EXPLORING THE POTENTIAL AND CHALLENGES OF CHATGPT IN LANGUAGE EDUCATION: AN INVESTIGATION INTO STUDENT TEACHERS' CRITICAL THINKING, INTERACTIONS AND TASK-BASED ACTIVITY DESIGN

Tema

L'introduzione di ChatGPT ha destato alcune preoccupazioni in numerosi ambiti, tra cui quello dell'educazione linguistica. Pertanto, è importante che studenti ed educatori acquisiscano le competenze necessarie per usare in modo appropriato gli strumenti di intelligenza artificiale (IA) nell'insegnamento e nell'apprendimento. Sebbene alcune ricerche abbiano esplorato le potenzialità ed i limiti di ChatGPT, oggigiorno gli studi riguardanti la sua applicazione nell'attività didattica e il suo impatto sulle capacità di pensiero critico sono ancora esigui. Questa ricerca esplorativa indaga l'utilità di ChatGPT nel realizzare delle attività di *Task-Based Language Learning* (TBLL), il suo impatto sul pensiero critico degli studenti aspiranti insegnanti e l'interazione dei partecipanti con lo strumento. Attraverso un approccio *mixed method,* i dati raccolti evidenziano che i partecipanti sono stati in grado di creare delle attività di TBLL di buona qualità, risparmiando del tempo con ChatGPT. Tuttavia, circa la metà dei soggetti coinvolti nella ricerca ha copiato buona parte degli output forniti dal modello e sono emerse delle difficoltà nell'interazione con ChatGPT. Sulla base dei risultati ottenuti, sono state elaborate alcune soluzioni e proposte per studi futuri in merito all'uso di ChatGPT nell'educazione linguistica.

Corinne Ca' Foscari Baradel
| University of Venice

Corinne Baradel holds a Master's degree in Language Sciences from Ca' Foscari University of Venice and a Master's degree in Teacher Development for Foreign Language Education from Universitat Autònoma de Barcelona. Her research interests focus on the integration of AI-driven tools and digital technologies in language education.

Introduction

Recently, integrating Artificial Intelligence (AI) technologies into daily life has transformed multiple sectors, including education. Current AI tools have enormous potential to enhance learning and teaching, make education globally accessible and tailor learning to students' needs. Among these tools, ChatGPT (OpenAI, 2022) has attracted significant attention from researchers and academies for its capabilities. In fact, this Large Language Model (LLM) chatbot can serve as a virtual assistant, generating human-like conversations, answering questions and clarifying complex concepts (Kasneci et al., 2023).

The educational potential of AI tools, however, is accompanied by security risks and ethical concerns. These technologies often collect sensitive data, necessitating responsible data management. Moreover, biases inherent in training data can lead to errors, while concerns about plagiarism, unethical behaviour, diminishing critical thinking and creativity also arise. Given ChatGPT's emergence and impact on education, this research[1] aims at investigating the following two research questions amidst university learners: to what extent is the usage of ChatGPT useful for student teachers to design Task-Based Language Learning (TBLL) activities (Willis, 1996) addressed to high-school language students, in terms of planning time, quality of the final products and development of their critical thinking? What are the strategies used by student teachers when interacting with ChatGPT to create a TBLL activity?

Literature review

ChatGPT is an AI chatbot (OpenAI, 2022) trained on extensive text and able to generate human-like responses in multiple languages (García-Peñalvo, 2023). Its outputs, however, are strictly influenced by the quality of user prompts (OpenAI, 2022; Jalil et al., 2023). Despite its potential, the AI chatbot may generate biased

1 This paper is based on the master's thesis by the researcher (Baradel, 2023). Some sections of the text may replicate the original content from her master's thesis.

content or provide incorrect information due to the nature of its pre-training (OpenAI, 2022; Jalil et al., 2023; Rudolph et al., 2023).

In the field of language education, researchers (Kasneci et al., 2023) asserted that, on the one hand, ChatGPT can be a valuable resource for students throughout their language learning process. In fact, the AI tool may improve both writing and reading abilities by assisting in grammatical and syntactic corrections, summarising complex texts, creating quizzes, organising thoughts during the writing process and providing feedback. On the other hand, in the teaching practices, the chatbot may aid teachers in organising lesson plans and creating engaging materials. Therefore, teachers could save both efforts and time. Furthermore, empirical studies (Ali et al., 2023; Muñoz et al., 2023) demonstrated ChatGPT's efficacy in increasing student engagement and motivation, particularly in subjects like languages.

While ChatGPT offers numerous benefits in language education, various challenges in this sector need consideration. According to Kasneci et al. (2023), one concern is the difficulty in distinguishing AI-generated texts from human writing, which may undermine academic integrity. Hence, using ChatGPT in class may boost plagiarism and unethical behaviour (Lin et al., 2023). A further risk is developing an overreliance on ChatGPT, neglecting to use critical thinking abilities (Kasneci et al., 2023), defined as the combination of higher-order thinking skills (deduction, analysis, reasoning, problem solving and evaluation) (Bellaera et al., 2021). These capacities allow human beings to make reasonable decisions, interpret and analyse large amounts of information rationally and objectively (Thornhill-Miller et al., 2023).

Given the aforementioned concerns, educators should integrate ChatGPT responsibly into educational environments by teaching students to critically evaluate ChatGPT's outputs in order to avoid possible false information or bias (Bitzenbauer, 2023). Moreover, teachers are advised to elaborate higher-order assignments that emphasise skills ChatGPT lacks, such as providing emotional and real-world practical situations (Liu & Wei, 2024).

> Current AI tools have enormous potential to enhance learning and teaching experiences, make education globally accessible and tailor learning to students' needs.

Besides ChatGPT and its implications on language education, for this research it is important to introduce the Task-Based Language Learning (TBLL) method proposed by Willis (1996). This framework for language education focuses on using meaningful tasks as the core of the learning process and includes three phases. In the pre-task, learners are introduced to the topic and related vocabulary, often participating in activities that activate their prior knowledge. During the task cycle, students in groups perform, plan and report the task and their results. Finally, the post-task refers to the language focus, in which learners not only analyse and reflect on the language used throughout the task, but also practice certain language features to enhance their understanding and accuracy.

Methodology

Considering both potentials and limits of ChatGPT in language education, an exploratory study was conducted to comprehend the usefulness of ChatGPT in designing TBLL activities (Willis, 1996) in terms of planning time and quality, the effects on student teachers' critical thinking skills and the interaction strategies between ChatGPT and student teachers. The current investigation involved 39 voluntary postgraduate participants (referred to as *student teachers*) studying the subject Instructional Design for Language Education. The participants in groups or individually had to design a TBLL activity with the usage of ChatGPT. It is important to underline that before this assignment, the student teachers attended a lesson where the TBLL model (Willis, 1996) was deeply explained and ChatGPT was briefly introduced.

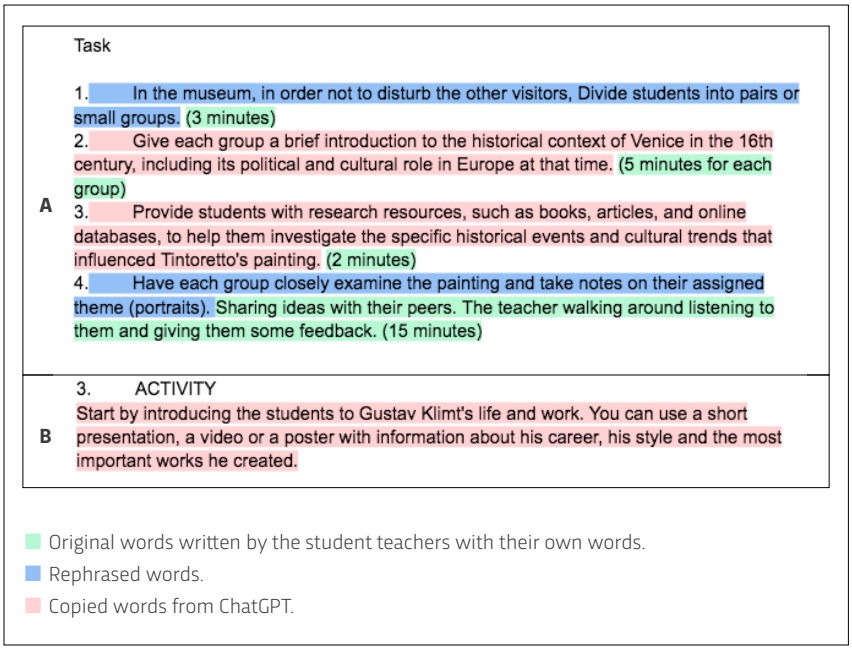

Figure 1

Illustration of original, copied, and rephrased words from two TBLL activity samples.

Note. Reprinted from *Interaction, Design, and Assessment: An Exploratory Study on ChatGPT in Language Education* by Baradel (2023). http://hdl.handle.net/10579/25711.

The assignment required designing a TBLL activity (Willis, 1996) based on a specific theme (food education, selfies and portraits or anything related to sustainability) for 16-to-18-year-old students with a B1 English level. The task had to be set in a museum, focussed on speaking skills and include learning aims, a detailed description of the phases, exercises, materials and ICT tools used.

In April 2023, data was collected anonymously through a mixed-method approach (Dörnyei, 2007), including 33 student teacher online surveys, 2 student teacher online interviews, 39 student teacher interactions with ChatGPT (based on the GPT-3.5 model) and a total of 14 final TBLL activities. A custom ChatGPT webpage was created to gather the chats between the student teachers and AI chatbot, enabling the researcher to directly analyse their interactions, minimising any self-report bias.

The participants were expected first to interact with the chatbot, second, to upload the TBLL activity on Moodle and finally, to fill in the online questionnaire. In the end, the participants were invited for an interview. The evaluation criterion for the TBLL activity required the participants to simply publish it on Moodle. If the activity adhered to the TBLL structure (Willis, 1996), the student teachers would obtain 0.3 points to add to their final grade.

After gathering the data, the information was organized to address the research questions. The quantitative findings from the online survey were integrated with qualitative insights into open-ended questions and online interviews. Content analysis (Dörnyei, 2007; Evans, 2017) was used to examine the responses to the open-ended questions in the online questionnaire. In order to compare student teachers' interactions with ChatGPT to their final TBLL activities, a quantitative content analysis using several categories (original, copied, rephrased and copied-exercise words) was conducted and word counts for each were calculated. To identify the copied sequences, each conversation and related TBLL activity were analysed by employing the *Similarity Texter* tool (Kalaidopoulou & Weber-Wulff, 2016) (see an example in Figure 1). No further tools were used to verify whether original words were copied from other sources. Regarding interactions between student teachers and ChatGPT, a further content analysis (Dörnyei, 2007; Evans, 2017) was performed categorising the different types of prompts provided by the participants to ChatGPT.

Research findings

The results consisted in examining to what extent ChatGPT was useful for student teachers to design TBLL activities in terms of planning time, quality of the final products and development of student teachers' critical thinking skills. It is important to mention that some student teachers (39.4%) did not know ChatGPT and the majority (78.8%) did not use the AI chatbot before this research. The first findings demonstrated that all student teachers (100%) considered ChatGPT as a time-saving AI tool to create TBLL activities, especially in the creation of exercises (84.8%), brainstorming ideas (72.7%) and structuring the task (66.7%).

The second outcomes suggest that most student teachers (97%) believed they had created a better or same quality TBLL activity thanks to ChatGPT. In fact, student

teachers' interviews revealed that original ideas for their classes came up and their self-confidence in teaching abilities increased when utilising ChatGPT.

The third results reveal a mixed experience among student teachers when using ChatGPT. A majority (61.3%) reported not encountering any difficulties or limitations with the AI tool. In contrast, 38.7% did experience challenges. Specifically, only 25.8% of all student teachers emphasised the importance of critically analysing ChatGPT's outputs and expressed dissatisfaction with the answers provided by the chatbot. Participants noted that they often needed to rewrite certain parts and that the AI sometimes hallucinated links or provided inaccurate information about museums. Furthermore, 19.4% encountered difficulties in formulating prompts to obtain meaningful answers or suggestions, particularly during the initial phase of use. In general, these outcomes may raise concerns about an excessive reliance on ChatGPT. In fact, researchers (Cooper, 2023; Jalil et al., 2023; Rudolph et al., 2023) underlined that human judgment is necessary when using ChatGPT to interpret results and make ethical decisions, ensuring the AI tool is applied appropriately.

Indeed, the overall findings indicate a pressing need to organise trainings for student teachers in proper ChatGPT use, outlining its limits to enhance critical thinking skills. This recommendation aligns with previous studies (Kasneci et al., 2023), supporting continuous training for teachers and students to keep updated on ChatGPT's latest advancements.

Furthermore, the comparison between student teachers' interactions with ChatGPT and their related final TBLL activities (Figure 2) reveals that, on the one hand, 6 out of 14 TBLL activities present minimal copying from ChatGPT (less than 14% of copied words). In fact, checking one by one those activities, it can be seen that the student teachers analysed, interpreted, elaborated, and modified ChatGPT's responses, demonstrating critical thinking skills. Moreover, in the two interviews, the student teachers explained that each member studied the structure and the characteristics of TBLL model before using ChatGPT. Then, they discussed the content and elements to include in their activity. Finally, the group members interacted with ChatGPT and collaboratively reflected on each of its responses.

The overall findings indicate a pressing need to organise training courses for student teachers in the proper use of ChatGPT, outlining its limits to enhance their critical thinking skills.

On the other hand, 6 out of 14 TBLL activities contain a high percentage (more than 45%) of copied text. This result may suggest a mix of psychological, pedagogical and practical factors that could explain the observed reliance on copied words. First, student teachers possibly faced time constraints, causing them to prefer the quicker approach to copy words directly. Second, the participants might lack confidence in their teaching abilities or knowledge about the TBLL methodology (Willis, 1996), which led to an increased tendency to trust ChatGPT's responses and opt for copying to avoid mistakes. Third, the instructions of how to perform the task could have been misunderstood. Fourth, a few student teachers may be primarily focussed on getting some points easily for their final grade in the university subject. Consequently, they copied ChatGPT's answers without applying critical thinking skills.

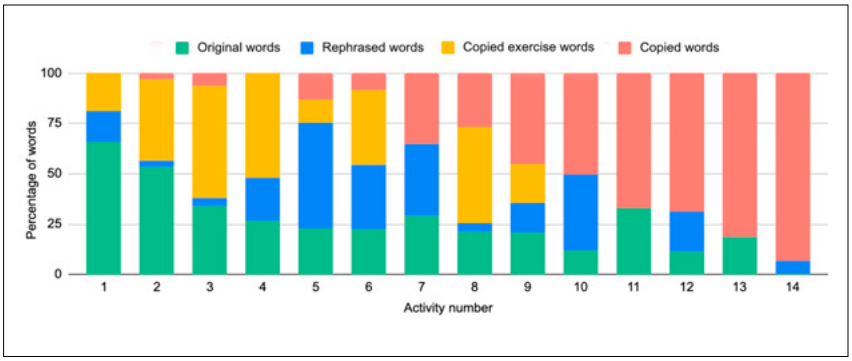

Figure 2

Bar chart displaying the percentages of original, rephrased, copied exercises and copied words from the student teacher ChatGPT conversations in the 14 TBLL activities.
Note. The analysis was performed solely using the Similarity Texter tool (Kalaidopoulou & Weber-Wulff, 2016).

These final outcomes may indicate that ChatGPT offers valuable ideas for designing TBLL tasks and demonstrates the potential to encourage student teachers to apply critical thinking when using the AI tool, as also indicated by Jiang et al. (2024). For student teachers who did not effectively use their critical thinking skills, these findings underline the need to improve assessment methods to foster intrinsic motivation and address plagiarism. This aligns with the recommendations of some scholars (Liu & Wei, 2024; Rudolph et al., 2023), advocating for assignments that require higher-order thinking and creativity.

The second research question aimed to examine student teachers' interactions with ChatGPT in designing TBLL activities, particularly exploring self-developed strategies without prior training. Outcomes reveal various interaction methods, including requests for general information (16.7%), material creation (8.3%), idea brainstorming (54.2%) and learning objective formulation (25%). Most student teachers asked for precise items, but nearly a third (29.1%) tried to generate the entire task with a single input and prompt quality often lacked critical details, affecting ChatGPT's accurate outputs. According to the scientists Fulford and Ng (2023), inputs have to follow some principles: they should be clear, detailed and include delimiters, when necessary. In addition, the prompts should be fragmented, giving ChatGPT time to process. This means that the correct design of a TBLL activity strongly depended on the detailed student teachers' inputs given to ChatGPT. Moreover, some requests exceeded ChatGPT's capabilities at that time like requesting for valid internet links, for instance, the student teachers asked: *Could you suggest a brief introductory video about Frida Kahlo?*, *Can you please link a website where I can find more information about the ongoing exhibition in Padua you mentioned?*, underlining the need for a training in effective interactions and understanding the chatbot's limitations.

For student teachers who did not effectively use their critical thinking skills, these findings underline the need to improve assessment methods to foster intrinsic motivation and address plagiarism.

Conclusions

Overall, the current exploratory study demonstrates that the student teachers could save time and create equivalent or better-quality TBLL activities using ChatGPT. Moreover, the chatbot was considered a useful AI tool, capable of giving a feeling of self-confidence in their teaching abilities. However, some participants seemed to fail to recognise ChatGPT's limitations, with about half copying outputs directly, indicating possible misunderstanding of assignment instructions or inadequate assessing methods for the subject. Although the student teachers employed various strategies to interact with ChatGPT, some participants faced challenges of writing prompts. To address these issues, the researcher recommends a training course to teach proper interactions with ChatGPT, comprehending its limits, avoiding plagiarism and boosting participants' critical thinking skills. Furthermore, reassessing assignment methods to emphasise critical thinking, problem solving and creativity is strongly advised.

However, the current investigation presents some limitations due to the specific sample and context for data collection; therefore, it would be worthwhile to replicate the study with a broader range of classes from different years or assigning other activities beyond TBLL. Furthermore, future research could delve deeper into a comparative analysis of the TBLL activities designed by student teachers and their application of critical thinking skills before and after a training in ChatGPT usage. Additionally, empirical investigations into ChatGPT's effectiveness in promoting student teachers' self-confidence would be of considerable interest.

Bibliography

Ali, J., Shamsan, M., Hezam, T., & Mohammed, A. (2023). Impact of ChatGPT on Learning Motivation: Teachers and Students' Voices. *Journal of English Studies in Arabia Felix, 2*(1), 41-49. https://doi.org/10.56540/jesaf.v2i1.51.

Baradel, C. (2023). *Interaction, Design, and Assessment: An Exploratory Study on ChatGPT in Language Education* [Master's Thesis, Università Ca' Foscari Venezia]. DSpace. http://hdl.handle.net/10579/25711.

Bellaera, L., Weinstein-Jones, Y., Ilie, S., & Baker, S. (2021). Critical thinking in practice: The priorities and practices of instructors teaching in higher education. *Thinking Skills and Creativity, 41*, 1-16. https://doi.org/10.1016/j.tsc.2021.100856.

Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to- implement activities. *Contemporary Educational Technology, 15*(3), 1-10. https://doi.org/10.30935/cedtech/13176.

Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology, 32*(3), 444-452. https://doi.org/10.1007/s10956-023-10039-y.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford University Press.

Evans, M. (2017). Analysing Qualitative Data. In E. Wilson (Ed.), *School-based Research: A guide for education students*. Sage publications.

Fulford, I., & Ng, A. (2023, July). *ChatGPT Prompt Engineering for Developers* [Online course]. DeepLearning.AI. https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/.

García-Peñalvo, F. J. (2023). The perception of Artificial Intelligence in educational contexts after the launch of ChatGPT: Disruption or Panic? *Education in the Knowledge Society (EKS), 24*. https://doi.org/10.14201/eks.31279.

Jalil, S., Rafi, S., LaToza, T., Moran, K., & Lam, W. (2023). ChatGPT and Software Testing Education: Promises & Perils. *2023 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, 4130-4137. DOI: 10.1109/ICSTW58534.2023.00078.

Kalaidopoulou, S., & Weber-Wulff, D. (2016). Similarity Texter [Computer software]. *Hochschule für Technik und Wirtschaft Berlin*. https://people.f4.htw-berlin.de/~weberwu/simtexter/app.html.

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., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 1-9. https://doi.org/10.1016/j.lindif.2023.102274.

Jiang, J., Li, J., & Chen, C. H. (2024). Enhancing Critical Thinking Skills with ChatGPT-Powered Activities in Chinese Language Classrooms. *International Journal of Chinese Language Teaching, 5*(1), 47-73. https://doi.org/10.46451/ijclt.20240105.

Lin, S. M., Chung, H. H., Chung, F. L., & Lan, Y. J. (2023). Concerns About Using ChatGPT in Education. In Y. M. Huang & T. Rocha (Eds.), *Innovative Technologies and Learning. ICITL 2023. Lecture Notes in Computer Science*, vol. 14099. Springer, Cham. https://doi.org/10.1007/978-3-031-40113-8_4.

Liu, Y., & Wei, L. (2024). A Study on Students' Critical Thinking Cultivation in English Teaching in the ChatGPT Era. In I. A. Khan et al. (Eds.), *Proceedings of the 2024 2nd International Conference on Language, Innovative Education and Cultural Communication (CLEC 2024)*, Advances in Social Science, Education and Humanities Research 853. Atlantis Press. https://doi.org/10.2991/978-2-38476-263-7_10.

Muñoz, S. A. S., Gayoso, G. G., Huambo, A. C., Tapia, R. D. C., Incaluque, J. L., Aguila, O. E. P., Cajamarca, J. C. R., Acevedo, J. E. R., Rivera, H. V. H., & Arias-Gonzáles J. L. (2023). Examining the Impacts of ChatGPT on Student Motivation and Engagement. *Social Space, 23*(1), 1-27.

OpenAI. (2022, November 30). Introducing ChatGPT. *OpenAI*. https://openai.com/index/chatgpt/.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 1-22. https://doi.org/10.37074/jalt.2023.6.1.9.

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., Hayek, S. E., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence, 11*(3), 54. 1-32. https://doi.org/10.3390/jintelligence11030054.

Willis, J. (1996). *A framework for task-based learning*. Longman.