



Foreign Language Teachers' Perceptions of the Use of a Generative AI Application in Designing Reading Classroom Assessments¹

Alexis A. López 

Southern New Hampshire University
Manchester, USA

Gabriel Cote Parra 

Universidad de Pamplona
Pamplona, Colombia

1/32

Abstract

Artificial intelligence (AI) has become an essential part of language assessment. AI tools are used for automated item generation, scoring, and feedback, improving the development, administration, and interpretation of large-scale language assessments. This study examines the use of an AI tool, ChatGPT, in classroom assessment and proposes a way for teachers to use the tool to simplify the language complexity of written texts (i.e., reading passage) and generate comprehension questions for basic English level (A1-A2). Seven English as a foreign language (EFL) teachers currently teaching basic users participated in one-on-one interviews to discuss their perceptions of ChatGPT, and to evaluate the quality and appropriateness of the simplified text and the questions. The study illustrates how ChatGPT was used to generate the assessment content and presents the teachers' perceptions, offering implications for using generative AI tools to design reading classroom assessments.

Key words: artificial intelligence; generative AI; ChatGPT; classroom assessment; teachers' perceptions.

¹ Research article

Resumen

Percepciones de los profesores de lenguas extranjeras sobre el uso de una aplicación de IA generativa para el diseño de evaluaciones de lectura en el aula

La inteligencia artificial (IA) se ha convertido en una parte esencial de la evaluación en lenguas extranjeras. Las herramientas de inteligencia artificial se utilizan para la generación, calificación y retroalimentación automatizadas de ítems, mejorando el desarrollo, la administración, y la interpretación de evaluaciones en lenguas extranjeras a gran escala. Este estudio examina el uso de una herramienta de inteligencia artificial, ChatGPT, en la evaluación en lenguas y propone una forma para que los maestros utilicen la herramienta para simplificar la complejidad del lenguaje de los textos escritos (es decir, pasajes de lectura) y generar preguntas de comprensión para estudiantes de nivel básico de inglés (A1-A2). Siete profesores de inglés como lengua extranjera, que actualmente imparten clases de inglés como lengua extranjera de nivel básico, participaron en entrevistas individuales para discutir sus percepciones sobre ChatGPT y evaluar la calidad y adecuación del texto simplificado y de las preguntas. El estudio ilustra cómo se utilizó ChatGPT para generar el contenido de la evaluación y presenta las percepciones de los profesores, lo cual devela implicaciones del uso de herramientas de IA generativa para el diseño de evaluaciones de lectura en el aula.

2/32 **Palabras clave:** inteligencia artificial; IA generativa; ChatGPT; evaluación en el aula; percepciones de los docentes.

Résumé

Perceptions des enseignants de langues étrangères sur l'utilisation d'une application d'IA générative pour la conception des examens de compréhension écrite en classe de langue

L'intelligence artificielle (IA) est devenue un élément essentiel de l'évaluation en langues. Les outils d'IA sont utilisés pour la génération automatisée d'items, la notation et le feedback, améliorant ainsi le développement, l'administration, l'évaluation et l'interprétation de l'évaluation en langues à grande échelle. Cette étude examine l'utilisation d'un outil d'intelligence artificielle, ChatGPT, dans l'évaluation en classe et propose aux enseignants un moyen d'utiliser l'outil pour simplifier la complexité linguistique des textes écrits (c'est-à-dire la lecture de passages) et générer des questions de compréhension pour les apprenants d'anglais de niveau élémentaire. Sept enseignants d'EFL pour des classes de niveaux A1 et A2 ont participé à des entretiens individuels pour discuter de leurs perceptions de ChatGPT et évaluer la qualité et la pertinence d'un texte et des questions simplifiées. L'étude illustre comment ChatGPT a été utilisé pour générer du contenu d'évaluation et présente les perceptions des enseignants, rapportant des implications pour l'utilisation d'outils d'IA générative pour la conception des évaluations de lecture en classe.

Mots-clés : intelligence artificielle ; IA générative ; ChatGPT ; évaluation en classe de langue ; perceptions des enseignants.

ABOUT THE AUTHORS

Alexis A. López

He earned a Ph.D. in Education from the University of Illinois at Urbana-Champaign and is currently a visiting professor at Southern New Hampshire University. His research focuses on developing personalized assessments for multilingual learners, formative assessments, and digital assessments, as well as understanding the classroom assessment practices of ESL/EFL teachers.

E-mail: a.lopez4@snhu.edu

Gabriel Cote Parra

He earned a Ph.D. in Education from the University of Nebraska at Lincoln and is currently a professor in the Bachelor of Arts in Foreign Languages, English-French program at Universidad de Pamplona. His research focuses on the teaching and learning of foreign languages, aligning with the research focus of the Foreign Language Teachers' Research Group (GRILEX).

E-mail: gcote@unipamplona.edu.co

3/32

HOW TO CITE THIS ARTICLE

López, A. A. & Cote Parra, G. (2025). Foreign Language Teachers' Perceptions of the Use of a Generative AI Application in Designing Reading Classroom Assessments. *Lenguaje*, 53(1S), e20214386. <https://doi.org/10.25100/lenguaje.v53i1S.14386>

INTRODUCTION

Artificial intelligence (AI) is a field of computer science focused on developing information-processing computer systems that can perform tasks traditionally requiring human intelligence (Morandín-Ahuerna, 2022; Wang, 2019). AI is significantly reshaping education and presents exciting opportunities for personalized learning, automated assessment, and enhanced teaching methods (Celik et al., 2022). While AI is presently mainly used for automated scoring of speaking and writing, as well as for remote proctoring, its potential applications are extensive. These include content generation, estimating item difficulty, and creating new types of tests, such as those supported by virtual reality.

Teachers play a vital role in integrating AI into education. They act as both mentors for students and evaluators of AI-driven content (Ramakrishnan et al., 2024). While many educators are open to utilizing AI technologies, it is important to recognize that some have concerns about potential misuse and cheating (Zormanová, 2024). AI can assist teachers in various tasks, including planning, implementation, and assessment (Celik et al., 2022). Additionally, it can help reduce workload and enhance information literacy (Chaudhry & Kazim, 2022). However, challenges such as ethical considerations and the need for professional development persist. As AI becomes more prevalent in education, teachers must adapt their roles and acquire new competencies to effectively utilize these technologies for improved learning outcomes (Ramakrishnan et al., 2024). In this study, we investigate the use of AI in classroom assessment and how teachers perceive the content generated by AI technology.

4/32

BACKGROUND

In this section, we provide information about generative AI and its applications in classroom and language assessment.

Generative AI

Generative AI is a type of AI that is used to generate new content such as text, images, audio, and videos (García-Peñalvo & Vázquez-Ingelmo, 2023; Kar et al., 2023; Ramdurai & Adhithya, 2023). Generative AI uses machine learning to train large language models from input data. By utilizing deep learning techniques, generative AI can analyze data patterns and create new content based on what it has learned from existing content (Kar et al., 2023). This technology has far-reaching applications in education and other fields (Satya, 2024; Warankar & Patil, 2024). However, its use also sparks concerns regarding copyright issues, privacy, and ethical considerations (Sidhu, 2024). As generative AI continues to advance, it holds the potential to transform industries and redefine how we

collaborate and communicate (Feuerriegel et al., 2023). Notable examples of generative AI tools include Bard, ChatGPT, Copilot, DALL-E, Gemini, and Midjourney, among others (Bengesi et al., 2023; García-Peñalvo & Vázquez-Ingelmo, 2023).

AI in Classroom Assessment

AI is increasingly used in educational assessment, which brings both opportunities and challenges (Gardner & Yuan, 2021; Owan et al., 2023). AI tools can enhance the accuracy and efficiency of assessments, help students with special needs, provide personalized feedback, and support adaptive assessment strategies (Owan et al., 2023). These tools include automated essay scoring, computerized adaptive tests, and the analysis of process data from digital assessments (Gardner & Yuan, 2021). AI can also assist with formative assessment practices such as self-assessment, peer assessment, and teacher feedback (Hopfenbeck et al., 2023). For instance, AI can provide instant feedback to students during self-assessment, or it can analyze peer feedback to identify common misconceptions. However, there are obstacles to overcome, including the need for a more robust integration of pedagogical principles (González-Calatayud et al., 2021) and addressing potential biases, ethical concerns, and the requirement for teacher training (Al Braiki et al., 2020; Chattopadhyay et al., 2018; Kamalov et al., 2023). As AI continues to advance, it has the potential to transform educational assessment, leading to improved learning outcomes and equipping students with essential 21st-century skills (Al Braiki et al., 2020; Owan et al., 2023).

5/32

AI has the potential to improve classroom assessment, but there are still challenges related to trust, fairness, and the redefinition of teacher and student roles (DiCerbo, 2021). While AI can help in test development, administration, and scoring, it should be used to support human judgment rather than replace it (Owan et al., 2023). Teachers play a crucial role in implementing AI, including training algorithms and verifying AI-generated assessments (Celik et al., 2022). However, to fully benefit from AI in education, stakeholders such as policymakers, school administrators, educators, and students need to work together to address challenges, minimize risks, and ensure that AI integration is trustworthy (Hopfenbeck et al., 2023; Rizvi et al., 2023).

AI in Language Testing

Recent research highlights the efficiency and potential of AI in language assessment (Xi, 2023). AI technologies, such as machine learning, natural language processing (NLP), and large language models, offer innovative methods to accurately and swiftly evaluate language proficiency (Voss et al., 2023). These advancements enable AI to be used for various purposes. For example, NLP, automatic speech recognition (ASR), and processing technologies empower AI to understand and interpret human language, facilitating the assessment of complex language skills like writing and speaking (Chapelle

& Chung, 2010). As a result, AI can be utilized for automated scoring and feedback (Cotos, 2023; Xi, 2010; Zechner & Hsieh, 2024). AI algorithms can analyze written and spoken responses to objectively assess grammar, vocabulary, fluency, and coherence, reducing human error and expediting the evaluation process. Several automated scoring tools are currently in use to score speaking assessments, such as ETS SpeechRaterSM (Chen et al., 2018), Interactional Competence Elicitor (ICE) (Chukharev-Hudilainen & Ockey, 2021), and Automated Pronunciation Evaluation for L2 Speech (Moxon, 2021). For writing assessments, a few automated scoring tools include ETS Writing E-rater© Scoring Engine (Attali & Burstein, 2006) and Pearson's Intelligent Essay Assessor (Folz et al., 2013).

AI has been utilized in computer adaptive testing to customize assessments based on the test taker's proficiency level, resulting in a more efficient and accurate evaluation (Dunkel, 1999; Mislevy et al., 2008). Some studies have investigated the use of keystroke log data, which automatically records all the keystrokes typed on a computer or other input device, to analyze aspects of writing performance (e.g., Choi et al., 2021). AI technology is also employed in remote proctoring and test security (Purpura et al., 2021; Voss et al., 2023). Remote proctoring utilizes technology such as facial recognition and keystroke analysis to detect potential cheating or suspicious behavior during the administration of the remote assessment. Some tests utilizing remote proctoring include the Duolingo English Test, TOEFL iBT Special Home Edition, and Versant.

Moreover, AI is being used for large-scale assessment by automatically generating many assessment items to ensure the reliability and validity of tests (Belzak et al., 2023; Shin & Lee, 2024). Test developers are utilizing newer technologies such as NLP and generative AI to create test items. For example, NLP techniques are used to extract and generate multiple-choice questions from existing reading passages. Generative AI is also employed to automatically develop reading passages and corresponding comprehension items (Attali et al., 2022). Further research is necessary to explore the use of generative AI tools for generating classroom assessments automatically.

AI has numerous advantages, but its integration with assistive technologies in language assessment raises important questions about how these technologies are defined, how valid they are, and the ethical implications (Voss et al., 2023; Xi, 2023). As AI continues to evolve, it has the potential to transform language assessment by moving beyond traditional rating scales and analyzing natural language communication (Kjell et al., 2023). Importantly, AI can also help ensure the fairness and integrity of language assessment, providing reassurance to all involved stakeholders.

GOAL OF THE STUDY

We have selected ChatGPT as the generative AI tool for our study among numerous available options. Since its launch in 2022, ChatGPT has garnered global attention and has served as the inspiration for various research studies and innovations in the AI field (Bengesi et al., 2023). Generative AI tools have a wide range of assessment applications, including generating assessment items, scoring responses, and providing feedback. However, our focus is to explore how ChatGPT can be used for automated item generation in classroom assessment. We focus on how to use ChatGPT to simplify texts (such as reading passages) and create multiple-choice questions to evaluate the reading comprehension abilities of basic users of English (A1-A2 level students). Previous research has emphasized the difficulty of finding appropriate reading materials for basic English learners, highlighting it as a crucial requirement for language teachers in the classroom (e.g., Cote Parra & López, 2024; Erbaggio et al., 2012). In this study, we aimed to address the following research questions:

- Research Question 1: What are the perceptions of language teachers regarding the simplified text generated by ChatGPT?
- Research Question 2: What are the perceptions of language teachers regarding the multiple-choice questions generated by ChatGPT?

7/32

METHODS

This research employed a qualitative approach to investigate the perceptions of EFL teachers regarding a simplified text and its associated comprehension multiple-choice questions. These questions were created using ChatGPT, a generative AI tool. The main objective was to comprehend how teachers viewed the use of ChatGPT in devising reading assessments for basic English learners. Through in-depth interviews with teachers, the research sought to obtain a comprehensive understanding of their viewpoints on this subject. The upcoming sections will provide detailed information about how the simplified text and the questions were generated using Chat GPT, the participants, data collection procedures, and data analysis methods.

Text and Questions Generated by ChatGPT

First, we used ChatGPT to create a reading passage intended for advanced-level learners (i.e., C1 and C2 CEFR levels). We chose a topic familiar to students, such as adapting to diverse climates. The prompt used to create the original text is shown in Appendix 1 and the advanced-level passage can be found in Appendix 2. Following this, we created two prompts for ChatGPT. A prompt refers to the instructions a user provides to a generative

AI tool. It could be a question, command, statement, or a series of guidelines used to get a specific response or output (e.g., a text, a test question). One prompt was employed to generate a simplified text of the original text C1-C2 level text for basic English learners (refer to Appendix 3), while the other was used to generate comprehension questions (multiple-choice questions and options) for the simplified text (refer to Appendix 4).

The process we used to develop the prompts in ChatGPT involved multiple iterations, which were crucial for the development of our simplified text and comprehension questions. This means that we went through several rounds in which we revised, deleted, or added information to the prompt to achieve the desired text. For example, the initial simplified text we received from ChatGPT did not include the headings found in the original text, such as Introduction, Understanding Climate, Biological Adaptations, Cultural Adaptations, Technological Adaptations, Urban Planning and Infrastructure, and Conclusion. Therefore, we added a command in the prompt to instruct ChatGPT to incorporate headings in the simplified text and to use simple everyday language. As a result, ChatGPT generated a simplified text using headings such as Weather Basics, Changing Our Bodies, Cultural Ways, New Technology, Building Smart Cities, and Conclusion. The final version of the simplified text that we used for this study is showcased below.

8/32

Adapting to Different Climates

Humans live in many different places with different weather. We are good at adapting to these different places. Let's see how we do it.

Weather Basics

Weather means what it's like outside in a place, like if it's hot, cold, rainy, or windy. Different places have different weather patterns. These patterns affect plants, animals, and people.

Changing Our Bodies

Humans change to fit where we live. For example, if it's cold, people might become shorter and stockier. If it's hot, they might be taller and thinner. Our bodies change to help us survive.

Cultural Ways

Besides body changes, we also have cultural ways to deal with different weather. People share knowledge about farming, hunting, and building that works best in their area. This helps communities thrive.

New Technology

Technology helps us adapt too. We use fire for warmth and cooking. We also have things like air conditioning to cool us down. Modern stuff like insulated clothes and desalination plants make life easier in tough climates.

Building Smart Cities

Cities are designed to work well with the weather. They have things like shaded streets and good drainage. In hot places, cities might use special materials on roofs to stay cool.

Conclusion

Humans are great at adapting to different climates. We use our bodies, culture, technology, and city planning to do it. Even with climate change, we can keep adapting to make life good for everyone.

Our process allowed us to have control over the structure of the text. We noticed that the paragraphs in the simplified text we created had different structures. Therefore, we revised the prompt to tell the generative AI tool to always start all the paragraphs with the main idea. After several rounds of review, we identified a few complex structures in the simplified text. To address this, we added a command to the prompt, instructing ChatGPT to only use simple sentences in the simplified text and to avoid including sentences with dependent clauses, prepositional clauses, and conjunctions.

9/32

Similarly, we took great care in creating the prompt for generating multiple-choice questions and options. Initially, we noticed that ChatGPT was generating questions with three options. To improve its capabilities, we added a command to create questions with four options. We also found that the tool consistently selected the first option as the correct response. To add variety, we adjusted the prompt to include instructions for scrambling the key and asked ChatGPT to include the correct answer. Additionally, we realized that some questions could be answered using general knowledge, meaning they did not require reading the text. To address this, we added a command to ensure that the tool created questions that could only be answered by reading the text. The final version of the questions and options are shown below.

1. What is the main idea of the text?
 - a) Humans live in places with the same weather.
 - b) Weather patterns affect plants, animals, and people differently.
 - c) Humans can't adapt to different climates.
 - d) Humans are great at adapting to different climates.Correct answer: d) Humans are great at adapting to different climates.

2. What do humans use to stay warm and cook food?
 - a) Insulated clothes and desalination plants.

- b) Fire and air conditioning.
 - c) Only wearing shorts.
 - d) Eating lots of ice cream.
- Correct answer: b) Fire and air conditioning.

3. How do people share knowledge about dealing with weather?
- a) By discussing movies and TV shows.
 - b) By sharing farming, hunting, and building techniques.
 - c) By playing sports.
 - d) By gossiping about celebrities.
- Correct answer: b) By sharing farming, hunting, and building techniques.

4. What helps make life easier in tough climates?
- a) Shaded streets and good drainage.
 - b) Buildings with no windows.
 - c) Only tall buildings.
 - d) Lots of snow.
- Correct answer: a) Shaded streets and good drainage.

5. What does weather mean?
- a) What it's like underwater.
 - b) What it's like inside buildings.
 - c) What it's like outside, such as hot, cold, rainy, or windy.
 - d) What it's like in outer space.
- Correct answer: c) What it's like outside, such as hot, cold, rainy, or windy.

10/32

Participants

For this study, we selected participants using a combination of purposive and convenience sampling. We specifically focused on current English as a Foreign Language (EFL) teachers who were teaching courses for basic users (A1-A2 CEFR levels) and had at least one year of experience working with basic English learners. These teachers were selected from two local higher education institutions. The first institution (Institution 1) is a public mid-size college, while the second institution (Institution 2) is a large public university. We chose teachers from these two institutions because they were easily accessible. The use of purposive and convenience sampling was suitable for our exploratory research study, as it helped us gather initial insights about the use of generative AI tools for classroom assessment purposes. Initially, we recruited eight teachers, but one dropped out in the middle of the study. Background information for the remaining seven participating teachers is provided in Table 1. Among them, two worked at Institution 1 and five at Institution 2. Their teaching experience ranged from

one year to 30 years, and they also varied in terms of years of experience teaching basic English learners, ranging from one to 15 years.

Table 1. Background information about the participating teachers

Teacher	Institution	Years of teaching experience	Years teaching basic level courses
T1	1	30	5
T2	2	10	5
T3	1	25	15
T4	2	2	2
T5	2	3	1
T6	2	6	3
T7	2	1	1

Data Collection Procedures

After obtaining the teachers' consent to participate in this study via e-mail in English, we organized an online meeting to introduce them to the study's objectives, the CEFR levels, ChatGPT, the prompts, the original text, the simplified text, and the questions generated by ChatGPT. The online meeting was conducted in English. We recorded the meeting, which lasted approximately 60 minutes, so teachers could refer back to it when reviewing all the materials. We provided the teachers with two weeks to thoroughly review the simplified text, the multiple-choice questions, and the prompts. We asked them to highlight aspects they liked and disliked about what was generated by ChatGPT and the prompts. Additionally, we requested the teachers to complete and submit an online pre-interview questionnaire. The purpose of this questionnaire was to gather additional background information from the teachers, such as their age, gender, the institution where they were currently teaching, total years of teaching experience, and years of experience teaching basic English learners.

We conducted individual online interviews with each participant. We used an interview protocol that had been tested with two EFL teachers. The protocol guided the interview process and helped us gather information about the reading materials used for teaching and assessment. We also asked about their preferences and any concerns they had regarding the simplified texts and questions. We also inquired about the suitability of the materials for their students in terms of grammar, vocabulary, clarity, organization, and length. Additionally, we asked for their opinion on how similar the materials were to the ones they normally use in class, as well as whether their students could comprehend the simplified text and answer the questions. We also sought input on how they would modify the simplified text, questions and prompts, and their thoughts on the prompts used to generate the text and questions. We recorded each interview, and they all lasted approximately 60 minutes.

Data Analysis

After each interview, we held a short debrief meeting to reflect on the session, identify any missing information, plan for subsequent interviews, cross-reference data, and start recognizing patterns and recurring themes. This thorough process enabled us to gain in-depth insights into the data (Spall, 1998). Additionally, we independently analyzed each interview transcript to identify patterns in how teachers perceived the simplified text, questions, and prompts. We used a two-step coding process. First, we carefully categorized interview sections based on their content, and then, we closely examined these categories to find recurring themes (Saldaña, 2009). We compared our findings to ensure consistency and resolved any disagreements through open discussion, as suggested by Strauss and Corbin (1990). Ultimately, we identified key themes related to teacher opinions about the text, questions, and prompts, including their strengths, weaknesses, potential uses, and suggested improvements.

FINDINGS

12/32

In this section, we present our findings on two research questions: the teachers' perceptions of the simplified text and the questions generated by ChatGPT. We discuss their likes and dislikes regarding both the text and the questions, how they plan to use them in their classrooms, and their recommendations for improving the outputs generated by ChatGPT.

Teachers' Perceptions of the Simplified Text Generated by ChatGPT

The teachers all had a positive view of the simplified text created by ChatGPT, although they had varying reasons for liking it. Specifically, all the teachers believed that ChatGPT was able to simplify the original text according to the guidelines provided in the prompt. The teachers were fond of the way in which ChatGPT simplified the text. The following interview excerpt from T7 illustrates how teachers viewed ChatGPT's ability to simplify the text: "I consider [ChatGPT] is almost perfect. Yeah, we ask that machine to have one expected result. And this machine does it without any problem" (personal communication, June 13, 2024). T5 highlighted how ChatGPT was able to implement the guidelines provided in the prompt. T5 commented:

If we see the rules, we can say that these rules were followed in the simplified text in one way or another. For example, 'climate' is a word for advanced learners, and 'whether' is the word A-1 students are familiarized with. (personal communication, June 12, 2024)

T1, T3, and T4 were particularly impressed by ChatGPT's efficiency in simplifying texts. They noted that finding appropriate texts for basic English learners is a time-consuming task, whether online or in textbooks. T4, for instance, found ChatGPT to be a time-saving tool, stating: "...I found that it was really useful. It was easier than looking on the Internet for something that was already done" (personal communication, June 13, 2024).

T3 and T5 liked how the AI tool was able to summarize the main ideas and remove unnecessary information from the original text. T5 commented the following about the simplified text: "The paragraphs are shorter in the simplified version, shorter than in the original text. Also, in the simplified text, each paragraph starts with the main idea. It's simple" (personal communication, June 12, 2024). Furthermore, all the teachers highlighted that the simplified text is suitable for assessing basic English learners' comprehension of the main idea and important information. They particularly liked that the simplified text incorporated basic vocabulary, verb tenses, and sentence structures suitable for A1 and A2 level students. T4 provided an explanation for why she believes the simplified text is appropriate for basic English learners. She stated:

Simple structure. It uses present simple. It doesn't have any other advanced grammar or tense. So, it is really easy to understand, from my perspective, as an A-1 English teacher. I would say, this is really appropriate for the A-1 students to read and to understand. For example, if I put this test to my students at this point, like a second term. I'm pretty sure most of them they would understand the whole text. (personal communication, June 13, 2024)

13/32

The teachers also mentioned that they liked the simplified text because it was well organized (five teachers), concise and direct (five teachers), and the content was familiar to their students (two teachers). Six of the teachers indicated that the simplified text generated by ChatGPT resembles the texts they typically use for teaching and assessment. T4 even commented that the generated simplified text is better than the texts she finds online in terms of length and grammar. However, T2 and T6 expressed a preference for using longer texts in their classes and assessments because these texts resemble what their students read outside the classroom.

All the teachers mentioned that they could use the simplified text for classroom instruction and assessment, but they would need to make a few adjustments. The teachers suggested different ways to improve the simplified text to make it more accessible to students with basic language proficiency. For instance, T3 and T6 noted that the simplified text contained some complex structures, such as passive voice and phrasal verbs, and unfamiliar vocabulary. T6 highlighted some of the complex structures included in the simplified text: "Regarding syntax, the Rule number 8, you said, for example, use only the active voice instead of the passive voice. And in this one is the

passive voice: Cities are designed to work well with the weather” (personal communication, June 17, 2024). Additionally, T4 suggested that adding linking words (transitions) could enhance the simplified text. Lastly, T2 observed that the simplified text generated by ChatGPT seemed somewhat artificial and lacking in human emotion. For instance, the text does not include personal opinions, and the tone is very neutral, which makes it difficult for students to engage with it.

All the teachers liked the specific and detailed prompt that was used to generate the simplified text. They found that the prompt met the requirements for creating appropriate texts for basic English learners in terms of language, clarity, organization, and content. The teachers also emphasized the importance of clear prompts when using AI tools. T5 shared her thoughts about the prompt, saying: “I think that when you presented the prompts to ChatGPT, AI simplified the text. You were so specific about the things you needed. It’s important to give clear prompts, clear instructions” (personal communication, June 12, 2024). However, six teachers mentioned that they could adjust the prompt to better suit their specific needs. For example, T1 and T5 wanted to add information about the learners, specifying that their students' home language is Spanish. T6 and T7 would change the topic or content of the text to align it with their instructions, and T4 and T6 wanted to specify which particular grammar structure to include in the text to match the grammar focus of their instruction.

14/32

Teachers’ Perceptions of the Questions Generated by ChatGPT

The teachers also had a positive view of the multiple-choice questions created by ChatGPT. They all appreciated that the questions were directly related to the information in the text, which required students to extract key information from the text to answer them. For example, T7 explained the following about the questions: “I think that other questions can be added. But the questions this platform generated are appropriate. They’re based on the text” (personal communication, June 13). T1 concurred by adding that “The students can identify the answers by scanning or skimming the text easily” (personal communication, June 10, 2024). As a result, the teachers felt that these questions could be answered by basic English learners and could provide useful insights into the students' ability to identify the main idea and important information.

However, some teachers provided recommendations to improve the questions. T3 and T5 felt that the first question, which asked students to find the main idea of the text, was too subjective or broad, and they preferred questions like the last four. To address this, these two teachers suggested that the AI tool could be adjusted to generate more specific questions, ensuring a clearer focus. In addition, these two teachers noted that the last question was ambiguous because it could be answered from personal experience. They suggest specifying that this question should be answered based on the text. This suggestion is highlighted in T5’s response:

Well, I think that the 5th question, the last one, lacks instructions because [the question is:] What does weather mean? But well, I don't know, with A1 students... sometimes they need everything in the questions we ask. With something like that, they'd say: 'but teacher, is about the text, or it's about me, or what do you mean by, what does weather mean?' So, we have to say, based on the text, or based on the information presented in the text. (personal communication, June 12, 2025)

Moreover, the teachers mentioned other positive aspects of the multiple-choice questions created by the AI tool. T4 was thankful that each question included the correct response, T6 liked that each multiple-choice question had four options, and T1 valued that the correct responses were scrambled, making it more challenging for students to guess the correct answer. All the teachers also agreed that the multiple-choice questions were similar to the ones they typically use in their teaching and assessments. However, six teachers preferred using other question types, such as open-ended questions or questions that require students to justify their answers. T2 talked about the type of questions she uses in class: "Personally speaking, I prefer open-ended questions, just to know [their] creativity, points of view, and to see some language production there" (personal communication, June 10, 2024). Similarly, T5 described his preferred reading comprehension questions: "I'd like to propose [using] more reflective questions because, apart from the reading itself, it's important that they reflect on the information for a situation related to the context" (personal communication, June 12, 2024).

15/32

Additionally, the teachers felt that the questions generated by the AI tool are suitable for basic English learners, especially those in the early stages of learning, as they are simple and clear and mostly use simple grammar and vocabulary. T1 highlighted that the questions were clear, saying: "...the instructions are clear. The questions are clear and simple to understand. They are aligned to the CEFR level. For example, the use of everyday language is another important aspect" (personal communication, June 10, 2024). Nonetheless, T3 and T4 pointed out that a few words in the questions could be challenging for basic English learners, such as the word "insulated," which may be beyond their current vocabulary.

The teachers all expressed appreciation for the questions created by ChatGPT, but they also mentioned a desire to customize the prompts to better suit their teaching styles and meet their students' needs. For instance, T1 recommended requesting the AI tool to provide test directions in Spanish to align with her classroom activities. The other six teachers suggested generating different types of questions to better align with their teaching methods. Lastly, T5 stressed the importance of teachers reviewing the questions generated by ChatGPT before using them in class.

DISCUSSION AND CONCLUSION

The study aimed to investigate the use of ChatGPT in developing simplified texts and multiple-choice questions for basic English learners. This research is significant as it can show how EFL teachers could employ AI tools to create materials for assessing the reading comprehension skills of students with basic English proficiency. The process began with a high-level text on climate adaptation, and then ChatGPT was used to create a simpler version along with multiple-choice questions. The prompts were refined through several iterations to achieve the desired text structure, sentence complexity, and question format. The final outputs included a simplified text with clear headings and straightforward sentences, and multiple-choice questions based solely on the text content.

Effective prompts are crucial for guiding AI systems to generate accurate, relevant, and creative outputs (Bozkurt, 2024; Clarisó & Cabot, 2023). Recent studies emphasize the importance of AI prompt engineering as an emerging digital skill that is essential for optimizing interactions with generative AI systems (Bozkurt, 2024). These studies highlight the growing significance of prompt engineering across various fields, indicating a need for further research to fully leverage the potential of generative AI technologies. Additionally, as educators increasingly incorporate generative AI tools into language teaching, it is essential to support them in developing prompt engineering skills to effectively integrate these tools into their classroom practices (Mena Octavio et al., 2024).

This study highlights the adaptability of generative AI tools to produce customized assessment content based on students' English language proficiency (Voss et al., 2023; Xi, 2023). It specifically focuses on reading passages and related questions (Attali et al., 2022) in the context of classroom assessments. Utilizing generative AI tools to adapt texts in classroom assessments to meet the needs of all students, particularly those with basic English proficiency levels, ensures the collection of valid assessment information about students' language skills (O'Sullivan, 2023; Youn, 2023). However, it is important to note that this study only included one reading passage and five multiple-choice questions.

Additionally, our research revealed that teachers responded positively to ChatGPT's ability to simplify texts and generate questions for students with basic language proficiency. They felt the AI tool effectively transformed complex texts into materials suitable for students at A1 and A2 levels. The teachers also expressed interest in using AI tools to develop educational materials. This is a significant finding, as previous studies have shown that teachers have been hesitant to incorporate AI tools into their teaching and assessment methods (e.g., Azzam & Charles, 2024). Teachers also valued the time saved and the tool's accuracy in following simplification guidelines.

Other studies have shown that using AI tools can lighten teachers' workloads (Chaudhry & Kazim, 2022). This allows them to utilize their preparation time more effectively, focusing on important issues rather than spending it on searching for, designing, or adapting teaching and assessment materials.

As teachers explore the use of generative AI to design reading comprehension activities, their feedback becomes essential to ensure the quality and pedagogical relevance of AI-generated outputs. In our study, the teachers were also pleased that the questions asked students to find main ideas and key information in a text. However, a few teachers felt that some questions were too broad or ambiguous. They also suggested refinements to the prompts for more tailored outputs. For example, teachers suggested reducing complex structures, adding linking words, and infusing more humanlike language in the text, and they wanted more specific questions and options that clearly required information from the text. They also preferred a mix of question types and expressed the need to adapt the questions to different student levels and teaching styles. These reflections illustrate how essential it is for teachers to remain actively engaged in evaluating and shaping AI-generated materials, a practical example of the “human-in-the-loop” approach. In this case, their involvement is crucial to identify and address limitations and to contextualize and adapt outputs to the needs of specific learners and teaching environments.

17/32

Selecting appropriate texts is essential for effective literacy instruction and assessment (Ebe, 2012; Grabe, 2009). For language learners, it is important to consider the complexity of the text (Bunch et al., 2014). Additionally, other factors such as the characteristics and needs of the learner, the context in which the text will be used, and the specific tasks assigned to students are critical when choosing texts (Bunch et al., 2014). When selecting texts, it is important to consider aspects such as vocabulary, fluency development, and student motivation (Hoffman & Schallert, 2004). Research on suitable texts for basic language learners highlights the significance of using simplified materials (Petersen & Ostendorf, 2007). Studies indicate that simplified texts are processed more quickly and lead to improved comprehension, especially for learners with lower language proficiency (e.g., Crossley et al., 2012; Crossley et al., 2014; Rets & Rogaten, 2020). Overall, texts for basic language learners should be carefully selected or adapted to align with their proficiency levels and support their language development.

The teachers found ChatGPT to be a valuable resource for simplifying reading passages and creating questions. However, they emphasized the need for human oversight and adjustments to optimize its effectiveness. This reinforces the crucial role of teachers when using generative AI to develop classroom assessments and can make them feel valued and essential to the tool's successful implementation. Several studies have discussed the importance of having humans in the loop (e.g., Wu et al., 2022; Xi, 2023), in this case, classroom teachers, to monitor the quality of AI-generated assessments.

'Human in the loop' refers to the active participation of individuals in the training, evaluation, or operation of AI tools (Wu et al., 2022). Their involvement provides valuable guidance and feedback to improve the accuracy, reliability, and adaptability of these tools (Wu et al., 2022). When using generative AI tools for classroom assessment, it is crucial for teachers to review the AI-generated content. This ensures that the material is accurate, sensible, and relevant to the intended learners. If necessary, teachers can make edits to the prompt, the output, or both. By doing so, teachers can ensure the text and the questions generated by the AI tool meet specific classroom needs and are linguistically responsive to the students' English language proficiency levels (Yang, 2024).

This research offers insights into how AI can be used to facilitate the assessment practices of EFL teachers by generating assessment materials and enhancing assessment accuracy. However, it is important to help teachers address potential challenges associated with integrating AI tools into classroom assessment. These challenges include ethical and responsible use of AI, such as preventing bias in AI-generated assessments (Al Braiki et al., 2020; Chattopadhyay et al., 2018; Kamalov et al., 2023; Xi, 2023), as well as the necessity of providing comprehensive teacher training for the implementation of AI technologies in the classroom (Azzam & Charles, 2024; Chattopadhyay et al., 2018).

Implications for Practice

AI is rapidly changing education, including classroom assessment (Owan et al., 2023) and language testing (Voss et al., 2023; Xi, 2023). Therefore, it is crucial to assist EFL teachers in integrating new technologies to improve teaching and learning. While there could be challenges in adopting AI tools for classroom assessment, the benefits outweigh them. The results of this study indicate that EFL teachers have a positive perception of using generative AI tools, such as ChatGPT, to create effective assessment tools. The use of these generative AI tools has the potential to enhance the development of assessment tools, save time, and reduce teachers' workload. Next, we outline various ways in which the study's findings can guide teachers in effectively incorporating AI tools in their classroom assessment practices.

Teachers require additional training opportunities to effectively, responsibly, and ethically use AI tools. These training sessions should introduce teachers to various AI tools and emphasize the importance of their ethical use in improving classroom assessment practices. The training should also include strategies for teachers to create effective prompts and establish a process for monitoring and reviewing the AI-generated content, instilling a sense of responsibility and conscientiousness in teachers. Additionally, publishers, test developers, and teachers can create flexible templates (i.e., prompts) for teachers to use in AI tools to generate assessment content, ensuring they are relevant to their unique classroom contexts and the needs of their students.

Limitations and Future Directions for Research

The study was limited because we only used one AI tool (ChatGPT) to simplify a text and create comprehension questions for basic English learners. Additionally, teachers were asked to evaluate the tool's performance, but they did not actually use it to create assessment materials for their classes. As a result, the generated assessment materials were not used with the students. Despite these limitations, the study's findings on using ChatGPT to create classroom assessment content and on teachers' perceptions of the AI-produced text and questions are valuable and offer insights into how EFL teachers can enhance their assessment practices using generative AI tools.

This might showcase the pressing need for additional research to verify the effectiveness of using generative AI tools in classroom assessment. Although the teachers in this study expressed positive views about the simplified text and the questions generated by ChatGPT, it would be interesting to explore various ways teachers can utilize generative AI tools to improve their assessment methods in the classroom. Further research could involve working with teachers to create assessment content for different language domains (such as listening, reading, speaking, writing, grammar, and vocabulary). Additionally, other studies could explore alternative ways teachers can utilize AI tools, such as scoring student responses and providing automated feedback.

19/32

In addition, upcoming research studies could examine how teachers incorporate these AI-generated assessments in their teaching, how they utilize the assessment data to enhance teaching and learning, and how students, who are the main beneficiaries of these tools, perceive the use of AI tools for classroom assessment. The potential for student empowerment using AI tools is especially intriguing and warrants further investigation. Furthermore, additional studies could explore validity and reliability issues associated with the use of AI tools in assessment, including how students can leverage AI tools in classroom assessment to demonstrate their language proficiency.

Final Thoughts

Incorporating artificial intelligence into foreign language assessment offers an excellent opportunity for both educators and learners. By utilizing AI tools, teachers can expand their assessment methods. For example, AI can design reading passages and comprehension questions. Using AI-generated assessment saves time and enables teachers to concentrate on more advanced language skills, such as critical thinking and creativity. Additionally, generative AI has great potential for enhancing the assessment of reading comprehension. For instance, it can create customized reading passages and questions tailored to each learner's proficiency level and interests. It also provides immediate feedback on comprehension questions and identifies areas that need improvement. AI can also generate personalized reading comprehension tests based on

a learner's past performance, ensuring the appropriate difficulty level and enabling better progress tracking. Furthermore, AI can create realistic reading scenarios—such as stories, news articles, blogs, or e-mails—allowing learners to demonstrate their understanding of authentic reading materials.

This research highlights the significance of integrating AI into classroom assessments. AI-powered assessments are highly adaptable and can be customized to address each student's specific needs through adaptive testing that aligns with their progress and areas for improvement. This personalized approach would guarantee that students are evaluated on their proficiency and receive focused feedback to support their development, which would help teachers be more flexible and responsive to the individual needs of each student.

Instead of viewing AI as a threat, educators should acknowledge its potential to improve assessment practices. By using AI, teachers can provide enhanced assessment experiences that cater to students' varied needs, ultimately resulting in more effective and meaningful language learning outcomes. This ability of AI to enhance assessment practices should instill optimism in educators about the future of language teaching, learning, and assessment.

20/32

Financing Data

The authors do not report having received institutional, governmental or private funding for the research, writing or publication of the article.

Authors' contribution

Alexis A. López and Gabriel Cote Parra: conceptualization, formulation of objectives, methodological design, data collection, formal analysis, validation, writing (original draft), writing (revision and editing).

Conflicts of interest

The authors state that they do not have any financial, professional, or personal conflicts of interest that would interfere with the submission or publication of the manuscript.

Ethical implications

The study obtained informed consent from the participants (English teachers) via e-mail. In addition, no minors were involved and no sensitive data were collected. Therefore, it is concluded that the study complies with accepted ethical standards in qualitative educational research.

Statement using artificial intelligence

Grammarly was used to review the style of the manuscript.

REFERENCES

- Al Braiki, B., Harous, S., Zaki, N., & Alnajjar, F. (2020). Artificial intelligence in education and assessment methods. *Bulletin of Electrical Engineering and Informatics*, 9(5), 1998–2007. <https://doi.org/10.11591/eei.v9i5.1984>
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. *Journal of Technology, Learning, and Assessment*, 4(3). <https://ejournals.bc.edu/index.php/jtla/article/view/1650>
- Attali, Y., Runge, A., LaFlair, G. T., Yancey, K., Goodwin, S., Park, Y., & Von Davier, A. A. (2022). The interactive reading task: Transformer-based automatic item generation. *Frontiers in Artificial Intelligence*, 5, e903077. <https://doi.org/10.3389/frai.2022.903077>
- Azzam, A. & Charles, T. (2024). A review of artificial intelligence in K-12 education. *Open Journal of Applied Sciences*, 14, 2088-2100. <https://doi.org/10.4236/ojapps.2024.148137>
- Belzak, W. C. M., Naismith, B., & Burstein, J. (2023). Ensuring fairness of human- and AI-generated test items. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial intelligence in education. Communications in computer and information science* (Vol. 1831). Springer. https://doi.org/10.1007/978-3-031-36336-8_108
- Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2024). Advancements in generative AI: A comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers. *IEEE Access*, 12, 69812–69837. <https://doi.org/10.1109/ACCESS.2024.3397775>
- Bozkurt, A. (2024). Tell me your prompts and I will make them true: The alchemy of prompt engineering and generative AI. *Open Praxis*, 16(2), 111–118. <https://doi.org/10.55982/openpraxis.16.2.661>
- Bunch, G. C., Walqui, A., & Pearson, P. D. (2014). Complex text and new common standards in the United States: Pedagogical implications for English learners. *TESOL Quarterly*, 48(3), 533–559. <https://doi.org/10.1002/tesq.175>
- Celik, I., Dindar, M., Muukkonen, H., Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, 66, 616–630. <https://doi.org/10.1007/s11528-022-00715-y>
- Chappelle, C. A., & Chung, Y. R. (2010). The promise of NLP and speech processing technologies in language assessment. *Language Testing*, 27(3), 301–315. <https://doi.org/10.1177/0265532210364405>

- Chattopadhyay, S., Shankar, S., Gangadhar, R. B., & Kasinathan, K. (2018). Applications of artificial intelligence in assessment for learning in schools. In J. Keengwe (Ed.), *Handbook of research on digital content, mobile learning, and technology integration models in teacher education* (pp. 185–206). IGI Global. <https://doi.org/10.4018/978-1-5225-2953-8.ch010>
- Chaudhry, M. A., & Kazim, E. (2022). Artificial Intelligence in Education (AIEd): A high-level academic and industry note 2021. *AI Ethics* 2, 157–165. <https://doi.org/10.1007/s43681-021-00074-z>
- Chen, L., Zechner, K., Yoon, S.-Y., Evanini, K., Wang, X., Loukina, A., Tao, J., Davis, L., Lee, C. M., Ma, M., Mundkowsky, R., Lu, C., Leong, C. W., & Gyawali, B. (2018). Automated scoring of nonnative speech using the SpeechRaterSM v. 5.0 engine. *ETS Research Report*, 18(10), 1–31. <https://doi.org/10.1002/ets2.12198>
- Choi, I., Hao, J., Deane, P., & Zhang, M. (2021). Benchmark keystroke biometrics accuracy from high-stakes writing tasks. *ETS Research Report*, 21(15), 1-13. <https://doi.org/10.1002/ets2.12326>
- Chukharev-Hudilainen, E., & Ockey, G. J. (2021). The development and evaluation of Interactional Competence Elicitor for oral language assessments. *ETS Research Report*, 21(6). <https://doi.org/10.1002/ets2.12319>
- 22/32 Clarisó R., & Cabot, J. (2023). Model-driven prompt engineering. In *ACM/IEEE 26th International Conference on Model Driven Engineering Languages and Systems Proceedings* (pp. 47–54). MODELS. <https://doi.org/10.1109/MODELS58315.2023.00020>
- Cote Parra, G., & López, A. A. (2024). Examining the assessment practices of foreign language novice teachers. *Profile: Issues in Teachers' Professional Development*, 26(1), 97–113. <https://doi.org/10.15446/profile.v26n1.106384>
- Cotos, E. (2023). Automated feedback on writing. In O. Kruse, C. Rapp, C. M. Anson, K. Benetos, E. Cotos, A. Devitt, & A. Shibani (Eds.), *Digital writing technologies in higher education: Theory, research, and practice* (pp. 347–364). Springer. https://doi.org/10.1007/978-3-031-36033-6_22
- Crossley, S. A., Allen, D., & McNamara, D. S. (2012). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, 16(1), 89–108. <https://doi.org/10.1177/1362168811423456>
- Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language*, 26(1), 92–113. <http://hdl.handle.net/10125/66686>
- DiCerbo, K. (2021). Why not go all-in with artificial intelligence? In R. A. Sottolare, & J. Schwarz (Eds.), *Adaptive instructional systems: Design and evaluation* (pp. 361–369). International Conference on Human-Computer Interaction. Springer. https://doi.org/10.1007/978-3-030-77857-6_25

- Dunkel, P. A. (1999). Considerations in developing or using second/foreign language proficiency computer-adaptive tests. *Language Learning & Technology*, 2(2), 77–93. <http://dx.doi.org/10125/25044>
- Ebe, A. E. (2012). Supporting the reading development of middle school English language learners through culturally relevant texts. *Reading & Writing Quarterly*, 28(2), 179–198. <https://doi.org/10.1080/10573569.2012.651078>
- Erbaggio, P., Gopalakrishnan, S., Hobbs, S., & Liu, H. (2012). Enhancing student engagement through online authentic materials. *The International Association for Language Learning Technology*, 42(2), 27–51. <https://doi.org/10.17161/iallt.v42i2.8511>
- Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, 66, 111–126. <https://doi.org/10.1007/s12599-023-00834-7>
- Foltz, P., Streeter, L., Lochbaum, K., & Landauer, T. (2013). Implementation and applications of the intelligent essay assessor. In M. D. Shermis, & J. Burstein (Eds.), *Handbook of automated essay evaluation, current applications and new directions* (pp. 68–88). Routledge.
- García-Peñalvo, F. J., & Vázquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in generative AI. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(4), 7–16. <https://doi.org/10.9781/ijimai.2023.07.006>
- Gardner, J., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?'. *Journal of Computer Assisted Learning*, 37(5), 1207–1216. <https://doi.org/10.1111/jcal.12577>
- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, 11(12), 5467. <https://doi.org/10.3390/app11125467>
- Grabe, W. (2009). *Reading in a second language*. Cambridge University Press.
- Hoffman, J. V., & Schallert, D. L. (Eds.). (2004). *The texts in elementary classrooms*. Routledge. <https://doi.org/10.4324/9781410611086>
- Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., & McGrane, J. A. (2023). Challenges and opportunities for classroom-based formative assessment and AI: A perspective article. *Frontiers in Education*, 8, e1270700. <https://doi.org/10.3389/feduc.2023.1270700>
- Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, 15(16), 12451. <https://doi.org/10.3390/su151612451>
- Kar, S., Roy, C., Das, M., Mullick, S., & Saha, R. (2023). AI horizons: Unveiling the future of generative intelligence. *International Journal of Advanced Research in Science*,

Communication and Technology (IJARSCT), 3(1), 387–391.

<https://doi.org/10.48175/ijarsct-12969>

Kjell, O., Giorgi, S., & Schwartz, H. A. (2023). The text-package: An R-package for analyzing and visualizing human language using natural language processing and transformers. *Psychological Methods*, 28(6), 1478–1498.

<https://doi.org/10.1037/met0000542>

Mena Octavio, M., González Argüello, M. V., & Pujolà, J.-T. (2024). ChatGPT as an AI L2 teaching support: A case study of an EFL teacher. *Technology in Language Teaching & Learning*, 6(1), 1142. <https://doi.org/10.29140/tl.tl.v6n1.1142>

Mislevy, R., Chapelle, C. A., Chung, Y.-R., Xu, J. (2008). Options for adaptivity in computer-assisted language learning and assessment. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 9-24). Iowa State University.

Morandín-Ahuerna, F. (2022). What is artificial intelligence? *International Journal of Research Publication and Reviews*, 3(12), 1947–1951.

<https://doi.org/10.55248/gengpi.2022.31261>

Moxon, S. (2021). Exploring the effects of automated pronunciation evaluation on L2 students in Thailand. *IAFOR Journal of Education: Language Learning in Education*, 9(3). <https://doi.org/10.22492/ije.9.3.03>

O'Sullivan, B. (2023). Reflections on the application and validation of technology in language testing. *Language Assessment Quarterly*, 20(4-5), 501–511.

<https://doi.org/10.1080/15434303.2023.2291486>

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Basse, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>

Petersen, S. E., & Ostendorf, M. (2007, October) *Text simplification for language learners: A corpus analysis* [Conference paper]. Speech and Language Technology in Education (SLaTE 2007), Farmington, Pennsylvania USA.

<https://doi.org/10.21437/SLaTE.2007-20>

Purpura, J. E., Davoodifard, M., & Voss, E. (2021). Conversion to remote proctoring of the Community English Language Program Online Placement Exam at Teachers College, Columbia University. *Language Assessment Quarterly*, 18(1), 42–50,

<https://doi.org/10.1080/15434303.2020.1867145>

Ramakrishnan, S., Bishnoi, M. M., Joghee, S., Jijitha, S., & Kumar, A. (2024). *Social engineering: Role of teachers in cohabitation of AI with education* [Conference paper]. 2nd International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates. <https://doi.org/10.1109/ICCR61006.2024.10532897>

- Ramdurai, B., & Adhithya, P. (2023). The impact, advancements and applications of generative AI. *SSRG International Journal of Computer Science and Engineering*, 10(6), 1-8. <https://doi.org/10.14445/23488387/IJCSE-V10I6P101>
- Rets, I., & Rogaten, J. (2020). To simplify or not? Facilitating English L2 users' comprehension and processing of open educational resources in English using text simplification. *Journal of Computer Assisted Learning*, 37(3), 705–717. <https://doi.org/10.1111/jcal.12517>
- Rizvi, S., Waite, J., & Sentance, S. (2023). Artificial intelligence teaching and learning in K-12 from 2019 to 2022: A systematic literature review. *Computers and Education: Artificial Intelligence*, 4, e100145. <https://doi.org/10.1016/j.caeai.2023.100145>
- Saldaña, J. (2009). *The coding manual for qualitative researchers*. Sage.
- Satya, C. B. V. V. (2024). Generative AI: Evolution and its future. *International Journal for Multidisciplinary Research (IJFMR)*, 6(1), 1–6. <https://doi.org/10.36948/ijfmr.2024.v06i01.12046>
- Shin, D., & Lee, J. H. (2024). AI-powered automated item generation for language testing. *ELT Journal* 78(4), 446-452. <https://doi.org/10.1093/elt/ccae016>
- Sidhu, B. K. (2022). Generative artificial intelligence: Unveiling the potential and challenges. *International Journal of Science and Research (IJSR)*, 13(4). <https://doi.org/10.21275/sr24414234432>
- Spall, S. (1998). Peer debriefing in qualitative research: Emerging operational models. *Qualitative Inquiry*, 4(2), 280–292. <https://doi.org/10.1177/107780049800400208>
- Strauss, A., & Corbin, L (1990). *Basics of grounded theory methods*. Sage.
- Voss, E., Cushing, S. T., Ockey, G. J., & Yan, X. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly*, 20(4–5), 520–532. <https://doi.org/10.1080/15434303.2023.2288256>
- Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2), 1–37. <https://doi.org/10.2478/jagi-2019-0002>
- Warankar, M., & Patil, R. (2024). Generative artificial intelligence. *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 8(4), 1–7. <https://doi.org/10.55041/ijrem31146>
- Wu, J., Huang, Z., Hu, Z., & Lv, C. (2022). Toward human-in-the-loop AI: Enhancing deep reinforcement learning via real-time human guidance for autonomous driving. *Engineering*, 21, 75–91. <https://doi.org/10.1016/j.eng.2022.05.017>
- Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing*, 27(3), 291-300. <https://doi.org/10.1177/0265532210364643>
- Xi, X. (2023). Advancing language assessment with AI and ML—Leaning into AI is inevitable, but can theory keep up? *Language Assessment Quarterly*, 20(4-5), 357–376. <https://doi.org/10.1080/15434303.2023.2291488>

- Yang, X. (2024). Linguistically responsive formative assessment for emergent bilinguals: exploration of an elementary teacher's practice in a math classroom. *International Multilingual Research Journal* 19(1), 67-90.
<https://doi.org/10.1080/19313152.2024.2339757>
- Youn, S. J. (2023). Test design and validity evidence of interactive speaking assessment in the era of emerging technologies. *Language Testing*, 40(1), 54–60.
<https://doi.org/10.1177/02655322221126606>
- Zechner, K., & Hsieh, C. N. (2024). Automated scoring and feedback for spoken language. In M. D. Shermis, & J. Wilson (Eds.), *The Routledge International Handbook of automated essay evaluation* (pp. 141–160). Routledge.
<https://doi.org/10.4324/9781003397618>
- Zormanová, L. (2024). The attitudes of Czech teachers towards the use of artificial intelligence in schools. *Horyzonty Wychowania*, 23(65), 31–41.
<https://doi.org/10.35765/hw.2024.2365.05>

APPENDIX

Appendix 1. Prompt Used to Create the Original Text in Chat GPT

Write an essay exploring the various ways humans have adapted to diverse climates through biological traits, cultural practices, technological advancements, and urban planning, highlighting people's resilience and ingenuity in thriving across different environments. The essay should be written at a C1 CEFR level and should be between 400 and 500 words. Include subheadings to organize the essay.

Appendix 2. Original Text Created by Chat GPT

Adapting to Diverse Climates: A Human Endeavor

Introduction

The world we inhabit is a tapestry of diverse climates, ranging from the icy tundra of the Arctic to the sweltering heat of the equatorial regions. As humans, we have not merely survived, but thrived in these varied environments, thanks to our remarkable ability to adapt. In this essay, we will explore how people adapt to different climates, examining the strategies and innovations that have enabled us to flourish across the globe.

Understanding Climate

Before delving into the ways in which humans adapt to various climates, it is crucial to understand the nature of climate itself. Climate encompasses long-term weather patterns characteristic of a particular region, including temperature, humidity, precipitation, and wind patterns. These factors exert a profound influence on the flora, fauna, and human inhabitants of a given area.

28/32

Biological Adaptations

One of the most remarkable aspects of human adaptation to different climates is our biological flexibility. Over millennia, populations have developed physiological adaptations to their local environments. For example, people living in colder climates tend to have shorter, stockier builds and higher metabolic rates to conserve heat, whereas those in warmer climates may have taller, leaner bodies to dissipate heat more effectively.

Cultural Adaptations

In addition to biological adaptations, humans have developed a rich array of cultural practices to cope with diverse climates. Traditional knowledge passed down through generations has enabled communities to thrive in their environments. This includes techniques for farming, hunting, fishing, and building shelter that are tailored to local conditions.

Technological Innovations

Advancements in technology have further enhanced our ability to adapt to different climates. From the invention of fire for warmth and cooking to the development of air conditioning systems for cooling, humans have continually devised new tools and technologies to mitigate the challenges posed by extreme climates. Modern innovations

such as insulated clothing, greenhouses, and desalination plants enable us to inhabit environments that would have been uninhabitable in the past.

Urban Planning and Infrastructure

The design of cities and infrastructure also plays a crucial role in how we adapt to different climates. Urban planners take into account factors such as temperature, precipitation, and natural disasters when designing buildings, roads, and drainage systems. Cities in hot climates may incorporate shaded walkways, reflective roofing materials, and efficient public transportation systems to minimize heat exposure and energy consumption.

Conclusion

In conclusion, the ability of humans to adapt to different climates is a testament to our ingenuity and resilience as a species. Through biological adaptations, cultural practices, technological innovations, and thoughtful urban planning, we have overcome the challenges posed by a wide range of environmental conditions. As we continue to face the impacts of climate change, our capacity to adapt will be more critical than ever. By drawing on our collective knowledge and creativity, we can build a more sustainable future for generations to come.

Appendix 3. Prompt Used to Create the Simplified Text in Chat GPT

We are going to simplify a written text. I will give you a written text at C1 CEFR level and need you to rewrite it at an A1 CEFR level. As you simplify the text, preserve the information in the original text.

You will generate a revised text that follows the rules below.

Rules

Organization

Rule 1: Make the text concise. Remove unnecessary or repeated information.

Rule 2: Avoid long paragraphs.

Rule 3: Always start the revised text with the main idea.

Rule 4: Sequence the information in the text in a logical order. Keep the headings and subheadings.

Rule 5: Structure critical pieces of information within close proximity to each other.

Rule 6: Use clear, meaningful, and properly nested heading and subheadings whenever needed.

Rule 7: If needed, break text into lists. Use numbers instead of bullet points for the lists.

Syntax

Rule 8: Use active voice instead of passive voice. For example, write “Peter fixed the car” instead of “the car was fixed.” Confirm that all sentences are in active voice.

Rule 9: Use simple sentences (subject + verb + object).

Rule 10: Avoid using complex and compound sentences. Instead, use simple sentences.

Rule 11: Use simple and commonly used verb tenses such as simple present, simple past, and simple future.

Rule 12: Avoid sentences with dependent clauses, prepositional clauses, and conjunctions.

Instead, reorganize the sentences to have only coordinating clauses.

Vocabulary

Rule 13: Avoid using pronouns. Use nouns instead of pronouns.

Rule 14: Omit unnecessary or excess words. Write "to" instead of "in order to," or write "difficult" instead of "particularly difficult."

Rule 15: Use short words from everyday language that are used in oral communication. For example, write "use" instead of "utilize," write "show" instead of "demonstrate," write "start" instead of "initiate," or write "buy" instead of "purchase."

Rule 16: Pair down redundant phrases. For example, write "now" instead of "at this point in time."

Rule 17: Avoid abbreviations and acronyms.

Rule 18: Avoid metaphors and figure of speech. If unavoidable, illustrate them with practical examples.

Rule 19: Avoid idiomatic expressions or idioms. For example, "once in a blue moon," "beat around the bush," or "under the weather."

Rule 20: Avoid using words from other languages. If unavoidable, explain them, except commonly used words (e.g., tacos).

Rule 21: Avoid legal and technical words (i.e., specialized vocabulary from a specific field).

Rule 22: Avoid abstract concepts.

Rule 23: Use high frequency words. These words should be in the first levels, A1 or A2, of the Common European Framework of Reference (CEFR) or in the first two bands of the Corpus of Contemporary American English (COCA). Use a band 1 word over a band 2 word when possible and do not use higher band words.

Rule 24: Write it so it can be understood by a 9th grader.

Appendix 4. Prompt Used in ChatGPT to Generate the Multiple-choice Questions

Create five multiple-choice questions for the text below.

Make sure the questions can only be answered by reading the text. Do not ask general knowledge questions.

Questions should be clear and simple to understand.

The language of the questions and options is for students at A1 CEFR level.

Use everyday language.

Each question has 4 options.

Only one of the options is correct.

For each question, indicate which one is the correct option.

Scramble the options so the correct response is not always in the same position or order.

The first question asks about the main idea of the text.

The other questions ask about information that is clearly presented in the text.

32/32