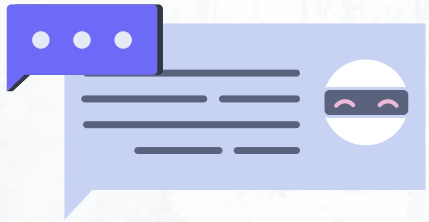
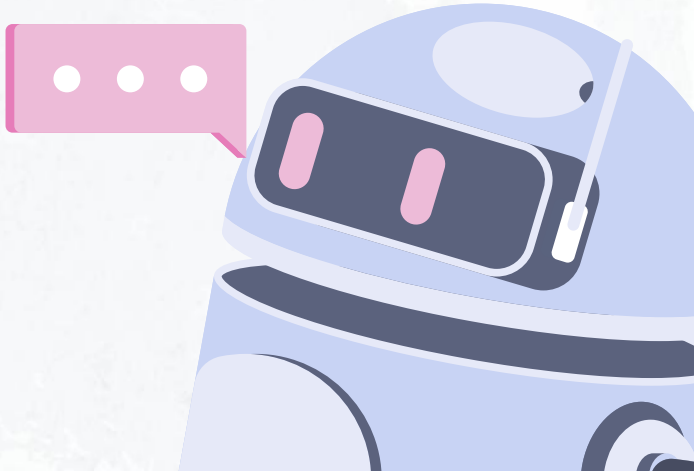
Generative AI, the teacher's best friend?

Evaluating GenAI for creating Chinese reading comprehension materials

Helena Casas-Tost, Isabel González-Torrents,
Sara Rovira-Esteva and Mireia Vargas-Urpí





GELEA2LT

UAB
Universitat Autònoma
de Barcelona

GELEA2LT's projects

Language and Translation: Literacy in Digital Environments and Resources

LT-LiDER (ref. 2023-1-ES01-KA220-HED-000161531) has been funded with support from the European Commission. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Description of **Generatese** in Second Languages and Translation

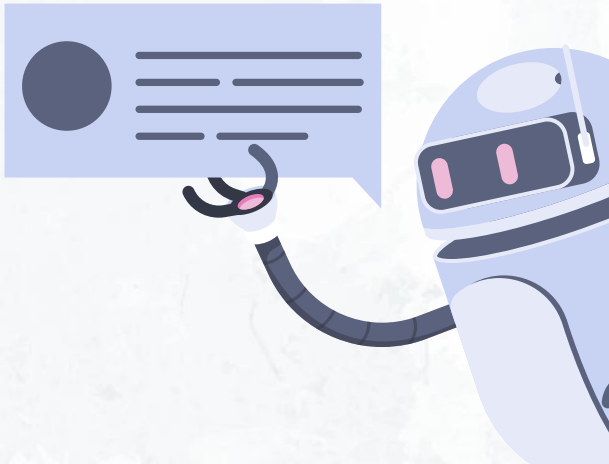
This work was supported by the Spanish Ministry of Science and Innovation under the project "Description of Generatese in Second Languages and Translation" 2025-2029 (PID2024-156763OB-100).

http://lt-lider.eu/

Contents

01 →

Previous research



Exponentially growing body of research

Literature Review

SAGE Open
January-March 2026: 1–17
© The Author(s) 2026
DOI: 10.1177/21582440261418315
journals.sagepub.com/home/sgo
Sage

Roles of Generative Artificial Intelligence (GenAI) in English as a Foreign Language (EFL) Instruction: A Systematic Literature Review

Luying Deng[1] and Khairul Azhar Jamaludin[1]

INTERACTIVE LEARNING ENVIRONMENTS

Routledge
Taylor & Francis Group

Interactive Learning Environments

ISSN: 1049-4820 (Print) 1744-5191 (Online) Journal homepage: www.tandfonline.com/journals/nile20

Generative AI (GenAI) in the language classroom: A systematic review

Seongyong Lee, Hohsung Choe, Di Zou & Jaeho Jeon

Contents lists available at ScienceDirect

ELSEVIER

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching

Belle Li[a,*], Victoria L. Lowell[a], Chaoran Wang[b], Xiangning Li[c]

ELSEVIER

Computers and Education Open

Volume 6, June 2024, 100174

Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review

Locky Law (Locky lok-hei Law)

GenAI's revolution in language learning?

GenAI shows great potential

Li Belle et al. (2024: 12) reviewed 36 articles on the use of ChatGPT in language learning and concluded that ChatGPT has the potential "to enhance teacher productivity and support the development of educational materials".

... but there is a lack of empirical studies

There is a difference between "research based on potential perceptions rather than actual measurable outcomes" (Stockwell, 2024: 5).

and a bias towards English

It is important considering "a wider range of language proficiency levels, learning objectives and educational levels" (Li Belle et al., 2024: 12-13).

Lee et al. (2026) – A systematic review of GenAI in the language classroom

(*) Studies largely rely on self-reported data

- Perception studies are more frequent than acquisition studies or behaviour studies which cover classroom interaction or interaction with GenAI

(*) Main research focus (n=46):

- Bias towards English as target language*
- An important bias towards writing (25 perception studies, 15 acquisition studies) with respect to other skills (e.g. 4+1 studies on reading)*
- Roles of GenAI: feedback provider (42), language learning tutor (26), resource provider (22), content generator (20), evaluator (14), affective supporter (12), conversation partner (12), interaction facilitator (10), cognitive simulator (8)

*Deng and Jamaloulin (2006) also note these biases

Studies focusing on reading comprehension tasks

(*) Drackert et al. (2025), German as an additional language

- ChatGPT-4: greater variation in content words, higher lexical density, denser concentration of information.
- ChatGPT-4 texts: significantly more adjectives, conjunctions, determiners, and nouns; benchmark texts have higher frequencies of adverbs, numerals, pronouns, proper nouns, and verbs.
- “AI-generated texts can be used as a starting point for creating new text input (...), however, these texts should always be refined by experts. “

(*) Jen et al. (2025), English as an additional language

- “While GenAI tools demonstrated efficiency and consistency in content generation, human teachers produced higher-quality materials, particularly in relevance, clarity, and alignment with educational objectives” (p. 122).

Studies focusing on reading tasks for Chinese as an additional language (CAL)

(*) Wu & Lin (2026) - ChatGPT for “text simplification”

- Comparison of L5 texts simplified by humans vs. same texts simplified by ChatGPT using three different kinds of prompting (zero-shot, few-shot, instruction-based)
- “The results indicate that ChatGPT-4o can produce systematic linguistic simplification, particularly in lexical and syntactic domains, and that few-shot and instruction-based prompting yield outputs closer to expert adaptations on several key measures”.
- Regarding content-level adaptation, “ChatGPT primarily reduced difficulty through compression and deletion, especially when faced with abstract or culturally embedded material. Expert teachers, however, more frequently employed **pedagogically oriented strategies** such as contextualization, conceptual reframing, and discourse-level restructuring to support learner comprehension.”

02 →

Objectives*

*Casas-Tost, Helena; González-Torrents, Isabel; Rovira-Esteva, Sara; Vargas-Urpí, Mireia (under revision).

Objectives

1. To assess the usefulness of GenAI as a teaching assistant for Chinese as an additional language.
2. To evaluate the potential of AI-generated texts in developing reading comprehension activities for learners at the CEFR A2 level.
3. To compare the performance of two GenAI systems (ChatGPT-4o and DeepSeek-R1) in generating texts and questions for reading comprehension.
4. To develop an evaluation tool for AI-generated reading comprehension activities to determine their cost-effectiveness in an educational setting.

03 Method

First step: prompt design
+ generation of activities

Generation of the activities (texts and questions)

- Design of four prompts covering four topics (daily routines, holiday plans, my Chinese family, looking for a flatmate) to generate texts + questions
- The same prompts were used for ChatGPT-4.o and DeepSeek-R1
- Prompts in Spanish
- Instruction prompting strategy (no example provided, but clear instructions about task definition, contextual information, output format specification, style)
- Texts for CEFR A2 or level 3 in HSK 3.0 (our previous research focused on A1).

Example of one prompt

You are a teacher of Chinese as a foreign language. Write a text in Chinese to create a reading comprehension activity at the basic level (corresponding to A2 of the CEFR or HSK level 3) about the daily routine of an exchange student studying Chinese at a university in Beijing. The text should include at least three paragraphs and contain enough content to develop a subsequent reading comprehension activity. Use only language appropriate for the basic level.

Original prompt in Spanish available in Casas-Tost, et al. (2025a)

03 →

Second step: assessment of chatbots outcomes

Assesment tool used for texts*

Text: Level

- ❑ Appropriate vocabulary and/or grammar
- ❑ Vocabulary and/or grammar too advanced and not acceptable
- ❑ Vocabulary and/or grammar too advanced but acceptable
- ❑ Vocabulary and/or grammar too basic

"Acceptable" means that the necessary changes are minimal or that the text includes words that, while not immediately understood, can be inferred from context.

*Casas-Tost, et al. (2025b)

Assessment tool used for texts*

Linguistic Accuracy (vocabulary, grammar, and/or punctuation)

- ❑ Correct
- ❑ Incorrect but easily correctable (examples)
- ❑ Incorrect and not acceptable due to the number of errors (examples)

Pragmatics

- Even if you understand the text, do you think there are elements that a native speaker would not phrase this way? If so, which ones, and how would you rephrase them?
- Even if you understand the text, do you think there are ideas that a native speaker would not include in such a text? If so, which ones?

Assesment tool used for texts*

Content

- Lexical diversity and richness: Yes/No
- Sufficient content to assess reading comprehension (RC): Yes/No

Overall Evaluation

- ❑ Acceptable without changes
- ❑ Acceptable with minor revisions
- ❑ Not acceptable

Total time spent on text revision: _____

Cathoven used as supplementary tool

−∞ HSK1 HSK2 HSK3 HSK4 HSK5 HSK6 HSK7-9 +∞

留学生的日常生活

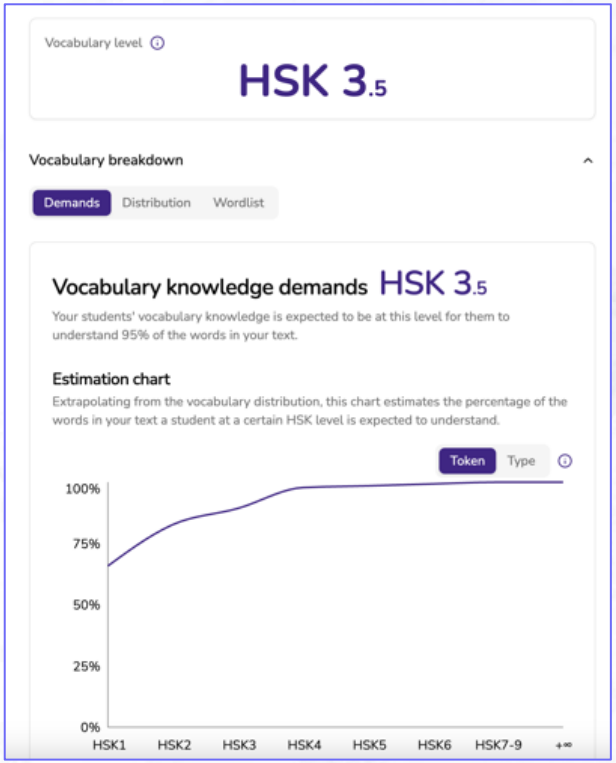
王明是一名来自西班牙的留学生，他在北京的一所大学学习中文。他每天早上七点起床，洗脸、刷牙，然后去食堂吃早饭。他喜欢吃包子和喝豆浆。早上八点，他去教室上课。他的老师很友好，课堂内容很有趣，大家一起学习生词、语法和汉字。

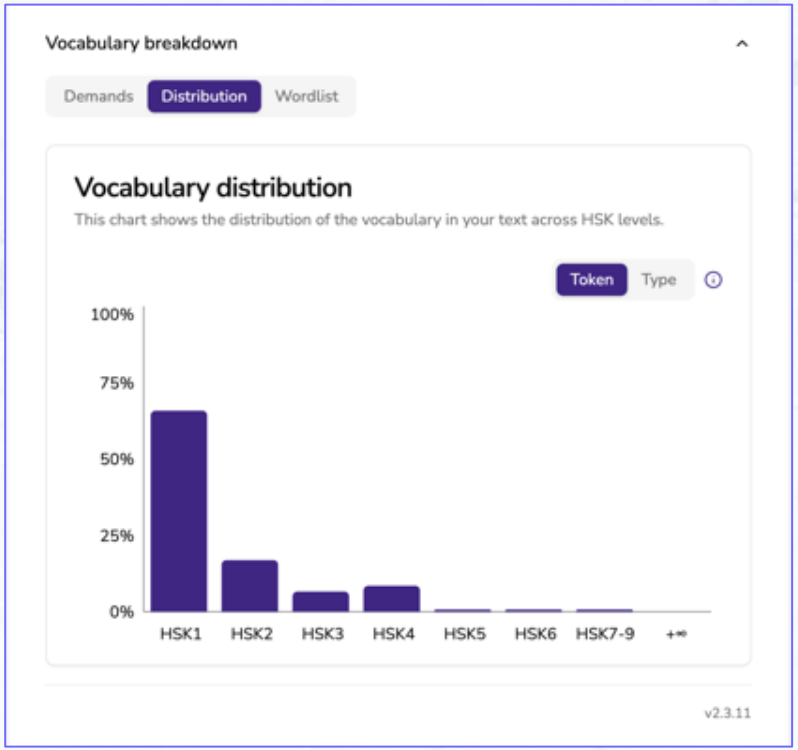
中午十二点，王明和他的同学们一起去学校食堂吃午饭。他们常常吃米饭、炒菜和汤。吃完饭后，他有时候去图书馆看书，有时候在宿舍休息。下午两点到四点，他继续上课。老师会教他们如何用中文交流，大家一起练习口语。

晚上，王明喜欢去学校的体育馆打篮球，也会和朋友一起去附近的公园散步。他晚上十点睡觉，每天都很开心。他觉得在北京学习中文很有意思，因为他不仅学到了很多新知识，还交了很多中国朋友。

https://www.cathoven.com/

Cathoven used as supplementary tool



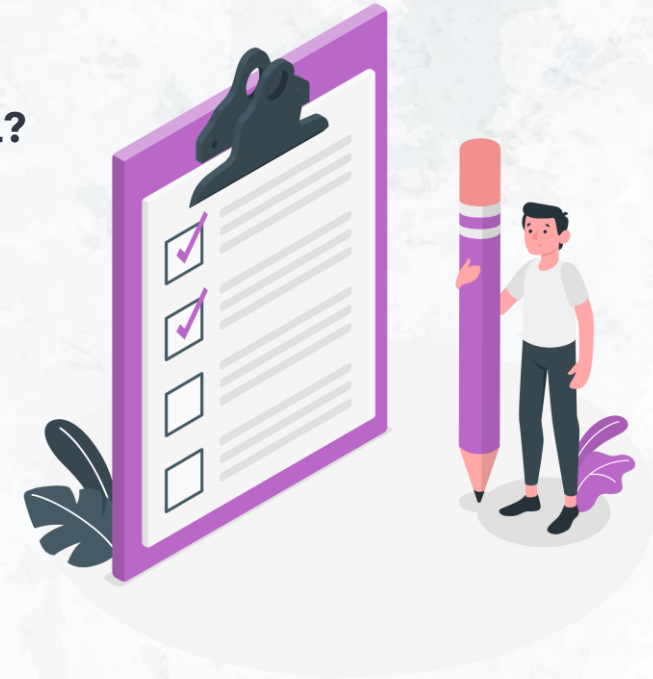


https://www.cathoven.com/

Assessment tool used for questions

1. Do the questions match the language level?

- ❑ Yes
- ❑ No, level too advanced
- ❑ No, level too basic



Assessment tool for questions

2. Are all the questions appropriate?

- ❑ Yes
- ❑ No

If not, highlight in colours the questions that you find inappropriate. Why are they not suitable? (Select all that apply)

- The answer is too explicit in the text (too obvious and/or identical wording as in the text). [yellow]
- The distractors in the answer choices are unrealistic (not present in the text). [orange]
- They are not relevant to the text because they are unrelated to its content. [red]

Assessment tool for questions

3. Overall evaluation

- ❑ Acceptable without changes
- ❑ Acceptable with minor changes
- ❑ Not acceptable

04 →

Results

(A)

Assessment of 8 activities using the tool

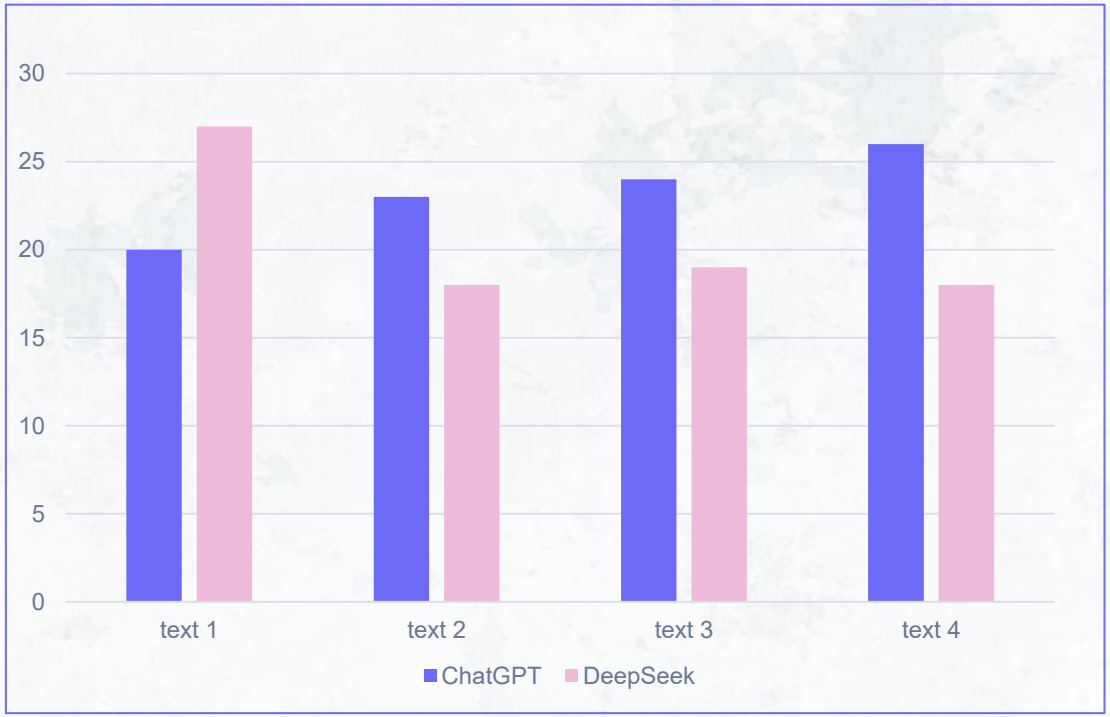
	ChatGPT-4o Text	ChatGPT-4o questions	DeepSeek-R1 Text	DeepSeek-R1 Questions
Activity 1. Routines	Acceptable, minor changes	Multiple choice: not acceptable T/F: Acceptable, minor changes	Acceptable, no changes needed	Multiple choice: not acceptable T/F: Acceptable, minor changes
Activity 2. Holidays	Acceptable, no changes needed	T/F: Acceptable, minor changes	Version 1: Not acceptable Version 2: Acceptable, minor changes	Based on version 2 T/F: not acceptable
Activity 3 – My Chinese family	Acceptable, minor changes	T/F: Acceptable, minor changes	Acceptable, minor changes	T/F: not acceptable
Activity 4 - Flatmate	Acceptable, minor changes	No consensus reached	Acceptable, minor changes	T/F: not acceptable

Analysis of 8 activities using Cathoven

	ChatGPT-4o	DeepSeek-R1
Activity 1	3.5	3.1
Activity 2	3.4	(6.0) 4.00
Activity 3	2.9	2.7
Activity 4	4.2	4.6
Average	3.5	3.6

Survey regarding texts' naturalness

- 30 native Chinese speakers were surveyed about the texts
- Likert scale (1-5).
- The graphic shows aggregated answers for 4 (rather natural) and 5 (very natural).



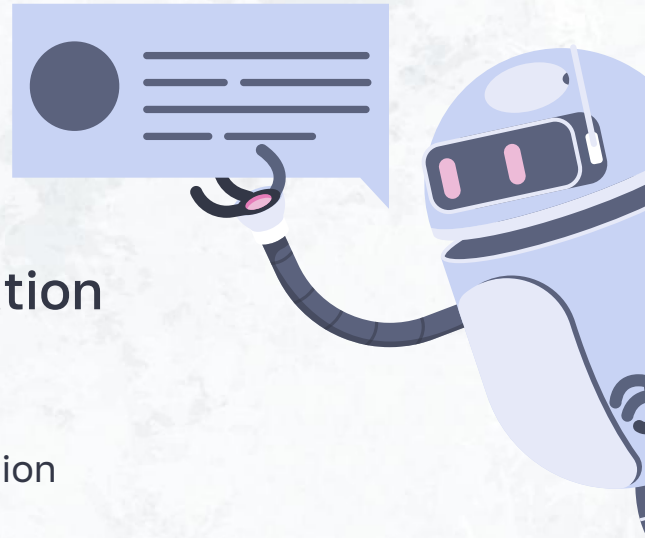
05 →

Concluding remarks

To what extent was GenAI useful?

(+) Good results in text generation

- Better results in A2/HSK3 than in A1
- Very minor changes needed in general



(-) Less efficient regarding question generation

- Questions require a pedagogical approach
- Questions must be useful to evaluate reading comprehension

Other conclusions

(*) ChatGPT vs. DeepSeek

- ChatGPT yielded better outcome than DeepSeek (according to the assessment tool and to the survey)

(*) Assessment process

The combination the ad hoc tool + Cathoven proved useful

(*) Work in progress and future research

- Definition of the “generatese” in texts produced for CAL
- Larger scale project, involving the assessment of more texts and comparing them with human-created texts
- Comparison with other GenAI systems (e.g. Doubao)

References

Casas-Tost, Helena, Isabel González-Torrents, Sara Rovira-Esteva & Mireia Vargas-Urpí. 2025a. Materiales de análisis de actividades de comprensión escrita generadas con inteligencia artificial. *GELEA2LT Working Papers.* https://ddd.uab.cat/record/313580.

Casas-Tost, Helena, Isabel González-Torrents, Sara Rovira-Esteva & Mireia Vargas-Urpí. 2025b. Herramienta de evaluación de actividades de comprensión escrita generadas con inteligencia artificial. https://ddd.uab.cat/record/313581.

Casas-Tost, Helena, Isabel González-Torrents, Sara Rovira-Esteva & Mireia Vargas-Urpí. Under revision. Using GAI for reading comprehension activities in the teaching and learning of Chinese as an additional language.

References

Deng, Luying, & Jamaludin, Khairul A. (2026). Roles of Generative Artificial Intelligence (GenAI) in English as a Foreign Language (EFL) Instruction: A Systematic Literature Review. *Sage Open*, 16(1).

Drackert, A., Horbach, A., & Peters, A. (2025). How good are LLMs in generating input texts for reading tasks in German as a foreign language? *Annual Review of Applied Linguistics*, 45, 222.

Jen, FL., Huang, X., Liu, X., Jiao, J. (2024). Can Generative AI Really Empower Teachers' Professional Practices? Comparative Study on Human-Tailored and GenAI-Designed Reading Comprehension Learning Materials. *Communications in Computer and Information Science*, vol 2330. Springer, Singapore.

Law, Locky (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6, 100174.

Lee, Seongyong; Hohsung Choe, Di Zou & Jaeho Jeon (2026) Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, 34:1, 335-359

Li, Belle, Victoria L. Lowell, Chaoran Wang & Xiangning Li. 2024. A Systematic Review of the First Year of Publications on ChatGPT and Language Education: Examining Research on ChatGPT's Use in Language Learning and Teaching. *Computers and Education: Artificial Intelligence* 7

Stockwell, Glenn. 2024. ChatGPT in Language Teaching and Learning: Exploring the Road We're Travelling. *Technology in Language Teaching and Learning* 6 (1): 2273.

Wu, J., & Lin, Y. (2026). ChatGPT for Text Simplification in Chinese as a Second Language Textbooks: A Comparative Study With Expert Adaptations. *International Journal of Applied Linguistics*. https://doi.org/10.1111/ijal.70147

Thanks! :)

mireia.vargas@uab.cat