Does using LLMs in daily life help or hinder learning a second language?

Wei Li, Andy Zhao, Adrian de Wynter,
Si-Qing Chen, Paul Karimov, Joshua Hartshorne

Introduction

There are two competing hypotheses about how AI tools affect language learning:

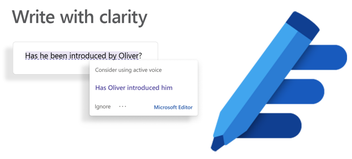
A. AI tools helps with the language learning, because users receive more assistances from AI *(Shaikh et al., 2023; Song and Song, 2023; Xiao and Zhi, 2023)*

B. AI hinders the language learning, because users over-rely on the AI (Kosmyna et al., 2025)

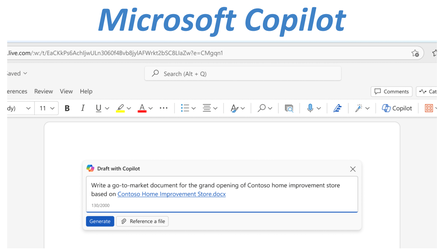
Both hypotheses sound reasonable, but…

- What dimensions in the language abilities?
- The way of using AI for writing matters
- The exact AI product (interactive pattern and the AI capacity) matters
  - In-text assistant (non-LLM powered) vs. ChatBot (LLM-powered)

Microsoft Editor



Microsoft Copilot



Method

6-month Experiment

(1) Recruitment

- We recruited 24 international college students
  - The English non-native speakers who just came to the US
  - Students have more usage scenarios for writing tasks
  - Ended up with 15 participants in our final survey

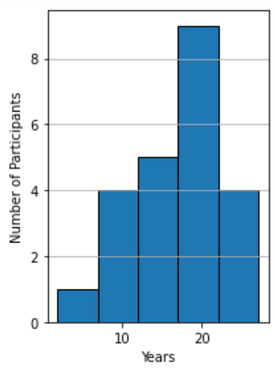


Figure 1. Distribution of participants' learning duration

(2) Procedure

- For every month in this 6-month study
  - Answer 60 lexical decision questions
  - Write an essay: prompts from TOEFL
  - Self-report their usages of AI tools: frequency & deepness

(3) Treatment

- Participants were randomly provided different levels of AI access in the MS Word
  - Both groups have AI access as AI is already everywhere.
  - We provided MS Editor access to all participants.
  - Treatment group receive a more advanced LLM-powered chatbot **MS Copilot** + nudge emails every month

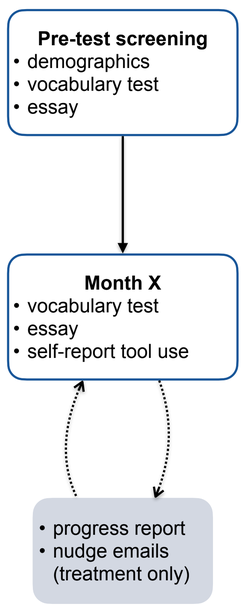


Figure 2. Data collection procedures

Tool Usage Measurement

Table 1 Metrics of tool usages

Score	Frequency of chatbot/in-text assistant tool use	Deepness of tool use
0	Not at all	I quickly skim the suggestions or generated text without much thought.
1	Rarely: 1-2 times	I briefly review the suggestions and sometimes make small edits.
2	Occasionally: 3-5 times	I carefully read the suggestions, try to understand them, and make thoughtful edits.
3	Frequently: 6-10 times	I analyze the suggestions in detail, compare them with my original writing, and think critically about how to improve.
4	Always: more than 10 times	I thoroughly evaluate the suggestions, research related concepts, and actively apply what I learn to future writing tasks.

Language Ability Measurement

- Vocabulary score: LexTale (Lemhöfer, K., & Broersma, M., 2012)
- Essays were graded by three professional annotators in three metrics (score range: 0-5)

Table 2 Grading guideline of essays

Metrics	Frequency of use chatbot/in-text assistant tool
Idea	how the idea is developed and elaborated by explanations, exemplifications, and details
Fluency	How the writing flows and phrase reads naturally throughout
Accuracy	How much lexical or grammatical errors are in the writing

Results

Is the manipulation effective?

- No significant group differences over time in Chatbot or in-text assistant

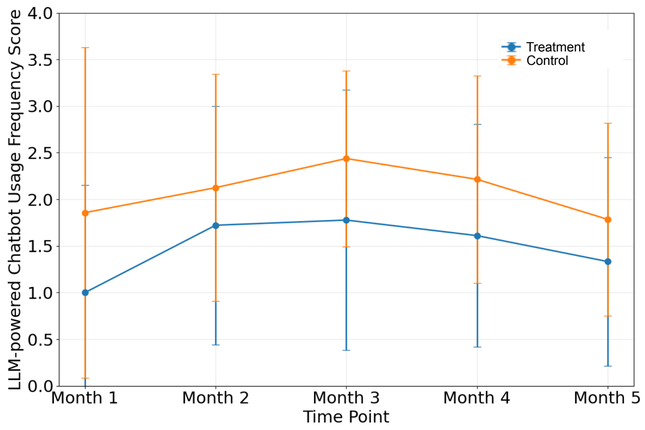


Figure 3. Chatbot writing tool use frequency over time

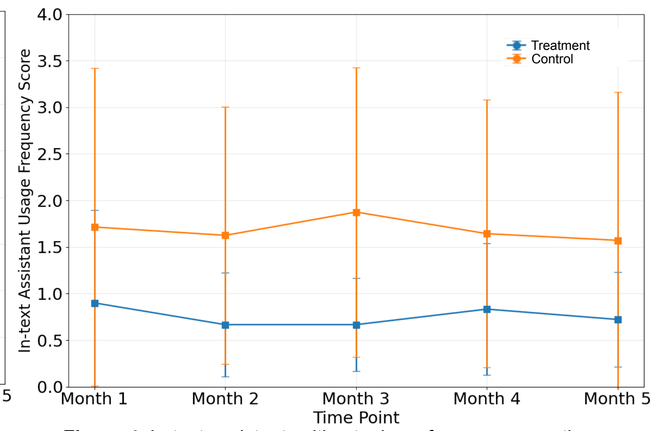


Figure 4. In-text assistant writing tool use frequency over time

Table 3 Regression table

Predictors	Effects on Vocabulary Score	Effects on Essay Composite Score	Effects on Essay Idea Score
Timepoints	-1.3	0.1**	0.15***
Frequency of Chatbot Use	-2.12	-0.21	-0.18
Frequency of In-text Assistant Tool Use	0.1	0.07	0.014
Deepness of Tool Use	5.12***	0.24	0.30*
Interactions	/	>.05	Chatbot_use*Deepness, -0.26*

Note: the best fitting model for vocabulary score has random slope without interactions; the best fitting model for essay score has fixed slope with interactions

How does the language ability change over 6 months?

- Vocabulary: no significant change over 6 months.
- Essay: the idea scores (p<.001***) and fluency scores (p<.05*) increases over 6 months, but not the accuracy score.
- Learners' second language abilities keep improving even after 20-year of learning, but most in how they use the language not the basic knowledge.

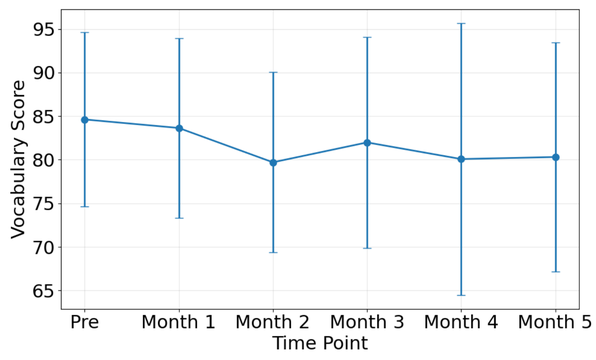


Figure 5. Vocabulary scores over time

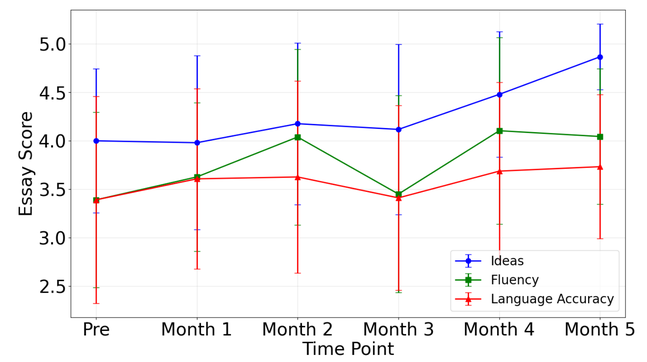


Figure 6. Essay scores over time

Is the use of writing tools correlated with language development?

- Frequency of AI use did not affect vocabulary scores or essay scores.
- The more deeply users processed AI feedback, the higher their scores (p<.05*).
  - The learners who analyze and evaluate tools' suggestions have higher scores than whose just skim and accept the suggestions.
  - The effect is weakened by the use frequency of chatbot in essay idea scores.

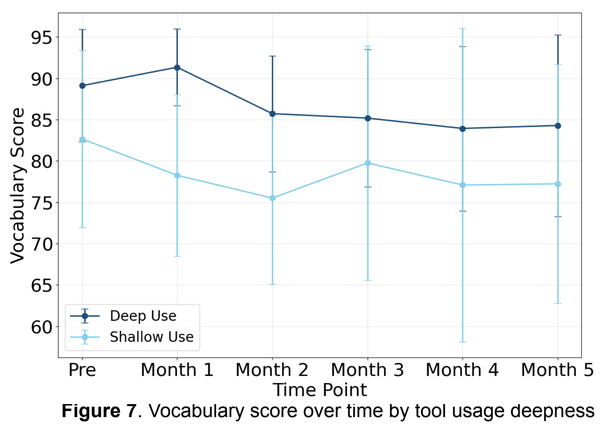


Figure 7. Vocabulary score over time by tool usage deepness

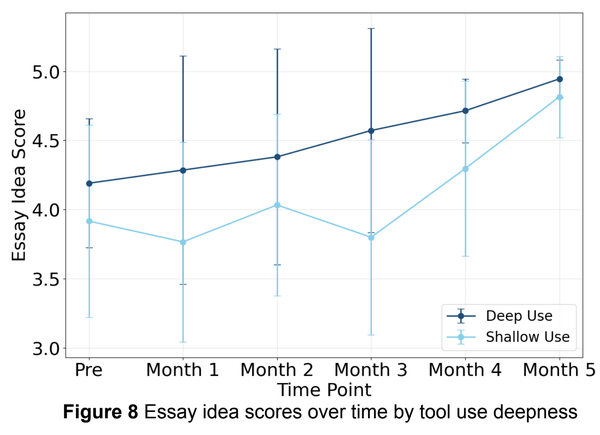


Figure 8 Essay idea scores over time by tool use deepness

Discussion

Take-aways

- Manipulating tool use is extremely hard!
- Advanced second language learners' language proficiency keeps improving.
- The way of learners using language tools affect their second language proficiency.

Limitation

- Small sample size and limited measurements of language ability
- Our provided AI is limited in user experience and not effective in manipulation
- Uncontrolled AI usage among participants in the real world

Reference

1.Shaikh, S., Yayilgan, S. Y., Klimova, B., & Pikhart, M. (2023). Assessing the usability of ChatGPT for formal English language learning. *European Journal of Investigation in Health, Psychology and Education*, *13*(9), 1937-1960.
2.Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology,* 14, 1260843.
3.Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), 212.
4.Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, *44*, 325-343.
5.Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., ... & Maes, P. (2025). Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. *arXiv preprint arXiv:2506.08872*.