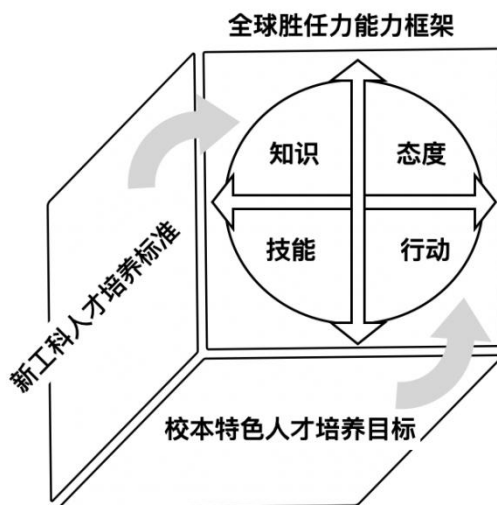
表 1 信息科技特色高校全球胜任力新内涵

维度	要求	二级指标
知识	掌握国内外文化科技知识、跨文化知识以及语言与专业知识	国内外文化科技知识 英语语言知识
态度	在自我文化认同的同时，了解、尊重、欣赏和包容多元文化	文化自信与开放包容 媒介素养与批判思维
技能	与不同文化背景的人进行国际学术交流与互动	跨文化交流与沟通能力 学术写作与表达能力
行动	面向未来产业发展、为参与全球治理采取行动	洞察科技前沿 参与全球治理

团队结合新工科人才培养标准及北邮校本特色，以全球胜任力四个维度为基础，搭建三阶段的全球胜任力英语课程体系，学生从了解中国、了解世界开始，逐步深入理解中国、交流互鉴，最终实现表达中国、沟通世界（表2）。

表2 信息特色高校全球胜任力课程设置

全球胜任力维度	新工科人才培养标准	教学阶段	教学主题及方式
知识	掌握工程及人文知识，了解可持续发展理念，考虑复杂工程问题解决	了解中国，了解世界（第一、第二	国内外文化科技知识 跨文化理论 英语语言知识

	学期）		
态度	职业伦理，社会责任感， 家国情怀及国际视野		
技能	工具使用，调研能力， 多学科、跨文化团队合 作及外部沟通，项目管 理，跨学科系统思维及 整合能力	表达中国， 沟通世界 （第二、第三 学期）	英语语言技能 跨文化交际策略 演讲表达技能 写作技能
行动	终身学习，创新及应用		

四、 全球胜任力英语课程体系实践反馈

（一）研究设计

本研究采用问卷调查、电子档案袋以及半开放式访谈等工具考察学生的全球胜任力提升情况。使用 SPSS 27.0，Rost Content Mining 6.0 和 Nvivo11.0 软件作为问卷数据和文本分析的工具。问卷主要参考 Dearoff 的成熟量表[15]，结合 Ortiz-Marcos 设计的工程师胜任力量表[16]和 PISA 全球胜任力评估框架增加了"理解可持续发展内涵"和"全球化的职业发展"两个维度，全球及中国知识内容主要根据刘扬[17]有关中国大学生全球素养的量表进行设计，并根据课程教学情况和当前全球热点议题进行了表述调整。问卷分为知识、态度、技能与行动三个部分，题项均采用李克特 5 点量表进行计分，知识部分的考察主要以学生自评的方式进行测量（要求学生想象自己是否能以英文为主的方式对所涉及的知识点进行解释和沟通），以电子档案袋内的阶段性小测和反思进行补充。

（二）数据收集

本课程体系在未来学院 2022 级计算机类和电子信息类专业的 100 位学生中（其中男生87 人，女生 13 人）完成为期一年半的实践。共 84 人参与问卷调查（其中得到 74 份有效问卷），87 人参与电子档案袋撰写。学生分别在第一学期前和第三学期后完成问卷，在前两学期完成的电子档案袋(共计 32354 字)。电子档案袋针对课程收获、新闻述评活动总结与

反思、跨文化交际反思、全球胜任力认知等模块进行情感分析，最终进行开放性编码与主轴编码，将文本内容概念化、数量化，从而检验学生在本课程体系培养后对于全球胜任力的认识以及课程体系实践成效。

（三）问卷数据分析

为了检验问卷设计的原有四个维度的稳定度和新增加的工程师胜任力和校本特色内容是否能符合维度条件，在试测时对数据进行了探索性因子分析，全球胜任力量表共 28 个题项，通过删去 3 个题项得到 KMO 指标为 0.778，删去题项后的 Cronbach's α 系数为 0.893。各因子的组合信度在 0.756-0.872 之间，总体解释比例达到 65%，各因子的负载量及描述见表 3。

表 3 全球胜任力量表各维度信效度分析

维度	因子	题项描述	特征值	方差百分比	累计百分比
知识	国内外文化科技知识	中国的国际角色，国际安全，气候问题，中国文化遗产，全球互联网(8 个题项)	4.437	17.748	17.748
	全球理解及跨文化知识	跨文化知识，全球化知识，理解可持续发展（6 个题项）	4.159	16.636	34.384
态度	国际化意识和价值认同	参与国际议题，服务和沟通意愿（5 个题项）	2.810	11.238	45.622
技能	工具使用和沟通能力	互联网工具使用，语言及非语言沟通能力（3 个题项）	2.642	10.570	56.192
行动	个人融入全球发展	个人的全球化发展（3 个题项）	2.249	8.997	65.189

表 4 显示了全球胜任力水平得分的配对样本 t 检验统计结果，各维度得分通过内部因子题项取均值得出，总分由各维度题项均分之和得出。第一学期课程开始前，学生的整体全球胜任力水平偏低，尤其在知识层面储备不足，技能维度次之，行动和态度维度处于中等水平，且学生内部全球胜任力的水平差距较大，总分偏低。经过本课程体系的实践，学生的知识维

度提升最大，且具备较高的一致性，佐证了教学措施的有效程度。学生的全球胜任力技能、行动和态度也有不同程度的显著整体提升。由于数据样本由学生自评得出，具有一定的局限性，因此本研究结合电子档案袋和访谈内容对结果进行进一步分析。

表 4　　全球胜任力各部分前后测得分

	前测得分(n=74)	后测得分(n=74)	t-Test
知识	2.398±0.66	4.322±0.368	-22.191***
态度	3.5±0.897	4.148±0.533	-5.417***
技能	2.861±0.797	3.919±0.506	-9.784***
行动	3.743±0.705	4.232±0.515	-4.998***
全球胜任力总分	57.203±10.919	79.446±7.028	-15.337***

*** $p < .0001$

（四）学生反思情感分析

本文通过 Rost Content Mining 6.0 软件对学生的电子档案袋文本内容进行情感分析，结果见附录 1。通过分析结果可知，84.21%的学生对课程反馈积极情感，认同课程体系的教学提升了个人能力和对未来学习的信心。极少数学生的文本呈现出消极态度，主要集中于自我能力认知与人才培养目标的差距。通过总体情感分析发现，学生对于本课程体系的反思大多呈积极态度，接受度良好。

（五）编码分析

在情感分析的基础上，本研究使用 Nvivo11.0 的自由编码功能，对采集的文本进行了编码和标记，共产生 9 个初始概念范畴，分别为"沟通""交流""全球""问题""世界""中国""文化""英语""英文"。研究者对自由编码节点进行同义词合并，最终提取 5 个主范畴，部分开放性编码的结果如附录二所示。

由编码结果可知，学生电子档案袋记录的思考内容主要关注沟通能力的提升、对本土文化与世界文化的认识、对全球问题的关注以及全球治理的责任意识以及对自身英语学习的反思与总结，其中跨文化交流与沟通以及以及英语语言知识的频次高达 100 以上，批判思维与

媒介素养、中外文化科技知识以及全球治理与胜任力的频次也高达 60 以上，这些发现呼应了全球胜任力的核心要素，见附录二。

在课程实践结束后一学期，本研究随机抽取了 100 位学生中的 10 位进行回访，重点探讨课程对他们全球胜任力的影响。多数学生表示，课程中学到的信息搜集和甄别能力对他们影响深远，尤其在分析中外报道差异方面。在认知层面，学生们普遍提升了对文化多样性和全球治理问题的理解，开始更多地关注相关信息。在技能层面，课程培养了他们处理跨文化信息的能力，包括有效信息搜集、分析以及独立判断的技能。由电子档案袋的文本分析以及半结构化访谈可以得出以下结论。首先，信息科技特色高校的胜任力人才培养需要建立在教学目标重构的基础上，符合校本特色的全球胜任力目标能够为教学设计以及课程体系构建提供指导。其次，学生能力及认知的逐渐提高印证了本课程体系的科学性，经过三个学期的实践，课程教学目标基本实现。此外，周期性的胜任力发展评估有助于课程体系及教学实践的完善，通过定期收集学生电子档案袋，有助于监测学生学习动态，开展形成性评价。

六、 结语

本文对全球胜任力英语课程体系的构建与实施进行了探讨，通过对北邮 2022 级未来学院 100 名学生在 2022-2023 年一年半的教学实践进行分析，本研究动态验证了该课程体系在促进学生主动学习、增强跨文化交流能力以及提升综合素养方面的有效性。学生通过参与新闻述评、跨文化案例分析等活动，以及在电子档案袋中记录学习过程和反馈，展现了对全球胜任力核心要素的深入理解和应用。此外，学生在各类英文竞赛以及创新创业活动中的积极表现也进一步验证了课程设计的成效。

然而，本研究也发现当前课程实施中存在的局限性，即对于教学团队的关注不足，未来的研究将着重于强化教学团队的理念一致性和提升教师的跨文化教学能力，同时进一步精炼课程架构，量化评价标准，优化教学实践，从而更有效地培育具备全球胜任力的未来人才。

Reconstruction and Practice of Global Competence of Foreign Language Course in Information Technology Characteristic Universities

Abstract: The current global governance system is undergoing profound changes, and the need for global talents is becoming more urgent. The foreign language teaching of new engineering talents should aim at cultivating engineering science and technology talents with national feelings and international vision. According to the information technology characteristic talents training of Beijing university of post and telecommunications, this study reconstructs the global competency teaching objectives, around knowledge, attitude, skills and action. After three-semester practice of the four dimensional global competency English teaching mode, students' global competence cognitive level is significantly improved through the effect test of electronic portfolio tools.

Key words: information technology characteristics; global competence; college English teaching

参考文献

[1] 徐坤. 建设教育强国：高水平研究型大学的使命担当和行动指向[J]. 中国高等教育. 2023(19): 12-15.

[2] 徐坤. 建构行业特色鲜明的卓越工程师培养体系　服务网络强国战略和数字经济发展[J]. 学位与研究生教育. 2022(07): 6-12.

[3] Jackson V, Anthony B M. Educating for Global Competence Preparing Our Youth to Engage the World[J]. 2011.

[4] 滕珺，杜晓燕. 经合组织《PISA 全球胜任力框架》述评[J]. 外国教育研究. 2018, 45(12): 100-111.

[5] ABET, Criteria for Accrediting Engineering Programs. (2015-2016 Accreditation Cycle)[Z].

[6] May D A T A. The globally competent engineer: What different stakeholders say about educating engineers for a globalized world[C]. Oxford, UK: 2014.

[7] Parkinson A. The Rationale for Developing Global Competence[J]. Online Journal for Global Engineering Education. 2009, 4.

[8] 张春良，刘长红，江帆，等. "多元协同、多维评价"工程人才培养模式探索[J]. 高等工程教育研究. 2022(03): 112-116.

[9] 张力. 新工科背景下未来技术人才培养目标调整与育人新范式[J]. 江苏高教. 2022(10): 74-79.

[10] 黄雯怡. 基于全球胜任力培养的大学英语教学路径探析[J]. 中国大学教学. 2023(12): 56-62.

[11] 赵双花，王雪梅. 全球胜任力视域下世界一流大学外语人才培养模式研究[J]. 山东外语教学. 2024, 45(01): 55-64.

[12] 姜霞，王雪梅. 全球胜任力导向的"综合英语" OBE 教学模式建构[J]. 中国外语. 2023, 20(06): 4-10.

[13] 洪军，王小华，王秋旺，等. 校企协同、产教融合卓越工程科技人才培养探索[J]. 高等工程教育研究. 2024(03): 37-41.

[14] 中国工程教育专业认证协会. 工程教育认证通用标准解读及使用指南[Z]. 2022.

[15] Deardorff D. Identification and assessment of intercultural competence as a student outcome of internationalization[J]. Journal of Studies in International Education. 2006, 3(10): 241-266.

[16] Ortiz-Marcos I, Breuker V, Rodríguez-Rivero R, et al. A framework of global competence for engineers: The need for a sustainable world[J]. Sustainability. 2020: 9568.

[17] 刘扬，孔繁盛. 大学生全球素养:结构、影响因素及评价[J]. 现代教育管理. 2018(01): 67-71.

附录一 学生课程情感分析结果

情绪类型	占比（%）	分段类型	占比（%）	举例
积极情绪	84.21	一般（0—10）	26.32	提高了英语水平和国际化视野
		中度（10—20）	36.84	期望未来在英语以及相关领域的学习上有更大的进步期望有更多的同学间互促活动
		高度（20 以上）	21.205	英语课让我敢于去开口说英语了老师和助教们对于英语作业的批改也非常的细致耐心我希望可以增加更多提升口语和听力的部分
中性情绪	10.53	-	-	我认为全球胜任力英语概括的说便是培养我们的综合能力，区别于传统的英语教学，全球胜任力英语培养的是具有国际化视野的多语言人才。
消极情绪	5.26	一般（-10—0） 中度（-20—-10） 高度（-20 以下）	0.00 5.26 0.00	我还有很多的问题急需改正，首先，我的口语能力欠佳，一些发音和节奏掌握的不好，…此外，我的临场能力较差，一旦出现了和我练习过程中冲突的情况，我就会磕巴、愣神，这也就导致我的观点不太具有说服力。

附录二 电子档案袋文本自由编码结果

开放式编码提取范畴	频次	部分文本内容
中外文化科技知识	71	英语课程在理解中国文化这个层面给了我未曾预料到的惊讶，在初级的理解英文词汇，读懂英文句子的基础上，让我可以用英文这个载体表达中国故事； 通过同学们的演讲，我领会到了世界万象，从生活到科技，从个体到人类，我的眼界和思维都极大地扩展开来； 当我们在学习英语时，我们明白我们不只是在摄取国外的文化，我们也在学习如何用外国人能接受的方式输出我们自己

的文化、讲好中国故事，因此我在平时里的学习生活中也要注重如何让中国走向世界。

本学期《英语品中国》教材补充了我许多我国科技方面的认识盲区，不局限于词汇，更多的是对于具体<u>中国文化、中国方式</u>的体会。领域涉及<u>在线教育、工程、智慧通信</u>等。

英语语言知识	149	通过一学期的学习，在基础方面我掌握了更多的英文单词，一些<u>地道的英语表达</u>，在其他方面，我明白了日常练习对于英语的重要性，<u>听写和口语表达</u>也是英语学习重要的一部分； 通过这学期的英语课程的学习，我的英语<u>听、说</u>能力有了较大提升，逻辑表达方面收获颇丰；在未来，我还要投入更多的时间，加强基础，学习单词，学习语篇，对英语中的各种语法尽可能多地了解并掌握，积极参加各种英语活动。
批判思维与媒介素养	68	<u>尊重他国文化，理解文化冲突，</u>借鉴其他文化的优点，才能具备胜任力；同学们进行的新闻述评让我看到了很多来自国内外的新闻，加强了我对当今世界格局以及各国之间文化的差异的认识，提高了我的认知水平。 通过利用新闻渠道，能够查询有关资料为自己提供案例的支撑，这是我认为大学英语学习与以往中学学习的一大不同。 <u>媒体的不同立场决定了他们对同一件事的不同解释。</u>我们应该注意我们自己的立场，从各个方面来看待这个问题，<u>有意识观察不同媒体的立场</u>，对立场有初步了解就能够理解行文的思路，也能<u>坚持自己的观点不被媒体影响</u>，而是根据事实去进行自己的分析和理解。
跨文化交流与合作能力	181	我也深刻意识到了<u>沟通的重要性</u>，勤加练习沟通技巧，为中国文化向世界<u>传播</u>出一份力； 在同学们的新闻述评以及自己阅读新闻的过程中，了解到许多中国文化、时事相关的英文表述，增强了<u>跨文化沟通交流</u>

能力，能更好地进行本土文化的传播与推广；

全球治理与胜任力	**64**	面对全球问题，全球治理，我们不能一副事不关己高高挂起地模样，甚至于幸灾乐祸看笑话，而是应该坚定人类命运共同体，为世界问题出谋划策，寻找良方，贡献中国智慧、中国方案； 全球胜任力是以开放的世界观与格局观察世界，在保留自身文化身份的基础上客观地、辩证地看待全球立场与跨文化现象，但是最重要的是能够恰当开放地进行跨文化互动与沟通，这一点是非常重要的，从某种意义上来说，全球胜任力简直是人类命运共同体理念在英语学科的生动体现。

Applied Linguistics

应用语言学

The Relationship between Linguistic and Ideologies in the Discourse of "Perfect Parenthood" from Parenting Guides in Thai Society

Rungtiva Supanunt

PHRANAKHON SI AYUTTHAYA RAJABHAT UNIVERSITY

Email: rungtiva.s@aru.ac.th

Abstract

This research aims to study the relationship between linguistic strategies and ideologies in the discourse of "perfect parenthood" as depicted in Thai parenting guidebooks. The study collected and analyzed specific discourse excerpts from parenting guides published between 2020 and 2024 using critical discourse analysis within a linguistic framework. The findings reveal that authors of "perfect parenthood" discourse employ various linguistic strategies such as lexical choice, quoting, asserting, denying assertions, and figurative language to construct ideologies. These ideologies include the notions that: 1. Perfect parents spend ample time with their children; 2. Perfect parents are creators of their children.; 3. Perfect parents are financially secure; and 4. Perfect parents possess desirable characteristics. These ideologies aim to shape perceptions and acceptance that successful parenthood hinges on adhering to these discourse-defined formulas, thereby serving marketing purposes and influencing societal norms.

Keywords: Parenthood; Ideologies; Linguistic Strategies; Critical Discourse Analysis

1. Introduction

"Family" consists of members: father, mother, and child. This is the primary and foundational institution where love and knowledge are cultivated for the child's future as a member of society. Therefore, learning from parenting guidebooks is crucial for parents. These guidebooks refer to books that advise on how to prepare for parenthood and propagate norms that shape parental roles in Thai society. They play a significant role in constructing and replicating ideas about "parenthood," as exemplified by the following passage:

Example (1) Children often misbehave towards their parents because they dislike their parents' attitudes. However, when the parents stop interfering, they misbehave even more to provoke their parents into re-engaging and behaving as they did before: overly strict, controlling, or overly protective, or showing that they do not love or want them. Parents are usually ready to retreat easily when their children misbehave. What parents need to do is show them that we love them and have changed. This work must be done to show, not just say, that we love our children. (Raise your child to be a smart person, 2024)

Researchers noted that the concept of being a parent in the text reveals the set of ideas the writer constructs for a perfect parent, such as loving and instructing children according to desired characteristics. If analyzing the language use method, it is found that the statement of "loving children" must show a change in behavior. If parents do bad things to their children, like being strict or saying that children will leave, it will not be the desired parent characteristics. Along with finding the set of ideas of being a parent who must perform a full role in the life of the couple, to nourish children together, and to show the love of children, with this reason. (Van Dijk, 1998)

The researcher suggests that the example cited from the book can be analyzed using the framework of Critical Discourse Analysis (CDA), which Fairclough (1992) proposed as a three-dimensional framework. CDA aims to reveal underlying ideologies embedded in discourse, exerting power over societal beliefs and thoughts. Related studies have explored the relationship between language and the notion of "parenthood" in social media contexts such as the "Little Monster" fan page on Facebook (Namsopa, P., 2019).

Parents are highly expected to serve as exemplary figures in behavior and teaching, making parenting guides significant for transmitting and influencing readers' thoughts, thereby shaping societal ideologies. Recognizing the importance of this relationship, the researcher is interested in studying the correlation between language and the concept of "parenthood" in child-rearing guides to understand discourse dynamics in this context.

2. Objective

The objective of the study is to investigate the relationship between linguistic devices and ideological content in the discourse of "perfect parenthood" as portrayed in a guidebook on child-rearing in Thai society.

3. Methods

The study focused specifically on linguistic expressions and analyzed ideological viewpoints presented in the guidebook on child-rearing. Drawing from Fairclough's framework (1992) for Critical Discourse Analysis (CDA), the study proposed a three-dimensional analytical framework for discourse analysis.

Data Scope:

For this study, the researcher collected data from popular child-rearing guidebooks authored by Dr. Prasert Palitponganpim. Three randomly selected volumes published within the past five years (2020–2024) were included:

1. "Raise your child to be a smart person" (2024),

2. "43 settler erases the messy one" (2022),

3. "Parent School" (2021).

4. Results and discussions

4.1. Lexical choices

Lexical choices refer to the selection of words or phrases used to convey thoughts on complete parenting, including selecting words that convey parents' responsibility to guide their children, show love, and devote time to them.

Example (1) The key issue of playtime reflects parents' concerns and whether parents have time to play with their children to be playmates or to care for them (Parent School, 2021).

Example (2) Therefore, how should parents stay with and use their judgment, especially when climbing high or swimming in constant danger? (Parent School, 2021).

Example (3) Regardless, we must take the best care of our children, giving them as much time as they need, without comparing our children to those of other households (43 settler erases the messy one, 2022).

Example (4) Whether parents are close to their children, they have raised and embraced them since birth until 8 months and continue to nurture and embrace them regularly until 3 years old. If they can do this, the image of parents will be truly clear and real, not disappearing when hidden. We will benefit from this throughout our lives (43 settler erases the messy one, 2022).

Example (5) illustrates that parents must have a consistent plan for raising their children, both knowingly and unknowingly (Raise your child to be a smart person, 2024). Examples (1), (2), (3), and (4) demonstrate the rhetoric that parents must dedicate "time" continuously, caring for their children to engage in various activities and always be close by. Example (5) shows that parents must have long-term goals set for guiding their children's upbringing.

4.2. Claiming refers to stating facts, reasons, and supporting arguments to define or guide the audience's thoughts. Using this method makes the discourse presented weighty and credible. Analysis shows speakers use claiming in various forms, as follows:

Example (1): Many parents are perfectionists, particularly successful middle to upper-middle class ones. ("Raise your child to be a smart person, 2024")

Example (2): Unpredictable behaviors of parents often pose more developmental challenges for children, as they do not know how to react. ("Raise your child to be a smart person, 2024")

Example (3): Today's society faces two main issues with parents: 1. Comparison, and 2. High expectations. ("Parent School, 2021")

Example (4): Many parents insist that their child "choose" once they know the truth, but the truth is the child did not choose to be anything. If they do become something, it's because they are themselves. ("43 settler erases the messy one, 2022")

Example (5): Every parent should aim to raise a child with skills for thriving in the 21st century, more than just knowing or excelling in one thing alone. ("Raise your child to be a smart person, 2024")

4.3. The use of presupposition refers to shared background knowledge between the speaker and the listener. Examples illustrating this include:

(1) In the past, parents didn't allow their children to talk to strangers. Nowadays, with the advent of WiFi, Facebook, and online games, parents can no longer prevent their children from talking to strangers. Therefore, children must develop conversational skills with strangers on their own. (Parent School, 2021)

(2) I think there could be many answers to this question. Every parent should sit down and talk. Why don't children listen to us, but there is one correct answer. That is, we give children too much time and hug them too much. (43 settler erases the messy one, 2022)

(3) In any family, parents must love each other in the beginning and love each other more than they love their children. Then both send love to their children in equal measure. Not to one side or the other. (Raise your child to be a smart person, 2024)

(4) The mother raised her son to be a good person, which is important. Sons will take off the moral template from their mothers. As for sons playing with older sisters often, it's good. There's nothing to worry about. (Raise your child to be a smart person, 2024)

(5) The child wants us to sleep together, live together, and share the same heart in the peace of mind for no more than ten years. Parents who know how to use this time will find happiness in another dimension. And when the time passes, they will find that it was not a waste of time, not regretful that they did not give them time, or give them time and each other. Above all else is the relationship or attachment that the child has with us every night for the first decade of life. It is safe, strong, and cannot be destroyed. This means that when he goes to find his conscience wherever he is, no matter how he is provoked by society, he still thinks of us. (Raise your child to be a smart person, 2024)

From examples (1), (2), (3), (4), and (5), it is evident that the discourse on child rearing differs between the past and now, showing the knowledge and experience of parenting in the past versus today.

Today, it is necessary to increase time, closeness, and love more than before, such as paying attention to children's play, spending bedtime together, which means having time with children throughout the day.

Examples (1), (3), (4), and (5) illustrate collective claiming, using phrases like "many parents," "our homes," and "every parent." Example (2) demonstrates claiming with reasoning, using "because." These examples aim to build credibility so that parents reading such discourse believe and act upon the writer's advice.

4.4. Negation refers to the use of negative statements to contradict previous ideas, beliefs, or events.

Example (1): The modern world has many stories where parents don't know what's right or wrong, and many things are neither right nor wrong. (Parent School, 2021)

Example (2): Parents without qualifications tend to create unavoidable problems for their children and perpetuate mental health issues from one generation to the next. (Raise your child to be a smart person, 2024)

Example (3): Parents who treat their children like dolls often neglect their safety and fail to nurture their abilities to become smart or good individuals. (Raise your child to be a smart person, 2024)

Example (4): No parent can consistently draw a stable line; some days they say one thing and do another. (Raise your child to be a smart person, 2024)

Example (5): These parents think that having money solves everything, which is wrong. Their children don't need as much money as they think; they need parents who spend more time with them. (Raise your child to be a smart person, 2024)

From example (1), it contradicts the previous idea that raising children is neither right nor wrong. Example (2) uses negation of "without qualifications" to emphasize that such parents cause ongoing problems for their children. Examples (3) and (4) illustrate that parenting involves "training" children to become smart and good, which depends largely on the parents. Example (5) negates the idea that money is everything to children, emphasizing instead their need for parental time. These examples also reveal undesirable parental characteristics.

4.5. The use of hedging language refers to using expressions that convey uncertainty or indicate a lack of desire to definitively assert something as true. This is done to lessen the forcefulness of statements (Hyland, 1998).

Example (1): Therefore, if we were to say that every child has a duty to be rebellious, we might find that he tests his father, tests his mother, tests his grandparents, and tests everyone in the family (Raise your child to be a smart person, 2024).

Example (2): Symptoms of indulgence often lead to the neglect of parents, causing a vicious cycle (Raise your child to be a smart person, 2024).

Example (3): Unpredictable behavior by parents often poses more developmental problems for children because they do not know how to behave (Raise your child to be a smart person, 2024).

Example (4): Young children who are disciplined will know when to eat, bathe, and play. What parents often do not know is that children who do this will be proud that they did (43 settler erases the messy one, 2022).

Example (5): Overprotective parenting often results in fearful children (Raise your child to be a smart person, 2024).

From examples (1), (2), (3), (4), and (5), it is evident that rhetoric is used to lessen the weight of the writer's advice, using words like "likely," "often," and "typically" to discuss parental and child behaviors, to present both causes and effects according to the writer's advice, but also to avoid fully committing to responsibility for such advice. For example, advising that "Overprotective parenting often results in fearful children. Therefore, the use of hedging language is a method that allows rhetoric to create a balance between facts and opinions.

5. Conclusions and recommendations

The research findings reveal that speakers on the topic of "ideal parenthood" from Thai parenting guidebooks employ various linguistic strategies such as including lexical choice, claiming, presupposition, negation and hedging to create sets of ideas:

1. The ideal parent is one who devotes time to their child consistently. This is because both parents, through their language choices, must dedicate time to their child in every activity and aspect of life, starting from the child's birth and continuing through their upbringing.

2. The ideal parent is one who shapes and plans their child's life path. This involves parents designing the blueprint and guiding their child's life journey.

3. The ideal parent is financially capable, nurturing their child in a quality social environment. This reflects the financial status of parents who consistently provide for their child in all aspects.

4. The ideal parent possesses desirable qualities because parents must exhibit good behavior as exemplary companions, both as a father and mother, avoiding any negative behaviors that could adversely impact their child.

These ideas contribute to fostering a comprehensive mindset of ideal parenthood in Thai society. They aim to convince and encourage acceptance that being a well-rounded parent means successfully raising a child according to the formulas outlined in parenting guides. The recipients of these messages are likely to trust the advice from the aforementioned speakers, thereby benefiting commercial interests and serving as a method to influence thought. This helps in creating target groups with successful market formulas that impact society. Thai society believes that children reflect their upbringing, as reflected in the Thai proverb "Look mai lhon mai klai ton" (Like father, like son) and the English proverb "The apple doesn't fall far from the tree," emphasizing that children often resemble their parents closely when raised well.

In analyzing Discourse Practice and Socio-Cultural Practice, it is found that the narrative in childcare discourse perpetuates various beliefs to the recipients. This includes creating a complete parental ideology that suggests parenting according to prescribed norms is the correct way to be a good parent. This aligns with the research by Namsopa (2019), which explains that defining the qualities of a family today may simply be creating a societal "ideology" that the mass media constructs, convincing people that conforming to these norms is natural for every family in Thai society to be considered "good" and ultimately accepted. However, in reality, the discourse of a "good family" is merely a "social ideology" constructed by Facebook creators, fan pages, and groups producing goods, solely for commercial gain. Therefore, continuous reproduction of the meaning associated with the aforementioned concept of family might ultimately influence social attitudes and become societal norms.

Beyond the aforementioned research, this study also demonstrates that a complete family must primarily consist of father, mother, and child, reflecting societal affluence and readiness in all aspects provided to the child. Parents who cannot fulfill or do not possess the characteristics of ideal parents as presented in the discourse are thus marginalized as outsiders (Out group) who lack the complete readiness for life. This situation reflects back onto the child, potentially leading to a lack of proper nurturing from a complete parent, thereby perpetuating inequality in Thai society. This discourse plays a significant role in establishing standards

for parents in Thai society, and this study aims to enlighten the audience about the intentions behind such discourse from another perspective.

7. References

Fairclough, N. (1992). *Discourse and Social Change*. Cambridge: Polity Press.

Hyland, K. (1998). Hedging in Scientific Research Articles. Amsterdam: John Benjamins.

Namsopa, P. (2019). THE PRESENTATION OF FAMILY IDEOLOGY IN FACEBOOK FANPAGE "LITTLE

MONSTER" THE DEGREE OF MASTER OF LIBERAL ARTS MAJOR IN THAI FACULTY OF LIBERAL

ARTS UBON RATCHATHANI UNIVERSITY.

Palitponganpim, P. (2021). Parent School. Bangkok: Amarin Kids.

Palitponganpim, P. (2022). 43 settler erases the messy one. Bangkok: Amarin Kids.

Palitponganpim, P. (2024). Raise your child to be a smart person. Bangkok: Amarin Kids.

Van Dijk. T.A. (1998). *Ideology: A Multidisciplinary Approach*. London, Thousand Oales, New

Delhi: SAGE Publications.

Transcending Gendered Language Stereotype:
Power Dynamics in an Interview Program by a Chinese Hostess

Cao Fan[1]

([1]Department of foreign languages and literatures, Tsinghua University, Beijing, China)

Abstract: It is assumed and concluded in divergent studies that women's languages are much more submissive in comparison to those of men. These studies have reconfirmed, accentuated, and even reinforced the stereotypes of the subservient female identity. This paper seeks to delineate the ways in which Yi Lijing, who is widely known as a trenchant and strident female journalist and interviewer in China, takes her powerful and dominant stance and position in her interviews, thereby revealing the strategies Yi and her male interviewees employ in the power dynamics of the interview. Data in the current study are transcribed from video clips of an interview program entitled *Yi's Time* by adopting fine-grained conversation analysis to present the ways in which Yi holds power over her guest male speaker and the latter concedes to the former. In conjunction with the discursive behaviors, concomitant paralanguages will also be analyzed and understood as constituting a part of the harmonious orchestration in manifesting herself as the aggressive turn-taker. As a micro examination of female languages with an individual case, this study intends to show that females can be as forceful and dominant in leading the turn and taking the floors as their counterparts in tandem with the orchestrating paralanguages. Analyzing power relations between the male interviewees and a female interviewer, this study may shed some light on gender studies and female identity construction studies by the transcendence of stereotypes in the female languages in a Chinese context. Moreover, this study may also furnish pragmatic strategies for gaining power in conversations.

Keywords: female language, power strategies, Yi Lijing, paralanguage, interview

1. Introduction

This paper investigates the power dynamics in the interaction between a hostess and her male interviewees in an interview program in the Chinese context through the approach of conversation analysis on discursive behaviors as well as concomitant paralanguages. Language has always been construed as relating to power, especially during the emergence and development of critical discourse analysis (CDA) (see for example, Wodak, 1997; Fairclough, 1995, 2003, 2015). While Giddens (1993) orients the operation of relations of power to one of the three elements in the production of interaction, Fairclough (2003) observes that power is the capacity to transform and intervene in the courses of the interaction and that social actors differ in these capacities (p. 41). Fairclough (2015) also distinguishes two aspects of power, namely power in discourse and power behind discourse. The former is that "powerful participants control and constrain the contribution of non-powerful participants" (p. 76) and the latter subsumes the standardization of language, discourse types, and access to power. With a microscopic lens through the approach of conversation analysis, this paper focuses on power in discourse.

This study draws on data from an interview program entitled *Yi's Time* hosted by Yi Lijing. According to Chinadaily, once as a senior journalist in the southern people weekly (http://www.chinadaily.com.cn/china/2011-11/09/content_14065828.htm), she has also been selected as the best journalist interviewer. Calling for "professional respect from the interviewees" (ibid.), Yi has formed her own interviewing style which is often known as being direct, professional, calm, and meticulous, thus forcefully challenging as well as countering the traditional idea of weak and less sure women's languages. Featuring interviews with both male and female celebrities and produced in 2015, these interviews of the first season with 24 celebrities (17 males and 7 females) have gained 260 million views online according to the Chinese dominant search engine *Baidu* (https://baike.baidu.com/item/易时间 /17928613?fr=aladdin). Yi has arguably impressed the Chinese audience with her relatively powerful position in her interviews, even with male guests who are traditionally deemed as inhabiting roles with higher power status in cross-gender communications. This paper will mainly focus on the discourses of Yi with her male interviewees as the contrast and conclusion will be more vibrant in cross-gender examinations of women's language. How does Yi manipulate her elicitation to occupy a more powerful position in the interview besides the control garnered from her role as the interviewer? How do the guest speaker and Yi struggle for power during the interview? Out of these concerns, this paper intends to address the power relations between Yi and her male guests and tries to engage with the specific patterns in their interviews by drawing on the approach of conversation analysis.

The remainder of this paper is organized as follows. In the next section, I'll elaborate on power in discourse as well as gender and language, after which data and the research method will be introduced. Then I'll deal with power relations and corresponding examples from the videos. The discussion section will present other modes of communication in the interview and the interplay of languages and paralanguages in power dynamics. In the last section, I'll conclude the study with its contribution and limitations.

2. Power and gendered language stereotypes

2.1 *language and power*

A body of research has addressed the relationship between language and power. For example, Tourangeau (2017) discusses the power dynamics in the discourses of law and policy debate, where he connotes that power can be articulated through discourses. Besides, Shabangu (2016) also acknowledges the power of literary language in an African novel by "laying bare the violent force of coloniality". In his seminal book, Fairclough (2015) discusses three aspects of power in discourse, namely contents, relations, and subjects. Contents are related to the things said and done; Relations denote the participants' relationships and subjects imply the positions people can occupy (P. 76). This paper focuses on the part of the content as the relationship and subject positions are relatively fixed in the context of interviews. The relationship is a professional one and the subject positions are interviewer and interviewees. This paper

will pay special attention to the contents of discourses that the interviewer and interviewees have made to examine the power relations between those discourses.

2.2 *Gendered Language and Stereotypes*

According to Mesthrie et al. (2009), there are mainly three models of gendered language. As early as 1975, Lakoff proposes the deficit model in which women's languages are marked ones compared with the unmarked men's languages. The second model is oriented towards the cultural differences between males' language and females' language which is represented by Tannen (1990). The last model is the dominance model by Zimmerman & West (1975) that the dominance of men's language influences the subservience of women's language. Though these models are different, they all entail the assumption that women's language is indirect, cooperative, hesitant, and uncertain. Later, there is a postmodern concept that gender is performed and performativity constructs the identity of one's gender. The present study subscribes to the most recent construal of gender as transcending the understanding of the overarching male language dominance.

3. Patterns of power dynamics

Three strategies of power negotiation are found. In the next part, each pattern will be exemplified. After the transcription within the tradition of conversation analysis, nine excerpts of interviews with male guest interviewees are selected and analyzed.

3.1 *Situated interrogation*

In this pattern, the interviewer situates her trenchant interrogation in a specific context related to the past experience of the male interviewee. Predominantly, these experiences are geared toward the shortcomings, weaknesses, and vulnerability of the interviewees. Through the situated experience of the interviewee, Yi "coerces" them into admitting and confessing.

In example 1, Yi is talking with Guo Jingming, a polemical writer, producer, and director with his works widely understood as being pompous and lacking literary features while claiming to be literary works. At the outset of the extract, Yi first situates the setting back to 2006's turbulence in Guo's career development, at which time he was sentenced to copying another's writer's idea into his own novel. Several years later Guo dedicated himself and his team to attempts of protecting intellectual property rights. Yi juxtaposes these two inconsistent acts and interrogates the possibility of contradictoriness. Obviously, Guo is reluctant to deal with this topic, thus demonstrating a great extent of anxiety, uneasiness, and also evasion. The backchanneling and also the self-reassurance signals such as "yeah" in lines (2), (4), and (10), "well" in lines (2) and (4), and "umm" in line (6) are all reflexive of the lower power status of Guo and his unwillingness of talking about his plagiarism history. With this situated interrogation in the context of Guo's prima fancie contradiction, Yi gains the upper hand in this power struggle. By withholding his bona fide feeling on that part of the experience, Guo utters line (4) which carries the overtone that this

experience is suffering to him by using the word "also" after Yi's demanding of whether Guo suffered from that experience.

Example 1. Interview with Guo Jingming Part 1¯

(1) Yi: In 2006, you were stuck in the disturbance of plagiarism, so is that contradictory to your current concern about protecting intellectual property rights? *(keep staring at the other and body forward)*

(2) Guo: *(drinking water while Yi is speaking)* It's not. **Well**, **yeah**.

…

(3) Yi: So, did you suffer from the plagiarism event?

(4) Guo: *(2 seconds' silence)* **Well**, I think it actually also made me grow up a lot. **Yeah**...

(5) Yi: Do you feel like it's a gift now?

(6) Guo: It's not necessarily a gift. I think it's an experience, **umm**. *(nodding head and avoiding eye contact)*

This is the same case in example 2 when Yi is interviewing Huang Xiaoming, a Chinese actor as well as a singer who was mocked nationwide because of his bad English accent of saying the phrase "not at all". Yi sets the background by deploying lines (1) and (3) to presuppose that Huang is an actor who has been trained professionally and acted for over 20 years. After situating this by two turns and Huang's confession of both being annoyed and embarrassed by the situation of denial by audiences, Yi puts Huang into question whether the declaration "I'm not a method actor" of the commercial is serious or just joking in line (9). Huang confirms and claims himself as an actor without method acting, which is ironic and frustrating to the extent that he is supposed to be quite experienced in acting after years of professional training and acting experience. Yi's strategy of situating interrogation invariably sets the clearest prerequisite for the audience to understand the situation and put the interviewees under significant pressure of dealing with things that happened years ago and have been registered in history without the possibility of erasure. In the interaction, Yi holds power over Huang discursively not only represented in leading the topic but also in the continuous and passive admittance of Huang in questions that embarrass him.

Example 2. Interview with Huang Xiaoming

(1) Yi: **You have received professional acting training?**

(2) Huang: Yes I have.

(3) Yi: **You have been an actor for more than twenty years**

(4) Huang: Yes that's right.

(5) Yi: Will you be embarrassed when others judge your acting as badly?

(6) Huang: Yes.

(7) Yi: Will you be annoyed?

¯ All examples are translated by the author from Chinese to English.

(8) Huang: Yes.

(9) Yi: When you got a commercial in 2011, there was a line in the commercial that said I*'m not a method actor*. Did you mean it when you said that, or was it just because of the need for lines of a commercial?

(10) Huang: I meant what I said at the time.

In example 3, Before asking Sun Honglei in line (2) about the low quality of his works, Yi contextualizes her statements in the fact that Sun failed in many of his attempts at roles in films, which can be a strike for an actor who strives for success in his career. Despite using the hedge "*it seems that*" to decrease the certainty of her statement, Yi still demands the feasibility of the sentence about Sun's bad works in a direct and hurting manner. Sun's paralanguages like rubbing hands in line (3) are obviously reflective of his reluctance to confess his weakness. Therefore, by using situated interrogation, Yi's grip on power in the dialogue appears more firmly.

Example 3. Interviewee: Sun Honglei

(1) *Voice-over*: (by Yi) As he (Sun) got older, he started to try out for various roles, notably in the film industry, but these endeavors did not provide the intended outcomes.

(2) Yi: But it seems that the quality of your work in recent years (*2s silence*) is low. Can I say it this way?

(3) Sun: Yes, absolutely. (*nodding head while grinning embarrassedly and rubbing hands*)

3.2 *Concession of interviewees*

Power relations also change when the interviewees step back and concede to Yi's discursive aggression. This can be manifested in terms of many aspects, such as backchanneling, implicature, also silence, pause, and mitigating devices. In example 4, when Yi is challenging Guo if he has been recognized and accepted by the professional writers' circle, Guo first back-channeled and said "Well…" in line (2). Interestingly, on the one hand, this exhibits his lack of words to express himself with a wise answer. On the other hand, he is out of his expectation that Yi would ask such a direct question. Without satisfactory responses, Yi continues to ask about the existence of this question in Guo's mind in line (3). Guo self-repairs and also stumbles to make feedback and then even pretends to show his indifference to the acceptance of the other professional writers. From the perspective of conversational maxims of Cooperative Principles by Grice (1975), Guo flouts the maxim of quantity by offering less than is expected which demonstrates the implicature of his fear of admitting that he is excluded as a non-member of the genuine writer's circle. This eschewing and uncertainty shows the concession of Guo in his power struggle with Yi.

Example 4. Interview with Guo Jingming part 2

(1) Yi: Has the writers' circle ever accepted you? In your mind.

(2) Guo: Well ......

(3) Yi: Have you ever considered this question?

(4) Guo: **I actually, I actually** don't mind this so much.

In example 5 with Chen Chusheng as the interviewee, Yi mentions two layers of information: the first one is the lawsuit that Chen involved with an entertainment company and another is the quantity of Chen's representative works. Chen uses several engagement resources (Martin & White, 2005)in his answers to Yi's question which shows subservience such as *entertain* resources "I think" in line (4), and "probably" in line (6) to lower the ascertaining tone of his statements. Hedge such as "a bit" is also employed to show the indirectness and uncertainty in Chen's concession to Yi's aggressive questions. While Yi quotes the "voice" (Bakhtin, 1981) of the third party, namely, other people, to criticize the unproductivity of Chen's song composition, he uses the *pronounce* resource (Martin & White, 2005) "actually" in line (8) to confirm and also uses the *denial* resource "not" in line (8) to admit his lack of talent in songwriting. Power relations in example 5 are obvious to the extent that Chen keeps conceding and accepting what Yi purports.

Example 5. interviewee: Chen Chusheng (a Chinese singer)

(1) Yi: who is the one that got eliminated in that lawsuit between Tianyu⎺ and you?

(2) Chen: **Well** ...... got eliminated?

(3) Yi: can I use this word?

(4) Chen: **I think** this word is **a bit** harsh.

(5) Yi: But is it the truth?

(6) Chen: It is **probably** also the truth.

(7) Yi: People are most familiar with one of your songs named *Did anyone tell you*.

(8) But what I want to say is that you have no other songs that impress people.

(9) Chen: I **actually** would like to say it is very normal. It means it's either that or I'm not good enough.

In the next example, Yi and Guo are still discussing the plagiarism turbulence and Guo seems to be unhappy to delve into that topic anymore so he exercises his power in the dialogue and proposes to change to the next theme by using the mitigating devices **let's** (Fraser, 1975; Ran, 2011) in line (1). However, Yi doesn't submit to this proposal and continues to ask if the event in 2006 has been a trauma. Guo submissively answers this question although in an equivocal way. Power transfers again to Yi when Guo tries to change the topic but fails. In Yi's question in line (2), she uses the *counter* resources "still" according to engagement theory (Martin &White, 2005) to admit her understanding of the trauma effect of the event on Guo's mind but she is curious to explore whether the trauma is still there. Guo doesn't answer it directly, instead, he uses the *entertain* resource "I think" in line (3) in the appraisal system by Martin ad

---

⎺ Tianyu is the name of an entertainment company.

White (2005) to milder his speech and also allow for spaces for Yi and other audiences' voices. This, too, is a submissive act by the interviewee: He abstains from direct response to the question that demands explicit "yes" or "no" answers. That is also to say, instead of using a categorical response, he chooses the modality to mitigate his attitude expression.

Example 6. Interviewee: Guo Jingming

(1) Guo: **Let's** not talk about that anymore. **Let's** go to the next topic.
(2) **Yi: Is it still a trauma in your mind now?**
(3) Guo: It's been many years in the past, and **I think** it doesn't make much sense to keep talking about a thing happening a decade ago.

3.3 *Provoking confirmation*

Yi's higher power can also be reinforced in her repetition of the others' words both in the statement and interrogative forms. By questioning the interviewee after repeating the utterance of the latter or the third party, Yi delivers great pressure on the addressees in explaining themselves for acts or utterances lacking logic or reason. In this way, Yi inhabits a higher power position than the addressees.

In the next example when Yi interviews Zhao Youting, a famous Chinese actor, Yi asks about his favorite role model. Zhao selects Leonardo Dicaprio to the extent that the latter can demonstrate the same acting skills in every play. Yi confirms with an interrogative sentence in line (2) by repeating Zhao's words with the implicature that it is silly and ridiculous to compliment an actor to be exactly the same in every character he plays. This kind of aggressive confirmation underscores the unreasonableness of the addressee to think in that way, thereby rendering Zhao appear silly and therefore in lower power status.

Example 7. Interviewee: Zhao Youting

(1) Zhao: Maybe my role model is Leonardo because I think he has never acted badly. Since he acts every single character in his plays exactly the same, I think it's pretty good.
(2) Yi: "To act exactly the same in every play", is it a compliment?

In example (8), Huang is coerced to accept the notorious comments when Yi uses the provoking confirmation. At the outset, Yi is asking Huang if he had any comments that he found unable to accept, then she quotes assumed netizens' comments which are typical of bad judgment. While it seems that Yi is confirming with Huang if he can accept other stigmatized expressions besides the exemplified ones like Jack Sue⁻, the underlying implicature is that Huang has already gotten accustomed to being judged especially with these typical expressions. Carrying all these labels of stigma, Huang seems to be powerless in the confirmation.

⁻ Similar to Mary Sue, a perfect and imagined male image who can get all the appreciation and love from other characters in the drama.

Example 8. Interview with Huang Xiaoming

(1) Yi: Do you have any unacceptable comments now? For example, laughing with evil impetuous, being like Jack Sue, being narcissistic, acting badly, and being silly.

In example 9, Yi and Wang⁻ are discussing Wang's attempt of protecting Chinese local dialects by organizing fieldwork and recording those varieties. In line (1), Yi holds that the purpose of Wang is to have his name remembered forever and she provokes that thought by confirming with a question. In line (2) From Wang's backchanneling such as "well" and self-repetition of the words "subjectively" and "objectively", we could see the tensity of Wang when he thought about the means for explaining this in a decent way.

Example 9. Interviewee: Wang Han

(1) Yi: what you're doing actually serves the purpose of leaving your name in history, can I understand it that way?

(2) Wang: **Well**, **subjectively** I don't feel this way that strong. **Subjectively** I just think it's interesting and fun, but **objectively, objectively** it certainly will, because no one has done this so far ...

4. Discussion

With situated interrogation in a context, the concession of the interviewee, and provoking confirmation by Yi, their power relations can be manifestly demonstrated. What deserves more attention is other modes of communication that also convey concomitant information as verbal messages. For example, in Guo's case, when he expresses himself he drinks water several times and avoids eye contact with Yi. Besides, he also touches his neck for assuring himself by saying the statement. This is related to kinetic languages or paralanguages which include gestures, facial expressions and etc. Alternatively, when Yi elicits the answers, she looks directly at the interviewees and moves her body forward to show an aggressive stance, which reinforces her verbal language, whether in the form of situated interrogation or provoking confirmation. On the other hand, the interviewees' avoidance of eye contact, neck touching, moving body backward, and frequent drinking water can all be seen as signs of curtailing the power of the speaker, thereby further indicating the concession of the interviewees.

These three processes as well as concomitant paralanguages interact with each other and reinforce the power relations.

⁻ Wang Han is a famous host in China.
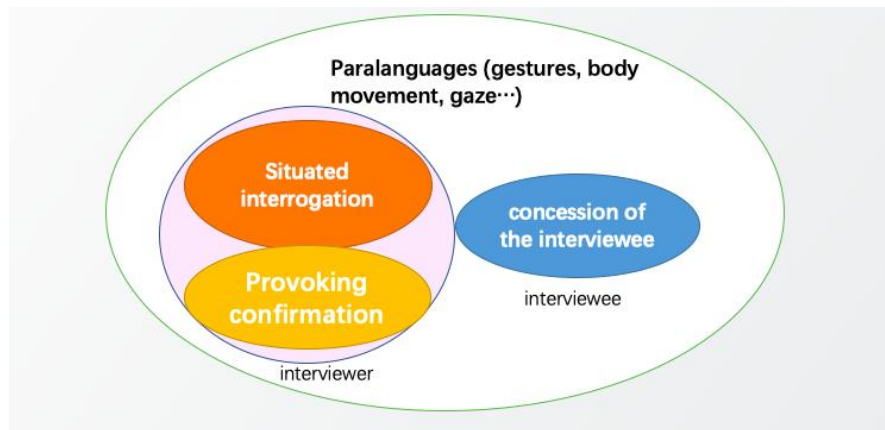
Figure 1. Power relation model in *Yi's Time*

As is shown in *Figure 1*, there is no specifically clear-cut demarcation between situated interrogation and provoking confirmation for the interviewer. We can arguably say that these two patterns work in tandem for the final aggressive impression. After the interviewee's concession, either by kinetic language or discourse, Yi will take advantage of the weak point and exercise control over the painful or negative experience of the interviewees. On the other hand, it is the situated contextualization of the negative experience of the interviewees that make them anxious and cognizant of the possibility of the revelation of being silly or embarrassed. Yi's confirmation will further strengthen her status as a powerful interviewer. To summarize, the whole process is in profound imbrication, during which other voices are entertained by Yi even though they are with no guarantee of veracity.

5. Conclusion

In this paper, I have articulated the power struggle between Yi and her guest speakers who are mainly famous male actors and singers working in the Chinese entertainment industry with an attempt to discover the strategies and processes of the powerful interview by Yi. During the analysis, I maintain that there are mainly three layers of features in this power game: situated interrogation in a context by Yi, the concession of the interviewee, and provoking confirmation by Yi. Besides these, concomitant modes of communication also contribute to the meaning-making process in the interview, which delineates varying degrees of power relations and augments with verbal language.

Being a female interviewer, Yi's language is no less powerful than her male interviewees. As we can see, women's language is not necessarily less forceful than men's language. Although the genre of figure interview is slightly different from everyday interaction, it is still a face-to-face spontaneous talk and interaction. Both the interviewer and the interviewees share the consensus of saving the other party's positive face, as a result, even though there are constant power struggles, there is no conflict formation, which is of course beyond the scope of this paper. Meanwhile, paralanguages are also implicated in the meaning-making of power relations, which enables the accentuation of the curtailing or strengthening of power processes.

This paper may shed some light on gender studies and power relation studies. The strategies discovered can also inspire further studies to have a more in-depth and wider range of interview studies. Due to the researcher's interest, only one perspective, that is, power relations regarding gendered languages, has been studied, which is far from enough. As a result, further studies can explore more aspects of the data from divergent perspectives, thus enriching the studies of the very characteristic interviews by female hosts. Meanwhile, there are also implications for researchers in the orchestration of various modalities and discourse strategy studies.

References

Bakhtin, M. M. (1981). *The dialogic imagination* (C. Emerson & M. Holquist, Trans.; M. Holquist, Ed.). The University of Texas Press.

Fairclough, N. (1995). *Critical discourse analysis*. Longman.

Fairclough, N. (2003). Analyzing *discourse: Textual analysis for social research*. Routledge.

Fairclough, N. (2015). *Language and power*. Longman.

Fraser, B. (1975). Hedged performatives. In P. Cole, & J. L. Morgan. (Eds.), *Syntax and Semantics, Vol. 3, Speech Acts* (pp. 187-210). Academic Press.

Giddens, A. (1993). *New rules of sociological method*. Polity Press.

Grice, H. P. (1975). Logic and Conversation. In P. Cole, & J. L. Morgan. (Eds.), *Syntax and Semantics, Vol. 3, Speech Acts* (pp. 41-58). Academic Press.

Lakoff, R. (1975). *Language and woman's place*. Harper & Row.

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave.

Mesthrie, R., Swann, J., Deument, A., & Leap, W. (Eds.) (2009). *Introducing* sociolinguistics, 2nd edition. Edinburgh University Press.

Ran, Y. (2012). Rapport orientation of mitigators and their interpersonal pragmatic functions. *Contemporary Foreign Language Studies*(11), 4-10.

Shabangu, M. (2016). The Outsider Within: An Analysis of the Representation of Power and Language in Tayeb Salih's Season of Migration to the North. *Eastern African Literary and Cultural Studies*, *2*(1-2), 45-53.

Tannen, D. (1990). *You just don't understand: Women and men in conversation*. William Morrow.

Tourangeau, W. (2017). Debating Bill C-18: An analysis of power and discourse in parliamentary proceedings on Canada's Agricultural Growth Act. *Journal of Canadian Studies*, *51*(2), 485-519.

Wodak, R. (1997). Critical discourse analysis and the study of doctor-patient interaction. In Gunnarsson, B., Linell, P., and Nordberg, B. (Eds.), *The construction of professional discourse* (pp. 173-200). Longman.

Zimmerman, D. H. and C. West. (1975). Sex roles, interruptions and silences in conversation. In Thorne, B. and N. Henley (eds.) (1975). *Language and sex: Difference and dominance*. Rowley, MA: Newbury House.

Analysis of Term Presentation Strategies and Influencing Factors in Chinese Translation Teaching Materials

YANJI CHEN

University of Pompeu Fabra

Abstract:

Nowadays, effective terminology translation has become one of the most popular topics in translation studies since it is essential for technology diffusion. However, Spanish undergraduate studies in China have included few subjects related to terminology (Liang, 2006:24). Therefore, this paper attempts to understand the representation of Spanish terminology and translation in Chinese universities by counting and evaluating the terminological content in the selected translation textbook with the aim of contributing to the future training in both terminology and translation for Chinese students. The study is based on the curriculum standards of foreign language university studies in China, the development of terminological competence of translators and the communicative theory of terminology. 《汉西翻译教程》(*hàn xī fānyì jiàochéng*, *Chinese to Spanish Translation Manual*) has been selected as the object textbook. We examined its content in order to create a terminology database. Based on this corpus, the analysis of this study reveals the insufficiency of terminological content in the Spanish translation textbook, leading to the reflection that, at the undergraduate level, we should introduce terminological concepts and their corresponding translations into the curriculum to ensure that students can systematically learn and apply relevant terms and develop their terminological competence.

Keywords：Terminology instruction; Translator training; Teaching materials

Table of Contents

2

1. Introduction

As a linguistic crystallization of knowledge, terms are fundamental expressions of knowledge in a specific field. Effective terminology translation has become one of the hot topics in translation studies because it is crucial for the diffusion of technology (Wang, 2011:68). Moreover, there are various ways to represent the real world in different languages and cultures, highlighting the growing importance of terminological translation in the international arena. Therefore, for translators, knowledge of terminology is essential for effective work.

However, in translator training in China, there are generally few courses related to terminology (Liang, 2006:24), and this pattern is more or less the same for Chinese-Spanish translation. For example, after reviewing the curriculum of Beijing University[1], no training related to terminology is mentioned, and learning of terms is usually carried out within the translation course. When undertaking Chinese-Spanish translation tasks, we have noticed problems due to the lack of terminology knowledge, such as difficulties in term management, inconsistencies in the use of the same term within the same document, and lack of precision in the translation of terms, among other aspects.

Therefore, this work focuses on the translation teaching materials used in China, seeking to delve into the term presentation strategies and analyze the factors influencing these choices. This area of study is not only vital for understanding the pedagogy of translation but also reflects cultural and linguistic trends in language teaching. Through this study, we aim to contribute to the development of more effective teaching materials and the enrichment of pedagogical practices in the field of translation.

---

[1] The website of Beijing University: <https://deyp.sfl.pku.edu.cn/jxpy/xbyyzyjxdg/1229131.htm>.

3

2. Literary Review

2.1 Translator Training and Terminology Teaching

The primary goal of translator training is to cultivate professionals with excellent competencies and professionalism to meet the societal demand for translation services. Cabré (2004:2) emphasizes the close relationship between translation and terminology in both theoretical and practical realms, noting that they share numerous commonalities. Consequently, the teaching of terminology and translation is intrinsically linked in academia. Also, Patricia García (2007) suggests that the purpose of using terminology in translator and interpreter training is for students to view it as a tool that enables them to perform three basic tasks.

In other words, terminology becomes an indispensable tool for translators and interpreters, allowing them to manage and use terminology effectively, ensuring terminological coherence in their work, and facilitating precise and appropriate communication in a professional setting. Therefore, it is crucial to establish a system of terminology teaching applicable to students and professionals in translation, which includes curriculum design, training of qualified teachers, and the selection of appropriate teaching materials and methods.

2.2 Educational Standards in China

The curriculum standards for foreign language studies in Chinese universities aim to provide a comprehensive framework for language instruction, including translation training. However, a detailed examination of these standards reveals that terminology instruction is often underemphasized. Previous research on terminology instruction and translator training has provided valuable insights into the current state of education in this field, particularly within the context of Spanish language programs in China. Liang (2006) conducted a seminal study that highlighted the lack of dedicated terminology courses in translation programs. Liang's findings suggest that the integration of terminological training into existing translation courses is insufficient for developing the necessary skills and knowledge. Subsequent studies,

4

such as those by Wang (2011) and Zhang (2015), have further explored the challenges faced by Chinese universities in implementing effective terminology instruction. These studies point to a range of issues, including inadequate teaching materials, insufficient instructor expertise, and a lack of standardized terminology resources. This body of research underscores the urgent need for a more structured and systematic approach to terminology instruction within translator training programs.

2.3 The Role of Didactic Translation Materials

Translation textbooks play a crucial role in teaching, not only for students but also for teachers. As significant support for the development of translation teaching, translation teaching materials are key to enhancing the training of highly competent translators (Long and Gong, 2023:93). The organization of content in translation learning manuals reflects the teaching philosophy of a given period (Wang, 2011:72). In their analysis of the current state of translation construction in China, Jingjing Long and Qi Gong (2023:94) mention:

> "Translation textbooks are not only a vehicle for implementing the objectives and program of a translation course but also an important means for students to understand and master the theory, practice, and technology of translation."

Therefore, it is necessary to observe the presentation of content in translation teaching materials to understand the characteristics of Spanish translation courses in Chinese universities. Examining the content organization and chapter structure of a teaching manual provides insight into how the content is organized and the areas of focus. The arrangement and sequencing of terminology teaching in a textbook can reveal the learning emphasis and difficulties students encounter at different stages of learning. Additionally, examining how terminology is taught and presented in the textbook, such as whether it contains glossaries, explanations and definitions, example

5

sentences, and contextual analyses, helps students understand and apply the terms effectively.

3. Methodology

3.1 Collection of Translation Teaching Materials

Teaching materials can be divided, based on the type of translation, into materials for interpreters and materials for translators. In this study, we focus on translation textbooks. It is noteworthy that Spanish courses in China are offered both in general universities and specialized institutions, such as business and law colleges. Thus, we chose to focus on translation materials for courses in general universities rather than specific sectors. Through a general search for such textbooks via popular online shopping applications in China like Taobao, Jingdong, and Dangdang, and consulting publishers, we discovered that most translation teaching materials are distributed and published by two main publishers: Foreign Language Teaching and Research Press (FLTRP) and Shanghai Foreign Language Education Press (SFLEP). We then conducted a specific search on the publishers' websites to collect information on translation textbooks.

On the publishers' websites, we recorded seven translation teaching materials, comparing the following aspects to select the most suitable for analysis:

1. Scope and Objectives of the Textbook: Whether the textbook is relevant to the needs and goals of Spanish students.

2. Structure and Organization: Whether the chapters and linguistic points are well-organized and user-friendly for teachers and students.

3. Content: Whether the chosen translation topics are comprehensive and include terminology translation.

4. Update and Development: The publication date of the textbook, whether it contains updated content, and if it keeps pace with the times and the development of the translation subject.

6

These are 7 textbooks:

Teaching Materials	Author(s)	Publisher	Year of Publication
《理解当代中国：汉西翻译教程》 *Course in Chinese to Spanish Translation (Understanding Contemporary China)*	Man Yu, Peng Zhang, Yaxing Tong, Xing Chen, Xiaoke Li, Lei Chen	FLTRP	2022-07
《理解当代中国：高级汉西翻译教程》 *Advanced Course in Chinese to Spanish Translation (Understanding Contemporary China)*	Man Yu, Fulaing Chang, Ke Zhang, Wei Cao, Zheng Ma, Menglu Zhang	FLTRP	2022-07
《汉西翻译教程》 *Chinese to Spanish Translation Manual*	Jianzhong Li, Ke Zhang	FLTRP	2021-09
《汉西互译教程》 *Course of Bidirectional Translation between Spanish and Chinese*	Jicheng Meng	SFLEP	2019-09
《西汉翻译理论与实践》 *Course of Spanish to Chinese Translation: Theory and Practice*	Kai Zhu	University of International Business and Economics Press	2013-09
《西汉翻译教程》 *Course of Spanish to Chinese Translation*	Li Sheng	FLTRP	2011-08
《新编西汉翻译教程》 *New Course in Chinese to Spanish Translation*	Shiyu Zhang	FLTRP	1999-07, new edition published in July 2010

Table 1: List of Translation Teaching Materials

3.2 Construction of the Terminology Corpus

In this project, our goal is to analyze the didactics of terms in translator training. After comparing the seven textbooks mentioned above, we selected the most suitable one

7

and extracted its terminological content to create a corpus of terminological units.

First, to create this corpus, we selected texts from the section of brief exercises for subsequent screening, aiming to select texts with a relatively high level of specialization for detailed analysis of the terminological content. For texts without a clear source, we manually evaluated, selected, and classified them based on the type of unit presented (based on the textbook's original five themes), the audience of the text (public or professional), the context in which it was used, and the degree of linguistic formality (legal contracts, scientific texts, press releases, popular science texts, literature, etc.).

Next, to identify and extract terms from the selected texts, we manually selected the terms, recorded information about the term in Chinese, its Spanish translation in the book, the literal translation of the term, its pinyin, context, source, whether there was an explanation in the text, and confirmed it on two terminology websites: 术语在线 (*shùyǔ zàixiàn, Termonline*) and 中国特色对外翻译标准术语库 (*zhōngguó tèsè duìwài fānyì biāozhǔn shùyǔkù, Base of Data Terminological Normalized for Foreign Language Translations with Chinese Characteristics*).

Beyond the information about the term in Chinese, we also classify them according to the discipline in which they are applied, the industrial sector, and the thematic unit (e.g., political unit) in which they appear to facilitate subsequent statistics and analysis. Additionally, to facilitate reading and understanding, we need to extract the Spanish translation of each terminological unit in the corresponding paraphrase for research and interlinguistic communication. Lastly, we also recorded the source of each term, i.e., the title of the article it came from, to facilitate tracking and information verification.

To explore the frequency of these terms and determine their prevalence, we consulted

8

them using the BLCU Chinese Corpus (BCC, where we can confirm the use of terms in Chinese based on real corpus data.

4. Analysis

In comparison with the seven textbooks that we collected, 《汉西翻译教程》(*hàn xī fānyì jiàochéng*, *Chinese to Spanish Translation Manual*) is the most recently published, covering various aspects and providing a correspondence between the original text and the corresponding translation. In this regard, we believe it can reflect the current changes in the translation of Chinese terminology to some extent. Therefore, this textbook is chosen for detailed statistical and terminological analysis in this study.

4.1 Corpus Statistics

In this section, we analyze the detailed content of the terminology from the second part of the textbook focusing on the 32 short texts which includes total 10,485 characters. For the final selection, we chose 15 articles, totaling 5,230 characters. Among them, we found 5 articles belonging to the political unit, 4 to the economic unit, 2 to the social unit, 3 to the cultural unit, and 1 to the literary unit, as these articles have a relatively high level of specialization and terminological content.

From the 15 articles, a total of 158 terminological units were extracted, of which 56 correspond to the political unit, 36 to the economic unit, 24 to the social unit (covering environmental and political contexts), 34 to the cultural unit (derived from philosophical, sports, and medical contexts), and 8 to the literary unit. Regarding the translation of terms, 129 terms are direct translations, and 29 are indirect translations.

4.2 Reclassification of Terms

To better analyze the extracted terms, we divided the terms in this didactic material into two groups based on their characteristics: original Chinese terms and non-original

9

Chinese terms. Specifically, the first group refers to terms that have meanings and concepts primarily derived from China's specific cultural, historical, and social background. These terms are widely used in the original Chinese text to convey specific Chinese ideas, values, and practices. For example, the term "和平共处五项原则" (hépíng gòngchǔ wǔ xiàng yuánzé, the Five Principles of Peaceful Coexistence). The second group refers to concepts that already exist in Western contexts and may already have equivalents in Spanish. These can be communicated through reciprocal translation or substitution of similar concepts, such as the term "温室效应" (Wēnshì xiàoyìng, the greenhouse effect).

Based on our observations, out of 158 terms, we found 46 terminological units that can be identified as original Chinese terms, representing approximately 30% of the total number of terms in the constructed corpus, of which 17 terms did not use the form of direct translation.

Among them, 33 terms can be identified as political terms, such as "一带一路" (yī dài yī lù, the Belt and Road Initiative), 2 terms from the environmental field, including "生态文明"(Shēngtài wénmíng, ecological civilization) and "落后产能" (luòhòu chǎnnéng, outdated production capacity), 4 terms specific to the context of Confucianism, for example, "取义舍利" (qǔ yì shě lì, uphold justice while disregarding personal gain), 1 term from the sports field, which is "体育强国" (tǐyù qiáng guó, a sports power), and the remaining 6 terms from the field of Chinese medicine, such as "气" (Qì, Qi/energy).

In general, the export of concepts related to China is significant in the translation textbook, with political terms occupying a larger proportion of Chinese terminological production. In this textbook, political terms are often related to governance, foreign relations, and legal systems, such as China's foreign policy and fertility policy. Moreover, as globalization continues, political exchanges and cooperation are becoming more frequent. In China, at the university level, courses related to the study

10

of Chinese politics are offered as part of the curriculum. The export of Chinese political terminology can meet the need of the Spanish-speaking world to understand Chinese political concepts and promote intercultural exchange and cooperation.

4.3 Case study: Translation of the Terms

Example 1: "风险" (fēngxiǎn)

In the economic unit, we find a translation mismatch in the text *Contracto* containing the term "风险" (fēngxiǎn), which is not translated directly but as "pérdida" (loss).

The term "风险" (fēngxiǎn) is defined in the field of economics as an event that may occur in the future and can be excluded through insurance and foresight. Its English translation, recorded in *Termonline* as "risk", corresponds to the Spanish word "riesgo", which denotes a potential danger or uncertainty that could lead to an unfavorable outcome or loss, emphasizing uncertainty. In contrast, the Chinese equivalent of "loss" is "损失" (sǔnshī), mainly referring to material or financial damage, which has a different meaning from "风险" (fēngxiǎn).

In this textbook, the context of this article is the conclusion of a commercial contract, where the term "风险" (fēngxiǎn) appears in the phrase "双方按其注册资本金额共同分享利润和承担风险" (ambas partes comparten ganancias y asumen pérdidas de acuerdo con la suma de su capital registrado, both parties share profits and assume losses according to the amount of their registered capital). In this case, "风险" (fēngxiǎn) can be understood as the loss of benefits and is used to correspond to gains.

Given the communicative nature of terminology, translating "风险" (fēngxiǎn) as "loss" for the economic context is feasible. However, this translation choice can cause confusion among students, as they might equate "风险" (fēngxiǎn) with "loss" instead of "risk" and apply it in other contexts. When the context changes, for example, in BCC, we find the word "风险" (fēngxiǎn) in the context of an environmental and

11

health article in the phrase "大气质量对人体健康的风险评估结果" (risk assessment results of air quality on human health). Here, "风险" (fēngxiǎn) is not equivalent to "loss" since it indicates uncertainty about possible health-related issues, so it should be translated as "risk".

Example 2: "针灸" (zhēn jiǔ)

The term "针灸" (zhēn jiǔ) in the text *Terapéutica acupuntural* has been translated as "acupuncture". However, according to Yao Yan and Minkang Zhou (2021), the term "针灸" (zhēn jiǔ) refers to two Chinese treatment techniques: "针" (zhēn, acupuncture) and "灸" (jiǔ, moxibustion). Specifically, acupuncture is a treatment method involving the application of various needles to meridians, acupuncture points, or other body parts, used for various painful conditions, sensory disorders, movement disorders, and dysfunctions. Conversely, moxibustion is a health method that uses burning moxa or other heat sources to burn or toast acupuncture points or lesions to warm meridians, harmonize qi and blood, and help dispel evil. In *Termonline*, we found that its corresponding English translation is consistently "acupuncture and moxibustion", so in Spanish, it should be translated as "acupuntura y moxibustion".

This example demonstrates that ignoring the true meaning behind the term "针灸" (zhēn jiǔ) leads to inaccurate transmission of its meaning. The lack of detailed explanations and descriptions of the two techniques included in the term "针灸" (zhēn jiǔ) in textbooks can also result in students learning the term without a clear understanding of what it entails, leading to translations that do not include both techniques.

5. Conclusion

This paper's analysis of the Chinese syllabus and the terminology in the translation teaching manual reveals that the translation materials contain limited and relatively simple terminological content, suitable for learners' levels and needs. However, the lack of attention to terminology in education is evident in the absence of

12

terminology-related courses in Chinese university curricula and the inadequate presentation of terminology in translation textbooks, which lack explanations, definitions, and examples.

The possible causes for this situation include that first, Chinese university curricula emphasize language skills such as grammar and reading, with less attention to terminology teaching. Many students in Spanish courses lack a basic level of Spanish, hindering their ability to translate and study professional texts, leading to inadequate terminology didactics and practical application. In addition, terminology is still an emerging discipline in China, resulting in a lack of systematic organization and updating of terminology in textbooks and teaching practices.

In conclusion, to improve translation teaching and translator training, as well as to promote communication and intercultural exchange, it is essential to emphasize terminology teaching and research. Introducing terminology concepts and translations into undergraduate curricula will help students systematically learn and apply relevant terms, enhancing their terminological competence. Textbooks should include more specialized articles with clear explanations and examples of terminology. Additionally, encouraging terminology research and organization will ensure that translation textbooks are scientifically and systematically updated.

Reference

Ailin Liang (梁爱林). (2006). 关于我国翻译教学中术语学培训体系的建设问题. (On the Problems of Establishing China's Terminology Training System within the Framework of Translation Teaching). *中国科技术语*. *(China Techology)*. 8(3), 24-27.

Base of Data Terminological Normalized for Foreign Language Translations with Chinese Characteristics 中国特色话语对外翻译标准化术语库 (zhōng guó tè sè huà yǔ duì wài fān yì biāo zhǔn huà shù yǔ kù). Retrieved from http://www.cnterm.com/index/index.jsp

Cabré, Maria Teresa. (2004). La terminología en la traducción especializada (Terminology in Specialized Translation). Gonzalo García, Consuelo; García Yebra, Valentín (eds.) *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros. p. 89-122. ISBN: 84-7635-578-5. (CL).

Ces, P. G. (2007). Terminología y terminótica en la formación de traductores e intérpretes en Argentina (Terminology and Terminotics in the Training of Translators and Interpreters in argentina). *Panace*, *9*(26), 158-161.

Española, R. A. (2014). Diccionario de la lengua española. Madrid: Espasa. Retrieved from https://www.rae.es/

Huang, W. (2014). Inicación a la traducción inversa: una propuesta comunicativa para la didáctica de la traducción del chino al español en China (Introduction to Reverse Translation: A Communicative Proposal for the Didactics of Chinese

14

to Spanish Translation in China). Universitat Autònoma de Barcelona.

Hui Zhang（张慧）. (2020). 西班牙语翻译教学在数字化时代的转型. (Transformation of Spanish Translation Teaching in the Digital Age). *中国科技翻译 (Chines Science & Technology Translators Journal)*, *33*(3), 27-30.

Jianzhong Li y Ke Zhang (李建忠&张珂). (2021). 汉西翻译教程（新经典高等学校西班牙语专业高年级系列教材），Chinese to Spanish Translation Manual). Beijing: Foreign Language Teaching and Research Press.

Jicheng Meng (孟继成). (2019). 西汉互译教程 (Course of Bidirectional Translation between Spanish and Chinese). Shanghai: Shanghai Foreign Language Education Press

Jingjing Long & Qi Gong (龙晶晶&宫奇). (2023).我国翻译教材建设反思：现状、问题与对策.（Reflections on Developing China's Translation Textbooks: Current Situations, Problems and Solutions) *外语电化教学 (Technology Enhanced Foreign Language Education)* (01),93-99+118. doi: CNKI: SUN: WYDH.0.2023-01-016.

Kai Zhu (朱凯). (2013). 西汉翻译理论与实践 (Course of Spanish to Chinese Translation: Theory and Practice). Beijing: Prensa de la Universidad de Negocios y Economías Internacionales.

Li Sheng (盛力). (2011). 西汉翻译教程（第二版）(Course of Spanish to Chinese Translation). Beijing: Foreign Language Teaching and Research Press.

Lin, J. (2022). Enseñanza de la lengua española en China: un análisis contrastivo de los manuales didácticos marco ELE (Teaching Spanish in China: A

15

Contrastive Analysis of ELE Teaching Manuals). *Revista de Didáctica Español Lengua Extranjera*, (35), 1-18.

Man Yu, Fulaing Chang, Ke Zhang, Wei Cao, Zheng Ma y Menglu Zhang (总主编：于漫；主编：常福良；副主编：张珂；编者：曹韦 马征 张梦璐). (2022). 高级汉西翻译教程（"理解当代中国"西班牙语系列教材）[Curso avanzado de traducción del chino al español (entender la China contemporánea), Advanced Course in Chinese to Spanish Translation (Understanding Contemporary China)]. Beijing: Foreign Language Teaching and Research Press.

Man Yu, Peng Zhang, Yaxing Tong, Xing Chen, Xiaoke Li y Lei Chen (总主编：于漫；主编：张鹏；副主编：童亚星；编者：陈星 李晓科 陈磊). (2022). 汉西翻译教程（"理解当代中国"西班牙语系列教材）[Curso de traducción del chino al español (entender la China contemporánea), Course in Chinese to Spanish Translation (Understanding Contemporary China)]. Foreign Language Teaching and Research Press.

Shaoshuang Wang (王少爽). (2011).面向翻译的术语能力：理念、构成与培养.(Translation-oriented Terminology Competence: Concept, Composition and Development). *外语界 (Foreign Language World)* (05),68-75. doi: CNKI: SUN: WYJY.0.2011-05-012.

Shiyu Zhang (赵士钰). (2010). 新编汉西翻译教程（新）Nuevo curso de traducción del chino al español (New Course in Chinese to Spanish Translation). Beijing: Foreign Language Teaching and Research Press.

16

Termonline 术 语 在 线   (shù yǔ zài xiàn) Retrieved from

https://www.termonline.cn/index

Universidad de Lengua y Cultura de Beijing (北 京 语 言 大 学). (2019). Beijing

Language and Culture University Corpus Center (BCC) 北京语言大学语料

库中心(BCC). Retrieved from http://bcc.blcu.edu.cn

Yao Yan & Minkang Zhou (严瑶&周敏康). (2021). 基于对《中医基本名词术语中

西对照国际标准》研究的中医术语西班牙语翻译分析 (Analysis of the

Spanish Translation of TCM Terms Based on the Study of the International

Standard for the Comparison of Basic TCM Terminology between Chinese

and Spanish). *中国科技术语 (China Terminology)* (03),68-74.

Develop a Second Language Grit Scale in a Group-Oriented Societal Context: Findings Based on Machine Learning Methods

Kenan Gao [1,2], Juan Zhang [1,2*], Yihui Wang [1,2], Chenggang Wu [3,4], and Wei He [5]

[1] *Faculty of Education, University of Macau, Macau, China*

[2] *Centre for Cognitive and Brain Sciences, University of Macau, Macau, China*

[3] *School of Education, Shanghai International Studies University, Shanghai, China*

[4] *Key Laboratory of Multilingual Education with AI, School of Education, Shanghai International Studies University, Shanghai, China*

[5] *Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Taipa, Macau, China*

Abstract

Previous studies indicated that in collective societies prioritizing group values (e.g., China, Japan, and Korea), the Triarchic Model of Grit, which includes perseverance of effort, consistency of interests, and adaptability to situations, is particularly applicable. This research examined the role of grit in acquiring a second language (L2) in such a societal setting (i.e., China) and its link to L2 learning factors among 459 undergraduate students learning English as an L2. An L2 grit measurement, the Second Language Grit Scale in a Group-Oriented Societal Context (L2GSGSC), was designed and validated to assess the L2 domain-specific grit of students. Results based on hierarchical regression and machine learning methods showed a positive association between L2 grit and students' enjoyment and engagement in L2 learning, surpassing the correlations observed with the Triarchic Model of Grit Scale, which is developed to examine general trait grit. The findings underscore the effectiveness of L2GSGSC for societies with collective values.

Keywords: L2 grit; Triarchic Model of Grit; Second language acquisition; Group-oriented societies; Machine learning

1 Introduction

Grit is defined as a fundamental and enduring trait reflecting a person's unwavering commitment and effort toward achieving long-term goals, despite obstacles, challenges, or slow progress (Duckworth et al., 2007). Grit is divided into two components: perseverance of effort (PE) and consistency of interest (CI) (Duckworth et al., 2007). PE refers to an individual's sustained dedication and effort in the face of challenges, while CI pertains to maintaining long-term enthusiasm and focus on a specific interest or goal.

Datu et al. (2017) expanded on this concept by introducing the Triarchic Model of Grit Scale (TMGS), which measures an individual's determination to pursue long-term goals within a collectivist context. In collectivist cultures, individuals prioritize group perspectives and objectives over personal ones, contrasting with individualistic cultures that emphasize personal achievement (Littlewood, 1999). The TMGS thus highlights not only PE and CI but also the adaptability to situations required to navigate varying social, situational, and environmental contexts.

While the general concept of grit has been extensively studied, some researchers suggest that grit should be viewed differently across specific fields rather than as a uniform trait (Duckworth & Quinn, 2009; Teimouri et al., 2020). In the context of second language (L2) acquisition, two L2-specific grit scales have been developed. Teimouri et al. (2020) validated a 9-item L2 grit scale in Iran, confirming the significance of PE and CI. Similarly, Alamer (2021) created a 12-item L2 grit scale in Saudi Arabia, affirming grit as a unified construct with distinct PE and CI components.

This study aims to conceptualize, develop, and validate a new L2 grit scale tailored to collectivist cultural contexts (L2GSGSC), based on the TMGS framework by Datu et al. (2017). It posits that L2 grit encompasses perseverance of effort, consistency of interest, and situational adaptability within second language acquisition. Additionally, this research will explore the impact of this nuanced L2 grit model on L2 learner behaviors.

2 Method of phase 1

Phase 1 was to identify the items of the L2GSGSC and assess its factor structure and score reliabilities.

2.1 Participants

The study utilized online surveys and convenience sampling to recruit 222 participants (79.7% male, 20.3% female) from two China's mainland universities. The majority of participants (99.1%) were aged 18 to 25, with 0.9% being older than 25.

2.2 Data collection procedure and analysis

A comprehensive review of general and language learning grit measures was undertaken. The final selection comprised 15 items from both the L2GSGSC and TMGS scales. The principal researcher selected these items, which were then reviewed by other researchers to ensure they accurately assessed L2 learning grit among native Chinese learners. Internal consistency was evaluated using Cronbach's alpha (α), with descriptive statistics and item-total correlations determining item reliability. An exploratory factor analysis (EFA) was performed using principal component extraction and promax rotation in SPSS 26 to identify items that effectively represented the intended constructs.

3 Results of phase 1

3.1 The preliminary scale

A preliminary 15-item scale was evaluated by an expert panel to ensure content validity. Inter-rater agreement was assessed using intraclass correlation coefficients (ICC; two-way mixed model, mean of 5 raters, absolute agreement). After removing two items for limited content relevance, the 13-item scale achieved an ICC of 0.87. The initial L2GSGSC used a 5-point Likert scale, ranging from 1 (Not like me at all) to 5 (Very much like me).

3.2 Factor structure and reliability

Initially, we calculated the mean, standard deviation, and correlations between individual items and the total score for the initial L2GSGSC to identify which items should be retained. The results showed that most items had similar means, appropriate

dispersions, and item-total correlations. Notably, none of the item-total correlation values were below 0.4. Therefore, no items were removed at this stage (Field, 2024).

Next, a factor analysis was conducted on the remaining 13 items in the L2GSGSC. The Kaiser-Meyer-Olkin (KMO) value was 0.79, indicating that the data was suitable for factor analysis as per the criteria set by Sofroniou and Hutcheson (1999). The examination of Eigenvalues revealed three factors, with values of 4.37, 1.99, and 1.18, respectively. Items on perseverance of effort, consistency of interest, and adaptability to situations in language learning constituted the first, second, and third factors, respectively. These factors collectively explained 58.01% of the L2 grit score variance, with all factors showing satisfactory loadings on their respective components.

We then calculated the Cronbach's alpha reliability coefficients for each subscale, resulting in α = 0.81 for perseverance, α = 0.65 for consistency, and α = 0.77 for adaptability. These values indicate robust internal consistencies, ranging from acceptable to good (Black et al., 2010).

4 Method of phase 2

Phase 2 was to further enhance the structural and construct validity of L2GSGSC and assess the external validity.

4.1 Participants

A new convenient sample of 237 university students (male: 75.9%, female: 24.1%) from China's mainland was utilized. The majority of participants (99.2%) were aged 18 to 25, with only 0.8% being over 25.

4.2 Data analysis

The goodness of fit for the 13-item version of L2GSGSC was evaluated using maximum likelihood estimation in CFA with AMOS version 23.

First, we examined the relationships between L2 grit subscales and general grit, conscientiousness, and language growth mindsets to gain preliminary insights into the nomological network of TMGS.

To assess the adaptability of L2GSGSC to the L2 learning environment, we conducted hierarchical regression analyses to evaluate how well the dimensions of L2

learning grit predicted relevant L2 variables, such as L2 learning engagement and enjoyment, using SPSS version 26. Additionally, we employed three different models using ML methods to further examine the adaptability of L2GSGSC to the L2 learning environment. ML technology excels in processing and analyzing large volumes of data, thereby enhancing prediction accuracy and learning from language-related data to forecast outcomes (Rezaii et al., 2019).

4.3 Measures

General grit. We used the Chinese adaptation of the TMGS developed by Datu and Zhang (2021) to evaluate individuals' general grit.

Conscientiousness. We used the conscientiousness aspect of Big Five personality traits by employing the NEO Five-Factor Inventory (NEOFFI) (Costa & McCrae, 1992) adapted for Chinese adult participants, as previously used (江逸凡, 2020).

Growth mindset. We used a six-item scale adapted from Dweck's seminal work (Dweck, 2013), as applied by Yao et al. (2021) to assess students' growth mindset toward their English learning.

Foreign language engagement. We used the Foreign Language Engagement scale (Li et al., 2023).

Foreign language enjoyment. We used Chinese Version of the Foreign Language Enjoyment Scale (CFLES) (Li et al., 2018).

5 Results of phase 2

5.1 CFA results

In the proposed measurement model, L2GSGSC was conceptualized with three latent factors representing the sub-dimensions of perseverance, consistency, and adaptability. These factors were indicated by 5, 4, and 4 observed indicators, respectively.

To evaluate the model fit, we applied established criteria for acceptable fit (Schermelleh-Engel et al., 2003): 1) a non-significant χ^2; 2) Comparative Fit Index (CFI), Goodness of Fit Index (GFI), and Tucker-Lewis Index (TLI) exceeding 0.90; and 3) Root Mean Square Error of Approximation (RMSEA) below 0.08. The initial CFA revealed suboptimal fit indices: $\chi^2 = 220.01$, df = 62, $p < 0.001$, CFI = 0.89, GFI

5

= 0.88, TLI = 0.89, and RMSEA = 0.10 (CI 90% = 0.089 - 0.12). A review of the modification indices (MI) suggested that including Item 9 in the consistency of interests was detrimental to the model fit (MI = 35 > 12.0).

Following the methodological framework presented by Steger et al. (2006), Item 9 was omitted, leading to another CFA with the revised 12-item version of the L2GSGSC. This revision yielded improved and acceptable fit indices: χ^2 = 148.25, df = 51, p < 0.001, CFI = 0.93, GFI = 0.91, TLI = 0.93, and RMSEA = 0.09 (CI 90% = 0.073 - 0.107) (Fig. 1).

Figure 1. Outcome of confirmatory factor analysis in Phase 2.
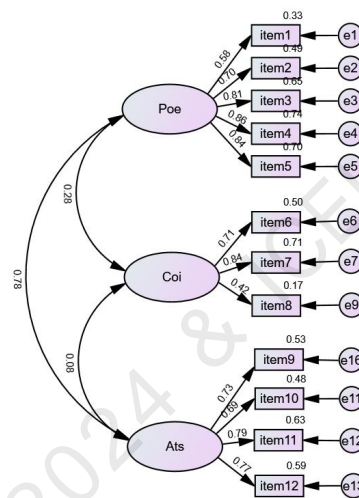
Following refinement, the reliability of the enhanced L2GSGSC is as follows: the overall scale exhibits α of 0.87, with sub-dimension α of 0.87 for perseverance, 0.66 for consistency, and 0.83 for adaptability. These values indicate an acceptable level of reliability for the scale.

Correlation analyses showed that perseverance and adaptability were positively associated with the perseverance and adaptability of general grit, growth mindset, and conscientiousness. However, consistency did not correlate with adaptability in TMGS or with the growth mindset.

5.2 Hierarchical regression results

The outcomes of hierarchical regression analyses demonstrated that L2 grit had a strong positive effect in both models, that is, it was an important predictor of L2 engagement and enjoyment. This connection persisted even when considering demographic characteristics and trait grit. In addition, consistency and adaptability

also show significant positive effects in Model 1. The introduction of L2 grit dimensions into the regression models highlighted additional variance (ranging from 4.7% to 29%) in the mentioned outcomes, emphasizing the substantial contribution of L2 grit in predicting L2 learning engagement and enjoyment.

5.3 Machine learning results

In two tasks of predicting L2 enjoyment and L2 engagement, employing L2GSGSC proves to be more advantageous than using TMGS, the specific results can be found in Table 1.

Table 1 Model Performance Comparison in Predicting L2 Enjoyment and L2 Engagement Using Data from L2GSC and TMGS

Model	Metric	L2 Enjoyment - L2GSC	L2 Enjoyment - TMGS	L2 Engagement - L2GSC	L2 Engagement - TMGS
RandomForest	Accuracy	**72.92%**	60.42%	**70.83%**	77.08%
	Precision	**75.50%**	58.89%	**75.20%**	73.51%
	Recall	**72.92%**	60.42%	**70.83%**	77.08%
	F1-score	**73.30%**	59.58%	**72.84%**	75.25%
XGBoost	Accuracy	68.75%	60.42%	72.92%	77.08%
	Precision	73.49%	60.84%	75.49%	73.51%
	Recall	68.75%	60.42%	72.92%	77.08%
	F1-score	70.09%	60.50%	74.08%	75.25%
LightGBM	Accuracy	68.75%	**66.67%**	70.83%	**77.08%**
	Precision	74.14%	**63.81%**	67.46%	**75.25%**
	Recall	68.75%	**66.67%**	70.83%	**77.08%**
	F1-score	69.40%	**65.20%**	69.11%	**76.16%**

6 Discussion

This research aimed to present initial evidence supporting the reliability and validity underlying a newly developed L2 grit scale for measuring L2 grit within a collectivist culture. Overall, the findings indicate that the L2GSGSC scores are psychometrically robust, offering a credible measure of L2 grit within the context of a collectivist culture (i.e., China in this study).

Our findings support the dual components of perseverance of effort (PE) and consistency of interests (CI) as defined by Teimouri et al. (2020) and align with the Triarchic Model of Grit by Datu et al. (2017), which introduces adaptability as a third dimension. The role of adaptability in L2 learning is highlighted, considering

7

that culture, as argued by Pettigrew (1979), influences self-perception and is shaped by human actions.

The study also explores the relationship between language learners' growth mindsets, conscientiousness, engagement, and enjoyment with L2 grit. Correlational analysis shows positive associations between PE, adaptability, and a language growth mindset, but not with consistency. The collectivist perspective emphasizes the importance of group and family relationships (Boucher, 2010), which may influence L2 learners' pursuit of language acquisition to meet societal expectations. Conscientiousness is found to be significantly associated with both PE and adaptability, but less so with CI, contrasting with Teimouri et al. (2020)'s findings. This discrepancy warrants further investigation into the relationship between conscientiousness and L2 grit within a collectivist culture.

Regression analysis indicates that students with higher L2 grit are more engaged in classroom activities and enjoy the language learning process more than those with lower grit. This supports the literature linking L2 grit with positive learning outcomes (Fathi & Hejazi, 2023; Sadoughi & Hejazi, 2023; Yang, 2021; Zhao, 2023), validating the L2 Grit Scale as a reliable indicator of L2 learners' behavior

ML analysis reveals that the L2 Grit Scale is more effective in predicting L2 enjoyment than TMGS, suggesting its strong predictive power for learners' behavioral trends. However, both scales perform similarly in predicting L2 engagement. The positive correlation between L2 enjoyment and vocabulary mastery underscores the importance of enjoyment in the learning process (Oyama, 2022) and highlights the predictive capabilities of ML methods for understanding language learning dynamics.

In brief, the research has elucidated several key findings regarding the characteristics and outcomes associated with L2 grit. First, students exhibiting grit can demonstrate sustained persistence in their L2 learning endeavors, maintain a passion for engaging with the target language, and show an increased awareness of and responsiveness to evolving opportunities and changes within their L2 learning context. Second, gritty students can experience a heightened sense of engagement and derive

8

more enjoyment from their L2 learning activities and may harbor the conviction that their language learning achievements can be significantly enhanced through increased effort. Therefore, it is clear that grit is not merely a peripheral trait but is central to fostering persistence, enthusiasm, and an adaptive response to the learning environment among students.

7 Conclusion

The present study developed and validated L2GSGSC, a tool designed to assess grit in the context of collectivist cultures specifically within SLA.

However, this study is not without limitations. The self-reporting way is subject to potential biases, including social desirability, where participants may portray themselves in a more favorable light, and memory recall bias, which can affect the accuracy of reported information. Additionally, the psychometric validation of the L2GSGSC has been restricted to a singular collectivist context, raising questions about its applicability in varied cultural settings. Future research should, therefore, deploy the L2GSGSC in a range of cultural environments to explore cross-cultural applicability and to verify if the correlations between L2 grit and psychological states are replicable across cultures.

Overall, our findings indicate that students exhibit higher grit levels in language learning scenarios than in general contexts when assessed with this scale. This suggests that a domain-specific measure of grit yields more definitive and significant outcomes. Building on prior studies, this work integrates the concept of L2 grit into the collectivist societal framework and investigates its impact on the learning behaviors of L2 learners. In short, the results suggested that grittier individuals tend to be more engaged and enjoyable in language learning compared to their less gritty peers.

9

References

Alamer, A. (2021). Grit and language learning: construct validation of L2-Grit scale and its relation to later vocabulary knowledge. *Educational Psychology*, *41*(5), 544-562.

Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective*. Pearson.

Boucher, H. C. (2010). Understanding Western-East Asian differences and similarities in self-enhancement. *Social and Personality Psychology Compass*, *4*(5), 304-317.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, *4*(1), 5.

Datu, J. A. D., Yuen, M., & Chen, G. (2017). Development and validation of the Triarchic Model of Grit Scale (TMGS): Evidence from Filipino undergraduate students. *Personality and Individual Differences*, *114*, 198-205.

Datu, J. A. D., & Zhang, J. (2021). Validating the Chinese version of triarchic model of grit scale in technical–vocational college students. *Journal of Psychoeducational Assessment*, *39*(3), 381-387.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, *92*(6), 1087.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of personality assessment*, *91*(2), 166-174.

Dweck, C. S. (2013). *Self-theories: Their role in motivation, personality, and development*. Psychology press.

Fathi, J., & Hejazi, S. Y. (2023). Ideal L2 self and foreign language achievement: The mediating roles of L2 grit and foreign language enjoyment. *Current Psychology*, 1-15.

Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.

10

Jiang, Y. (2020). A study on the multi-level impact of personality traits and organizational innovation situations on employee career growth. [Doctoral dissertation, University of Science and Technology of China]. China.

Li, C., Jiang, G., & Dewaele, J.-M. (2018). Understanding Chinese high school students' foreign language enjoyment: validation of the Chinese version of the foreign language enjoyment scale. *System*, *76*, 183-196.

Li, C., Wei, L., & Cai, J. (2023). Life satisfaction and L2 engagement in adolescents. *ELT Journal*, ccad054.

Littlewood, W. (1999). Defining and developing autonomy in East Asian contexts. *Applied linguistics*, *20*(1), 71-94.

Oyama, R. (2022). Exploring the Relationships between Enjoyment, Self-efficacy, Engagement, and Vocabulary Learning in Japanese Learners of English. *Journal of Asia TEFL*, *19*(4), 1163-1180.

Pettigrew, A. M. (1979). On studying organizational cultures. *Administrative science quarterly*, *24*(4), 570-581.

Rezaii, N., Walker, E., & Wolff, P. (2019). A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ schizophrenia*, *5*(1), 9.

Sadoughi, M., & Hejazi, S. Y. (2023). Teacher support, growth language mindset, and academic engagement: The mediating role of L2 grit. *Studies in Educational Evaluation*, *77*, 101251.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, *8*(2), 23-74.

Sofroniou, N., & Hutcheson, G. D. (1999). The multivariate social scientist: Introductory statistics using generalized linear models. *The Multivariate Social Scientist*, 1-288.

Teimouri, Y., Plonsky, L., & Tabandeh, F. (2020). L2 grit: Passion and perseverance for second-language learning. *Language Teaching Research*, *26*(5), 893-918.

11

https://doi.org/10.1177/1362168820921895

Yang, P. (2021). Exploring the Relationship Between Chinese EFL Students' Grit, Well-Being, and Classroom Enjoyment. *Frontiers in psychology*, *12*, Article 762945. https://doi.org/10.3389/fpsyg.2021.762945

Yao, Y., Wang, W., & Yang, X. (2021). Perceptions of the inclusion of automatic writing evaluation in peer assessment on EFL writers' language mindsets and motivation: A short-term longitudinal study. *Assessing Writing*, *50*, 100568.

Zhao, Y. L. (2023). On the relationship between second language learners' grit, hope, and foreign language enjoyment. *Heliyon*, *9*(3), Article e13887. https://doi.org/10.1016/j.heliyon.2023.e13887

Acknowledgements

语篇阅读过程词汇推理研究：理论基础、研究回顾及展望

郭琳 北京外国语大学英语学院

摘要：词汇推理指读者综合利用背景知识和语篇信息来推理非熟悉单词意义的一种认知策略，是读者阅读过程中不可或缺的认知过程。词汇推理不仅对语言学习者词汇知识发展有着重要作用，更能促进其对语篇意义的理解。本文回顾词汇推理相关理论基础和该领域开展的实证研究，分析这一领域存在的争议，并对词汇推理未来研究方向进行展望，以期为研究者在该领域开展更为深入的相关研究提供有意义的借鉴。

关键词：词汇推理；语篇阅读；认知加工；影响因素

Research on Lexical Inference in the Process of Discourse Reading: Theoretical Foundations, Research Review, and Prospects

Guo Lin

Beijing Foreign Studies University, International School of English Study

Abstract：Lexical inference refers to a cognitive strategy where readers integrate background knowledge and discourse information to infer the meanings of unfamiliar words, and it is an indispensable cognitive process in the reading activity of readers. Lexical inference not only plays a crucial role in the development of vocabulary knowledge for language learners but also enhances their understanding of the overall meaning of the text. This paper reviews the relevant theoretical foundations and empirical research conducted in the field of lexical inference, analyzes the controversies within this domain, and forecasts future research directions in lexical inference. The aim is to provide meaningful insights for researchers to carry out more in-depth studies in this field.

Key words: lexical inference; discourse reading; cognitive processes; influencing factors

1.引言

词汇推理（lexical inferencing)是指读者利用语篇提供的语言线索，并结合其自身世界知识、语境意识及相关语言知识，推测不熟悉词汇意义的过程（Haastrup，1991）。读者在阅读中难免会遇到不熟悉单词，为了顺利阅读也就需要推断词汇的意义。词汇推理对读者词汇知识发展和语篇意义理解具有重要作用，这方面的研究也引起了国内外专家的普遍重视（范琳，王珍，2016)。国内外研究者针对词汇推理开展了基础理论研究，并进行了有关实证研究。本文旨在阐明词汇推理的理论基础及相关实证研究，剖析该领域面临的问题，并展望词汇推理未来研究走向，希望给国内外学者展开更为深入的研究提供参考与借鉴。

2.词汇推理理论模型

研究者提出了若干相关理论模型，用以阐释词汇推理加认知过程。这主要包括心理模型、二语词汇加工模型、二语词汇推理认知加工模型。

2.1 心理模型

心理模型（mental model）是心智对客观与主观世界的心理表征，它是所有推理活动的认知与心理基础（王震，范琳，2012）。读者在阅读中，往往会试图构建一个相对完整的心理模型，以反映语篇所描述的情境和事件，而推理过程实际上是创建和验证这一心理模型的过程（Johnson-Laird, 1983）。心理模型所包含的信息不限于语篇本身提供的信息，还可以通过推理加工获取语篇隐含信息（van Dijk & Kintsch, 1983）。阅读过程中，当读者需要推测不熟悉词汇意义时，就会启动相关心理模型对这个单词进行投射映现（王震，范琳，2012）。此外，当构建的词汇心理模型不再需要某一心理模式所携带信息时，该心理模式的激活会受到抑制；而携带信息相关时，则该心理模式的激活程度将会增加（Gernsbacher, 1990；罗迪江，2008）。可以看出，读者推测生词意义的过程就是彼此存在联系的心理模型之间在线、动态作用的认知加工过程。

2.2 二语词汇加工模型

在 Flores d'Arcais 和 Schreuder（1983: 9）对交互模型（interactive model）中概念知识和刺激信息区分的基础上，Haastrup（1991:124）提出了二语词汇加工模型（model of L2 processing at word level），包括了两种类型的加工（见图 1）。分析加工（analytic processing）涉及上层（语境和语义线索）、下层两个线索层级，而整体加工（holistic processing）则只包括上层线索，仅涉及基于语境的自上而下的认知加工（参见魏蓉，范琳，2013）。
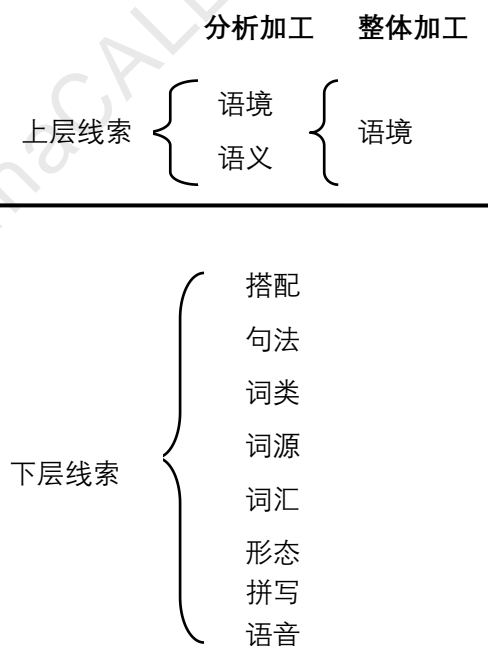
图 1. 线索层级（Haastrup, 1991: 124）

1	2	3	4	5	6
纯下层加工	下层线索支配	不可调和加工	上层线索支配的无融合加工	上层线索支配的局部融合加工	上层线索支配的整体融合加
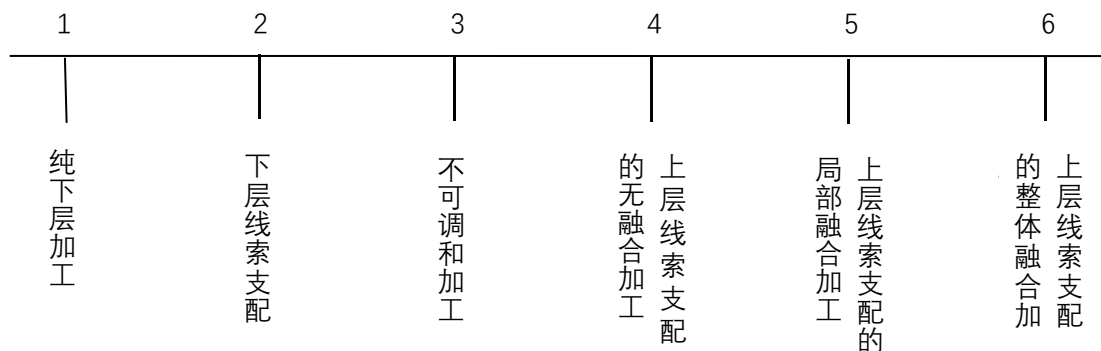
图 2. 跨话语连续体（Haastrup, 1991: 129）

除去纯上层加工线索，其他加工构成了加工深度逐步增加、效率渐渐提高的一个连续体，Haastrup（1991: 129）提出了跨话语连续体（the cross-talk continuum）（见图 2）。随着加工深度的增加，词汇加工过程被激活的相关信息也相应增多。上层线索也随着这些下层线索的激活而得以激活，词汇的整个加工过程也更为高效地进行。因此，读者成功推断出非熟悉单词意义的概率，也随之增大。

2.3 二语词汇推理认知加工模型

Huckin 和 Bloch（1993: 169-172）提出了一个结合串行加工（serial processing）和并行加工（parallel processing）的二语词汇推理认知加工模型。该模型阐释词汇推理加工动态的假说-生成-验证加工过程，包括生成/评估成分（generator/evaluator component）和元语言控制成分（meta-linguistic control component）。生成/评估成分包括语篇图式、句法形态、语篇表征、世界知识和长时记忆模块等大量基础知识模块。阅读过程中碰到不熟悉词语时，读者会开启二语词汇推理认知加工模式，通过调用其中的相关基础知识模块，来推测这些非熟悉词汇的意义（张姣，范琳，2018）。元语言控制成分指的是第二语言学习者尝试产出和验证其词汇推断是否正确时，使用的认知策略及决策策略。上述策略以串行方式在非熟悉单词意义推理加工中对各种可能的假设进行取舍，能够帮助语言学习者决定何时以及以何种方式来利用知识源、语境知识来推测词汇意义（Huckin & Bloch, 1993）。该模型的生成/评估成分和元语言控制这两个子成分各自独立又交互作用，动态作用于语言学习者词汇推理加工的整个流程。
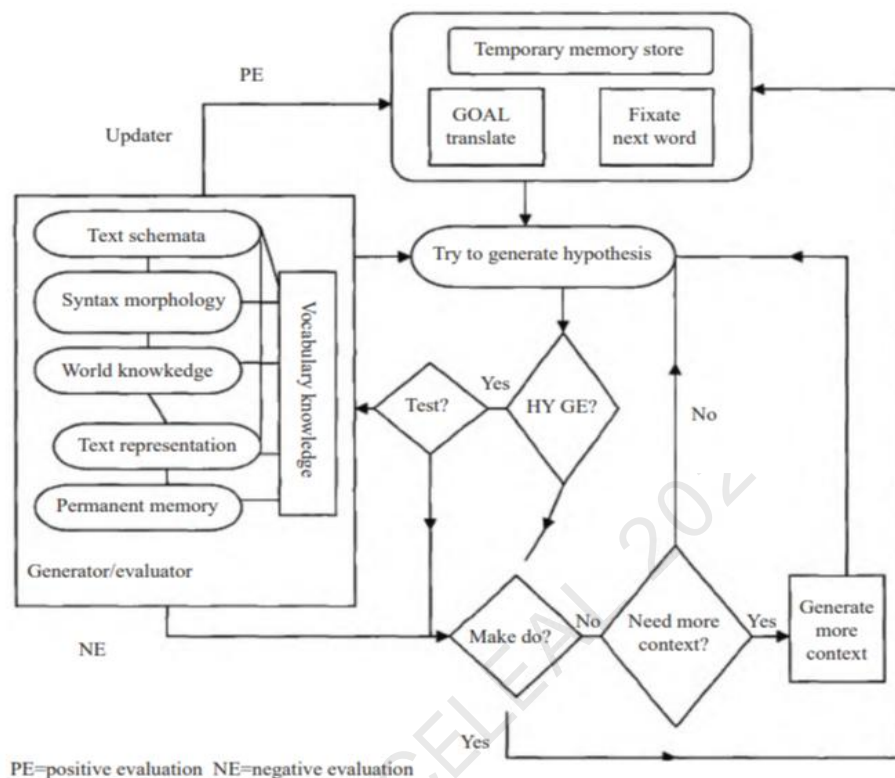
图 3. A cognitive model of L2 lexical inferencing (Huckin & Bloch,1993)

3.词汇推理加工实证研究

研究者进行了词汇推理认知加工过程的实证研究,考察了读者词汇推理认知机制及其影响因素。此外,研究者也实施了旨在提升二语学习者词汇推测能力的词汇推理策略训练研究。

3.1 学习者个体影响因素

语言学习者个体因素影响其词汇推理加工过程,本节主要阐述读者词汇知识水平、二语语言水平、语篇主题熟悉度、认知风格对这一认知过程的影响。

以往研究考察了词汇知识水平,包括词汇广度和深度对词汇推理的作用。多项研究证实,词汇知识深度对词汇推理的成功率有显著影响(如 Nassaji, 2006; Marzban & Hadipour, 2012; 寻阳,孙丽,2006)。其中,Marzban 和 Hadipour(2012)的研究还显示了词汇知识广度与推理成功率的正相关,但其显著程度不及词汇知识深度。另有研究发现,在特定阅读中,词汇知识阈限以上,词汇水平与推理成功率的关联性并不显著(王改燕,万霖,2011)。Zhang 等人(2019)考察了词素知识(词素意识和词素辨识)和词汇知识(字符知识和定义知识)在国际中文学习词汇推理中的作用,发现词素知识和词汇知识影响中文学习者的词汇推理过程。

学习者二语水平对词汇推理加工有着重要影响。以往研究发现高中生的英语水平与其词汇推理能力存在显著相关(孙兵,周榕,2005);较之低语言水平学习者,高语言水平者能够利用多种知识源进行推理(Kaivanpanah,Alavi,2008)。后续研究进一步确认了二语水平与词汇推理能力的正相关关系(Alavi & Kaivanpanah, 2009;Hamada, 2009;王瑛,黄洁,

2014）。Tavakoli 和 Hayati（2011）研究发现，中高水平学习者倾向于利用语篇知识进行推理，而中低水平者则更多依赖句子层面的语法知识。最近，采用反应时方法的研究也发现，学习者的二语水平影响其词汇推理中知识源的使用、推理反应时及其成功率（郭瑞，2021）。

语言学习者对语篇主题的熟悉度（topic familiarity）也会对其词汇推理认知过程产生影响。Pulido（2003, 2007, 2009）开展的一系列研究表明，语言学习者对所阅读语篇的主题熟悉程度越高，就越能成功推理出词汇意义。Alimorad、Ghaleb 和 Soozandehfar（2010）以及 Atef-Vahid、Maftoon 和 Zahedi（2012）研究也表明，较高的主题熟悉度有助于学习者建立相关的认知图式，从而提高其词汇推理成功率。对听力理解过程的相关研究也发现，语言学习者的词汇推理成功率与话题熟悉度相关（王怿旦等，2018）。最近，Zhang (2023) 采用问卷调查和回归分析，再次证实了较高主题熟悉度对词汇推理的积极影响，表明读者对主题越熟悉，其词汇推理的成功率也越高。

相关研究也发现学习者认知风格（cognitive style）影响其词汇推理。认知风格是指个体在知觉、思维、学习、问题解决和人际关系等方面稳定的个体差异（Witkin，Goodenough，& Karp，1967）。Alavi 和 Kaivanpanah（2009）考察了场独立型（filed-independent）和场依存型（filed-dependent）两种不同认知风格的语言学习者词汇推理加工情况，发现只是在阅读长语篇时，两种认知风格学习者的词汇推理能力才存在差异，即场独立型学习者表现出更强的推段生词意义的能力。Shen（2010）也发现学习者的认知风格对其二语词汇推理成功率产生影响；还发现词汇推理指导（训练）对听觉（auditory）和视觉（visual）学习者的推理能力提升有更好的效果。Bedir 和 Onkuzu（2014）也发现不同认知风格对学习者词汇推理成功率带来差异。然而，王瑛（2013）则表明，场独立型和依存型学习者在推测生词时使用的推理线索的种类、采用推理策略的数量以及推理成功率，均没有统计学意义上的显著差异。

3.2 语篇影响因素

以往研究者也考察了影响词汇推理加工的语篇因素,这主要涉及语义透明度、语篇语境、语篇难易度等因素。

语义透明度也称词汇透明度或形态透明度（semantic/morphological transparency），指不依赖或较少依赖语境，仅依据单词组成成分就能推测单词意义（参见范琳，吕新博，2018）。语义透明度是一个从透明到不透明逐渐变化的连续系统，不同词素构成的复合词呈现出不同程度的语义透明度（Auch, Gagné, & Thomas, 2020）。研究者开展了针对语义透明度对语篇阅读中词汇推理的影响的研究，但尚未取得一致结论。Nagy 和 Anderson（1984）对儿童受试词汇学习的研究发现，他们可以借助单词的语义或形态关联推测非熟悉单词的意义，说明词汇语义透明度对读者词汇推理加工产生影响。然而，也有研究得出了不一致的结论。Nagy，McClure 和 Mir（1997）、Shu 等（1995）对儿童一语词汇推理的研究、袁淑娟（2011）对中国大学阶段英语二语学习者词汇推理的研究,均没有发现词汇透明度对词汇推理有显著影响。究其原因可能是由于受试并没有关注词汇的结构，就对词汇意义做了主观判定（Bensoussan & Laufer, 1984; Frantzen, 2003）或者他们已关注到单词词缀，但未能利用这些线索来推测词汇意义和促进其词汇学习（Drum & Konopak, 1987）。

语境信息对读者成功推理生词意义有着重要影响。能够有效利用语义和语言线索是读者

5

进行词汇推理的前提（Haynes, 1993; Paribakht & Wesche, 1997）。二语阅读过程中，学习者需要具备识别和利用这些语境线索来推测生词意义的技能（Pritchard, 1990）。以往研究考察了语言学习者听力和阅读理解过程中的词汇推理加工。研究者发现英语听力理解中，学习者不仅依赖上下文线索，还借助背景知识进行推测（蔡薇，吴一安，2007）。针对语言学习者阅读过程中词汇推理的研究也发现，受试能够借助语篇提供的语境信息推理非熟悉词汇的意义（Kaivanpanah, 2008; Kanatlar & Peker, 2009），且语境线索的数量和质量，对推理成功率有不同影响（王改燕，万霖，2011）。然而，研究也显示，在限制性语境中，仅依靠上下文提供的语境线索不足以使读者推测出词汇的正确意义（Kelly, 1990）。还值得注意的是，尽管有可用的语境线索，这些信息有时可能不可靠，甚至误导学习者（Matsumura, 2010）。最近，Izmalkova 和 Blinnikova（2024）采用眼动技术开展的二语词汇推理加工的研究发现，词形和语境在词汇推理任务中有重要作用。

研究者也开展了语篇难易度对读者词汇推理影响的研究。Sternberg（1987）发现，阅读语篇中的生词比率不高于 5%时，有利于学习者进行有效词汇推理，Nation（2001）也发现超过 5%则会对学习者造成困难。Paribakht 和 Wesche（1999）、Fratzen（2003）等研究表明，句子长度、关联度、抽象词汇、专有词汇和低频词汇等这些构成语篇难易程度的特征，对读者词汇推理均有显著影响；他们还发现，语篇难度太大会导致读者无法有效利用语境线索等推理词汇意义。包含高密度生词的语篇可能导致学习者选择放弃，而不是尝试利用上下文推测生词的意思（Koda, 2005）。

3.3 词汇推理策略训练研究

读者在阅读中遇到不熟悉的单词时，会采用一系列的策略推测单词的意义，使阅读能够顺利进行。通过对学习者进行词汇推理策略训练，可以有效提升其策略意识。这种提升不仅有助于促进学习者的词汇知识水平发展，还能促进其整体阅读能力的提高。目前，针对读者二语词汇推理策略训练及其效果的实证研究仍较为缺乏。张华（2007）、顾云青（2008）对非本科生和高职生的研究表明，词汇语义推理训练后，受试词汇推理水平有了明显提高，尤其对低语言水平者的提升效果更为显著。近年来，邓倩倩（2022）开展了针对不同语言水平的初中生词汇推理策略训练研究，发现经过策略训练，受试者利用语境线索推理词汇意义的能力和语言水平都有明显提高，尤其是中、低水平学习者的进步更为显著。基于其国际中文学习者汉语词汇推理策略的研究发现，刘颂浩（2001）认为非常有必要对二语学习者进行词汇推理策略训练，这能够有效提升其利用相关知识推测不熟悉单词意义的能力。未来也需要更多地开展旨在提升读者词汇推理策略的干预训练研究。

4. 思考与展望

由以上论述可知，国内外研究者在词汇推理研究领域开展了一系列研究，并取得了较为丰富的研究成果。然而，部分研究结论还存在不一致，因此该领域仍需开展更多理论和实证研究，以更深入地揭示学习者词汇推理认知加工机制。

第一，进一步扩展多语者跨语言词汇推理认知机制和发展模式研究。目前，读者跨语言词汇推理加工研究，还较为有限（如 Albrechtsen et al., 2008; Wesche & Paribakht, 2009；范琳，

王震，杨赛，2013)，特别是对语言类型距离遥远的语言间跨语言词汇推理发展模式研究，也尤为欠缺（如范琳，何漂飘，魏泓，2015)。所以，未来应该更多地开展语言类型相距遥远的语言之间—汉语与拼音文字间—跨语言词汇推理加工过程以及发展模式的深入研究。

第二，开展更多针对不同群体的词汇推理加工研究。以往研究多关注大学本科生词汇推理加工，针对其他群体的研究较少，尤其是缺乏中小学生和来华留学生词汇推理的相关研究（范琳，王珍，2016）。中小学的基础教育阶段是二语言学习的关键时期，也是其词汇学习的关键期。随着国际中文教育的日益发展，来华国际中文学习者日益增多。词汇学习是语言学习的重要基础（范琳，王珍，2016）），未来需要进一步开展中小学生和国际中文学习者英语和汉语词汇推理加工的相关研究，为有针对性地提升这两个群体的英语或汉语词汇知识水平以及整体语言水平提供有意义的借鉴。

第三，研究者对影响词汇推理的读者个体因素（如语言水平、词汇水平、工作记忆）和语篇因素（如语境、主题熟悉度、语义透明度）进行了考察。然而，目前尚较为缺乏对个体和语篇因素交互影响的研究，未来可更多地开展词汇推理过程中读者个体因素和语篇因素相互作用的研究。

第四，进一步开展语言学习者词汇推理策略训练研究。自主学习能力及策略使用能力都是有意识的技能，通过培训可以得到提升（华维芬，2002)。早在上世纪九十年代，研究者就对二语学习策略训练给予了充分关注（如 Oxford, 1990; Cohen, 1998)。所以，实施对语言学习者，尤其是对低水平语言学习者的词汇推理策略训练尤为重要，可以使其掌握并能有效运用词汇推理策略，提升词汇学习的效率。另外，需要进行对各种训练方法效果的对比研究，具体可考察词汇推理策略的单独训练与整体训练、显性训练与隐性训练的不同效果，从而探索出适合不同二语水平学习者的有效词汇推理策略训练模式（范琳，王珍，2016）。

第五，采用更多元化的研究方法考察读者词汇推理认知机制。现有的词汇推理实证研究多采用问卷调查和测试等方法，较少采用有声思维和定性访谈等方法，也非常缺乏个案研究和跟踪研究（范琳，王珍，2016）。未来该领域的研究的方法应该更为多元，更多地开展学习者阅读及听力理解过程中实时定量及跟踪研究。具体研究方法可包括有声思维（也与回溯相结合）、反应时、眼动追踪（eye-tracking）及事件相关电位（Event-related potentials, ERPs)方法和技术，用以揭示读者词汇推理的实时认知过程。

5. 结语

综上所述，词汇推理在语篇阅读过程中扮演着至关重要的角色，是语言学习者理解不熟悉单词及整体语篇意义的重要认知策略。通过对词汇推理理论基础和实证研究的梳理，可以看出该领域的研究已经取得一定成效，尽管仍然存在一些争议。然而，这些争议也为未来研究提供了继续探索的空间。展望未来，研究者可以更多采用实时研究研究技术，如眼动跟踪、脑电和脑成像技术，深入揭示读者实际阅读过程中的词汇推理认知机制。随着词汇推理研究的不断深入，不仅能够深化我们对这一认知过程的理解，也能够为词汇和阅读教学提供更具意义的借鉴和启示。另外，也希望本文的讨论能够激发更多学者关注词汇推理这一重要认知过程，为提高语言学习者的词汇知识水平、阅读能力和整体语言水平做出有意义的贡献。

参考文献

1. Alavi, S. M. & Kaivanpanah, S. 2009. Examining the role of individual differences in lexical inferencing [J]. *Journal of Applied Sciences* 9 (15): 2829-2834.

2. Albrechtsen, D., Haastrup, K. & Henriksen, B. 2008. *Vocabulary and Writing in a First and Second Language: Processes and Development* [M]. New York: Palgrave Macmillan.

3. Alimorad, Z., Ghalebi, S. R. & Soozandehfar, S. M. A. 2010. The role of grammar in L2 lexical inferencing [J]. *AzerELTA Online Journal* 5: 1-12.

4. Atef-Vahid, S., Maftoon, P. & Zahedi, K. 2012. Topic familiarity, passage sight vocabulary, and L2 lexical inferencing: an investigation in the Iranian EFL context [J]. *International Journal of Research Studies in Language Learning* 2 (4): 79-99.

5. Auch, L., Gagné, C.L. and Thomas, L. 2020. Conceptualizing Semantic Transparency: A Systematic Analysis of Semantic Transparency Measures in English Compound Words. *Methods in Psychology* 3（3）:1-39.

6. Bedir, G. & Onkuzu, E. 2014. The relationship between learning styles of university preparatory class EFL learners and their success in L2 lexical inferencing[J]. *International Journal of Academic Research* Part B 6 (2): 81-87.

7. Bensoussan, M. & Laufer, B. 1984. Lexical guessing in EFL reading comprehension [J]. *Journal of Research in Reading* 7: 15-32.

8. Cohen, A. D. 1998. S*trategies in Learning and Using a Second Language* [M]. London: Longman.

9. Drum, P. & Konopak, B. 1987. Learning word meanings from written context [A]. In M. Mckeown & M. Curtis (eds.), *The Nature of Vocabulary Acquisition* [C]. London: Lawrence Erlbaum, pp. 73-87.

10. Flores d'Arcais, G. B. & Schreuder, R. 1983. The process of language understanding: a few issues in contemporary psycholinguistics [A]. In G. B. Flores d'Arcais & R. J. Jarvella (eds.), *The Process of Language Understanding* [C]. John Wiley and Sons, Inc, pp. 1-41.

11. Frantzen, D. 2003. Factors affecting how second language Spanish students derive meaning from context [J]. *Modern Language Journal* 87 (2): 168-199.

12. Gernsbacher, A. 1990. Language Comprehension as Structure Building [M]. Hillsdale, NJ: Erlbaum.

13. Haastrup, K. 1991. *Lexical Inferencing Procedures or Talking about Words: Receptive Procedures in Foreign Language Learning with Special Reference to English* [M]. Tubingen: Gunter NarrVerlag.

14. Hamada, M. 2009. Development of L2 word-meaning inference while reading [J]. *System* 37 (3): 447-460.

15. Haynes, M. 1993. Patterns and perils of guessing in second language reading [A]. In T. Huckin, M. Haynes & J. Coady (eds.), *Second Language Reading and Vocabulary Learning* [C]. Norwood, NJ: Ablex Publishing Corporation, pp.46-64.

16. Huckin, T. & Bloch, J. 1993. Strategies for inferring word meaning in context: a cognitive model [A]. In T. Huckin., M. Haynes & J. Coady (eds.), *Second Language Reading and Vocabulary Learning* [C]. Norwood, NJ: Ablex Publishing Cooperation, pp. 153-178.

17. Izmalkova; A. & Blinnikova, I. 2024. Cognitive strategies of second language vocabulary inferencing: An eye tracking study[J]. 14 (5): 490-…

18. Johnson-Laird, P. N. 1983. *Mental Models* [M]. Cambridge, MA: Harvard University Press.

19. Kaivanpanah, S. & Alavi, S. M. 2008. The role of linguistic knowledge in word-meaning inferencing [J]. *System* 36 (2): 172-195.

20. Kanatlar, G. M. & Peker, B. G. 2009. Guessing-words-in-context strategies used by beginning and upper-intermediate level EFL students [J]. *International Journal of Learning* 16 (5): 435-444.

21. Kaivanpanah, S. & Alavi, S. M. 2008. The role of linguistic knowledge in word-meaning inferencing [J]. *System* 36 (2): 172-195.

22. Kelly, P. 1990. Guessing: No substitute for systematic learning of lexis [J]. *System* 18: 199-207.

23. Koda, K. 2005. *Insights into Second Language Reading: A Cross-linguistic Approach* [M]. New York: Cambridge University Press.

24. Matsumura, Y. 2010. Factors influencing the lexical inferencing of Japanese EFL learners [D]. Tokyo: Temple University.

25. Marzban, A. & Hadipour, R. 2012. Depth versus breadth of vocabulary knowledge: assessing their roles in Iranian intermediate EFL students' lexical inferencing success through reading [J]. *Procedia-Social and Behavioral Sciences* 46: 5296-5300.

26. Nagy, W. & Anderson, R. 1984. How many words are there in printed school English? [J]. *Reading Research Quarterly* 19: 304-330.

27. Nagy, W. E., McClure, E. F. & Mir, M. 1997. Linguistic transfer and the use of context by Spanish-English bilinguals [J]. *Applied Psycholinguistics* 18: 431-452.

28. Nassaji, H. 2006. The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success [J]. *The Modern Language Journal* 90: 387-401.

29. Nation, I. S. P. 2001. *Learning Vocabulary in Another Language* [M]. Cambridge: Cambridge University Press。

30. Oxford, R. L. *Language Learning Strategies: What Every Teacher Should Know* [M]. Boston: Heinle & Heinle Publishers,1990．

31. Pulido, D. 2003. Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading [J]. *Language Learning* 53 (2): 233-284.

32. Pulido, D. 2007. The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading [J]. *Applied Linguistics* 28 (1): 66-86.

33. Pulido, D. 2009. How involved are American L2 Learners of Spanish in lexical input

processing tasks during reading? [J]. *SSLA* 31 (1): 31-58.

34. Paribakht, T. S. & Wesche, M. 1997. Vocabulary enhancement activities and reading for meaning in second language vocabulary development [A]. In J. Coady & T. Huckin (eds.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy* [C]. Cambridge, UK: Cambridge University Press, pp. 174-200.

35. Paribakht, T. S. & Wesche, M. 1999. Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing [A]. In M. Wesche & T. S. Paribakht (eds.), *Incidental L2 Vocabulary Acquisition: Theory, Current Research and Instructional Implications* [C]. Special Issue: Studies in Second Language Acquisition, pp. 195-224.

36. Pritchard, R. 1990. The effect of cultural schemata on reading processing strategies [J]. *Reading Research Quarterly* 25 (4): 273-295.

37. Shen, M. Y. 2010a. Effects of perceptual learning style preferences on L2 lexical inferencing [J]. *System* 38 (4): 539- 547.

38. Shu, H., Anderson, R. & Zhang, H. C. 1995. Incidental learning of word meanings: A Chinese and American cross-cultural study [J]. *Reading Research Quarterly* 30: 76-96.

39. Sternberg, R. J. 1987. Most vocabulary is learned from context [A]. In M. G. Mckeown & M. E. Curtis (eds.), *The Nature of Vocabulary Acquisition* [C]. Hillsdale NJ: Lawrence Erlbaum Associates, pp. 89-105.

40. Tavakoli, M. & Hayati, S. 2011. The relationship between lexical inferencing strategies and L2 proficiency of Iranian EFL learners [J]. *Journal of Language Teaching and Research* 2 (6): 1227-1237.

41. van Dijk, T. A. & Kintsch, W. 1983. *Strategies of Discourse Comprehension* [M]. New York: Academic.

42. Witkin, H. A., Goodenough, D. R. & Karp, S. A. 1967. Stability of cognitive style from childhood to young adultHOOD[J]. J*ournal of Personality and Socia* 7（3）：291-300.

43. Wesche, M. & Paribakht, T. S. 2009. *Lexical Inferencing in a First and Second Language: Cross-linguistic Dimensions* [M]. New York: Multilingual Matters.

44. Zhang, Y. 2023. The influence of topic familiarity on lexical reasoning of English learners [J]. *Applied and Educational Psychology* (4) 3:

45. Zhang, H., Koda, K., Han, Y. & Li, J. 2019. Word-specific and word -general knowledge in L2 Chinse lexical inference: An exploration of word learning strategies [J]. *System* 87: 102-146.

46. 蔡薇，吴一安．2007．第二语言听力理解中的词义推断研究[J].《外语与外语教学》(7)：1-5．

47. 邓倩倩. 2022.语境线索推理教学对初中生英语阅 读词汇推理的影响[D].北京外国语大学.

48. 范琳，杨赛，王震. 2013.中国英语学习者的汉英词汇推理加工模式研究[J].《现代外语》(3)：269-277．

49. 范琳，何漂飘，魏泓.2015.中国学生汉英跨语言词汇推理发展研究[J]．《外语研究》(4)：51-56．

50. 范琳，王珍. 2016. 我国二语词汇推理研究：20 年统计分析与展望[J].《外语界》（6）：61-69.

51. 范琳 吕新博.2018.跨语言视角的汉英词汇推理加工机制研究[M].北京大学出版社.

52. 郭瑞.2021.英语水平对英语学习者词汇推理加工过程的影响研究[M].山东大学.

53. 顾云青.2008. 运用猜词策略提高学生英语阅读能力[J].《语言与文化研究》(1) : 180 -186.

54. 华维芬.2002. 外语学习者策略训练刍议[J].《外语界》(3)：2-7.

55. 刘颂浩. 2001. 语境中猜测词义的调查[J].《汉语学习》(1) : 45-49.

56. 罗迪江.2008. 基于心理模型的词汇推理[J].《宜宾学院学报》(9)：83-86．

57. 孙兵，周榕. 2005. 高二学生利用语境线索猜测词义的研究报告 ［J］. 外语教学理论与实践（2）：49-54.

58. 王改燕，万霖．2011．二语阅读中语境线索水平对词义推测的影响[J].《外语学刊》(6)：94-97．

59. 王瑛. 认知风格对二语词义猜测影响的实证研究[J]. 外语研究，2013（3）：59-64.

60. 王瑛，黄洁芳. 2014. 二语水平与二语词义猜测的关系研究 ［J］. 外语与外语教学（1）：41-47.

61. 王怿旦，金叶，杨卉卉.二语听力中话题熟悉度与线索类型对词义推理的影响分析[J].《鲁东大学学报(哲学社会科学版)》，2018（2）：44-48.

62. 王震，范琳. 2012.语篇阅读过程词汇推理研究的进展[J].《外语教学》(3)：56-60.

63. 王震，郑志恒. 2016.国内外词汇推理研究统计与分析[J].《外语研究》157（3）：38-44.

64. 魏蓉，范琳. 2013. 第二语言词汇推理加工研究述评[J].《当代外语研究》13（7）：30-34.

65. 寻阳，孙丽.L2 读者词汇知识深度与词义推测策略的成功运用[J]. 外语界，2006（4）：41-47.

66. 袁淑娟. 2011. 中国英语学习者词汇推理研究[D].山东大学.

67. 张华. 2007. 词汇认知宽度与词义联想之间相关性初探[J].《大学英语(学术版)》(2) :24-28.

68. 张姣，范琳.2018.一语词汇化对二语词汇推理加工时间进程的影响:一项跨语言研究[J]．《外语教学理论与实践》161(2): 31-36.

The Southwestern Mandarin /n/-/l/ Merger: Effects on Production of Syllable Initial /n/ and /l/ in Standard Mandarin and English

Li Lulu
Beijing Foreign Studies University

Abstract: The current study investigated the acoustic properties of prescriptive N and L at syllable initial position in Southwestern Mandarin, aiming to offer solid evidence for the /n/-/l/ merger in SWM. The acoustic cues including F2-F1 spacing and F3 were carefully examined to run a parallel comparison across Standard Mandarin, English, and Southwestern Mandarin. In addition, informed by previous studies, this empirical study looked into various phonetic factors like the rhyming system in Mandarin and vowel quality in English as well as the presence of competitor sounds to see the potential impact factors in pronouncing [l] and [n] in speaker's L2 and L3. It was found that for the speaker a high level of language proficiency in both languages, the competitor sound emerged as the main culprit of the mispronunciation.

Keywords: /l/-/n/ merger, Southwestern Mandarin, Standard Mandarin, English production

1

1 Introduction

/l/-/n/ phonemic contrast is prevalent in many world languages and this contrast usually carries a heavy functional load which is a way to measure error gravity in regard to segmental mispronunciations. Phonemic contrast like /l/-/n/ of high functional load in a language will generate a great number of minimal pairs of high word frequency. Different languages may assign different weightings to the same phonemic contrast. Generally speaking, /l/-/n/ phonemic contrast has created more meaning contrast in English than in Standard Mandarin. In other words, the mispronunciation of [l] and [n] in English will take a heavier toll on the listener's comprehension than in Standard Mandarin. To be noticed, the non-distinction between the two phonemes in the onset position is evaluated at the highest level of the functional load scale in English (Catford, 1987; Brown, 1988), a measure of the likelihood of confusion for listeners. It is claimed by Bent et al. (2007) that segmental pronunciation accuracy of vowels and word-initial consonants is strongly correlated with judgments of intelligibility. Naturally, the accurate production of word-initial /l/-/n/ in English provides the most important cues for listeners to retrieve the corresponding words in their mental lexicon. Contrasted with phonemic contrast carrying such heavy functional loads in a certain language in which any substitution of [l] for [n] may bring about communication barriers, mispronunciation of sound contrasts of low level of functional load will not result in too many misinterpretations. Even when there is a pronunciation error, listeners still can successfully decode the utterance, for example, words like foot-boot.

Although this pair of phonemic contrast has been endowed with varying levels of functional load in many languages, Southwest Mandarin, among some other Chinese dialects (e.g., Fuzhou, Nanjing, Chengdu, Wuhan), is known to lack the /l/-/n/ contrast. In these languages, /l/-/n/ seems to be in free variation and substitution of one phoneme for another does not incur any difficulties in comprehension. There has emerged an ongoing or complete /n/-/l/ merger in these languages. While studies on the /l/-/n/ merger in Cantonese have been insightful and fruitful, this merger is still

2

understudied in Southwestern Mandarin despite its prevalence among Southwestern Mandarin speakers. It is also less well-known than some other pronunciation problems, in particular the /l/-/ɹ/ contrast for Japanese English learners, who are struggling to perceive and produce the /l/-/ɹ/ contrast. Brown (1998) suggested that the difficulties with nonnative phonemic contrasts are not equally distributed. When features in learners' L1 phonological system match the L2 features, L2 perception does not pose any challenge to the learners. However, when this feature responsible for differentiating important phonemes in L2 is absent in L1, L2 perception might be impaired despite the increased language proficiency. That's why speakers of Southwestern Mandarin might have persistent trouble in differentiating the alveolar nasal from the alveolar lateral both in their Mandarin production and perception.

A non-distinction between /l/ and /n/ in L1 may generate difficulties or even misinterpretations in L2 speech understanding. What's more, the absence of /l/-/n/ contrast in L1 may lead to the failure of /l/-/n/ differentiation in their L3 English language acquisition. This can be illustrated by the confusion between the English word "walnuts" and "wallets" by Southwestern Mandarin speakers. It is well justified that there are three distinct phonological systems stored in the minds of Southwestern Mandarin speakers who learn English as their first foreign language. How these three phonological systems interact with each other when speakers activate a particular language is of great interest and deserves thorough scrutiny. In this study, three phonological systems will be examined carefully to reveal the intricate interrelationship between them.

2　Literature Review

2.1　The Merger of /l/-/n/ in Cantonese and Min Dialect

/l/ and /n/ are two contrastive phonemes in many languages and they primarily differ in whether the airflow comes out of the nasal cavity or from the sides of the tongue. That is to say, they differ in both nasality and tongue blade involvement. Acoustically speaking, the F2 of Standard Mandarin (n) is weak and sometimes even

disappears, while the F2 of (l) sometimes shows only a low F2 value (Lin et al., 2013) but is higher when preceding a high vowel and lower when preceding a low vowel (Bao and Lin, 2014). However, not all Chinese languages differentiate the two phonemes so distinctly and there are at least three Chinese dialects in which /l/ and /n/ do not always contrast. The first is Southwestern Mandarin, the second is Cantonese and the third is Fuzhou Min. Historically, /l/ and /n/ are two distinct phonemes in Cantonese but they gradually merged into one phoneme with two allophones in complementary distribution (Zee, 1999). It is well acknowledged that [n]→[l] is one of the three long-standing mergers in Cantonese, partly resulting from the complex transformation of the socio-political landscape in Hongkong which has witnessed the interactions of diverse cultures (Cheng et al., 2022). According to Cheng et al., over the last several decades, the /l/-/n/ merger in Cantonese is close to completion. In other words, an eventual overlap of /l/ and /n/ in phonetic space takes shape. Before the complete overlap is a near-merger stage in which there are subtle but statistically significant acoustic differences between the two phonemic categories even though the speakers can't perceive the nuanced difference. Hay et al. (2013) suggested that the acoustic difference reflects distinct abstract representations that are pre-lexical. Now, this merger is prevalent among Cantonese speakers, both in the younger generation and the older generation. To et al. (2015) reported that nearly all the participants, regardless of their age and gender, produced the innovative form [l] to substitute for the historical /n/ (94.2% of children versus 94.6% of adults). In addition to these external social factors that drive the merging of /l/-/n/, speech style also impacts Cantonese speakers' production of the two consonants. It has been further investigated that [n]→[l] varies by speech register such that [n] is associated with formality (Pan, 1981) and careful speech (Zee, 1999). This is also supported by Bourgerie' findings (1990) that impromptu, conversational speech included the most [l] realizations (80.5%), followed by interview speech (58.2%) and public speech (38.6%). Despite the established rich and substantial investigations into mergers in Cantonese, there still lacks the acoustic analysis of [n] and [l] to reveal the acoustic cues to [l~n]

merger.

Likewise, the merger in Fuzhou Min is also modulated by both linguistic and social factors. In order to examine the phonetic space of L and N in Fuzhou Min, Cheng et al. (2023) conducted an acoustic analysis and the results revealed that none of the six acoustic cues indicated any difference across prescriptive L and N. It bears testimony to the [l~n] merger in Fuzhou Min production. Besides, he also suggested that vowel height and nasal coda can impact the listeners' discrimination accuracy of the two consonants. In regard to the speakers' production, it can be noted that there exists the vowel height effect with [n] more likely to be followed by low vowels.

2.2   The Merger of /l/-/n/ in Southwestern Mandarin

While the merger of /l/ and /n/ in Cantonese and Fuzhou Min has been extensively examined, the confusion of the two phonemes by Southwestern Mandarin speakers has drawn less attention from researchers. It has been noted that the historical loss of a distinction between /l/ and /n/ sounds in SWM is likely associated with a historical process of denasalization (Soejima et al., 1990; Hu, 2007). Different sub-dialects in SWM have different realizations of the historical change, leading to a complex landscape of the /l/-/n/ merger. In some dialects these phonemes coalesced into /n/ (e.g., Changsha in Hunan, Chengdu in Sichuan), and in some other dialects into /l/ (e.g. Nanjing in Jiangsu). According to Mou (2016), there are in total three types of /l/-/n/ production in Southwestern Mandarin, subject to a bunch of phonetic factors. Type one refers to the correct and clear production of /l/ and /n/ in varying phonetic environments, exempt from the influence of neighboring rhymes. In other words, speakers can differentiate the use of /l/ and /n/ regardless of their phonetic environment. The second type has partially mixed the use of /l/ and /n/, aligning the production of consonants with the quality of neighboring rhymes. Both /n/ and /l/ are likely to be pronounced when followed by /u/ and non-high vowels. In this case, /n/ and /l/ are in free variation. However, if the neighboring rhyme is /i/ or /y/, /n/ will turn into /ŋ/ and /l/ remains unchanged. The third type entirely confuses the use of /l/

5

and /n/ in any phonetic environment. In this case, /l/ and /n/ are in free variation, too. Despite the rich documentation of the /l/-/n/ merger in SWM, very few acoustic analyses have been conducted to examine their acoustic properties.

Investigation into the alternation of /l/ and /n/ by southwestern speakers has been conducted by many scholars from different perspectives. It has been found by Zhang (2007) in her spectrogram analysis that southwest Mandarin speakers tend to generate different realizations of /n/ in the onset, that is /l/, /n/, and /nl/. Wang (1994) also points out the nasal consonant [n] of Sichuan dialect is a nasalized lateral approximant. The option for either /l/ or /n/ seems to correlate with the height of the subsequent vowel as the production data show that other nasals in the onset pronounced as [l] are not followed by a high vowel. In terms of the /l/ production, Zhang (2007) suggested that /l/ can only be realized into [n] in English but not in Southwest Mandarin and Mandarin. The tendency of coalescing /n/ into /l/ rather than /l/ into /n/ seems to indicate that the two phonemes are not in free variation. But this awaits further investigation as Zhang (2007) only collected data from a single case. A more insightful and penetrating research, built upon this single case study, was initiated by Zhang & Levis in 2021. She collected linguistic data from 25 native speakers of southwest Mandarin who were divided into three groups according to their SWM dialect varieties and their first phonological system. Varying linguistic factors, including tones, rhymes, and competitor sounds have been examined to reveal the variables exerting impacts on the realization of /l/ and /n/. The results reject the effect of the four tones of Standard Mandarin but assert the role played by competitor sounds in /l/ and /n/ production. The presence of competitor sounds can increase the production accuracy of /n/ but undermines that of /l/. Besides, the data cannot confirm the effect of vowel height and no salient phonetic environment has been identified to account for the alternation of /l/ and /n/. Furthermore, the lack of L1 production data makes it unclear how the L1 phonological system is mapped onto the L2 and L3 phonological systems. Another drawback of Zhang & Levis's study is the absence of speakers' free speech production, which can better represent their language

6

performance. It is worthy of further investigation whether speech style (informal and formal) plays a role in [n]-[l] production. However, due to the time constraints pressuring data processing and analysis, informal speech production, originally designed as part of the data collection, is eventually abolished. As speakers' language proficiency acts as a crucial variable in mediating the potential cross-linguistic influence from the prior language, different levels of English proficiency also need to be taken into account. Considering the presence of rich scrutiny of speakers with low and intermediate levels of English proficiency, looking into the advanced learners' language performance can offer a more comprehensive picture of phonological transfer across Southwestern Mandarin, Mandarin, and English.

3 Research Methodology

3.1 Research Questions

Based on the gap extracted from the literature review, this paper aims to answer the following questions:

1) To what extent is there evidence of an [l~n] merger in southwest Mandarin production?

2) What linguistic factors impact advanced speakers' (l) and (n) production in Mandarin Chinese?

3) What linguistic factors impact advanced learners' (l) and (n) production in English?

In regard to the first research question, acoustic analysis is conducted to examine the production of [l~n] merger in southwest Mandarin. In regard to the second and third questions, word list reading is assigned to all the participants. Error rate analysis is rigorously implemented to figure out the linguistic factors that impact /l/-/n/ production. As for the fourth question, error analysis across Standard Mandarin and Mandarin are combined to shed light on the source of transfer in English production.

3.2 Data Collection

3.2.1 Participants

30 native speakers of Southwestern Mandarin, aged between 18 and 26, including undergraduate and postgraduate students in Beijing, were recruited in this research. All grew up in the southwestern part of China before attending college in Beijing, and they were reported to speak SW Mandarin as their first oral language, which dominates their daily communication at home. But in college, standard Mandarin has been spoken most of the time in their interaction with peers and teachers. In addition, they are all advanced English learners. Some are English majors, already passing TEM-8 and others have passed CET-6 with 600 scores above.

3.2.2 Material

Materials are composed of two parts: Chinese and English word lists. The Chinese word list comprises 46 monosyllabic words and 48 disyllabic words (two monosyllabic words that function as a single word) with nasals and laterals being put in different phonetic environments. The word-initial nasals and laterals are followed by all four Sihu rhymes. In the first word type, the vowels in pinyin were <a> (a), <e> (ɤ), <o> (o) or rhyming units that began with these vowels, such as <ai>, <ei>, <ou>, etc. In the second word type, the high front vowel (i) or rhyming units begin with <i>, such as <ie> (iɛ), <ing> (iŋ), etc. The third word type refers to those words with initial nasal or lateral followed by the high back vowel (u) or rhyming units beginning with (u), such as <uo> (uo), <uan> (uan), etc. In the last word type, /n/ or /l/ preceded the high front vowel (y) or rhyming units starting with <ü>, like <üe> (yɛ). As for those disyllabic words, sound competitors are introduced to some words, like niu lang (牛郎) while other words' phonetic environment remains simple without any distracting sound segment like huang li(黄鹂). The same principle also applies to the English word list. There are in total 28 English words with syllable initial /n/ and /l/ followed by different vowel qualities: high back vowel, low back vowel, mid back vowel, high front vowel, low front vowel, mid front vowel, and mid-central vowel. In addition to

8

the /l/ only group and /n/ only group, English words with /l/ and /n/ in medial positions are also provided for the reading test. The words with /l/ and /n/ in medial environments included words with <nl> or <ln> (only or walnut), words with <ll> (swallow), and words with <nn> (winner). Apart from the single English word list, English phrases are purposefully included to examine whether the presence of a competitor sound across word boundary would affect the accuracy production of word-initial/l/ or /n/.

3.2.3 Procedure

Speech samples were elicited in the following tasks: Chinese word list reading in Southwestern Mandarin and Standard Mandarin as well as English word list reading. At the very beginning, participants were asked to pronounce some words in southwestern Mandarin. These words are presented in a specific context to naturally elicit the speaker's first language. Then, they read Chinese monosyllabic and disyllabic words in Standard Mandarin English, followed by the English word list reading. The productions were recorded to a computer using praat software, with a sampling rate set to 44.1 kHz at 16 bits per sample on one channel. The mono recordings were saved as individual WAV files for evaluation. Before reading, participants were given sufficient time to familiarize themselves with the reading materials. They were asked to read aloud all the test items two times each.

3.3   Data Analysis

SPSS was used to run a Normality Test and an Independent Samples t-test, comparing the F2-F1 spacing and F3 of the prescriptive L and N groups in order to determine whether there is statistical evidence that the means across the two groups are significantly different. L and N are used as the prescriptive labels of the phonemes that map to an unmerged system, a distinction that is maintained in the hanzi writing system. The error rate of /n/-/l/ production in Standard Mandarin and English was calculated via Excel.

4 Results

F1, F2, and F3 at the midpoint of the word-initial consonants were extracted from the spectrogram. F2-F1 spacing and F3 frequency are among the most important acoustic correlates based on the acoustic properties of [l] and [n] as documented in previous studies. F2-F1 spacing is the frequency difference between F2 and F1. For [l], both F2 and F1 are low, while F3 is high (Prahler, 1998; Stevens, 1998). According to perturbation theory, the complete anterior constriction of [n] leads to a slight decrease in F1 and an increase in F2, contributing to greater F2-F1 spacing (Chiba & Kajiyama, 1941). As for F3, the length of the primary resonance tube (L) of [n] is longer than that of [l], the formula $F3 = 3(v/4L)$ predicts that [n] has a lower F3 than [l] when the volume of the resonance tube (v) is constant (K. Johnson, 2012). Some researchers did conduct acoustic analysis to probe into the features of different formant of [l] and [n]. Tabain et al. (2016a, 2016b) reported higher F3 and smaller F2–F1 spacing for [l] (F3 = 2839Hz, F2–F1 = 1245Hz) than [n] (F3 = 2761Hz, F2F1 = 1361Hz) across three Central Australian languages. In this study, F2-F1 spacing and F3 frequency of southwestern Mandarin have been measured from the spectrogram (See Figure 1 and Figure 2).

Figure 1

F2-F1 Spacing for Prescriptive N and L in Southwest Mandarin

10
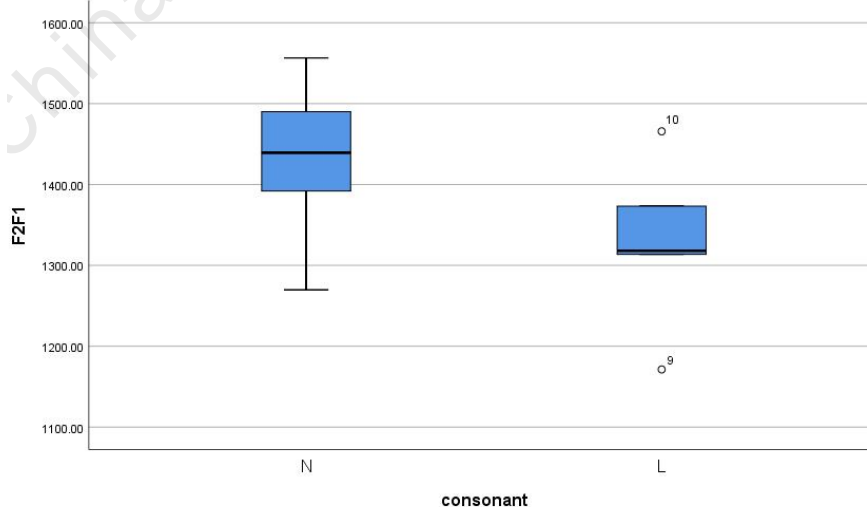
Figure 2

F3 for Prescriptive N and L in Southwest Mandarin
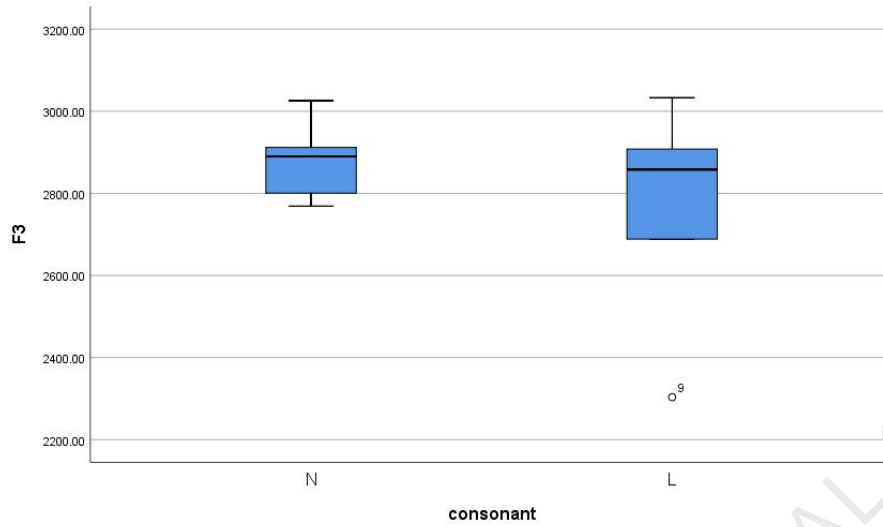
As a result of the absence of native English speakers and Mandarin speakers who speak the standard variety of Mandarin (Hou, 1986) as their native language, the data analysis will refer to the corresponding values obtained by the previous study. In Southwestern Mandarin, there was no significant difference between L and N for either F2–F1 spacing (SE = 80.56, t = 1.486, p = .176) or F3 frequency (SE = 111.93, t = 0.903, p = .408), providing solid evidence for the non-distinction between [l] and [n] (See Table 1 and Table 2).

11

Table 1

Comparison of the Acoustic Correlates of Nasal and Lateral Consonants in English, Mandarin, and Southwest Mandarin

Acoustic Parameter		English		Mandarin		Southwest Mandarin	
		[l]	[n]	[l]	[n]	L	N
F2–F1 spacing (Hz)	M	748	1,297	1,061	1,108	1328	1429
	SD	262	290	302	403	107	108
F3 frequency (Hz)	M	2,886	2,628	2,710	2,623	2,758	2,879
	SD	296	229	431	279	101	282

Note. The acoustic values of English and Mandarin are cited from "Production and Perception Evidence of a Merger:[l] and [n] in Fuzhou Min" by Cheng, R., Jongman, A., and Sereno, J. A. , 2023, *Language and Speech*, *66*(3), 533-563.

Table 2

Independent Samples Test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
F2F1	Equal variances assumed	.021	.888	1.486	8	.176	101.12200	68.07194	-55.85217	258.09617
	Equal variances not assumed			1.486	7.999	.176	101.12200	68.07194	-55.85497	258.09897
F3	Equal variances assumed	3.208	.111	.903	8	.393	121.21034	134.27103	-188.41921	430.83989
	Equal variances not assumed			.903	5.010	.408	121.21034	134.27103	-223.73433	466.15501

As for the second research question, more attention needs to be directed to the speaker's pronunciation of the Chinese disyllabic words. In this part, there are in total 96 tokens, with each word being repeated two times. Figure 3 shows the percentage of errors in Standard Mandarin speaking by SWM speakers. In this study, only one participant's speech production has been sampled due to time-constraint and hence more speech samples need to be collected in the future investigation. It is noticeable that no pronunciation error is detected in the speaker's Chinese monosyllabic words reading. In the disyllabic words group, pronunciation errors only occur in the #n+#l and #l+#n words. 50% of those tokens have been mispronounced in both the two types of words, indicating that the presence of competitors adds greater difficulty to the learners' [l] and [n] pronunciation. A closer look into these errors reveals that

12

substitution of [n] for [l] occurs more frequently than substitution of [l] for [n], accounting for 75% of the total errors (See Figure 4). This result is quite counter-intuitive as the prior studies indicate that the assimilation of [n] into [l] is trendy in southwestern Mandarin. It is hypothesized that SWM speakers will continue to confuse the [l] and [n] in their Mandarin speaking and mispronounce [n] as [l].

Figure 3

Error Rate of Chinese Disyllabic Word Production
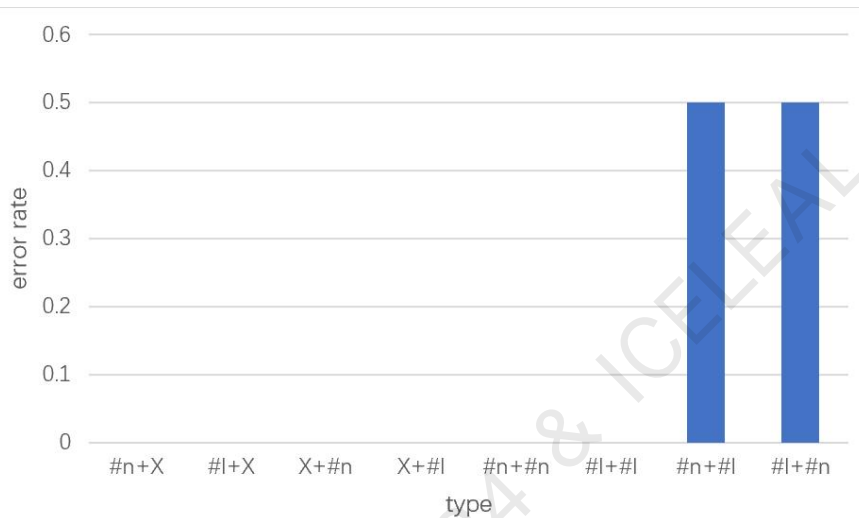
Figure 4

Error Percentage
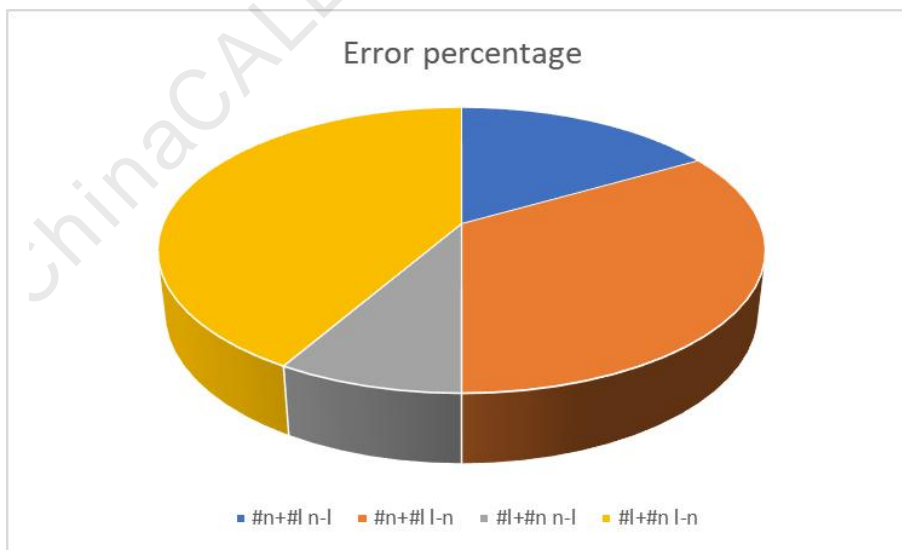
In the English production, the learner has produced 96 tokens of the single English word and 32 tokens of the English phrase in total. It is interesting to notice

13

that the English pronunciation errors are much fewer than expected, with 100% accuracy of single-word pronunciation and only 18.75% mispronunciation in the English phrase production. The errors are only produced in the #n+#l phrase pattern (See Figure 5). In other words, in the presence of sound competitors, the advanced English learner is still struggling to accurately pronounce each [n] and [l]. Figure 5 shows that 33.3% of the #n+#l English phrases have been mispronounced, with substitution of [n] for [l], while in the #l+#n pattern, delayed production of the word 'neighborhood' following the word 'lively' can be identified in the spectrogram, with a pause of 0.88s, although no error occurred in this phrase group.

Figure 5

Error Rate in English Phrase Production

Figure 6

Spectrogram of phrase "lively neighborhood"
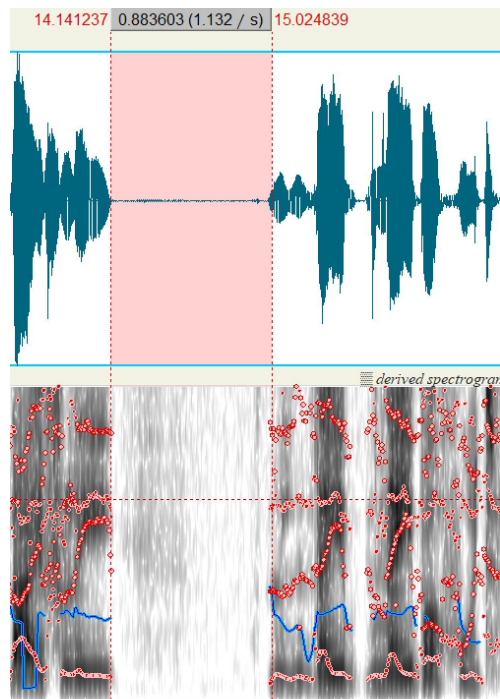
5   Discussion

The first research question looked at the acoustic properties of /l/-/n/ and aims to offer evidence to [l~n] merger in southwestern Mandarin by resorting to two acoustic cues as documented in previous studies, namely F2-F1 spacing and F3. The two acoustic cues have been reported to signal the contrast between nasal and lateral consonants in both Mandarin and English. To run a parallel comparison between the three languages, the acoustic values of English and Standard Mandarin, absent in this empirical study due to time limits, have been cited from Cheng et al.'s (2022) research. The results show that none of these acoustic parameters significantly differed across prescriptive L and N in Southwestern Mandarin. The phonetic space of the two phonemes overlapped to a large extent, serving as solid evidence of a merger in Southwestern Mandarin. In particular, the F3 frequency of both L and N is around 3000 Hz, similar to that of [l] in English and Mandarin, but F2-F1 spacing for N is

15

much larger. These figures are combined to indicate that in the speaker's dialect, /n/ is coalesced into /l/. Compared to Zhang's (2007) findings that there are different realizations of /n/ in the onset position, the acoustic analysis of this study shows that only one type of realization is produced by the SWM speakers. In Zhang's study, the speaker has adopted different approaches to /l/-/n/ alternations, suggesting that he can't disambiguate the alveolar nasal /n/ and the lateral approximant /l/ clearly. This way of /l/-/n/ alternation might fall into the second type of /l/-/n/ production in Southwestern Mandarin, partially mixing the use of /l/ and /n/ (Mou, 2016). This hypothesis needs to be tested via further investigation to see whether the quality of neighboring rhymes impacts the pronunciation of /l/ and /n/. In contrast, [l~n] merger in this study belongs to the third type of /l/-/n/ production as the speaker tends to confuse the use of /l/ and /n/, regardless of their phonetic environment. As is noted in prior research, [l~n] merger in Southwestern Mandarin takes different forms in different sub-dialects of SWM. In this study, the speaker's speech dialect belongs to Chengyu sub-variety of the Chuanqian variety and the acoustic analysis indicates /n/ is merged into /l/, contradicting Zeng's (2009) claim that only the alveolar nasal exists in the Chengyu sub-variety of the Chuanqian variety. But to make it more convincing, more participants from this group should be recruited in the future study.

The second question asked about the effects of the phonetic environment on /l/ and /n/ production in Standard Mandarin. The phonetic environments examined here are the effects of Sihu rhyming system in Mandarin and the impacts of competitor sounds. Standard Mandarin has a Sihu rhyming system with four kinds of rhymes, called Kaikou Hu, Hekou Hu, Qichi Hu, and Cuokou Hu. Linguistically speaking, Kaikou Hu is the rhyme that begins with non-high vowels, that is, not /i u y/, Hekou Hu begins with the /u/ sound, Qichi Hu begins with /i/, and Cuokou Hu begins with /y/ (Zhang & Levis, 2021). The data reveal no noticeable effect of the rhyming system on the speaker's sound production but show the role played by a competitor sound in disturbing the speech production. The findings also disagree with what has been concluded in Zhang's study (2021) in which the consonant's neighboring rhyming

16

system affects the speaker's production and the presence of a competitor means better accuracy for both /n/ and /l/. She suggested that for a rhyme beginning with (i), both sounds had similar error rates [18.05% for (n), 21.45% for (l)], but for the other three rhymes, the error rates for (n) production were considerably higher than for (l). The absence of the rhyming system's effect can be explained by the speaker's mature mastery of Standard Mandarin after long-term exposure to authentic input in Beijing and the regular practice in daily communication with her peers. The sharp contrast between the two empirical studies points to the significance of language proficiency in determining the possible phonetic factors. However, the presence of a competitor sound is found to negatively impact the accurate production of [l] and [n], in particular [l]. 75% of the pronunciation errors are committed, wrongly substituting [n] for [l] rather than substituting [l] for [n]. Such results are quite counter-intuitive as this SWM speaker tends to merge /n/ into /l/ in her dialect but it turns out to be the opposite in her Standard Mandarin speaking. It can be partly attributed to the speaker's acute awareness of /l/ and /n/ production in this test, and she wants to use some strategies to avoid the mispronunciation of [n] as [l]. However, the extreme focus on the [n] pronunciation may lead the speaker to mispronounce [l] as [n]. Besides, it is interesting to notice that many substitutions of [n] for [l] occur when the word-initial [l] is preceded by a word-final nasal in a disyllabic word, such as "年龄". The carryover of the nasal feature to the following sound constitutes a typical assimilation in speech production. In summary, despite the disappearance of vowel quality's effect on consonant production as indicated in the prior research, the neighboring sounds, in particular the nasals and laterals are still active in influencing the speaker's [l] and [n] production. They even prove to be the main culprits of the [l] and [n] mispronunciation in the speaker's Standard Mandarin production.

The third research question asked about the phonological environments most likely to elicit /n/ and /l/ pronunciation confusion in English. It is expected that different neighboring vowel quality can affect the accuracy of /l/ and /n/ pronunciation. However, the results show that vowel quality does not affect the

speaker's pronunciation. This may result from the participant's advanced English proficiency and she has already mastered the correct pronunciation of each single word, along with her Standard Mandarin speaking. To be noticed, this SWM speaker did not demonstrate any difficulty when both nasal and lateral sounds occurred together in the medial position, as opposed to the language performance by the speaker in Zhang's (2021) empirical study. The cooccurrence of both nasal and lateral sounds in the medial position is supposed to pose a great challenge to many Southwestern Mandarin speakers as they tend to produce a single nasal or lateral rather than two distinct sounds. The phonotactic constraints from Mandarin regulate that [l] is not allowed in a coda position and the two sounds can't appear together. The great ease with pronunciation of such type of words leads us to assume that the speaker has spent much time practicing them in her English learning and she has overcome these pronunciation difficulties. Mispronunciations mainly stem from the presence of the competitor sound, in particular in the #n+#l English phrase groups. It should be noted that Mandarin syllabic structure differs from English syllabic structure. The Coda position is usually not filled in Chinese and thus the time span across word boundary is shortened considerably compared to that of multisyllabic English words. The longer time it takes between two adjacent [l] and [n] productions, the less negative impact the competitor sound will exert on pronunciation. If the speaker's Standard Mandarin and English proficiency have both reached a high level, the sound competitor will be most likely to disturb a speaker's accurate production of /n/ and /l/. It is common across both languages that the target consonant in the first word is more likely to be mispronounced than in the second word. Different from the /l/-/n/ alternation in Standard Mandarin, in English production, /n/ is more likely to be mispronounced as /l/ rather than the opposite, aligning with the pronunciation habits in southwestern Mandarin. The speaker's opposite treatment of /l/-/n/ production in Mandarin and English is hard to explain based on the relatively small amounts of data. Future study needs to collect more speech data from a larger group of participants. Although no pronunciation error is detected in the #l+#n phrase pattern, the delayed

production of the second word with word-initial nasal in the phrase "lively neighborhood" also points to the increased pronunciation difficulty in the presence of a competitor sound.

The error analysis of the small data sample can't offer penetrating insights into the mechanism of phonological transfer but it still demonstrates the L1 phonological transfer to L3 pronunciation. As both Southwestern Mandarin and Standard Mandarin hold an equal typological distance to English, typological proximity no longer plays a role in the cross-linguistic influence. The language triplet examined in this study is largely similar to L1 English-L2 French-L3 Japanese adopted by Tremblay (2007) who strove to eliminate the typological proximity in his L3 phonological acquisition research. An in-depth comparison of the pronunciation errors across Standard Mandarin and English demonstrates the presence of a sound competitor constitutes the major challenge to the speaker's accurate pronunciation of [l] and [n] in both L2 and L3 production. The speaker's good mastery of Standard Mandarin and English pronunciation has helped her navigate a complex network of phonetic environments. However, the transfer from Southwestern Mandarin (L1) is still more evident than the transfer from Standard Mandarin in her English production, even though she has a very good command of Standard Mandarin, knowing /n/ and /l/ are two distinct phonemes.

6 Conclusion

The lack of contrast between /l/ and /n/ in SWM has caused much trouble in SWM speakers' accurate production of [n] and [l] in Standard Mandarin and English speaking. For the participant in this empirical study, /n/ is merged into /l/ in her SWM speaking, which is illustrated by the acoustic analysis based on the spectrograms of these consonants. The F2-F1 spacing and F3 between prescriptive N and L are not statistically significant, pointing to the collapse of the two phonemes in the speaker's linguistic repertoire. A closer look at the acoustic values reveals that the produced consonants are laterals. It is expected that the speaker might transfer the pronunciation

19

habits to the subsequent language learning based on the phonological acquisition models and [n] might be mispronounced as [l] in both Standard Mandarin and English. However, the results go beyond our expectations due to a lot of reasons. In the investigation into the speaker's /l/-/n/ production in Standard Mandarin and English, the neighboring rhymes or vowel quality seem to have no impact on the /l/-/n/ production by the advanced Mandarin speaker and advanced English learner. The most prominent phonetic factor is the presence of competitor sound as well as the time gap between the two competitors. The narrower time gap between alveolar nasals and laterals will lead to greater difficulties in pronouncing [l] and [n]. The fact that pronunciation errors in English are mostly substitution of [l] for [n] drives us to conclude that the transfer from Southwestern Mandarin is more pronounced than that from Standard Mandarin. As all the discussions are built upon a single case study, future research needs to recruit a larger number of participants, adding credibility to the final results.

References

Bao, H. Q., and Lin, M. C. (2014). *Laboratory phonetics* (2nd ed.). Beijing: Peking University Press.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, *23*(4), 459-484.

Barkley, S. (2010). *The acquisition of orthographic-phonological correspondence rules in L2 and L3 Portuguese: Error resolution, interference, and generalizability*. University of Florida.

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech. *Language experience in second language speech learning*, 331-347.

Bourgerie, D. S. (1990). *A quantitative study of sociolinguistic variation in Cantonese* (Doctoral dissertation, The Ohio State University). ProQuest Dissertations & Theses Global. http://search.proquest.com/docview/303872748/

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL quarterly*, *22*(4), 593-606.

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2), 136-193.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation: A systemic description of English phonology. *Current perspectives on pronunciation: Practices anchored in theory*, 87-100.

Cheng, L. S., Babel, M., & Yao, Y. (2022). Production and perception across three Hong Kong Cantonese consonant mergers: Community-and individual-level perspectives. *Laboratory Phonology*, *13*(1).

Cheng, R., Jongman, A., & Sereno, J. A. (2023). Production and perception evidence of a merger: [l] and [n] in Fuzhou Min. *Language and Speech*, *66*(3), 533-563.

Chiba, T., & Kajiyama, M. (1941). *The vowel: Its natural structure*. Tokyo-Kaiseikan.

De Angelis, G. & Selinker, L. (2001). Interlanguage transfer and competing linguistic systems in the multilingual mind. In *Crosslinguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*, J. Cenoz, B. Hufeisen & U. Jessner (eds), 42–58. Clevedon UK: Multilingual Matters.

Deterding, D. (2010). ELF-based Pronunciation Teaching in China. *Chinese Journal of Applied Linguistics (Foreign Language Teaching & Research Press)*, *33*(6).

Hammarberg, B. & Hammarberg, B. (1993). Articulatory re-setting in the acquisition of new languages. *Phonum* 2: 61–67.

21

Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition. In *Crosslinguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives,* J. Cenoz, B. Hufeisen & U. Jessner (eds), 21–41. Clevedon UK: Multilingual Matters.

Hay, J., Drager, K., & Thomas, B. (2013). Using nonsense words to investigate vowel merger1. *English Language & Linguistics*, *17*(2), 241-269.

Hou. (1986). 北京话音档 [Beijing Phonemic Inventory and vocabulary]. 上海教育出版社 [Shanghai Education Publishing House].

Hu, F. (2007). Post-oralized nasal consonants in Chinese Dialects—Aerodynamic and acoustic data. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1405-1408).

Lin, T., Wang, L. J., and Wang, Y. J. (2013). YuYinXue JiaoCheng (Revised Edition). Beijing: Peking University Press.

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, *7*(1), 39-57.

Llisteri, J. & Poch-Olivé, D. (1987). Phonetic interference in bilingual's learning of a third language. In *Proceedings of the XIth International Congress of Phonetic Sciences (Volume 5)*, 134–147. Austria, Tallinn, Estonia: Academy of Sciences of the Estonian SSR.

Pan, P. G. (1981). Prestige forms and phonological variation in Hong Kong Cantonese speech (Doctoral dissertation, The University of Hong Kong). DOI: https://doi.org/10.5353/th_b4389299

Prahler, A. A. M. (1998). Analysis and synthesis of the American English lateral constant [Master's thesis]. Massachusetts Institute of Technology.

Ringbom, H. (2002). Levels of transfer from L1 and L2 in L3-acquisition. In *Proceedings of the Second International Conference on Trilingualism*, J. Ytsma & M. Hooghiemstra (eds), Leeuwaarden: Fryske Adademie.

Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and high/low attachment preference in relative clause interpretations in L3 Brazilian Portuguese. *IRAL, International Review of Applied Linguistics in Language Teaching* 48(2/3): 245–273.

Rothman, J. (2011). L3 Syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research* 27(1): 107–127.

Rothman, J., & Cabrelli Amaro, J. (2010). What variables condition syntactic transfer? A look at the L3 initial state. *Second Language Research*, *26*(2), 189-218.

Soejima, K., Imagana, H., & Kiritani, S. (1990). Alternation between Stop Nasal and

22

(Nasalized) Flap or Lateral. *Ann. Bull. RILP*, *24*, 131-144.

Stevens, K. N. (1998). Acoustic phonetics. MIT Press.

Tabain, M., Butcher, A., Breen, G., & Beare, R. (2016a). An acoustic study of nasal consonants in three Central Australian languages. The Journal of the Acoustical Society of America, 139(2), 890–903.

Tabain, M., Butcher, A., Breen, G., & Beare, R. (2016b). An acoustic study of multiple lateral consonants in three Central Australian languages. The Journal of the Acoustical Society of America, 139(1), 361–372.

To, C. K., McLeod, S., & Cheung, P. S. (2015). Phonetic variations and sound changes in Hong Kong Cantonese: Diachronic review, synchronic study and implications for speech sound assessment. *Clinical Linguistics & Phonetics, 29*(5), 333–353.

Tremblay, M. C. (2007, August). *L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals* [Conference presentation]. In *Satellite Workshop of ICPhS XVI, Freiburg, Germany*.

Wang, W.H. (1994). Study of Phonetic Characteristics of Sichuan-accent Standard Mandarin. *Journal of Sichuan University*, *3*, 56-61.

Williams, S., & Hammarberg, B. (1998). Language switches in L3 production: Implications for a polyglot speaking model. *Applied linguistics*, *19*(3), 295-333.

Wrembel, M. (2012). Foreign accentedness in third language acquisition. *Third language acquisition in adulthood*, 281-310.

Zee, E. (1991). Chinese (Hong Kong Cantonese). *Journal of the International Phonetic Association*, *21*(1), 46-48.

Zee, E. (1999). Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics, 27*(1), 120–167. JSTOR. http://www.jstor.org/stable/23756746

Zhang, W. (2007). Alternation of [n] and [l] in Sichuan dialect, Standard Mandarin and English: A single-case study. *Leeds Working in Linguistics and Phonetics*, *12*, 156-173.

Zhang, W., & Levis, J. M. (2021). The Southwestern Mandarin /n/-/l/ merger: effects on production in standard mandarin and English. *Frontiers in Communication*, *6*, 639-390.

23

Appendix A

Reading materials in Southwestern Mandarin

男生和女生
过年啦
喝牛奶
牛郎织女
蓝方
可怜巴兮
流浪地球
来年再说

Appendix B

Chinese word list

/i/ rhyme	你	李
	尿	料
	扭	柳
	年	连
	凝	玲
	酿	量
	孽	列
	您	林
	泥	离
/u/ rhyme	怒	路
	暖	卵
	诺	落
/y/ rhyme	虐	略
	女	吕
Non-high vowel rhymes	那	辣
	讷	乐
	耐	赖
	内	累
	脑	老
	难	兰
	能	棱
	农	龙

#n+X	南方	男生	#l+X	蓝方	懒惰
	扭曲	泥土		驴肉	理智
	腻歪	怒气		旅客	落花
X+#n	黄泥	黄牛	X+ #l	黄鹂	经历
	收纳	刁难		波兰	可怜
	怀念	过年		怀恋	吉利
#n+#n	牛奶	扭捏	#l+#l	琉璃	流浪
	牛腩	男女		流利	理论
	泥泞	娜娜		力量	历练
#n+#l	牛郎	牛栏	#l+#n	流年	留念
	努力	男篮		李娜	岭南
	能力	年龄		来年	理念

25

Appendix C

English word list

onset	
not	lot
nothing	laugh
need	lead
navigate	lateral
nurture	alert
new	
noodle	loop
net	lesson
nasty	last
noise	loiter
no	low
near	
now	loud
name	late
night	light

medial	
walnut	vulnerable
swallow	pillow
unless	only
winner	beginner

phrase
language learning
little nonsense
last night
lucky number
noble leader
new lighthouse
natural light
national library
a new language
lively neighbourhood
long nap
lunar new year
next level
narrow lane
noble lady
neuron layers

26

Analysis of China's National Image Construction: A Critical Discourse Analysis Based on the Swahili News Corpus

Zulong Qin   Beijing Foreign Studies University

Abstract

With the increasing frequency of China-African exchanges, the importance of Swahili in China-African relations is steadily growing. However, there is still a notable lack of systematic research on Swahili news corpora and discourse analysis within China. This study aims to contribute to this area by constructing a Swahili news corpus and conducting a critical discourse analysis to reveal the characteristics of China Radio International (CRI) Swahili news in shaping China's national image and its role in China-African relations. Using web crawling technology, the study collects Swahili news texts from the "China/World" (China/Dunia) section of the CRI website spanning from 2019 to 2024 to establish the corpus. It then performs word frequency and concordance analyses on the news texts, combined with discourse analysis theory, to deeply explore their linguistic features and discourse structures. The findings indicate that the media generally holds a positive attitude in their reports. By covering China's efforts in promoting China-Africa relations, enhancing international economic cooperation, and diplomatic interactions between nations, the media successfully portrays China as a peaceful, friendly, open, and inclusive country. This aims to provide direction and reference for optimizing the construction of China's image in Africa.

Keywords

Swahili; News Corpus; Discourse Analysis; China-African Relations

Introduction

Swahili, an important indigenous language in Africa, is widely used in education, media, and government affairs in East Africa. In recent years, with the acceleration of globalization and the increasing frequency of international exchanges, China-Africa relations have become increasingly close, and Swahili is becoming increasingly significant in relating China and Africa, especially in news and media. As a medium for national information dissemination, official media plays a significant role in external communication and national image construction. China Radio International (CRI), as one of China's major international media outlets, has played an important role in shaping China's national image and disseminating China's voice through its Swahili news channel.

However, despite the growing importance of the Swahili language in China-Africa relations, systematic research based on Swahili corpora remains relatively scarce in China. This study uses Swahili news texts as samples, collecting news texts from the "China/World" (China/Dunia) section of the CRI website from 2023 to 2024 through web scraping technology to establish a specialized small news corpus. From the perspective of critical discourse analysis, the study explores the linguistic features and discourse structures of these news reports in constructing China's image. By employing corpus-based methods such as word concordance, this research aims

to elucidate the media's attitudinal tendencies when reporting on China-Africa relations and to grasp the national image of China in these interactions.

The study provides a new perspective on Swahili discourse analysis and aims to offer discourse optimization strategies to enhance the construction of China's image in Africa, thereby promoting the sustainable development of China-Africa relations and the construction of China's discourse system in Africa.

Theoretical Framwork and Literature Review

Fairclough, in his book *Language and Power*, established a three-dimensional model of "text, interaction, and context" to reveal how discourse practices are shaped by social realities and how they influence social realities. While critical discourse analysis (CDA) has achieved significant results in revealing the relationship between power and discourse, it has also faced some controversies within the academic community. For example, Fowler (1996) mentioned that the textual samples in CDA are too few to be representative, and the interpretation of discourse lacks objectivity and systematicity. Stubbs (1996) also argued that CDA only targets small excerpts of text, thus lacking sufficient credibility. In response to these criticisms, some scholars have begun to incorporate corpus linguistics methods into CDA to enhance its scientific rigor and empirical basis, leading to the development of the "corpus-based critical discourse analysis" research branch. By relying on large-scale authentic language data, corpus-based methods strengthen the empirical nature and comprehensiveness of the research. Additionally, the efficiency and reproducibility of this approach overcome the limitations of traditional discourse analysis, which often deals with small corpora and strong subjectivity, thus aiding in the discovery and verification of new linguistic phenomena in CDA. This enriches and deepens the theoretical and practical aspects of the field, becoming a new path in the development of discourse analysis.

Applying computers to conduct empirical analysis of media language patterns has made large-scale analysis possible. In recent years, more and more scholars have begun to apply corpus-based research methods to CDA. These studies, based on English news corpora, primarily analyze lexical collocations and grammatical structures from aspects such as word frequency, concordance, keyword in context, collocates and other methods to examine textual attitudes.

Putri and Mardiah (2024) explored the differing reporting attitudes of CNN and Aljazeera regarding the murder of George Floyd. Their study revealed how media, through discourse choices, convey ideological leanings, further influencing public perception and attitudes toward social events. Alluhaidah (2023) noted that although some media reports portrayed a positive image of Muslim women, overall, Western media still exhibit significant negative bias in their depiction of Islam and Muslims. Yang and Wang (2023) examined how The New York Times constructed China's national image in environmental news reports. They found that, economically, China was depicted as a country that pursues economic growth at the expense of the environment; politically, it was portrayed as demonstrating determination in governance but with poor governing capability.

Compared to Western countries, there is less attention paid to Africa. Safa Attia and Jesus Romero-Trillo (2022) analyzed media coverage of the 2011 Egyptian revolution, examining reports from Al Jazeera, Al Arabiya, BBC, and CNN in Arabic and English. They highlighted the extensive use of negative frames and violent imagery in media channels, pointing out that American media discourse often negatively depicts Syrian refugees and contextualizes their impact. Ndlovu and Nikabs (2019) explored how African news media reported and constructed

Africa's image during the COVID-19 pandemic. Analyzing reports from ten news agencies in Ghana, Tanzania, Zimbabwe, South Africa, and Egypt, the study found that African media played a complex role in shaping public perceptions of Africa and the pandemic. On one hand, they provided opportunities to counter Western negative narratives about Africa; on the other hand, some reports still perpetuated negative stereotypes about Africa. Although African media have the opportunity to challenge and change stereotypes about Africa, many reports inadvertently reproduced some negative biases.

In China, there were scholars who have used corpus research methods for CDA of Chinese and Western media news. Li (2024) systematically analyzed the portrayal of Beijing in mainstream British newspapers *The Guardian* and *The Daily Telegraph* during the COVID-19 pandemic, which revealed a relatively positive image of Beijing, but some negative reports still exsisted. Qian and Wang (2023) explored *People's Daily*'s discourse construction on the topic of "carbon peak and carbon neutrality" over the past 15 years. They found that *People's Daily* not only enriched the context of carbon neutrality but also enhanced public awareness and acceptance of carbon neutrality by introducing new discourse frameworks. Han and Chen (2018), based on transitivity system theory, conducted analysis of news texts about the "Chinese Dream" in Chinese and Western media, exploring their discourse differences in constructing China's image. Wang (2019) discussed the portrayal of Chinese enterprises in English news from countries along the "Belt and Road" initiative, finding that media held an objective and neutral attitude towards the development dynamics of enterprises.

Regarding African studies, Xie and Xia (2023) analyze reports on Sino-African relations on the English website of the Chinese Ministry of Foreign Affairs. They investigated China's national image and diplomatic strategies in China-African relations, discovering that the Chinese perspective on China-African relations primarily focused on cooperation and development, constructing a sincere and friendly brotherly partner image. Sheng (2022) analyzed Nigerian English media reports on China from 2015 to 2021, reveling that China's national image in Nigeria is positive with its contributions in infrastructure, education, agriculture, and public health recognized by Nigerian media.

In summary, although corpus-based CDA research in China has made certain progress, it still has some shortcomings. First, in terms of research subjects, domestic and foreign scholars mainly focus on Western countries and their related media, often concentrating on major news, with insufficient attention to African regions. Additionally, these studies prefer selecting English or Chinese, resulting in few studies on media using local languages. Second, regarding research samples, domestic corpus studies mainly focus on a few media outlets, and the representativeness of corpus sources need further improvement. Therefore, there is an urgent need to conduct corpus-assisted CDA on African and Swahili media, an area that has not yet started in China.

Methodology
Data Collection
The corpus used in this study comprises news reports from the "China/World" (China/Dunia) section of the Swahili channel of China Radio International (CRI), representing China's official stance. These reports cover interactions between China and African countries as well as other countries worldwide. As a significant platform for China's external communication, CRI uses

Swahili broadcasts with new media services to convey China's voice to Africa, particularly East Africa, and to shape a positive national image. The corpus collected includes a total of 1,132 Swahili news reports from the CRI website from March 2023 to July 2024, forming a small specialized corpus. To exclude reports that merely mention China without relevance to China, the frequency of "China" appearing in each report was set to at least three times. Images and image captions were removed, resulting in a total of 572 reports with a total token count of 93,739 words. The token counts for "China" and "Afrika" in the corpus are 3,132 and 298, respectively.

Procedure

First, this study employs the corpus analysis software AntConc 3.5.9 to generate a list of high-frequency lexical bundles for the word "China." This helps identify the frequently occurring lexical bundles in the reports, determining the countries or organizations closely associated with China. Next, based on the results of the high-frequency lexical bundles, the search term "China na" (China and...) is used to retrieve a list of high-frequency collocates. The collocates of the "China na" phrase, presented in a table format, reflect the discourse differences in news reports about various countries and help reveal CRI's tendencies, attitudes, and strategies in constructing China's image in Africa. Then, "China" is used as the search term in AntConc to generate all concordance lines related to China. Analysis of these c

Data Analysis
Word Frequency

By using AntConc 3.5.9 to analyze the corpus, high-frequency possessive markers and conjunctions in the Swahili reports were excluded, retaining only content words. Based on the frequency of occurrence, the top 30 content words most commonly found in the news reports are arranged and listed.

Table 1: High-Frequency Word List

Words	Freq.	Words	Freq.	Words	Freq.
China	3132	zaidi	430	kuhimiza	264
nchi	1021	hizo	400	dunia	260
rais	861	pande	392	mkuu	258
xi	747	kimataifa	367	uchumi	221
kati	741	jinping	355	kuhusu	216
maendeleo	633	mkutano	318	hali	205
ushirikiano	628	bw	313	umoja	201
mbili	522	marekani	312	muhimu	198
pamoja	505	afrika	298	mazungumzo	189
uhusiano	498	mpya	279	waziri	189

It can be learned from the high-frequency word list that "maendeleo" (development) and "ushirikiano" (cooperation) appear frequently, indicating China's active role in promoting China-Africa cooperation and global development. A collocation search using AntConc shows that "China na Afrika" (China and Africa) appears 79 times, while "China na" (China and...) with other African countries appears 62 times. The main countries reported include Kenya, Tanzania, Democratic Republic of Congo and etc. Additionally, CRI expresses an attitude towards China-Africa cooperation by frequently using verbs with positive connotations, such as "kuhimiza" (promote), "kusaidia" (help), and "kuendeleza" (develop). These verbs are prominently featured in phrases like "kuhimiza maendeleo na ustawishaji wa nchi hizo mbili" (promote the prosperity and development of the two countries), "kusaidia maendeleo ya Afrika" (help Africa to develop), "kuimarisha ushirikiano chini ya mfumo wa Baraza la Ushirikiano kati ya China na Afrika (FOCAC)" (strengthen cooperation under the framework of the Forum on China-Africa Cooperation).

These expressions reflect CRI's positive evaluation and supportive attitude towards China-Africa cooperation, accurately conveying China's commitment and determination to seek joint development with Africa. By depending the Forum on China-Africa Cooperation as a crucial mechanism, China implements various aspects of cooperation with African countries.

Concordance Analysis

Corpus linguistics focuses on language usage. Firth (1957) emphasized the importance of context in studying word collocations, believing that the meaning of a word is determined by its context. Therefore, to examine CRI's construction of China's image in Africa, observing and analyzing the concordance lines of "China na Afrika" (China and Africa) is important. The concordance feature of the corpus facilitates large-scale lexical co-occurrence analysis. The table below, generated by AntConc 3.5.9, shows the concordance lines in the corpus for the of words "China na Afrika"(China and Africa).

Table 2: The Representative Concordance of "China na Afrika"(China and Africa)

1	zimeonyesha kuwa, biashara kati ya	China na Afrika	**imeongezeka** kwa utulivu katika miezi
2	uwezo wa **ushirikiano** kati ya	China na Afrika	katika **kuboresha** mnyororo wa viwanda
3	**ushirikiano wa uwekezaji** kati ya	China na Afrika	**umenufaisha** pande hizo mbili,
4	Anaona kuwa **ushirikiano** kati ya	China na Afrika	**umeleta faida kubwa** na **utaimarishwa**
5	**Ushirikiano wa kilimo** kati ya	China na Afrika	umepata **matokeo makubwa** ukiwa na
6	**mawasiliano kati ya watoto** wa	China na Afrika	umezinduliwa mjini Beijing.Sherehe ya

7	na mazungumzo ya viongozi wa	China na Afrika	**yalifanyika kwa mafanikio**, ambapo pande
8	ni umwamba wa habari. Amesema	China na Afrika	zinatakiwa **kuimarisha ushirikiano** wa karibu
9	yake, Bibi Peng Liyuan amesema,	China na Afrika	zote **zinatilia maanani kuhimiza maendeleo**
10	Baraza la **Ushirikiano** kati ya	China na Afrika	(FOCAC) katika utoaji wa mikopo
11	wa Uchumi na **Biashara** ya	China na Afrika	Ikiwa na kaulimbiu ya "Kuongeza
12	**juhudi za pamoja** kati ya	China na Afrika	ili kuimarisha **ushirikiano** na kulinda
13	ya **ngazi ya juu** ya	China na Afrika	Kusini yenye **mustakabali wa pamoja**.
14	**ushirikiano wa kivitendo** kati ya	China na Afrika	Kusini katika **nyanja mbalimbali** umekuwa
15	wa muda mrefu kati ya	China na Afrika	na **kufunzana kati ya staarabu**

From the concordance lines in the table, it is evident that CRI's reports on China-Africa relations emphasize the achievements in economic and trade cooperation as well as cultural exchanges, and they express optimistic prospects for future cooperation. Generally, the attitude is positive. For instance, CRI describes aspects such as "biashara" (trade), "uwekezaji" (investment), and "kilimo" (agriculture), intending to convey that investment cooperation (ushirikiano wa uwekezaji) between China and Africa is mutually beneficial (umenufaisha pande hizo mbili) and that agricultural cooperation (ushirikiano wa kilimo) has yielded significant results (umepata matokeo makubwa). These frequently co-occur with "China na Afrika" (China and Africa), aligning with the main areas of cooperation between China and Africa, reflecting China's emphasis on these fields.

Moreover, CRI often uses verbs that semantically indicate promotion, showing the achievements of China-Africa cooperation and providing public support for strengthening this cooperation, such as "ongezeka" (increase), "tilia maanani" (pay great attention to), "imarisha" (consolidate), "himiza" (promote), and "funzana" (learn from each other). Adjectives and complements like "yenye mustakabali wa pamoja" (with a shared future), "wa kivitendo" (practical), and "nyanja mbalimbali" (various fields) demonstrate China's respect for and inclusiveness towards Africa as the foundation for China-Africa cooperation. This underlines China's determination to pursue comprehensive and in-depth cooperation with Africa, establishing an inclusive and responsible partnership image.

In addition to "China na Afrika" (China and Africa), the discourse of reports on China and other African countries is also worth discussing. The table below shows some concordance results for "China" with Kenya and Tanzania.

Table 3: The Representative Concordance of "China" and Other African Countries

1	**kibalozi**.Bw. Wang amesema tangu	China na Kenya	zianzishe **uhusiano wa kibalozi** miaka 60
2	mbili.**Rais Xi** amesema, tangu	China na Kenya	zilipoanzisha **uhusiano wa kidiplomasia** miaka 60
3	**Mmoja, Njia Moja**" kati ya	China na Kenya	umepata **maendeleo makubwa**, ukiongoza katika
4	**ushirikiano wa kiuchumi** kati ya	China na Kenya	kuendana na **Pendekezo la Ukanda**
5	China inauchukulia **uhusiano kati ya**	China na Kenya	kwa **jicho la kimkakati**, ikiongozwa
6	Amesema China inapenda **kushirikiana** na	China na Tanzania	hadi kuwa **ushirikiano wa kimkakati**
7	mnara wa **urafiki** kati ya	China na Tanzania	mnara wa urafiki kati ya
8	zaidi ya **uhusiano** kati ya	China na Tanzania	na China na Afrika.Hayo
9	"Huu ni mwaka wa 60 tangu	China na Tanzania	zianzishe **uhusiano wa kibalozi**, ikiwa
10	kuanzishwa kwa **uhusiano wa kibalozi,**	China na Tanzania	zimekuwa **marafiki** na **wenzi** wazuri
11	inatilia maanani **uhusiano** kati ya	China na DRC	na anapenda kufanya **juhudi** pamoja
12	(DRC).**Bw. Li Qiang** amesema,	China na DRC	ni **wenzi wa ushirikiano** wa
13	wa pande zote kati ya	China na DRC	**yawanufaishe** zaidi watu wa nchi
14	kati ya vyombo vya habari vikuu vya	China na DRC	na **kupanua mabadilishano** ya watu

From the table, it can be observed that "kibalozi" (diplomatic) and "uhusiano" (relations) appear frequently, indicating that China highly values maintaining and developing diplomatic relations with East African countries. CRI describes the relationships between China and Tanzania, China and the Democratic Republic of Congo as partnerships (wenzi) and friendships (marafiki),

emphasizing China's long-term and stable relationships with these countries on a macro level, symbolizing friendly relations between the peoples of these countries on a micro level.

Furthermore, by emphasizing the "Belt and Road Initiative" (Pendekezo la Ukanda Mmoja, Njia Moja), CRI highlights the strategic cooperation (ushirikiano wa kimkakati) between China and Tanzania and between China and Kenya, demonstrating the depth of cooperation between China and East African countries. By using collocates such as "nufaisha" (benefit), "panua" (expand), and "maendeleo makubwa" (great achievements), CRI stresses the mutual benefits and development opportunities brought by the cooperation. CRI's focus on portraying China's positive attitude towards developing friendly relations with African countries, which makes China's image more comprehensive and multidimensional in Africa.

Conclusion

This paper examines content related to China-Africa relations in the Swahili broadcasts of China Radio International (CRI) and initiates a corpus analysis from a critical discourse perspective. The analysis of corpus frequency and collocation lines reveals how CRI constructs discourse around China's active role in fostering China-Africa cooperation. CRI's reports use positive vocabulary to express favorable evaluations and supportive attitudes towards China-Africa cooperation, showcasing China's positive image as a responsible global player strengthening international ties. Further indexical analysis indicates that economic trade, cultural exchange, and political diplomacy are focal areas of China-Africa cooperation. The reports often describe the relationship between China and Africa as mutually beneficial and accurately convey the mainstream ideas and guidelines of this cooperation, successfully shaping an image of China that is mutually respectful and collaboratively developing with African countries. However, there are deficiencies in the diversity and balance of CRI's reporting. Firstly, the coverage predominantly highlights positive cooperative relationships, with insufficient reporting on challenges and controversies in the process. Secondly, the choice of language and vocabulary tends to describe things positively, and an overemphasis on China's role in China-Africa cooperation could lead to negative perceptions of China's image. Thirdly, reports often focus on the macro aspects of China-Africa interactions, with fewer detailed explanations of specific cooperation policies and mechanisms at the micro level. Finally, the reports are largely from a Chinese perspective, with less attention to African local voices and positions.

Overall, CRI's Swahili reports play a positive role in shaping China's national image. From a Chinese viewpoint, China aims to construct a peaceful, friendly, open, and inclusive image in Africa. However, from a media perspective, there is a need for further improvement in objectivity and comprehensiveness. Media must not only have a deep and comprehensive understanding of China's policies towards Africa and clearly define the specific contours of China's national image construction in Africa, but also delve into the African continent, paying more attention to African voices to enhance Africa's understanding of China. Additionally, the influence of Chinese Swahili media in Africa is far less than that of Western media, making the expansion of China's media discourse power in Africa an urgent task, with a long road ahead for Chinese journalists and language workers.

References

Alluhaidah, M. (2023). Representations of Islam and Muslims in Media Addressed by

Corpus-Based Approaches: Systematic Literature Review. *International Journal of Linguistics and Translation Studies, 4*(4), 187-198.

Attia, S. & Romero-Trillo, J. (2022). A corpus-assisted Discourse Study of the Media Language in the Egyptian Revolution. *Language, Discourse & Society, vol. 10, no. 2*(20), 105-127.

Elsoufy, M. A. (2024). Media Bias Through Collocations: a corpus-based study of Egyptian and Ethiopian news coverage of the Grand Ethiopian Renaissance Dam. *Humanities and Social Sciences Communications*, 11(681), 1-15.

Fairclough, N. (1989). *Language and Power*. London: Longman.

Firth, J. (1957). A Synopsis of Linguistic Theory, 1930-55. *Studies in Linguistic Analysis*. Special Volume of the Philological Society. Oxford: Blackwell, 1-32.

Fowler, R. (1996). *On critical linguistics*. In C. R. Caldas-Coulthard & M. Coulthard (Eds.), Texts and practices: Readings in critical discourse analysis. London: Routledge.

Ndlovu, M., & Nikabs, M. (2023). COVID-19 and the constructions of Africa in African news media. *Journal of African Media Studies*, 15(2), 179–198.

Stubbs, M. (1996). *Text and Corpus Analysis*. Oxford: Blackwell.

van Dijk, T. (1988). *News as Discourse*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Yang, M., & Wang, Z. (2023). A corpus-based discourse analysis of China's national image constructed by environmental news in The New York Times. *Humanities and Social Sciences Communications*, 10(545), 1-11.

陈建生 & 陈香君.(2017).西方媒体"眼"中的中国国家形象变化——基于语料库的研究.广东外语外贸大学学报(05),104-109+117.

钱毓芳 & 王琮涵.(2023).《人民日报》对"碳达峰碳中和"的话语建构——基于语料库的话语研究.中国石油大学学报(社会科学版)(04),69-77.

钱毓芳.(2010).媒介话语研究的新视野:一种基于语料库的批判话语分析.广西大学学报(哲学社会科学版)(03),80-84.

盛浩.(2022).基于语料库的尼日利亚英文媒体的中国国家形象建构研究硕士(学位论文,吉首大学).硕士 2022.000283.

李翼.(2024).新冠疫情背景下英国主流报刊中的北京形象研究——基于语料库的研究.传媒论坛(07),46-53.

王琦.(2019)."一带一路"英文媒体中的中国企业形象——语料库辅助下的批评话语分析.当代外语研究(03),99-113.

谢文琦 & 夏云.(2023).中非往来中的中国国家形象建构——基于语料库的报道话语分析.江苏外语教学研究(02),70-74.

平行文字与相交文字

—— 新视角下浅论世界文字区别

涡阳县城西学区宋庄中学　　杨柳

摘要：

　　语言学研究一个重要任务就是通过发现语言中客观存在的规律，让人们对语言有更加理性的认识，本文首次成功地揭示了世界文字区别的客观规律，并发现了世界不同文字的基本笔画结构具有平行排列和相交排列的规律，从而得出世界文字分为平行文字和相交文字的新认识，因此使人们摆脱了对于文字区分的认知困惑，也为语言学的研究注入新的活力。

关键词：

　　文字　区别　笔画　表音　表意　平行　相交

序言

众所周知，不同语言文字是不同地区人们智慧的结晶，也是民族文化的载体和不同民族文化产生的标志之一。世界上的文字是多元化的发展趋势，文字的出现，可以让世界的文化得以传播和发展，也方便人们记载历史文化等。所以文字是一个不可缺少的文化载体，也是世界文化博大精深的体现。但是不同国家有不同的文字表达方式，也是世界文化多样性的体现，随着近代以来的交流加快，人们对文字区分的认识也不断加深，也出现了不同的区分方式，例如表音文字和表意文字的区分、东方文字与西方文字的区别、拉丁语系、印欧语系、亚非语系等，人们通过比较不断加深对本民族文字的认同感。

从发展的观点看，语言的认知中会不断有新的认知出现，因此继承前人的看法，也要用批判辩证的观点加以理性科学分析，不断发现新的认知规律，这是语言学研究的共识，文字学的研究也是如此。

一、中国文字的起源和发展变化

周有光教授把文字分为原始文字、古典文字、字母文字，通常认为，原始文字起源于象形图画，例如刻符、岩画、文字画等。（世界文字发展史：3）文字的发展变化是人们社会实践的产物，是人类文明进步的需要。文字发展演化的规律需要人们通过不断探索、不断求证去发现。

我国古代神话关于文字的传说有许多，比如伏羲造字说和仓颉造字说。（中国文字起源与变迁：6）。另外，《尚书·序》中说："古

者伏羲氏之王天下也，始画八卦，造书契，以代结绳之政，由是文籍生焉。"（中国古代甲骨文 5）由于结绳的区分度很低，所以记录的功能很弱，并不具备文字的要素，只是一种辅助记忆的手段。（中国文字发展史：14）这些关于文字的传说虽然没有充分的证据去证实，但是这也一定程度上反映了人们对文字的理解与使用。

甲骨文是迄今所知的最古老而且已经比较成熟的汉字形态。（中国文字发展史：17）甲骨文的图像性强，象形意味很浓，是一种非常古拙的汉字早期形态。（中国文字发展史：17）

人们把从甲骨文到小篆的阶段称为古文时期，这一时期的汉字和现代汉字的差别较大，而从隶书开始，再到现代汉字，形体上没有太大变化，称为文字阶段，因此隶书成为古文字和现代文字的分水岭。（中国文字发展史：31）

二、世界其它文字的起源与发展

世界其它地方的文字发展历史悠久，据丹尼丝教授提到，文字起源于公元前 4 千世纪晚期的美索不达米亚，即现在的伊拉克地区，继而从那里传播到埃及、埃兰和印度和流域，据考古发现楔形文字起源于陶筹，它是一种用于商品计数和记账的工具（文字起源：2）。

陶筹大约起源于公元前 8000 年前，最初是为了记录农产品的数量，并且陶筹是楔形文字的前身。（文字起源：13）陶筹是小的、由黏土制成的物品，形状各异，有圆锥体、球体、圆柱体、卵形体等，并且大小不一。（文字起源：26）

有文字学者认为，字母文字是平民文字，它在实用中生长，为商

旅所掌握，字母文字取代丁头字，是在中东地区的经济和文化进入全新时期才发生的。（世界文字发展史：179）

人类尝试把语言写成文字，语言是一个个语词链接起来，但是语词数量极多，以语词为单位，写成的文字十分繁难，应用不变。后来发现语词是由音节组成的，如果每一个音节设计一个符号，符号数量会大大减少，文字进入"音节阶段"。（世界文字发展史：173）再后来，人们发现音节还可以再分，从表意兼表音的"语词-音节文字"发展为音节性的"辅音字母"，进一步又过度为"音素字母"。（世界文字发展史：174）

后来，字母发展为四个主要的字母系统，即迦南字母系统、阿拉马字母系统、南方闪米特字母系统、希腊字母系统。（世界文字发展史：182）希腊字母又演化为拉丁字母和斯拉夫字母，随着信息化的发展，拉丁字母成为国际交流的主要字母。（世界文字发展史：294）

总之，字母后来传播四方，其中最为重要的一条传播路线是，从腓尼基传到希腊，由希腊补充了元音字母，成为"音素字母"，再传到罗马帝国时期，成为罗马字母，最初只有 21 个，（世界文字发展史：381）后来又加上"J、U、W"，（改变西方世界的 26 个字母：202）到中世纪时才成为 26 个字母，这就是传到今天的现代罗马（拉丁）字母。（世界文字发展史：381）

三、 文字的分类

从文字的起源探究中，我们发现文字发展的历程，文字进化论认

为人类文字在不自觉中有一种共同的进化规律。（比较文字学初探：9）但是，这一规律究竟是什么？人们对东西方文字的差异产生了诸多困惑，汉字在东方使用了 3000 多年，跟西方的文字风马牛不相及，汉字没有演变为字母文字，因此也就认为所谓的从表形到表意到表音的文字发展规律缺乏事实佐证。（比较文字学初探：26）

不难看出，人们不断尝试解释文字类型，寻找文字区别的奥秘，这一过程显得辛酸，所谓的发现规律，却又被否决，这就是发现过程的神秘之处，也指引学者不断探索发现，希望揭示人类文字区别的神秘面纱。

一次偶然的机会，我不断的重复着语言学文字的音形意三要素，同时融合了那时困窘的处境，我脑海里浮出一个概念"平行和相交文字"，然后我进行长达十余年的揣摩，无数次的和一些摸不着头脑的人唠叨着"我发现了文字的区别"，每次也总是忐忑地怀疑自己的发现，同时也在不断寻找"证据"去支持这一发现。

1、平行文字

从新的角度观察，文字的形状尽管也存在变化，但是比较语音的变化显得比较小，人们从刻符到现代成熟的文字体系发展中，也发现文字的形状相对稳定。

平行文字是基于数学中平行线段概念，尝试用这种角度对文字的笔画架构进行分析，例如英语中的 26 个基本字母笔画，它们就是英语文字的最小切分单位，类似物质的分子成分。下面我们从几类有代表性的语言中去验证这一新的发现。

（1）日耳曼语族

分为北日耳曼语支（包括瑞典语、丹麦语、挪威语、冰岛语等）、西日耳曼语支（英语、德语、荷兰语等）、东日耳曼语支（现已灭绝，唯一已知文字的是哥特语。）

从平行线的图画中受到启发，如图： | | | | | ，每一条之间存在平行关系，再联系英语字母书写顺序"able"，字母笔画是一个接着一个按照平行趋势排列，我们不难发现每一个字母之间也存在平行的关系，因此我们大胆猜出结论，英语文字笔画之间存在平行关系，也可以依此类推日耳曼的其它语言的笔画这种排列规律。

语种	文字	笔顺	笔画关系	备注
英语	book	b-o-o-k	平行	
德语	flumen	f-l-u-m-e-n	平行	
荷兰语	vrouwen	v-r-o-u-w-e-n	平行	
瑞典语	stad	s-t-a-d	平行	
丹麦语	skole	s-k-o-l-e	平行	
挪威语	elver	e-l-v-e-r	平行	
冰岛语	eyja	e-y-j-a	平行	
哥特语				灭绝

（2）罗曼语族

印欧语系的罗曼语族一般包括中部罗曼语，例如法语、意大利语、加泰罗尼亚语等；西部罗曼语，例如西班牙语、葡萄牙语等；东部罗

曼语，例如罗马尼亚语等。下面我们分别举例说明这些语言文字笔画之间的平行特点。

语族	文字	笔顺	笔画关系	备注
法语	pommes	p-o-m-m-e-s	平行	
意大利语	scuole	s-c-u-o-l-e	平行	
加泰罗尼亚语	escola	e-s-c-o-l-a	平行	
西班牙语	fruta	f-r-u-t-a	平行	
葡萄牙语	cavalos	c-a-v-a-l-o-s	平行	
罗马尼亚语	ploaie	p-l-o-a-i-e	平行	

不难发现，这些语言的文字也可以划分为平行文字。

（3）印欧语族

印欧语族的两个语支：印度语支和伊朗语支。代表性语言分布于中东和印度次大陆。两个语支都可以被划分为三个时期：古代、中古和现代。古代印度语支以梵语教义为代表，最早的形式是吠陀语，可追溯到公元前两千年。伊朗语的最早阶段以公元前 6 世纪的《阿吠斯陀》和古波斯语为代表。

现代波斯语是伊朗的官方语言，印地语是印度共和国的官方语言之一。

语族	文字	笔顺	笔画关系	备注
波斯语	تفاحة .	ت - فا -ح - ة .	平行	
印地语		– –	平行	
加泰罗尼亚语	escola	e-s-c-o-l-a	平行	

2、相交文字

我国很久以前就对文字的结构有认知，根据《说文》："文，错画也，象交文。"，另据《易经·系辞》："物相杂故曰文"。（文字学常识 7）从中我们可以推测出"文"的含义有交错的意思。 周有光教授认为，单体符号称为"文"，复合符号称为"字"，后来统称为"文字"。）他还提到：如果把复合符号分开来作线性排列，不叠成"方块字"，那么书写现代汉语只要 800 来个符号就够了。叠成"方块字"使字数增加，造成字无定量。（比较文字学初探：115）

我们发现古时候人们就对汉字有了一定的认知，通过近现代学者的研究，也已经发现我国文字的雏形具有相交的特征，而西方文字是线性排列。因此，从我国文字的起源发展可以发现，我国与世界其它文字有不同之处，却又找不到一个解释的平台，恰恰作为一对范畴，平行和相交使我们对文字的这种区分找到了一个方向指引。

下面我们通过举例证明这一观点，

语族	文字	笔顺	笔画关系	备注
汉语	天	一，一，丿，丶	相交	
韩语	쌀	ㅆ-ㅏ-ㄹ	相交	
日语	人	丿--丶	相交	
越南语	Nước	N-ư-ớ-c	平行	

日本正式学习汉语汉字文言文，大致始于中国晋朝时候（公元

3-4 世纪），日在本熟悉了汉字之后，开始借用汉字书写日语，其中《万叶集》标志着借用汉字书写日语的成熟。（比较文字学初探：328）

同时我们发现日语文字又是是一个大杂烩，日语文字是由汉字、假名和罗马字构成的。我们通过对比分析，日语文字有不少受到中国文字的影响，采用了中国的文字作为本国文字的基础，同时又受到西方文字结构的影响，文字中掺夹着西方文字，因此总体上来说，日语文字作为一种特例，也应该认为文字笔画之间关系属于相交文字。

类似的还有韩语文字，在公元 3 世纪左右，汉字传入朝鲜半岛，后来又采用汉字的音和意来记录朝鲜语，即"吏读文"。1446 年，朝鲜王朝世宗大王李祹正式公布新创制的文字，诏书称为"训民正音"，意即教百姓以正确字音。正式名称为"谚文"，和日语的"假名"意义近似（日本称汉字为"真名"）。朝鲜（韩）文有 21 个元音和 19 个辅音，任意元音字母与辅音字母均可单独成字或组合成字。因此，韩语文字的笔画结构是相交关系。

另外，在越南，文字大约从东汉开始，汉字（越南人称之为"儒字"）开始有系统和大规模的传入越南。到了越南陈朝以后，汉字已经成为越南政府以及民间的主要文字，此时大量的汉字著作开始出现，最著名的就是 15 世纪编撰的《大越史记全书》。

在使用汉语汉字大约 1000 年之后，越南利用汉字书写本民族的语言，并以汉字为基础发展南字(越南文:喃字)（文言文的写字顺序从右到左:字喃）。越南有几个短暂的时期，以南字(喃字)为正式文字，跟汉字并行，多数时期只用于民间，不作为正式文字。喃字的发展包

括逐步发展时期（十三世纪到十八世纪上半叶）和繁荣时期（十八世纪上半叶到十九世纪上半叶）。

然而，越南在 1885 年后，法国神甫罗德设计了拉丁化的拼音文字，废弃了喃字。（比较文字学初探：222）因此，现在的越南语文字属于平行文字。

四 结语

我们通过新的视角，对平行文字和相交文字的分类进行了初步的论证，进而发现了世界文字区分的客观事实，这也是对国内外语言学研究中关于文字分类理论的继承和发展，我们有理由对这一发现做进一步的研究，期待能得到国内外语言学相关研究专家的支持。

参考文献

中国文字起源与变迁　吴贯因　山西人民出版社　2014

文字学常识　胡朴安著 王其和整理 山东画报出版社 2017

文字起源　丹尼斯 •施曼特-贝瑟拉 【美】著 王乐洋译 商务印书馆 2021

古埃及象形文字 佩内洛普• 威尔逊 【英】颜海英译　海南出版社 2021

人类文字简史 【英】爱德华• 克劳德著 周挺译华文出版社　2021

世界文字简史 （第三版） 周有光著 上海教育出版社　　2018

古文字学初阶　李雪勤著　　中华书局　　　2019

中国文字发展　谢芳芳编著 山东科学技术出版社　2017

中国古代甲骨文 张湘江著　中国商业出版社　　　2022 年

改变西方世界的 26 个字母【英】约翰•曼著　江正文 译三联出版社 2018

比较文字学初探　　周有光著　语文出版社 2012 年

字母表效应：拼音文字与西方文明

古汉语语音：孙常叙著　上海古籍出版社　2018 年

Translation Studies

翻译研究

Vulgarization in the English-Chinese Translation of Swearwords in Subtitles of TV Series

Abstract: The study investigates the translation of swearwords from English to Chinese in TV series, focusing on how swearwords of varying vulgarity levels are translated in Chinese. Results data analysis reveals a mixed employment of translation strategies, with intensification being the most common and omission the least favored. With Correspondence Analysis, it is generalized that different swearwords are translated on the basis of their vulgarity categories. The overall translation outcome adopts a moderate approach, intensifying less vulgar swearwords and softening more vulgar ones. These findings are attributed to internet censorship and the pursuit of humor. The Vulgarization Hypothesis is partially validated, indicating cultural dependence in AVT. The study provides an indication by filling a gap by examining the relationship between swearword categories and translation strategies, offering insights into the translation of taboo language in AVT.

Key words: swearwords, audio-visual translation, vulgarity, English-Chinese translation, correspondence analysis

作者：申健

单位：北京外国语大学

Chapter 1 Introduction

Pragmatics, as a discipline, concerns itself with how language is utilized in specific contexts and how speakers create meaning in communicative acts (Guillot, 2017). Consequently, the intersection of pragmatics and AVT research offers valuable insights into the presentation of fictional exchanges on screen and the reception of information by viewers, considering both the auditory and visual dimensions of the audiovisual product (Chaume, 2012). Research on swear words, vulgar, offensive, and taboo language in audiovisual translation (AVT) has traditionally not received significant attention in academia; however, this trend appears to be shifting with the recent surge in studies and publications in this field over the last few decades (Avila-Cabrera, 2023).

Swearwords, originating from desemantised taboo expressions used to express emotions and convey attitudes, inherently possess offensive potential that varies in intensity and can be modified through interaction (Hughes, 2006; Stenstrom, 2020). Swearing is context-sensitive and primarily associated with informal registers of language (Stapleton, 2010, 2020; Stapleton et al., 2022). In the portrayal of daily interactions in both original and translated films, swearwords serve both intradiegetic and extradiegetic functions, depicting characters' emotional responses and stylistic choices, contributing to narrative progression, and influencing the audience's reception (Baines, 2015).

Swearing encompasses semantic, sociolinguistic, pragmatic, and cross-cultural dimensions (Stapleton, 2003, 2010). While the primary taboo domains are consistent across Western cultures and primarily include deities, sexual practices, body parts, and fluids, individual linguacultures exhibit unique preferences regarding the semantic domains targeted by interdiction, as well as the pragmatics, frequency, and syntactic constraints of swearing expressions (Hughes, 2006; Hoeksema, 2018; Pavesi and Zamora, 2022). Despite these cross-linguistic differences, AVT research has often overlooked them, instead focusing on censorship-driven sanitation strategies and the apparent lack of rationale behind translational choices (Bucaria, 2009; Beseghi, 2016). In contrast, it is argued that examining the grammatical, textual, and pragmatic functions of swearwords is crucial for understanding the coherence in translational decisions (Pavesi and Formentelli, 2023). Therefore, the present study is placed at the intersection of the pragmatics of swearing and AVT research by

researching into the correspondence between swearwords and the respective translation strategies in terms of English TV series translated into Chinese.

Chapter 2 Literature Review

2.1 Taboo Words, Swearwords and Impoliteness

According to Allan and Burridge's (2006) definition, taboo refers generally to the prohibition of certain behaviors within a specific community or group of people, during particular times and in identifiable contexts. The definition was further elucidated by stating that taboos are restrictions on behavior that stem from societal norms, where certain actions are seen as potentially causing discomfort, harm, or injury to others, and has been broadened to encompass political and social matters, and expanded to include the prohibition of using or engaging in anything, particularly expressions or topics deemed offensive and thus avoided or banned according to social norms (Allan, 2019). The concept of taboo is a complex and multifaceted one (Sabri et al., 2010), encompassing a range of meanings that have been ascribed by scholars across various disciplines within the social sciences and humanities. In everyday language, taboo is often associated with forbidden words or topics, such as swearing, as well as subjects that are typically avoided in open discussion, particularly those related to death, sexuality, and menstruation (Rossolatos, 2017). While the original conceptualization of taboo was relatively narrow, its interpretation has evolved over time, with its scope expanding through recontextualizations in scholarly discourse and its application in different cultural and communicative contexts. Taboos are seen as mechanisms of social control, reflecting class, gender, and racial dynamics, cultural norms, legal values, as well as the attitudes and beliefs of various subcultures and countercultures (Boker, 2010).

In the discussion among literature, the use of taboo languages and/or words is related to swearing. Swearing is a linguistic behavior rooted in taboo, referring to actions or expressions that are considered forbidden or prohibited (Stapleton, 2010). According to Love (2021) swearing is often categorized as a form of "bad language" (pp. 742), directly linked to taboo subjects such as sex, bodily functions, and religion. However, it can also be employed figuratively to serve various purposes, such as expressing abuse, humor, or emotional intensity. Due to its taboo nature, swearing seldom goes unnoticed in interactions, providing speakers with a means to clearly convey their stance in noticeable and often assertive manners. This enables them to assert a dominant position within a group, as these strong expressions of stance can be seen as justified and significant to be

expressed and shared (Debray, 2023). Swearing is subject to censure both at individual and societal levels (Jay, 2009), imbuing it with a unique interpersonal and psychological impact. At the individual level, swearing is often linked with the expression of intense emotions (Dewaele, 2016; Stephens et al., 2009). However, due to its taboo status and complex social norms regarding appropriateness, swearing serves various interactional and identity functions not easily fulfilled by other linguistic means (Stapleton, 2020). These functions include expressing both positive and negative emotions, generating humor or emphasis, fostering social cohesion, and constructing or displaying identity (Stapleton, 2010). Additionally, swearing plays a role in conversational narratives (Norrick, 2012), aiding in obtaining the floor, justifying narrative tellability, marking climaxes and closings, and evaluating narrative action. Despite its association with impoliteness and aggression (McEnery, 2006; Culpeper, 2011), swearing can also convey positive politeness and a concern for maintaining face (Daly et al., 2004; Christie, 2013). This distinction hinges on the hearer's perception of the speaker's intentions in a given context, particularly whether offense was intended (Culpeper, 2011).

The social acceptance of swearing is growing due to its widespread use. This suggests that people are becoming less sensitive to and likely to be offended by the use of swear words, even though they remain taboo. Dynel (2023) studies the widespread use of the hashtag #FuckPutin swearing on Twitter and suggested that the tabooed expression, that is "Fuck", has been embraced by many users, indicating that it may no longer be considered shocking or inappropriate. However, the presence of the swear word ensures that the hashtag retains its attention-grabbing quality, as swearing still carries taboo connotations despite its positive interactional effects. In addition, the utilization of taboo words in communication serves to foster and sustain contact between speakers, while also bolstering feelings of camaraderie, thus highlighting the interconnectedness of phatic and interactional functions (Stenstrom, 2014).

In the realm of pragmatics, the incorporation of swearwords has been associated with instances of impoliteness (Valdeon, 2020). Impoliteness, defined as a negative stance towards specific behaviors within particular contexts, stems from societal norms and expectations, shaping how individuals or groups are perceived during interactions. Behaviors are deemed impolite when they diverge from expected or desired standards, often leading to emotional repercussions, particularly

causing offense for at least one participant (Culpeper, 2011). Impoliteness is a dynamic concept necessitating a process-oriented perspective on conversation, as it extends beyond single utterances to encompass extended discourse (Mills, 2003: 38; Bousfield et al., 2007; Culpeper et al., 2003; Terkourafi, 2005). While impoliteness is occasionally presented as a deliberate strategy to disrupt social norms, it typically results from a misalignment between the speaker's intentions and the expectations of the audience, especially in cross-cultural communication contexts (House, 2012; Culpeper et al., 2010). Taboo language represents a form of offensive language use that exacerbates impoliteness (Leech, 2014: 299). It functions as a positive impoliteness strategy, intensifying linguistic impoliteness (Culpeper, 1996; 2019). Additionally, taboo language serves as evidence of religiously aggravated hate crime language (Culpeper, 2017).

2.2 Audio-Visual Translation and Taboo Words/Swearing

Audiovisual Translation (AVT) has undergone significant changes since its inception as a new research field in the mid-90s. Initially, it aimed to establish itself within Translation Studies, documenting primary concerns and limitations across different modes, including subtitling, dubbing, audio-description, and others. The goal was to classify issues and strategies for addressing them within their multimodal contexts (Guillot, 2020). Over time, AVT has broadened its disciplinary scope and now plays a more prominent role in pragmatics and intercultural communication. The expansion of AVT into the realm of pragmatics has led to a diversification of research topics, including communicative practices, (im)politeness, taboo language, dialects, and language impairments (Bruti and Zanotti, 2016; de Pablos-Ortega, 2019; Ghia, 2019). A key question for pragmatics is to understand AVT's cross-cultural or intercultural impact on its audiences. This inquiry motivates further collaboration between pragmatics and AVT research from various perspectives. Specifically, mainstream film subtitling offers insights into the interplay of different cultural and linguistic elements within AVT (Guillot, 2020).

The depiction of swearing in audiovisual dialogue differs significantly from face-to-face conversation, with the former often carrying negative connotations and associations (Pavesi and Formentelli, 2023). Swearing in audiovisual texts serves various functions intrinsic to the genre, such as creating realism, humor, and conveying ideologies while controlling viewers' emotions.

Additionally, it plays a crucial role in pushing the narrative forward in the fictional world portrayed on the screen (Beine, 2015; Bednarek, 2019). Scriptwriters utilize swearing to construct characters' social and identity features, often associating it with specific sociolinguistic features and stereotypes, such as age, gender, and social class (McEnery and Xiao, 2004; Drummond, 2020). Within the realm of AVT, extensive research has focused on translation strategies and the degree to which taboo expressions are rendered across different language and cultural contexts. These strategies range from omission to intensification, including loss of taboo components, mitigation, compensation, literal translation, etc.(for more information, see Khoshsaligheh and Ameri (2014), Avila-Cabrera (2015), Díaz-Perez (2020), Valdeon (2020) and Formentelli and Ghia (2021))However, studies have shown a prevalent trend of omitting or toning down taboo expressions in AVT, impacting characterization and causing pragmatic and stylistic loss (e.g., Dobao and Maria, 2006; Formentelli and Monti, 2014). The phenomenon of reducing swearing in AVT applies to both subtitling and dubbing, despite the absence of space or time constraints in the latter mode that would justify such omissions. This significant decrease in the depiction of taboo language in AVT has often been attributed to the widespread censorship of forbidden meanings and expressions. Translators typically adhere to the restrictions imposed by clients and production companies, leading to the suppression or toning down of swearwords and other taboo elements in the translated audiovisual content (Soler Pardo, 2013). This has led to the perception of taboo expressions as a homogeneous phenomenon within translation practices, suggesting a degree of arbitrariness in the choices made by translators. There are no consistent patterns indicating a specific rationale for either omitting or toning down swearwords or other potentially offensive elements in the translated content (Bucaria, 2009).

Amid the general tendency to standardize and cleanse audiovisual language by diminishing the use of swearwords in translation, there are differences observed among particular language pairs, highlighting that the presence of vulgar language in media largely depends on cultural factors (Dynel, 2012). The vulgarization hypothesis, as outlined by Valdeon (2020: 261), suggests that contemporary American and British programs, when dubbed into European Spanish, tend to feature an increased use of swear words. This phenomenon is notable not only in translation in general but also specifically in AVT (Santaemilia, 2008). This finding is significant as it can inform translator

training programs and offer insights into the evolution of AVT practices and the incorporation of swear words in mass media (Valdeon, 2020). Research on the vulgarization phenomenon in translation of swearwords has been expanded to other European languages. For example, Giampieri (2017) delved into the translation of English swearword into Italian but the results were inconsistent with the vulgarization hypothesis; rather, the findings suggested that swearword translation from English to Italian tends to reduce or tone down the sense of vulgarity, as are other similar cases, including Avila-Cabrera's (2016) research into English-European Spanish translation of swearwords and Diaz-Perez's (2020) analysis of the English-Galician Veiga translation of 'shit' and 'fuck'.

However, most research in AVT and swearword translation has been conducted between English and other European languages, while Asian languages remain under-surveyed. And current research has been conducted based on the translation strategies of swearwords but neglected how translators might deal with different swearwords in accordance with the pragmatic circumstances. Therefore, the present study shifts its research focus on swearword translation from English to Chinese and further classifies how swearwords of different vulgarity levels in English TV series are translated into Chinese accordingly.

Chapter 3 Research Methods

3.1 Research Questions

(1) Is there any difference in English-Chinese taboo translation in terms of vulgarity? If so, does the vulgarity embedded in source texts tend to be softened or intensified or in a mixed manner?

(2) Are English swearwords translated differently in audiovisual translation to Chinese in terms of their vulgarity?

3.2 Research Procedure

The present study delves into the English-Chinese translation of swearwords in English TV series. The data for the study were derived from an American TV comedy named *Two Broke Girls*, where taboo language and swearwords are profoundly noticed, and what was collected was subtitles with English swearwords or taboo languages along with the translated Chinese subtitles. The data were collected from the subtitles of the series from Tencent Video, one of the leading online video streaming platforms in China, offering a wide range of premium content including movies, TV dramas, variety shows, and original productions. After collecting the subtitles, the data were coded according to their translation strategies, which are based on Valdeon's (2024) research with the classification of omission, preservation and intensification. In order to be more specific in terms of the effect of swearwords and their impoliteness (or vulgarity, as will be used in the following part), toning down is added to the strategies and the subtitles were coded accordingly. For a clear demonstration, examples of the translations of *ass* are given below.

> Omission: …, get up off your <u>ass</u> and cook the damn fish.
>
> ……，定吼他们把鱼给我煮熟了。
>
> Preservation: Feel like Han's more of an <u>ass</u> man.
>
> 憨是爱屁股型的男人吧。
>
> Toning down: Yeah, a connection to her babies' <u>asses</u>.
>
> 对啊，直通她宝宝的"后门"
>
> Intensification: …right up Snooki and her cousin's <u>asses</u>.
>
> ……给山寨斯努奇和她表姐的狗眼瞧瞧

For omission, the sense of vulgarity of *ass* is lost by translating the phrase into "shouting angrily", which enhances the emotional effect rather than vulgar effects. For preservation, the translation of *ass* remains in its commonsense use, that is the rear end, and vulgarity remains unchanged in this manner. For toning down, *ass* is translated by metaphorizing an ass into a back door, the sense of which is similarly consistent with the human body, and more sense of humor rather than that of vulgarity is added. For intensification, *ass* is translated in an impolite manner in that in Chinese *gouyan* (dog's eyes) is regarded as a rude expression that causes insults. Therefore, this translation intensifies the vulgarity of *ass*. The rest of data will be coded according to it. In addition, as a result of internet censorship, some episodes were edited by deleting the sensitive contents or deleted as a whole. In this sense, the present study sourced back to the original video series and regarded the translation strategy in those "lost" contents as omission because of deletion.

In terms of the vulgarity level of swearwords, McEnery (2006) has concluded 'a scale of offence' (pp. 31) to categorize the swearwords according to the extent of offence. The present study will employ the categorization with proper adaptations, as in Table 3.2. By coding the translation strategies and categorizing the swearwords to the scale, the results will be uploaded to R Studio for Correspondence Analysis, in which manner whether swearwords of different vulgarity level are translated differently will be answered.

categorization	Words in the category
very mild	bird, bloody, crap, god, hell, hussy, idiot, pig, pillock, sod, son-of-a-bitch, tart,holy, nerd
mild	arse (ass), balls, bitch, bugger, christ, cow, dickhead, git, jesus, jew, moron, piss off, screw, shit, slag, slut, tit, tosser
moderate	arsehole (asshole), bastard, bollocks, gay, nigger, piss, paki, poofter, prick, shag, spastic, twat, wanker, whore, prostitute, hooker, sucker
strong	fuck
Very strong	cunt, motherfucker

Table 3.2 a scale of offence of swearwords adapted from McEnery (2006:31)

Chapter 4 Data Analysis and Discussion

As is mentioned in the research method, the first step of the analysis was the classification of the translations using the three categories abovementioned. Table 4.1 illustrates the statistical results of translation strategies of English swearwords into Chinese in terms of vulgarity (or impoliteness, in a very general sense) in *Two Broke Girls*.

translation strategies	number	percentage
Intensification	12485	42.15%
Preservation	9499	32.07%
Toning down	6160	20.80%
Omission	1476	4.98%
Total	29620	100.00%

Table 4.1 Translation Strategies in *Two Broke Girls*

Based on the results, it is clear that intensification is the most frequently used translation strategies when English swearwords in English TV series are translated into Chinese, and it is followed immediately by toning down. Omission is the least frequently used translation strategy among the four. The results show that in English-Chinese ATV, English swearwords tend to be translated with intensified vulgarity, which is consistent with Santaemilia (2008) and Valdeon (2020), and it can provide evidence for Vulgarization Hypothesis in English-Chinese ATV. For a comedy like Two Broke Girls, the intensification of vulgarity prevails with the function of creating a humorous effect (Valdon, 2020), though it might be a dark one. For example, in the following subtitles:

e.g.:   This one says "Screw U"   这个写着"滚你妈蛋"

And really let us bitches have it!   专门为我们的贱人朋友准备

The two examples manifest an intensification of vulgarity in English-Chinese swearword translation, and a sense of (dark) humor has been enhanced. The two subtitles are taken from the same episode and the same scene where the two main characters, Max and Caroline, try to learn how to decorate

cupcakes with cream flowers from two Italian unreasonable bakers. The first one is uttered when Max lost her patience in making a perfect cream rose on the cupcake and decided to make insulting verbal decorations on the cupcake, and *Screw U* was written on it surface, which is translated into a more insulting expression in Chinese. Originally, the English expression is to show Max's personality of being positively cynical and her unique pursuit of beauty that beauty is not defined by others but by ourselves. Therefore, Max says "Screw U" to the bakers, making a (dark) humorous effect, and the intensification of vulgarity enhances the effect. Similarly, the second subtitle is said by other two gay bakers who will hold a bachelor party before their wedding and are in need of cupcakes. They love the insulting cupcakes and wish to get them for their friends. Calling good friends as bitches expresses their familiarity and creates a sense of humor. From bitches to *jianren pengyou* (bitchy friends), the vulgarity is intensified for the former refers only to friends, which makes the utterance more interesting.

However, vulgarity that is toned down in the translation process is unneglectable, and this might be explained by Internet Censorship. Internet censorship is the approach of governors or regulators to deciding what information is legitimately allowed to be published, accessed, and spread on the Internet and this method is prevailing within countries of both democratic and authoritarian ideologies (Chang and Lin, 2020). In China, the Internet censorship system has experienced an expeditious development and is renowned historically and internationally (King et al, 2014). Its filtering system is regarded "the most sophisticated" (Zittrain, 2017: 8) tool that surveil Internet information containing sensitive words or contents about the Communist Party of China (CPC). The criteria of censorship are multifarious among platforms, but the main orientation is towards more meticulous manipulation (Sun and Zhao, 2022). For a TV series abundant in expressions that are not in conformity with Chinese ideologies and for a video platform like the Tencent video to publish the TV series, carefully translated subtitles play a crucial role. From this point of view, the toned-down vulgarity of English swearwords is employed in light of national, political and ideological inclinations. Internet censorship in China also contributes to the omission of vulgarity in translation, for it cut off scenes and episodes that are unsuitable against a Chinese backdrop.

The second step of data analysis will delve into whether swearwords in different vulgarity categories are inclined to be translated in different translation strategies, and it requires a calculation of every swearword and the respective translation strategies. Table 4.2 illustrates the swearwords that are uttered on the screen and statistic results of the translation strategies. In order to be more succinct, Table 4.2 lists the lexemes of swearwords based on the scale of offence in Table 3.2, and the lexemes are ordered according to their vulgarity categories, from very mild to very strong.

Swearwords	Omission	Toning down	Preservation	Intensification
crap	31	96	217	455
damn	43	142	344	746
god	89	253	345	637
hell	72	142	295	485
son of a bitch	54	192	389	860
idiot	66	96	237	325
bloody	44	89	129	193
holy	25	102	189	301
nerd	34	87	131	224
ass	67	347	603	943
ball	63	139	427	625
bitch	106	313	539	812
Christ	43	87	306	488
cow	11	31	65	77
Jesus	23	98	282	303
moron	12	36	155	147
piss off	55	139	315	573
screw	44	358	633	509
shit	82	332	553	784
slut	80	302	653	378
asshole	28	169	317	429
bastard	29	108	255	218
piss	98	370	256	319
prick	9	73	69	53
wanker	11	59	67	44
whore	45	351	446	299
prostitute	23	156	237	372
hooker	43	316	473	413
sucker	14	58	75	67
fuck	65	772	353	308
motherfucker	67	347	144	98
total	1476	6160	9499	12485

Table 4.2 swearwords and the translation strategies applied

Based on the data presented in Table 4.2, correspondence analysis was conducted and it obtained the correspondence results between swearwords and the translation strategies (see Table 4.3). According to Table 4.3, dimensions 1 and 2 together explain 92.4% of the data distribution, while dimension 3 only explains 7.6% of the data. Increasing dimension 3 does not significantly improve the explanation of data variance. The adequacy of the correlation analysis indicates a good fit of the data. Therefore, the relationship between variables can be adequately represented using a two-dimensional coordinate system, as shown in Figure 4.1.

	Dimension 1	Dimension2	Dimension 3
Singular value	0.174	0.069	0.014
Inertia explained ratio	0.817	0.107	0.076
cumulative inertia ratio	0.798	0.893	1.000

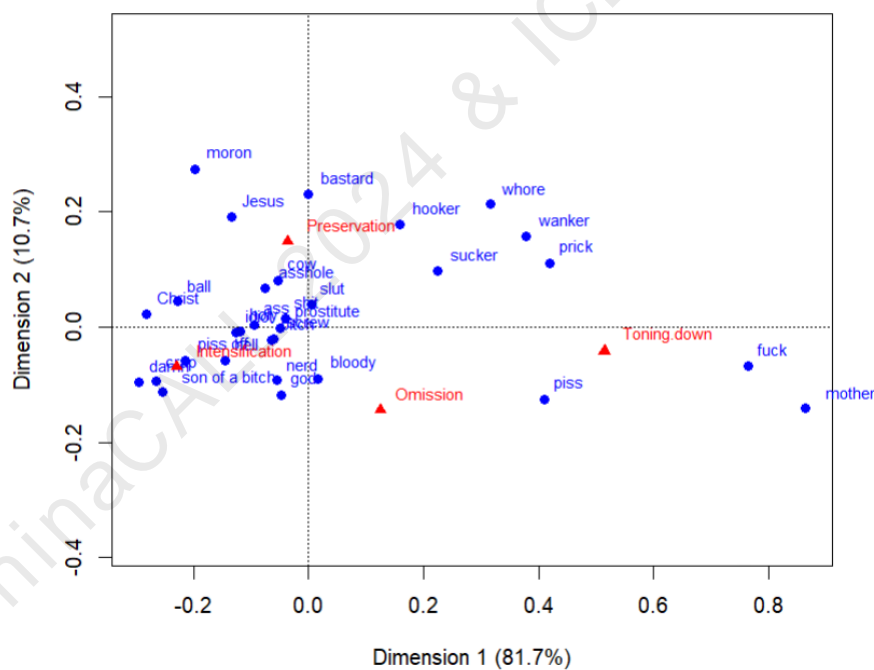Table 4.3 Result of Correspondence Analysis Between Swearwords and Translation Strategies

Figure 4.1 Figure of Correspondence Analysis Between Swearwords and Translation Strategies

According to Figure 4.1, Dimension 1 clearly divides the translation strategies based on the vulgarity: preservation and intensification entails a maintenance and increase in vulgarity while toning down and omission softens vulgarity in translating the swearwords. Dimension 2 divides the choice of translators in terms of whether to change the extent of vulgarity in subtitles or not. Therefore, through the annotation and statistical analysis of translation strategies for

swearwords, it can be observed that translators follow certain patterns in translating such terms. These patterns can reflect the translators' understanding of the two language cultures, as well as the corresponding national language policies and ideological orientations.

Correspondence Analysis was conducted in order to testify how different swearwords are translated based on their vulgarity, and therefore research question 2 will be answered. In terms of intensification, swearwords including *nerd, bloody, piss off, crap, god, Christ, ball, ass*, etc., are translated. Those words, according to Table 3.2, belong to the categories of *very mild* and *mild*, which means those swearwords in English subtitles with a lower level of vulgarity tend to be intensified in vulgarity when translated into Chinese. For two words with the highest vulgarity level, that is, *fuck* and *motherfucker*, translators choose to tone down the vulgarity when translating. It is worth noticing that *cunt* does not appear in the data. This word, which bears a very strong vulgarity, is very likely to alleviate the sense of humor in *Two Broke Girls*, a comedy TV series. Likewise, words with high levels of vulgarity are toned down in case that those words go so far that the (dark) humorous effects are ruined. Words that are translated by means of preservation belong to the categories of *mild* (e.g., *Jesus, cow, moron*) and *moderate* (e.g., *bastard, hooker, whore, slut*). In the Figure, *piss*, though in a moderate level of vulgarity, tends to be translated in the manner of toning down. It is interesting in that the data show that *piss* is translated as a word with a strong emotion of anger, but Chinese expression of anger is not inclined to bearing vulgarity, as in the following example.

Example of translating piss:

I'm so pissed! How could you give my number to that crazy nerd?

我真要气死了！你怎么就把我的电话号码给那个神经呆子了呢？

With respect to the results above, there is an inclination of translating English swearwords in subtitles into Chinese; that is, the vulgarity of swearwords translated experiences a shift from the two ends of continuum to a moderate. Mild and very mild swearwords are vulgarized by adding vulgarity, and strong and the effects of very strong swearwords are softened by toning down or even omitting vulgarity. Therefore, the Vulgarization Hypothesis (Valdeon, 2020; 2024) stands partially true in terms of strong swearwords, though in generally, more cases of intensification and preservation were observed. The strategy orientation of translating

swearwords also reflects, to some extent, the Chinese philosophy of the Doctrine of the Mean (Zhongyong). By maintaining a moderate level of vulgarity, it is possible to preserve the vulgar meaning of the original text, enhance its humor as a comedy, and avoid the impact of plot deletion or inability to release due to online censorship.

In summary, for English swearwords in English TV series to be translated into Chinese in terms of audiovisual translation, translation strategies about vulgarity are employed in a mixed manner, with intensification used the most frequently and omission the least. For a final translation result of the thorough series, swearwords at different vulgarity categories are translated with a moderate outcome; that is, swearwords in low vulgarity level are vulgarized and those in high vulgarity level are softened in their vulgarity. The findings can be explained by internet censorship and the creation of humorous effects. The Vulgarization Hypothesis, in the situation of English-Chinese AVT, stands partially true, and since previous studies into the Vulgarization Hypothesis have not specifically delved into the association between the categories of swearwords and their translation strategies, the present study opens a new perspective into swearwords and taboo language translation in the background of ATV.

Chapter 6 Conclusion

The current study redirects its research attention towards the translation of swearwords from English to Chinese, specifically examining how swearwords of varying degrees of vulgarity in English TV series are translated into Chinese in a manner that corresponds to their respective levels of vulgarity. Based on the data analysis, the findings demonstrate that the translation of English swearwords from English TV series into Chinese through audiovisual translation employs a combination of strategies regarding vulgarity, with intensification being the most commonly utilized and omission being the least preferred. The resulting translation of the entire series demonstrates a moderate approach to swearwords of varying degrees of vulgarity; specifically, those with lower levels of vulgarity are intensified while those with higher levels are softened. These findings can be attributed to internet censorship and the desire to create humorous effects. The Vulgarization Hypothesis holds partially true in the context of English-Chinese AVT. Moreover, as previous studies have not specifically explored the relationship between swearword categories and translation strategies, this study offers a new perspective on the translation of swearwords and taboo language within the realm of audiovisual translation.

Certainly, this study also has several limitations. Firstly, regarding the study of swearwords, the data selected for this paper were from an American comedy. Although this TV series often uses swearwords and taboo language, their primary function is to create a humorous atmosphere and achieve comedic effects. Therefore, the translation of swearwords in comedies is more inclined towards this function, leading to a bias introduced by the genre, and the pragmatic function of the translation results may not be entirely objective. Thus, future research could broaden the range of genres in the corpus, such as including soap operas, documentaries, and science fiction films, to study the relationship between swearword translation strategies and the degree of vulgarity in various genres of audiovisual content, thereby ensuring the objectivity of the research. Secondly, due to the insufficient depth of research on Chinese swearwords in the current Chinese academic community, in lack of studies similar to McEnery's (2006) classification of English swearwords, the coding of the data in this paper is more subjectively judged. In the translation process, translators' emphasis on the changes and orientations of

swearword vulgarity relies more on the author's personal emotional judgment gap between the original text and the translation. Therefore, the data section lacks credibility to some extent. In future research, it is necessary to delve deeper into the classification and ranking of vulgar or impolite Chinese swearwords, relying on big data and corpora for a more comprehensive study of Chinese swearwords, thus promoting the development of research on swearwords in the Chinese pragmatics academia.

References

Allan, K. (2019). *The Oxford handbook of taboo words and language* (First).

Ávila-Cabrera, José Javier. (2023). Once Upon a Time in Hollywood: Analysis of Dubbed and Subtitled Insults into European Spanish. *Journal of Pragmatics*, 217, 188-198.

Baines, Roger, 2015. Subtitling taboo language: using the cues of register and genre to affect audience experience. *Meta* 60 (3), 431e453.

Bednarek, Monika, 2019. The multifunctionality of swear/taboo words in television series. In: Lachlan, J., Alba, L. (Eds.), *Emotion in Discourse*. John Benjamins, Amsterdam, pp. 29e54.

Boker, Uwe. 2010. Taboo and transgression: A socio-historical and socio-cultural perspective. In Stefan Horlacher, Stefan Glomb, & Lars Heiler (eds.). *Taboo and transgression in British literature from the renaissance to the present*, 23–48. London: Palgrave.

Bucaria, Chiara, 2009. Translation and censorship on Italian TV an inevitable love affair? VIAL, *Vigo International Journal of Applied Linguistics* 6, 13e32.

Chaume, Frederic, 2012. *Audiovisual Translation: Dubbing*. Routledge, London and New York.

Christie, C., 2013. The relevance of taboo language: an analysis of the indexical values of swearwords. *Journal of Pragmatics*. 58, 152-169

Culpeper, J., 2011. *Impoliteness: Using Language to Cause Offence.* Cambridge University Press, Cambridge and New York

Daly, N., Holmes, J., Newtown, J., Stubbe, M., 2004. Expletives as solidarity signals in FTAs on the factory floor. *Journal of Pragmatics*. 36, 945-964

Dewaele, J.-M., 2016. 'Thirty shades of offensiveness: L1 and LX English users' understanding, perception and self-reported use of negative emotion-laden words'. *Journal of Pragmatics*. 94, 112-127.

Guillot, Marie-Noelle, 2017. Cross-cultural and audiovisual translation. Pragmatics and the translation of fiction. In: Locher, M.A., Jucker, A.H. (Eds.), *Pragmatics of Fiction*. De Gruyter Mouton, Berlin, pp. 397-424.

Guillot, Marie-Noëlle. (2020). The pragmatics of audiovisual translation: Voices from within in film subtitling. *Journal of Pragmatics*, 170, 317-330.

Guillot, Marie-Noëlle. (2023). Doing swearing across languages – The curious case of subtitling. *Journal of Pragmatics*, 215, 41-54.

Hoeksema, Jack, 2018. Taboo terms and their grammar. In: Allan, K. (Ed.), *The Oxford*

Handbook of Taboo Words and Language. Oxford University Press, Oxford, pp. 160-179.

Hughes, Geoffrey, 2006. *An Encyclopedia of Swearing: the Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-Speaking World*. ME Sharpe, London.

Jay, T., 2009. The utility and ubiquity of taboo words. *Perspective of Psychological Science*. 4 (3), 153-161

Love, R., 2021. Swearing in informal spoken English: 1990s-2010s. *Text Talk* 41 (5e6), 739-762.

McEnery, T., 2006. *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. Routledge, London.

Norrick, N.R., 2012. *Swearing in literary prose fiction and conversational narrative*. Narrat. Inq. 22 (1), 24-49.

Pavesi, Maria, & Formentelli, Maicol. (2023). The pragmatic dimensions of swearing in films: Searching for coherence in dubbing strategies. *Journal of Pragmatics*, 217, 126-139.

Pavesi, Maria, Zamora, Pablo, 2022. The reception of swearing in film dubbing: a cross-cultural case study. *Perspective of Studying Translation Theory Practice*. 30 (3), 382-398.

Rossolatos, George & Margaret K. Hogg. 2013. *Fetish, taboo, simulacrum: An applied psychoanalytic/semiotic approach to the experiential consumption of music products*. Proceedings of the 12th International Marketing Trends Conference, Paris, France, 17–19 January.

Sabri, Ouidade, Delphine Manceau & Bernard Pras, 2010. Taboo: An underexplored concept in marketing. *Recherche et Applications en Marketing* 25(1), 59–85.

Stapleton, K., 2010. Swearing. In: Locher, M., Graham, S.L. (Eds.), Interpersonal Pragmatics (*The Handbook of Pragmatics*, Vol. 6). Mouton de Gruyter, Berlin, pp. 289-305.

Stapleton, Karyn, 2003. Gender and swearing: a community practice. *Women Language*. 26 (2), 22-33.

Stapleton, Karyn, 2020. Swearing and perceptions of the speaker: a discursive approach. *Journal of Pragmatics*. 170, 391-395.

Stapleton, Karyn, Kristy, Beers Fagersten, Richard, Stephens, Catherine, Loveday, 2022. The power of swearing: what we know and what we don't. *Lingua*, 277, 1-16.

Stenstrom, Anna-Brita, 2020. English- and Spanish-speaking teenagers' use of rude vocatives. In: Nassenstein, N., Storch, A. (Eds.), *Swearing and Cursing: Contexts and Practices in a Critical Linguistic Perspective*. De Gruyter Mouton, Berlin, pp. 281-302.

Stephens, R., Atkins, J., Kingston, A., 2009. Swearing as a response to pain. *Neuroreport*, 20 (12), 1056-1060

Valdeón, Roberto A. (2020). Swearing and the vulgarization hypothesis in Spanish audiovisual translation. *Journal of Pragmatics*, 155, 261-272.

Valdeón, Roberto A. (2024). The vulgarization hypothesis and the translation of swearwords by male and female translators in AVT in Spain. *Journal of Pragmatics*, 219, 1-11.

归化和异化视阈下看《习近平谈治国理政》中"精神"一词的翻译策略

魏鑫 青岛科技大学

摘要：在翻译时，是遵循"归化"还是"异化"策略，一直存在诸多争论。但在涉及跨文化交往翻译时，要特别注意对方语言的表达习惯及价值观，因此应尽量遵循"归化"策略。在全球化背景下中西合作不断深入，中西跨文化交往越来越频繁，在翻译实践中应尽可能接近英语的表达习惯与价值观念，可以有效避免文化冲突与误解。本文将通过《习近平谈治国理政》中"精神"一词在不同语境下的翻译对此加以探讨。

关键词：归化理论；异化理论；跨文化交往；外宣文本

Translation Strategies for the Term "精神" in *Xi Jinping: The Governance of China* from the Perspective of Domestication and Foreignization

Abstract: There has been ongoing debate on whether to follow the "domestication" or "foreignization" theory in translation. However, when it comes to cross-cultural communication translation, it is important to pay special attention to the expression habits and values of the target language, thus adhering to the "domestication" theory as much as possible. In the context of deepening Sino-Western cooperation and increasing cross-cultural communication between China and the West under globalization, it is crucial to approach English expression habits and values as closely as possible in translation practice, which can effectively avoid cultural conflicts and misunderstandings. This article will explore this through the translation of the term "精神" in Xi Jinping: The Governance of China in different contexts.

key words: domestication; foreignization; cross-cultural communication; political discourse

1. 引言

"有效传播中国故事，并广泛传递中国声音"这一策略突显了近年来我国在文化领域所采取的"走出去"战略（汪宝荣、陈宇哲、江海燕，2021：17）。在此背景下，中国重要的政治宣传文本之外译工作已成为学界研究的焦点。然而，考虑到中国悠久的历史背景和丰富的文化传统，外宣文本中的语言使用显得尤为复杂，一个词在不同的语境下可能承载多种含义（吴晓静，2014:139）。当前，国际受众对我国外宣文本的接受程度有限，这主要体现在语言表达上的地方色彩以及沟通效果的不理想，这种情况在一定程度上阻碍了中国形象的有效传递。为了有效传播中国形象，建立一个既体现中国特色又能与国际接轨的外语表达系统变得尤为重要。在这个系统中，中央文献的对外译介也愈发重要，未来对《习近平谈治国理政》的海外传播与接受研究必将愈加丰富和深入（杨安文、张文莉、谭瑶，2023:9）。若在缺乏深入理解的情况下将中文中的"精神"直译为英文中的"spirit"或"mind"，无疑会导致语义模糊，给外宣文本的英译带来挑战。本文通过整理《习近平谈治国理政》多语数据库综合平台中"精神"一词的多种译法，结合归化、异化的翻译策略，归纳分析出以下几种译法。

2. 归化与异化的特征

2.1 归化策略的特征

尤金·A·奈达作为归化翻译的代表人物，在其"动态对等"理论中提出，译者应致力于"提供最接近的自然表达方式，从而使读者能够在目标语言文化中找到最合适的理解。"从跨文化交流的视角来看，一种文化中有效的交流方式可能在另一种文化中并不适用。1999年，郭建中率先在《当代美国翻译理论》专著中专节介绍韦努蒂异化翻译策略，指出归化、异化各有使命，两种翻译将永远并存，采用何种翻译策略，最终取决于目的语文化的意识形态、赞助人的力量以及翻译诗学，取决于翻译的目的和读者对象（郭建中，1999: 199）。归化的特征主要体现在让作者或原文内容靠近读者，使外语文本符合译语的文化价值观，将原作者带入译语文化中。归化翻译旨在减小目标语读者的陌生感，使译文透明、流畅，易于理解和接受。这种策略要求译者在翻译时考虑到原作及原作作者，同时也要考虑到译文的读者，尽可能地把原文的内容和风格准确而生动地表达出来，以便于读者理解和接受。归化翻译的核心在于将异国文化特色、语言形式以及作者的写作手法本土化，使译文符合目标语读者的文化背景和语言习惯。通过归化，译文能够更好地融入目标语文化，减少文化障碍，提高译文的可读性和可接受性。

2.2 异化策略的特征

劳伦斯·韦努蒂（Lawrence Venuti）认为，翻译的目的在于呈现两种文化之间的差异，而非消除它们。因此，他主张采用异化翻译策略，即保留源语言的表达形式，以作者为中心，将读者引入源语言文化，使其更贴近作者思维（Venuti，1992）。

外宣文本中的中英文翻译旨在让目标读者在获取信息的同时了解中国文化的魅力。在外宣文本的中英文翻译过程中，译者应尽量保留中文文本的词汇和表达方式，使翻译文本更贴近中国文化，让英文读者感受到中国文化的魅力，深化他们对中文文本的理解。然而，如果英文读者对中国文化不熟悉，采用异化策略有时可能使翻译难以理解，无法准确传达政治话语的原始信息，这可能导致误解和严重的沟通错误。外宣文本具有专业性、术语众多，并受语言风格中文化因素的极大影响。因此，外宣文本的中英文翻译是一项高度挑战性的任务，译者需要具备强大的语言技能和广泛的专业知识。在翻译过程中，译者需要将这两种翻译策略相结合，使译文流畅自然、为英文读者所理解，从而实现两种文化之间的交流。

3. 《习近平谈治国理政》中"精神"一词的译法

笔者通过在《习近平谈治国理政》多语数据库综合平台检索"精神"一词，发现全文中"精神"一词共出现 391 次，占比 0.14%。其中主要搭配有"十八大精神""改革创新精神""民族精神""精神追求""精神文明""精神财富""爱国主义精神"等，通过对原文和译文的对照，结合上下文语境、固定搭配，将"精神"一词的主要翻译分为以下六类，以供参考。

3.1 guiding principles

"精神"一词可用于指"会议精神"，即会议的指导原则和工作重点等内容时，可译作"guiding principles"。这个词语在中文政治语境下通常指的是一系列指导思想、原则和政策方针的集合体，体现了大会的核心价值观和战略部署。在翻译成英文时，"精神"被译为"guiding principles"，这在英文中通常指代一套用于指导行为或决策的原则或法则。

例 1

原文：紧紧围绕坚持和发展中国特色社会主义学习宣传贯彻党的<u>十八大精神</u>

译文：Study, Disseminate and Implement the <u>Guiding Principles of the 18th CPC National Congress</u>.

例 2

原文：我们在 学习宣传<u>全会精神</u>上还要下细功夫、苦功夫、深功夫 ，夯实全面深化改革的思想认识基础。

译文：We should meticulously and strenuously study and promote <u>the guiding principles of the Third Plenary Session of the 18th CPC Central Committee</u>, and gain a solid understanding of continued reform.

3.2 spirit

《牛津词典》中对"spirit"的解释之一是："the part of a person that includes their mind, feelings and character rather than their body 精神；心灵"。"精神"在中文中可以指的是人的思想、意志、情感等心理方面的活动和感受，与英文中的"spirit"含义相近。

例 3

原文：全国各族人民一定要弘扬伟大的<u>民族精神</u>和时代精神

译文：Our people of all ethnic groups must foster the great <u>national spirit</u> and follow the call of the times.

例 4

原文：没有理想信念，就会导致<u>精神</u>上"缺钙"。

译文：Without ideals and convictions one's <u>spirit</u> becomes weak.

3.3 faith

在某些情况下，"精神"特指党员对党和国家的忠诚和信仰，所以可以翻译为"faith"，即"信仰"。

例 5

原文：现实生活中，一些党员、干部出这样那样的问题，说到底是信仰迷茫、<u>精神</u>迷失。

译文：This has been proved true by the cases of some Party members and officials who acted improperly due to lack of ideals and confused <u>faith</u>.

这里的

例 6

原文：坚定理想信念，坚守共产党人<u>精神</u>追求，始终是共产党人安身立命的根本 。

译文：It has always been the foundation for the lifeline and pursuit of all Communists to have full confidence in ideals and firm <u>faith</u> in communism .

3.4 cultural、cultural-ethical 或 intellectual and cultural

当我们表达与"物质"层面相对的"精神"层面时，可译为 cultural-ethical 或 intellectual and cultural，这一译法已成为"精神文明/精神文化生活"这一类词的准确译法。早些时候，"精神文明"一词也曾被译为 cultural and ideological progress，但 ideological 一般用来指意识形态，容易引起误解，因此十九大之后，党政文件中一般均以 cultural-ethical 来翻译精神文明。

例 7

原文：满足人民日益增长的<u>精神</u>文化需求。

译文：We will be well-placed to meet people's ever-growing intellectual and <u>cultural</u> needs.

例 8

原文：中国式现代化是物质文明和<u>精神</u>文明相协调的现代化。

译文：Chinese modernization is the modernization of material and <u>cultural-ethical</u> advancement.

例 9

原文：丰富人民<u>精神</u>世界

译文：enriching the people's <u>cultural</u> lives

3.5 inspiring 或 inspiration

"精神"作为一种内在力量，其主要作用就是激励人、鼓舞人，因此，在有些语境中可以从其作用发挥的角度，将其译为 inspire 一词的其他形式。

例 10

原文：弘扬以伟大建党精神为源头的中国共产党人精神谱系。

译文：We will carry forward the long line of <u>inspiring</u> principles for the Chinese Communists that originated with the great founding spirit of the Party.

例 11

原文：丰富了民族精神和时代精神的内涵，是我们极为宝贵的<u>精神</u>财富。

译文: This has enriched the ethos of our nation and the times, and become an invaluable source of <u>inspiration</u>.

3.6 不译

某些情况下，"精神"一词已经成为一种范畴词，通常情况下可以省译，只翻译其所修

饰的中心词即可。此外，特定群体的职业精神或专业精神常常用以特定词缀结尾的词来翻译，如企业家精神 entrepreneurship。工匠精神 craftsmanship，爱国精神 patriotism 等。

例 12

原文：第一，希望大家坚守爱国主义精神。

译文：First , I hope you will adhere to patriotism.

例 13

原文：第四，必须大力弘扬劳模精神、发挥劳模作用。

译文：Fourth, we must emulate model workers.

例 14

原文：在迈向未来的征程上，我们必须充分发挥我国工人阶级的重要作用，焕发他们的历史主动精神，调动劳动和创造的积极性。

译文：As we forge ahead, we must give full rein to the vital role of the working class and its historical initiative, and arouse its motivation for work and creation.

5. 结语

《习近平谈治国理政》作为重要政治文献，具有显著的权威性和严肃性，因此对其忠实度标准的要求远高于一般翻译。政治文章的翻译应该紧扣"政治"主题，如资深政论文翻译家程镇球（2003）所强调，这意味着翻译应与原文密切相关，不得随意删除或添加文字，亦不得偏离原意自行发挥。应审慎权衡词语的政治含义和影响，确保表达恰如其分。换言之，重要政治文件的翻译不仅应忠实准确，还应坚持正确的政治立场，否则将可能引发严重后果，并导致政治错误的发生。

近年来，中国重要政治文件的外文翻译研究，包括《习近平谈治国理政》已成为研究热点，取得了诸多成果。然而，尽管翻译实践不断推进，翻译研究却显然滞后于此。未来应更加重视基于大数据的实证研究和翻译研究，包括对首席翻译和资深翻译的深度访谈，以及针对目标读者的问卷调查，以获取翻译决策过程和翻译接受与反馈的实证数据，从而提升研究结果的可靠性和可信度。这将为提高外宣文本的翻译量和传播效果提供智库支持。

6. 参考文献

[1] Nida, Eugene A,Charles R.Taber.The Theory and Practice of Translation[M].New York:E.J. Brill,Leiden.1969.

[2] Venuti L.Rethinking Translation: Discourse,Subjectivity, Ideology[M].London and New York: Routledge,1992.

[3] 吴晓静.浅谈"辛苦"一词在不同语境下的翻译策略[J].神州旬刊, 2014, 000(007):139-139.

[4] 邱大平.国家形象视阈下政治话语外宣翻译策略研究[D].华中师范大学,2019.

[5] 汪宝荣;陈宇哲;江海燕.政治文献中的中国特色词翻译探析——以2021年《政府工作报告》英译为例[J].亚太跨学科翻译研究,2023,(01):16-28.

[6] 杨安文;张文莉;谭瑶.党的十八大以来《习近平谈治国理政》翻译研究综述[J/OL].西南交通大学学报(社会科学版), 2023,(11):1-18

[7] 朱耀武.话语权视角下中国特色政治词汇的翻译策略——以《习近平谈治国理政》为范本[J].唐山师范学院学报,2023,45(04):26-29.

[8] 郭建中.当代美国翻译理论[M].武汉: 湖北教育出版社,1999.

[9] 程镇球.政治文章的翻译要讲政治[J].中国翻译,2003,(03):20-24.

The Neglected Factor: Prompts in Evaluations of ChatGPT's Translation Performance

ZHANG Yu   Shanghai International Studies University

Abstract: Prompts have a significant influence on ChatGPT's performance, which differentiates ChatGPT from traditional machine translation tools such as Google Translate and DeepL. In light of the growing interest in exploring ChatGPT's translation capabilities, this paper examines how prompts are considered in related studies. 32 articles were selected using PRISMA methodology and analyzed through a set of standardized questions. The results indicate that prompts are generally underemphasized in evaluations of ChatGPT's translation performance, highlighting the need for greater attention to this aspect in future research, and more broadly, for a deeper understanding and more comprehensive consideration of ChatGPT's unique characteristics in order to explore its implications.

Keywords: ChatGPT; Prompts; Translation Performance; Evaluation

1. Introduction

Prompts are the instructions or queries that users provide to ChatGPT to elicit responses. Numerous studies have demonstrated that the performance of ChatGPT can be significantly enhanced through effective prompting strategies, such as providing additional examples (few-shot prompting) (Brown et al., 2020; Zhong et al., 2023), and instructing it to break down complex problems step-by-step (chain-of-thought prompting) (Wei et al., 2022; Kojima et al., 2022; Zhong et al., 2023). OpenAI (2023), the developer of ChatGPT, has emphasized this by publishing *Prompt Engineering Guide* to help users achieve better results from ChatGPT.

Given the critical role of prompts, they should be a key consideration in evaluating ChatGPT's translation performance. Firstly, prompts directly influence the generated content, thereby affecting the perceived translation quality. Secondly, including the prompts used in evaluations is essential for readers to fully understand the evaluation process and contextualize the results. Researchers should therefore explicitly detail the role and application of prompts in their studies, otherwise the validity and repeatability of the studies would be compromised.

This leads to a natural question: Do studies evaluating ChatGPT's translation performance adequately address prompts? This paper aims to investigate how prompts are taken into account in such studies and evaluate their sufficiency. The structure is as following: It starts by outlining the importance of prompts and their relevance in performance evaluations. Subsequently, based on the PRISMA guideline, 32 articles were collected and analyzed using a set of standardized questions to assess their consideration of prompts. Then the findings derived from the analysis are discussed.

2. Method

2.1 Eligibility criteria

Studies focusing on evaluating the translation performance of ChatGPT, or containing such evaluation as a major part were eligible, regardless of their publication discipline, evaluation method and outcome. Measures were taken during the selection process to ensure the inclusion of

high quality papers, which will be detailed in the following section.

2.2 Paper collection and selection

The papers were collected following the four phases of the PRISMA methodology: (i) identification, (ii) screening, (iii) eligibility, and (iv) inclusion (Moher, 2009). The entire process is documented in Figure 1.

(i) Identification: Relevant literature was searched across two databases—WOS and CNKI—to include studies in and out of China. For WOS, the search scope was the Web of Science Core Collection (all editions), using the search string "Topic = 'ChatGPT' AND 'translat*'", with a wildcard character to capture different forms of the word "translate". For CNKI, the search scope included CSSCI, Peking University Core Journals, and CSCD databases, which are considered high-quality sources. The search string used was "Topic = 'ChatGPT' AND '翻译[1]'". Additionally, a full-text search was conducted in CNKI[2], with the search string "Topic = 'large language model' + 'artificial intelligence' AND '翻译'; Full text = 'ChatGPT' in case relevant papers use general terms like "large language model" or "artificial intelligence" but actually evaluate ChatGPT's translation performance. The initial search, conducted on June 12, 2024, generated a total of 232 records (CNKI: 88; WOS: 144). 25 duplicate articles were removed before screening, leaving 207 to be screened.

(ii) Screening: Titles and abstracts were screened, resulting in the removal of 170 records for reasons including: (1) they were reviews or commentaries; (2) they were actually unrelated to translation as meant by this study, with "translation" referring to other concepts such as translational medicine or code conversion; (3) they were related to translation but did not include an evaluation of ChatGPT's translation performance. An additional 13 records were identified through backward citation, leading to a total of 50 records for further analysis.

(iii) Eligibility: Full texts of the 50 records were retrieved and assessed for eligibility. 18 records were excluded for the following reasons: (1) they only mentioned or briefly discussed ChatGPT's translation ability without in-depth evaluation; (2) they used "translation" in a broader sense, referring to activities such as paraphrasing, rather than the intended definition, such as paraphrasing.

(iv) Inclusion: Ultimately, 32 studies were included for final analysis.

---

[1] 翻译 means translation/translate in Chinese.

[2] WOS doesn't support full text search.
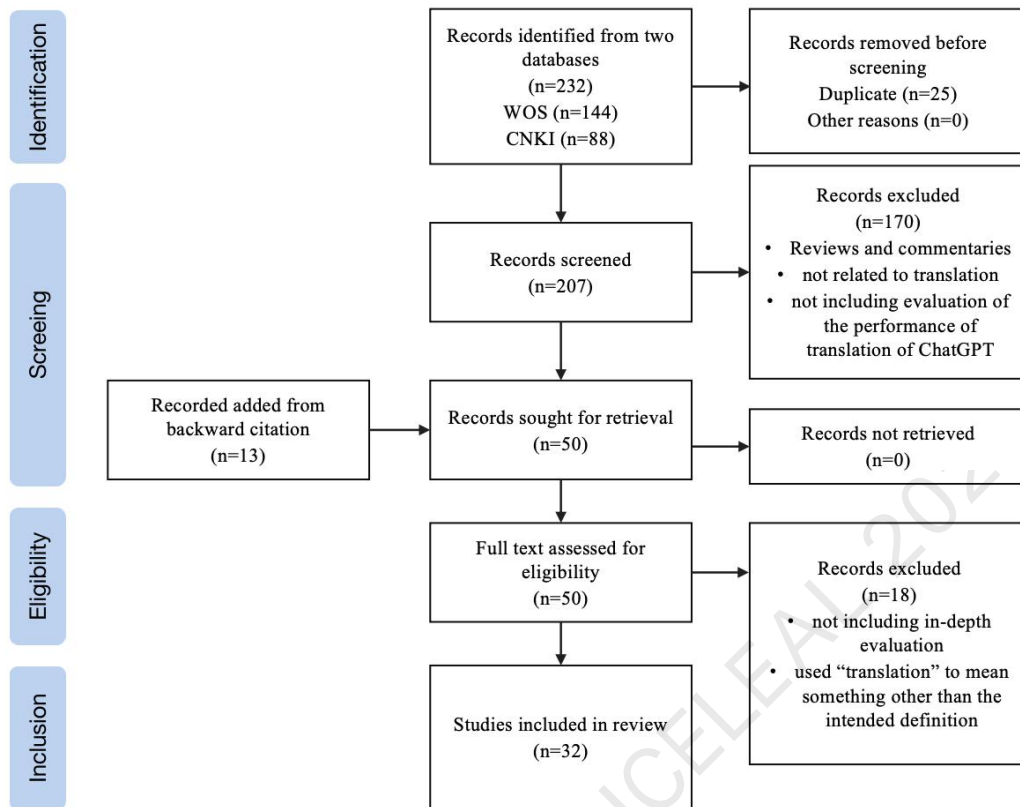
Fig. 1. PRISMA flow diagram

2.3 Data Analysis

First, an analysis was conducted to summarize the basic characteristics of the reviewed studies, including publication information and evaluation design. This provided an overall picture of the articles and laid the foundation for further analysis.

Then, each article was assessed in terms of their consideration of prompts. This study takes reference to the works of Hemkens (2018) and Munkholm (2020) who used standardized questions to assess whether a given factor (confounding bias) that was supposed to have impact on the study result had been adequately considered in existing studies. Nevertheless, the questions were modified to meet the need of this research. Five questions (Table 1) were formulated to systematically evaluate the consideration of prompts in the included studies.

Question 1 inquires whether the concept of prompts was mentioned at all at any point in the paper, indicating the minimal level of consideration given to prompts in the evaluation. Question 2 examines whether the role of prompts in utilizing ChatGPT was explicitly introduced or reviewed, for it underscores the authors' recognition of prompts as a pivotal aspect of ChatGPT which deserves explicit illustration. Question 3 aims to find out whether the authors perceived prompts as having a substantial impact on the performance of ChatGPT and therefore should be examined in the evaluation. Question 4 examines whether the authors specified how they used prompts in their evaluation, as this is essential for readers to understand the conditions under which the research was conducted and its findings. Additionally, the question seeks to determine whether the authors made efforts to use optimal prompts to fully elicit ChatGPT's translation strengths. The

last question aims to determine whether the authors referenced prompts in their conclusions, where all important implications of the research are stated.

Table 1. Evaluation of prompts consideration

Question
1. Did the authors make any mentions of prompts? Example for "yes": *"ChatGPT is an intelligent chatbot developed by OpenAI that builds on InstructGPT, a model designed to provide detailed responses to prompts."* (Calvo-Ferrer, 2023)
2. Did the authors give any introduction to or review of the role of prompts in utilizing ChatGPT? Example for "yes": *"The style of prompts may affect the quality of translation outputs. For example, how to mention the source or target language information matters in multilingual machine translation models, which is usually solved by attaching language tokens."* (Jiao et al., 2023)
3. Did the authors examine the impact of prompts on the translation performance of ChatGPT in their evaluation? Example for "yes": *"We observe that the translation quality with few-shot in context learning can surpass that of strong encoder-decoder MT systems, especially for high-resource languages."* (Moslem, 2023)
4. Did the authors specify the prompts they used in the evaluation? If yes, (1) did they provide the rationale for the prompt design? (2) did they leverage any prompting tactics? Example for "yes": *" In the case of CGPT-3.5, the prompt "Traduce de español ainglés" [Translate from Spanish into English] was used for it to act as a translator without finetuning the results through a more accurate prompt—a path currently under study in the wider project where this experiment develops."* (Sanz-Valdivieso & López-Arroyo, 2023) Example for " rationale given": *"We consider the following three prompting strategies for GPT-3.5 that allow us to compare the model's abilities to translate with and without discourse-level context (see Table 3 for templates and Appendix B for the exact prompts)."* (Karpinska & Iyyer, 2023) Example for "prompting tactics leveraged": *"We use few-shot prompting, in which a model is provided with a prompt consisting of five demonstrations. We manually curate the five demonstrations from literary texts for each of the 18 language pairs, resulting in 90 total demonstration examples. "* (Karpinska & Iyyer, 2023)
5. Did the authors reference prompts in their conclusions? If yes, (1) did they discuss ChatGPT's translation performance in relation to the prompting strategies used? (2) did they highlight any limitations of the evaluation due to the prompting? (3) did they offer insights for future research or applications concerning prompting?

Example for "yes": *"We further explore an interesting strategy named pivot prompting that can improve the translation performance of distant languages noticeably."* (Jiao et al., 2023)

Example for "discussed in relation to the prompting": *"As a result, our findings demonstrate that GPT systems can produce highly fluent and competitive translation outputs even in the zero-shot setting especially for the high-resource language translations. By utilizing the in-context learning capability of GPT models with few-shot examples, we were able to further improve translation quality."* (Hendy, 2023)

Example for "limitation highlighted": *"we only compare GPT translations in the standard zero-shot and few-shot settings and it is quite conceivable that more specific & verbose instructions could steer the LLMs to produce translations with different characteristics."* (Raunak et al., 2023)

Example for "insights for the future offered": *"It would also be interesting to explore how different prompts may increase its effectiveness in terms of accuracy, speed, and ease of use, both at the professional level and in translator training."* (Calvo-Ferrer, 2023)

3. Results

3.1 Description of studies

The basic characteristics of the 32 articles are summarized in Table 2. These articles span various disciplines, indicating that the interest in ChatGPT's translation performance in not limited to the field of translation but extends across a wide range of disciplines including translation, linguistics, computer science, information technology, medicine, chemistry, environment and multidisciplinary studies. While the largest portion of studies come from translation & linguistics, they still constitute less than half (46.88%). This underscores the relevance of the topic across multiple fields and the potential for new findings to benefit them all.

It is also noteworthy that even though all these articles aimed to evaluate ChatGPT's translation capabilities, they differed significantly in their approaches. They looked at different aspects of performance, examined different language pairs, and employed different methods. For example, in terms of evaluation factor, half of them focused on specific text types such as legal text (Moneus & Sahari, 2024; Song, 2024), literal text (Zhao et al., 2023; Wang, 2024), and political text (Wen & Tian, 2024), etc., while the other half looked into overall translation performance, translation characteristics such as syntactic complexity and lexical diversity (Yu, 2024), and translation performance on specific languages including ancient languages (Ross, 2023; Klamra, 2023) and low-resource languages (Wu et al., 2023; Hou et al., 2024).

Table 2. Characteristics of studies

Study characteristics			N	(%)
Publication information	Publication year	2023	20	(62.50)
		2024	12	(37.50)
	Source type	Journal	23	(71.88)
		Preprint	5	(15.63)
		Conference	4	(12.50)
	Publication discipline [1]	Translation & linguistics	15	(46.88)
		Computer science & information technology	9	(28.12)
		Others [2]	8	(25.00)
Evaluation design	Evaluation factor	Translation performance on specific text types	16	(50.00)
		Overall translation performance	7	(21.88)
		Translation characteristics	5	(15.62)
		Translation performance on specific languages	4	(12.50)
	Comparative or standalone evaluation	Comparative	25	(78.13)
		Standalone	7	(21.87)
	Language pair	One pair	21	(65.63)
		Multiple pairs	11	(34.37)
	Evaluation method	Human evaluation	17	(53.13)
		Hybrid evaluation	9	(28.13)
		Automated evaluation	6	(18.75)

[1] Publication discipline refers to how journals, preprint platforms, or conference categorize articles by discipline.

[2] Others include medicine, chemistry, environment, and multidisciplinary studies.

3.2 Consideration of prompts

Table 3. Statements on prompts

Question		N	(%)
1	Did the authors make any mentions of prompts?		
	Yes	22	(68.75)
	No	10	(31.25)
2	Did the authors give any introduction to or review of the role of prompts in utilizing ChatGPT?		
	Yes	10	(31.25)
	No	22	(68.75)
3	Did the authors examine the impact of prompts on ChatGPT's translation performance in evaluation?		
	Yes	7	(21.88)
	No	25	(78.12)
4	Did the authors specify the prompts they used in the evaluation?		
	Yes	19	(59.38)
	...(1) Did they provide the rationale for the prompt design?		
	...rationale given	10	(52.63)
	...not given	9	(47.37)
	...(2) Did they leverage any prompting tactics?		
	...prompting tactics leveraged	11	(57.89)
	...not leveraged	8	(42.11)
	No	13	(40.62)
5	Did the authors reference prompts in their conclusions?		
	Yes	10	(31.25)
	...(1) Did they discuss ChatGPT's translation performance in relation to the prompting strategies used?		
	...performance disscussed in relation to the prompting	6	(60.00)
	...not discussed	4	(40.00)
	...(2) Did they highlight any limitations of the evaluation due to the prompting?		
	...limitations highlighted	3	(30.00)
	...not hilighted	7	(70.00)
	...(3) Did they offer insights for future research or applications concerning prompting?		
	...insights offered	6	(60.00)
	...not offered	4	(40.00)
	No	22	(68.75)

3.2.1 Mere mentioning of prompts

Prompts were not mentioned in 10 articles, which means 31.25% of the articles showed no consideration of prompts. In the remaining 22 articles, prompts were mentioned at least once, but the extent of consideration given to prompts in these mentions requires further examination.

3.2.2 Introduction to or review of the role of prompts in utilizing ChatGPT

Of the 22 articles that mentioned prompts, only 10 provided an explicit introduction to or review of the role of in eliciting ChatGPT's capabilities. Very few of these articles offered a comprehensive introduction; most only touched upon that with one or two sentences and did not delve deeper into the issue. For example, in *Lexical Diversity and Syntactic Complexity in ChatGPT Translation* (Yu, 2024), the author mentioned in the introduction that although ChatGPT was not specifically designed for translation, its translation abilities can be brought out with accurate prompts. However, this point was not further elaborated upon in the subsequent sections.

3.2.3 Examination of the impact of prompts on the translation performance of ChatGPT

Only 7 articles (21.88%) investigated the impact of prompts on ChatGPT's translation performance. These articles looked at how change of prompt strategies lead to different translation results, such as utilizing different zero-shot templates (Jiao et al., 2023), providing more context information and specific instructions (Wang & Mao, 2023; Siu, 2023; Karpinska & Iyyer, 2023), shifting from zero-shot to few-shot prompts (Hou, 2024; Hendy, 2023; Moslem et al., 2023), incorporating chain-of-thought guidance (Hou, 2024), and applying other self-developed prompting methods (Jiao et al., 2023). The consensus across these articles was that prompt variation influences translation quality, sometimes markedly so. It is worth noting that while some articles explicitly mentioned the role of prompts in utilizing ChatGPT (as noted in Question 2), they did not evaluate this aspect in their research.

3.2.4 Specification of the prompts used in the evaluation

Out of 32 articles, 19 (59.38%) specified the prompts used in their evaluations, while the remaining 13 presented the evaluation results without detailing the prompts, making it difficult for readers to fully understand the evaluation setup or replicate the results. Among the 19 articles that specified prompts, only 10 gave a rationale for their prompt selection; the other 9 did not explain why specific prompts were chosen and how this might impact the evaluation results. Furthermore, only 11 articles employed any kind of prompting tactics, and most of them used only a single tactic rather than combining different approaches, or dynamically adjusting the interaction strategy based on ChatGPT's responses to elicit optimal performance. The remaining 8 articles only used plain and straightforward prompts such as "translate from language A to language B" (Sanz-Valdivieso & López-Arroyo, 2023; Yu, 2024; Zhang & Zhao, 2024), even though some of these articles acknowledged in earlier sections that effective use of prompts can enhance ChatGPT's performance.

### 3.2.5 Reference to prompts in conclusions

Most of the articles did not address prompts when summarizing the implications of their studies. Only 10 articles (31.25%) referenced prompts in their conclusions, focusing on different aspects:

- Six articles interpreted the translation performance of ChatGPT in relation to the prompting strategies used, describing the performance as under specific prompting conditions;
- Three articles acknowledged limitations of their studies due to possibly biased or sub-optimal prompting;
- Six articles offered insights for future research or applications concerning prompting, calling for deeper exploration of prompting and more effective use of prompts.

Only one article covered all three aspects, providing a comprehensive conclusion of the study closely related to prompting.

4. Discussion

The results presented above indicate that the consideration of all five aspects, as outlined by the five questions in Table 1, is generally inadequate. 31.25% of the authors completely ignored prompts throughout their articles. Significantly fewer than half of the articles introduced or reviewed the role of prompts in utilizing ChatGPT (31.25%), examined the impact of prompts on ChatGPT's translation performance in evaluation (21.88%), or referenced prompts in their conclusions (31.25%), suggesting that most articles did not address these aspects of prompt consideration. Although the majority of articles (59.38%) specified the prompts they used in evaluations, many failed to explain the rationale behind the choice of specific prompts, or utilized any prompting tactics to elicit the translation abilities of ChatGPT.
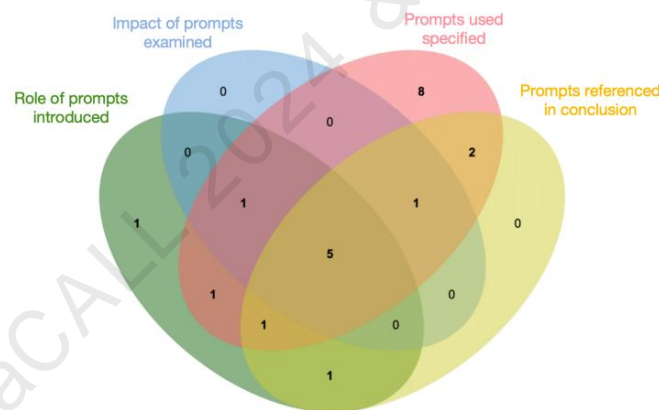
Fig. 2. Venn diagram on different aspects of consideration of prompts. Each ellipsoid represents one aspect as outlined by Question 2-5. Question 1 is excluded as it serves as the minimal threshold for consideration and is a superset of the other four aspects. The numbers in the diagram denote the number of studies sharing the characteristics in the overlapping area.

Figure 2 presents a Venn diagram illustrating the intersection of various aspects of prompt consideration. It shows minimal overlap across these aspects, indicating that while some articles touched upon prompt considerations, they typically focused on a limited range of aspects. Notably, eight articles (25%) only specified the prompts used, neglecting other aspects, whereas only five articles (15.62%) covered all five aspects of prompt consideration (regardless of the depth of their analysis). None of these comprehensive articles were published in mainstream translation and language journals: four appeared on preprint platforms such as arXiv and SSRN, and one was

from a journal on ceramics. Additionally, three of the five articles were categorized under computer science & information technology and authored by researchers from institutions such as Microsoft, Tencent, and the University of Massachusetts Amherst's Manning College of Information and Computer Sciences, rather than by translation scholars.

This trend highlights the general neglect of prompt consideration in mainstream translation and linguistics studies. Although nearly half of the articles analyzed (46.88%) come from the fields of translation and linguistics (as shown in Table 2), none provided a thorough examination of prompts. Conversely, authors from computer science and information technology are more likely to stress the importance of prompts, for 1/3 of them covered all 5 aspects of consideration of prompts, possibly due to their deeper understanding of the working mechanisms and technical characteristics of ChatGPT as a generative and conversational AI. Most authors tended to use ChatGPT similarly to traditional machine translation tools, which could be because they intuitively regarded ChatGPT as merely an advanced version of Google Translate, without recognizing its distinctive features. However, ChatGPT is designed to interact with users in natural language, generating and continually refining content rather than returning fixed translations based on static input.

Overall, the emphasis on prompts remains insufficient. while it is encouraging that many researchers have recognized ChatGPT as a revolutionary technology poised to bring significant changes and stressed the importance of studying it, greater efforts are needed to fully understand how ChatGPT differs and its broader implications. Studies examining the capabilities of ChatGPT should take prompts seriously, for example, by giving thought to the aspects outlined in this study. However, prompts are just one example; and there're numerous other defining attributes of ChatGPT that need exploration. Moreover, researchers across different disciplines should enhance their communication and exchange of insights, fostering a more holistic and in-depth understanding of ChatGPT.

This study has several limitations. The primary one is its relatively small sample size of only 32 papers, partly due to the short interval between ChatGPT's public release and the completion of this research. Additionally, the articles were primarily sourced from CNKI and the WOS core collection, potentially excluding valuable papers from lower-impact journals. Including more works might reveal additional insights regarding prompt consideration. Furthermore, this paper looks at how prompts are considered by relevant articles as a whole, with very limited attention to disciplinary differences. It would be valuable to examine these differences in greater detail across various disciplines.

5. Conclusion

The consideration of prompts is generally inadequate in evaluations of the translation performance of ChatGPT. Very few articles showed comprehensive consideration of prompts throughout their studies; in most cases, such consideration was superficial, incomplete, or entirely absent,. There is a critical need for researchers to place more attention on prompts in analysis of ChatGPT, and more broadly, to fully understand and consider the defining characteristics of ChatGPT to achieve deeper insights.

References

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems, 33,* 1877-1901.

Calvo-Ferrer, J. R. (2023). Can you tell the difference? A study of human vs machine-translated subtitles. *Perspectives,* 1-18.

Hemkens, L. G., Ewald, H., Naudet, F., Ladanie, A., Shaw, J. G., Sajeev, G., & Ioannidis, J. P. (2018). Interpretation of epidemiologic studies very often lacked adequate consideration of confounding. *Journal of clinical epidemiolog*y, *93,* 94-102.

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., ... & Awadalla, H. H. (2023). How good are gpt models at machine translation? a comprehensive evaluation. *arXiv preprint arXiv:2302.09210.*

Hou, Y. T., Abulizi, A., Shi, Y. Q., Musideke, M., & Abudukelimu, H. (2024), Mianxiang yidaiyilu de diziyuanyuyan jiqifanyi yanjiu [Research on low-resource language machine translation for the "Belt and Road"]. *Jisuanji gongcheng, 50*(04), 332-341.

Jiao, W., Wang, W., Huang, J. T., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv preprint arXiv:2301.08745.*

Karpinska, M., & Iyyer, M. (2023). Large language models effectively leverage document-level context for literary translation, but critical errors persist. *arXiv preprint arXiv:2304.03245.*

Klamra, C., Kryńska, K., & Ogrodniczuk, M. (2023, November). Evaluating the use of generative LLMs for intralingual diachronic translation of Middle-Polish texts into contemporary Polish. In *International Conference on Asian Digital Libraries* (pp. 18-27). Singapore: Springer Nature Singapore.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *Advances in neural information processing systems, 35,* 22199-22213.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group, T. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine, 151*(4), 264-269.

Moneus, A. M., & Sahari, Y. (2024). Artificial intelligence and human translation: A contrastive study based on legal texts. *Heliyon, 10*(6).

Moslem, Y., Haque, R., Kelleher, J. D., & Way, A. (2023). Adaptive machine translation with large language models. *arXiv preprint arXiv:2301.13294.*

Munkholm, K., Faurholt-Jepsen, M., Ioannidis, J. P., & Hemkens, L. G. (2020). Consideration of confounding was suboptimal in the reporting of observational studies in psychiatry: a meta-epidemiological study. *Journal of Clinical Epidemiology, 119,* 75-84.

OpenAI. (2023). Prompt engineering. OpenAI Platform. *https://platform.openai.com/docs/guides/prompt-engineering.*

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj, 372.*

Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., ... & Tao, D. (2023). Towards making the most of chatgpt for machine translation. *arXiv preprint arXiv:2303.13780.*

Raunak, V., Menezes, A., Post, M., & Awadalla, H. H. (2023). Do gpts produce less literal translations?. *arXiv preprint arXiv:2305.16806.*

Ross, E. A. (2023). A new frontier: AI and ancient language pedagogy. *Journal of Classics Teaching, 24*(48), 143-161.

Sanz-Valdivieso, L., & López-Arroyo, B. (2023). Google Translate vs. ChatGPT: Can non-language professionals trust them for specialized translation?. In *International Conference Human-informed Translation and Interpreting Technology (HiT-IT 2023)* (pp. 97-107).

Siu, S. C. (2023). Chatgpt and GPT-4 for professional translators: Exploring the potential of large language models in translation. *Available at SSRN 4448091.*

Song, L. J. (2024). Falv fanyi de shuzi renwen zhuanxing yanjiu -- yi zhuanti shujuku yu ChatGPT wei zhongxin [Innovative practices of digital humanities technology in legal translation: A case study on the construction and application of the BRI legal translation data platform]. *Waiyu xuekan, 2,* 51-57.

Wang, C. T. (2024). Rengong zhineng shidai de fanyi, chuban yu chuanbo: wenti yu zhanwang [Issues and prospects of translation, publication and communication in age of AI]. *Xiandai chuban, 04,* 79-91.

Wang, Z. Y., & Mao, C. (2023). ChatGPT yiwen zhiliang de pinggu yu tisheng--yi taoci lei wenben hanying fanyi weili [Evaluation and enhancement of translation quality of ChatGPT: A case study of Chinese-to-English translation in ceramic texts]. *Shandong taoci, 46*(04), 20-27.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems, 35,* 24824-24837.

Wen, X., & Tian, Y. L. (2024). ChatGPT yingyong yu zhongguo tese huayu fanyi de youxiaoxing yanjiu [The effectiveness of chatGPT in translating China-specific discourse text]. *Shanghai fanyi, 02,* 27-34+94-95.

Wu, J. H., Zhao, Y., & Zong, C. Q. (2023). ChatGPT nengli fenxi yu weilai zhanwang [Analysis of ChatGPT's capabilities and future prospects]. *Zhongguo kexue jijin, 37*(05), 735-742.

Yu, L. (2024). ChatGPT fanyi de cihui duoyangxing he jufa fuzadu yanjiu [Lexical diversity and syntactic complexity in Chat GPT translation]. *Waiyu jiaoxue yu yanjiu, 56*(02), 297-307+321.

Zhang, W. Y., & Zhao, B. (2024). Shengchengshi rengongzhineng kaichuang jiqifanyi de xinjiyuan le ma?--yixiang zhiliang duibi yanjiu ji dui fanyijiaoyu de sikao [Has generative AI opened a new era for machine translation? -- a contrastive quality study and reflections on translation education]. 2024. *Beijing dier waiguoyu xueyuan xuebao, 46*(01), 83-98.

Zhao, X., Zhao, Z. X., Sun, F. L., & Wang, D. B. (2023). Mianxiang yuyan wenxue lingyu de dayuyan moxing xingneng pingce yanjiu [Performance evaluation study of large language models for the field of language and literature]. *Waiyu dianhua jiaoxue, 06,* 57-65+114.

Zhong, Q., Ding, L., Liu, J., Du, B., & Tao, D. (2023). Can chatgpt understand too? a comparative study on chatgpt and fine-tuned bert. *arXiv preprint arXiv:2302.10198.*

Conference Website