English Language Teaching; Vol. 19, No. 2; 2026
ISSN 1916-4742 E-ISSN 1916-4750
Published by Canadian Center of Science and Education

Learning English in the Digital Age: The Application of ChatGPT and MagicSchool at Istinye University

Lucía Sánchez-Bejerano[1]

[1] Universidad Europea del Atlántico, Santander, Spain

Correspondence: Lucía Sánchez-Bejerano, Universidad Europea del Atlántico, Santander, Cantabria, 39011, Spain. E-mail: lucia.sanchez@uneatlantico.es

Received: October 7, 2025 Accepted: January 10, 2026 Online Published: January 28, 2026

doi: 10.5539/elt.v19n2p44 URL: https://doi.org/10.5539/elt.v19n2p44

Abstract

In the current context of globalization, proficiency in English has become an indispensable skill for academic and professional success. Aware of the challenges in acquiring a new language, Istinye University has adopted innovative approaches to improve English learning. This article focuses on the implementation of review sessions for English students, using artificial intelligence tools such as ChatGPT and MagicSchool during the 2023/2024 academic year. Adequate preparation for level exams, aligned with the Common European Framework of Reference for Languages, is essential to ensure the acquisition of linguistic competencies. AI tools allow for personalized learning, facilitating the design of materials tailored to levels A2, B1, and B2. This approach not only enhances the effectiveness of learning but also fosters student autonomy and engagement. The use of AI offers benefits to teachers, such as reducing administrative burdens and the ability to provide immediate feedback. However, the adoption of these technologies also presents challenges, such as the need for proper training for educators. The findings suggest that artificial intelligence can transform language teaching, improving the learning experience. Recommendations for future research are proposed, emphasizing the importance of exploring both the benefits and limitations of using AI in educational settings. This study opens new perspectives for English teaching, highlighting the relevance of combining traditional methods with advanced technological tools.

Keywords: artificial intelligence, tesol, chatgpt, magic school, teaching innovation.

1. Introduction

In contemporary higher education, English proficiency is a key requirement for academic mobility and professional participation. Universities increasingly face the challenge of supporting learners with diverse linguistic backgrounds while ensuring alignment with standardized proficiency frameworks such as the Common European Framework of Reference for Languages (CEFR).

Preparation for CEFR-aligned proficiency examinations requires students to develop integrated skills across grammar, vocabulary, reading, listening, and written production. For instructors working in multilingual classrooms, this process is particularly demanding, as students' needs, learning trajectories, and levels of autonomy vary considerably. Traditional exam-preparation practices—typically reliant on standardized textbooks, printed mock tests, and teacher-led correction—often lack flexibility and personalization.

Recent advances in artificial intelligence (AI) offer new possibilities for addressing these challenges. Tools such as ChatGPT and MagicSchool allow instructors to design customized learning materials, adapt tasks to specific proficiency levels, and provide scaffolded practice opportunities aligned with examination formats. While existing research highlights the efficiency and versatility of AI in educational design, fewer studies examine how these tools concretely influence student engagement, autonomy, and learning outcomes in exam-oriented EFL contexts.

This study investigates the implementation of AI-supported review sessions for English language students at Istinye University during the 2023–2024 academic year. By integrating ChatGPT and MagicSchool into CEFR-aligned exam preparation seminars at A2, B1, B2, and B2+ levels, the study aims to analyze how AI-mediated task design affects learner engagement, perceived usefulness, and instructional practices in comparison with more traditional approaches.

2. Literature Review

In the field of foreign language teaching, the use of technology has been extensively researched over the past decades. Various tools and resources have been employed to provide both students and teachers with opportunities to experience language use in a more authentic, native-like manner. Examples include radio, podcasts, television, cinema, and music, which are frequently used to develop oral skills. Similarly, technology enables learners to engage with authentic written texts such as newspapers, magazines, and books, thereby enhancing reading comprehension. Moreover, technological tools have facilitated the production of oral output through video calls and written output through chats and emails. Access to authentic materials has never been as widespread and immediate as it is today.

The use of authentic materials in the classroom ensures a closer alignment with students' real-life interests. However, in most cases, such materials require adaptation before being implemented in a pedagogical context. Authentic resources are typically not tailored to the specific linguistic content intended for instruction, nor do they necessarily align with an established curriculum. Consequently, some of the students' learning needs may not be fully addressed—particularly when they are preparing for standardized assessments designed to certify their level of proficiency in a foreign language.

Therefore, it becomes essential to employ resources that help students prepare for specific types of examinations by engaging them in activities similar to those they will encounter during testing. Such resources have been widely developed by publishing houses in various formats, often integrating digital platforms and technological tools. Nevertheless, teachers also have the option to create these resources themselves through the use of artificial intelligence, which fosters greater autonomy and independence from prefabricated materials. This approach promotes an ecological design of teaching materials—one that is adaptable, context-sensitive, and responsive to both learners' needs and pedagogical objectives.

2.1 The Common European Framework of Reference for Languages (CEFR) and Proficiency Examinations

The Common European Framework of Reference for Languages (Council of Europe, 2020) establishes descriptors for each communicative language activity: comprehension, production, interaction, and mediation. For each of these communicative activities, the CEFR provides a categorization of descriptors corresponding to levels A1–C2 and according to specific communicative situations. The following table presents the descriptors related to listening and reading comprehension for selected proficiency levels (see Table 1).

Table 1. CEFR Descriptors

	CEFR Descriptors	
Level	Listening Comprehension	Reading Comprehension (general)
A2	Understands enough to satisfy concrete needs, provided that speakers articulate slowly and clearly. Understands phrases and expressions related to areas of immediate personal relevance (e.g., very basic personal and family information, shopping, local geography, employment), provided that speech is slow and clear.	Understands short, simple texts on familiar topics containing high-frequency everyday or work-related vocabulary. Understands short, simple texts that include a large proportion of high-frequency words, including internationalisms.
B1	Understands straightforward factual information on familiar topics or related to work, identifying both the main message and specific details, provided the speech is clear and delivered in a familiar variety. Understands the main points of clear standard speech or a familiar variety on familiar matters regularly encountered in work, school, or leisure, including short narratives.	Reads straightforward factual texts on subjects related to their field of interest with satisfactory comprehension.
B2	Understands standard speech or a familiar variety, whether face-to-face or broadcast, on both familiar and unfamiliar topics of personal, social, academic, or professional relevance. Only excessive background noise, poor discourse organization, or idiomatic usage may impede comprehension. Understands the main ideas of complex speech on both concrete and abstract topics in standard language or a familiar variety, including technical discussions within their field of specialization. Understands extended speech and complex lines of argument provided the topic is reasonably familiar and the discourse is clearly structured with explicit markers.	Reads with a high degree of independence, adapting reading style and speed to different texts and purposes, and selectively using appropriate reference sources. Possesses a broad active reading vocabulary but may experience some difficulty with infrequent idiomatic expressions.

Note. Based on Council of Europe (2020, pp. 60-67).

The B2+ level, on the other hand, is defined by the CEFR (Council of Europe, 2020) as a "high degree of the Advanced level" (p. 194). Although specific descriptors for each communicative activity are not explicitly detailed for this level, the CEFR recognizes it in relation to aspects such as plurilingual competence and mediation.

Regarding grammar and vocabulary study, the CEFR promotes an action-oriented approach that emphasizes language learning through communicative tasks that require the use of language but do not focus exclusively on its formal study. Thus, the CEFR (2020, p. 36) "represents a shift from syllabuses based on a linear progression through linguistic structures."

Language proficiency examinations designed to certify foreign language competence are aligned with the CEFR and its descriptors. These examinations typically include similar types of tasks across levels A2, B1, and B2, encompassing reading comprehension, listening comprehension, and written and oral production as common components (Cambridge University Press & Assessment, 2025; Oxford University Press, 2025; ETS TOEFL, 2025; IELTS, 2025). However, some examinations also assess additional aspects such as grammatical and lexical competence. For instance, the APTIS tests developed by the British Council (2025) include specific components for grammar and vocabulary.

Based on the CEFR (Council of Europe, 2020), its certification and promotion examinations include mediation as an assessable skill. This mediation is evaluated through both oral and written activities, which are integrated with the assessment of reading, listening, and speaking comprehension and production skills.

2.2 Technological Innovations in Language Education

The integration of technology into language education has opened new possibilities for both students and educators, facilitating access to a wide range of resources and tools that enrich the learning process.

ChatGPT employs advanced natural language processing (NLP) capabilities to interact fluently with human users, generating coherent and contextually relevant responses from textual input. Its ability to autonomously

and comprehensibly produce text makes it a versatile resource for a variety of educational applications. Within the field of language teaching, ChatGPT can be employed in multiple ways to enhance learning outcomes.

Ruiz Mendoza (2023) examines how ChatGPT can be used to design examinations efficiently and flexibly. The study investigates best practices for formulating specific prompts that enable ChatGPT 4.0 to generate generic assessment designs, which can then be customized by instructors according to their specific needs. This approach significantly streamlines traditional exam-creation processes by leveraging advanced NLP capabilities. While the findings emphasize the satisfactory performance of ChatGPT 4.0 in responding to user prompts, they also highlight the importance of managing interaction duration to prevent inconsistencies during extended sessions.

Similarly, Ulloa et al. (2024) explore how ChatGPT can be applied in educational contexts to foster skills and knowledge development, particularly in the creation of children's music and other educational materials. This quantitative study reveals that participants responded positively to the experience of engaging in educational content creation using ChatGPT, and expressed interest in its integration as an instructional tool within the learning process. These findings underscore the potential of AI-based tools such as ChatGPT to strengthen constructivist approaches by actively engaging learners in the production of relevant and personalized educational materials.

Furthermore, Al-Zoubi (2024) highlights how ChatGPT's natural language processing capacities can be used to design educational content, learning objectives, interactive materials, and assessment tasks in academic courses—including the emerging field of Internet of Things (IoT) Engineering. This approach not only promotes more dynamic and engaging learning experiences by incorporating real-world examples and case studies but also optimizes the time and effort required by instructors to design and deliver effective courses.

MagicSchool (2025) is a specialized platform for teaching and for designing tasks, instructional units, and differentiated activities that address learner diversity. Among the features that may be particularly appealing to educators are the creation of multiple-choice or open-ended questions based on YouTube videos; the adaptation of tasks or texts through scaffolding strategies; the generation of vocabulary lists, review games, and thematic songs; feedback aligned with assessment criteria; and the development of rubrics, units, or project-based learning materials. In addition, MagicSchool offers specialized tools for subject-specific areas such as generating mathematical problems or designing training exercises for particular sports.

Tilak et al. (2024) present this platform as a valuable tool for course design, emphasizing its efficiency in reducing the time required to develop instructional tasks, although they note that ChatGPT may be more effective for the design of entire instructional units. Indeed, Ruksana (2024), in her study of various AI-based tools and platforms used in education, highlights that MagicSchool could save teachers up to ten hours per week in task design and lesson planning.

3. Method

The general objectives of this study were to design and implement exam preparation seminars for students of English as a Foreign Language (EFL) at various proficiency levels (A2, B1, B2, and B2+), integrating artificial intelligence into the instructional process. To this end, a design was developed based on tasks modeled after those found in English language proficiency certification exams such as APTIS (British Council, 2025), the Oxford Test of English (Oxford University Press, 2025), and the various certificates issued by Cambridge University Press & Assessment (2025).

The target participants for these activities were undergraduate students enrolled in different degree programs at Istinye University in Istanbul, Turkey. Throughout the academic year, these students attended modules aimed at achieving a CEFR-aligned language proficiency certification developed by the institution itself. The modules were divided into four groups according to students' proficiency levels. The A2 group consisted of forty-three students (twenty-five male and eighteen female). The B1 group included forty-eight students (fifteen male and thirty-three female). The B2 group comprised twenty-six students, fourteen male and twelve female. Finally, the B2+ group consisted of seven students, four male and three female.

Each group was allocated ninety minutes to complete the seminar. Consequently, each seminar was designed following the formats established in the APTIS, Oxford Test of English (hereafter OTE), Oxford Test of English Advanced (hereafter OTE Advanced), and Cambridge English Qualifications exams: A2 Key, B1 Preliminary, and B2 First.

The following section provides a detailed breakdown of how the activities were structured for each proficiency level, in accordance with the Common European Framework of Reference for Languages (CEFR) and taking into account the institutional specifications.

3.1 A2 Group

For the group with an A2 proficiency level, the university evaluated linguistic activities focused on grammar, vocabulary, and reading comprehension. To design tasks addressing these areas, exercises were generated using ChatGPT.

Regarding grammar and vocabulary exercises, it was observed that the APTIS examination includes fifty multiple-choice tasks to be completed within twenty-five minutes. Consequently, ChatGPT was prompted to create fifty exercises covering grammatical and lexical aspects appropriate for the A2 level.

For reading comprehension, ChatGPT was asked to design two tasks, each based on a cultural text—one about the United Kingdom and another about the United States. These tasks were modeled after Part 1 of the APTIS, Oxford Test of English (OTE), and Cambridge A2 Key exams, which involve reading a short text and answering five or six multiple-choice questions, each with three possible options. In this case, five questions were included per task. Given that the APTIS and OTE tests allocate thirty-five minutes for four tasks, and the A2 Key allows one hour for seven tasks, it was proposed that the two reading comprehension tasks be completed within a maximum of thirty minutes.

Additionally, for students who completed the tasks early, a supplementary activity was designed using ChatGPT. This consisted of a matching exercise in which students were asked to pair ten terms with their corresponding definitions, two of which served as distractors.

3.2 B1 Group

For the group with a B1 proficiency level, the university likewise assessed linguistic activities involving grammar, vocabulary, and reading comprehension. To design appropriate tasks for these areas, ChatGPT was again employed.

For grammar and vocabulary exercises, following a similar approach to that used for the A2 level, ChatGPT was prompted to generate fifty multiple-choice tasks encompassing grammatical and lexical features suitable for B1 learners.

The reading comprehension component consisted of three tasks. First, ChatGPT was asked to design a task similar in structure to that used for the A2 group. The second task was modeled after Part 6 of the Cambridge Preliminary B1 exam and required students to complete a cloze text, filling in missing words based on contextual understanding as well as grammatical and lexical knowledge.

Finally, drawing upon Part 3 of the APTIS exam and Part 2 of both the OTE and Preliminary B1 exams, ChatGPT was instructed to design a text describing four different personal profiles, each outlining the individual's main interests and activities. Students were then required to match six items to these profiles. Each item described a specific activity or preference, and students had to determine which profile best corresponded to each item. Since each item matched only one profile, students needed to analyze and interpret the information provided in the descriptions to make the correct associations.

3.3 B2 Group

For the B2 proficiency level, the university sought once again to assess grammar, vocabulary, and reading comprehension. To this end, tasks were designed using ChatGPT, following a structure similar to that employed in the A2 and B1 seminars. However, in this case, the reading comprehension section included a fourth task modeled after Part 4 of the Cambridge B2 First examination. This task, known as key word transformation, requires reformulating a given sentence using a specified key word while maintaining the original meaning. The student is presented with a base sentence and must rewrite it, incorporating the key word within a new grammatical structure that conveys the same meaning. This exercise assesses the learner's ability to comprehend the underlying sense of the sentence and to apply grammatical and lexical rules appropriately to produce an equivalent sentence, thereby demonstrating flexibility and accuracy in language use.

The sequencing of the sessions for this level, as well as for levels A2 and B1, was structured as shown in Table 2 below.

Table 2. Sequencing of A2, B1, and B2 Seminars

Task	Sequencing of A2, B1, and B2 Seminars Minutes
1. Seminar introduction	5
2. Grammar and vocabulary tasks	25
3. Reading comprehension tasks	30
4. Task correction and seminar conclusion	30

Note. Description of tasks sequence for the A2, B1, and B2 seminars.

3.4 B2+ Groups

For the group with a B2+ proficiency level, the university designed a final assessment that included listening and reading comprehension, as well as written production activities. To select and design tasks addressing these skills, the tools ChatGPT and MagicSchool were employed.

For the listening comprehension component, two videos were selected. The first, produced by Financial Times (2014), lasted approximately four minutes and discussed the economic situation of a community in the United Kingdom. The accompanying task was created using MagicSchool, a platform that allows users to generate questions for YouTube videos. The available formats include open-ended, multiple-choice, and true-or-false questions, with options to generate three, five, seven, or ten questions depending on the learner's academic level. In this case, five open-ended questions at the university level were generated.

The second video, produced by National Geographic (2019), was a thirteen-minute documentary focusing on science and technology topics. To design a task similar to those found in examinations such as the OTE Advanced, MagicSchool was instructed to create ten multiple-choice questions, also at the university level.

For reading comprehension, both AI tools were used collaboratively. First, ChatGPT was asked to produce an extended text of more than 500 words on the topic of technology use in education, appropriate for the B2+ level. Subsequently, MagicSchool was used to generate five multiple-choice questions to assess comprehension of the text. The design of this task drew upon Part 4 of the OTE Advanced exam and Part 5 of the Cambridge C1 Advanced exam, both of which include long reading passages of approximately 750 words accompanied by multiple-choice questions.

In the written production section, one of the main challenges was the limited time available to complete and thoroughly review each student's writing. Therefore, an approach based on the transformation of informal into formal texts was adopted. Activities were designed in which students worked with texts generated by ChatGPT, following a two-step process. First, an informal text was provided, and students were instructed to transform it into a formal version. This activity allowed students to become familiar with the conventions of formal English writing while practicing strategies for adapting style, tone, and vocabulary. A model version of the text in formal register was then provided, enabling students to compare and evaluate their transformations. This approach facilitated both the recognition of stylistic differences and the active production of formal written discourse.

Table 3. Sequencing of B2+ Seminar

Task	Sequencing of B2+ Seminar Minutes
1. Seminar introduction	*5*
2. Listening comprehension tasks	*30*
3. Reading comprehension tasks	*25*
4. Writing production tasks and seminar conclusion	30

Note. Description of tasks sequence for the B2+ seminar

4. Results

The results of this study are presented with a focus on how the integration of artificial intelligence tools influenced learning processes and instructional practices, rather than on the technical features of the tools themselves. The analysis is organized around learner engagement, task performance, differentiation, and comparison with traditional exam-preparation methods.

4.1 Impact on Task Performance and CEFR Alignment

Across all proficiency levels (A2, B1, B2, and B2+), students were able to complete the AI-generated tasks within the timeframes established by standardized CEFR-aligned examinations. This suggests that the tasks designed using ChatGPT and MagicSchool were appropriately calibrated in terms of linguistic complexity, cognitive demand, and exam relevance.

In contrast to traditional methods—typically based on fixed textbook exercises or previously published mock exams—the AI-supported tasks allowed instructors to fine-tune grammatical focus, lexical range, and text length to the specific cohort. For example, A2-level grammar tasks emphasized high-frequency structures and concrete contexts, while B2 and B2+ tasks incorporated abstraction, inference, and extended discourse. This differentiation is more difficult to achieve using static materials and represents a key advantage of AI-supported design.

4.2 Learner Engagement and Autonomy

One of the most significant outcomes observed during the review sessions was increased learner engagement, particularly in reading and listening comprehension activities. Students demonstrated sustained attention and higher participation when interacting with AI-generated or AI-mediated materials, especially when these materials were linked to contemporary or authentic topics.

Compared with traditional review sessions—often characterized by teacher-led correction and uniform task distribution—the AI-supported sessions encouraged greater learner autonomy. Students progressed through tasks at their own pace and were able to focus on areas of difficulty without disrupting the group. This shift aligns with learner-centered pedagogical principles and supports the development of self-regulated learning strategies.

However, engagement was not uniform across all participants. While many students benefited from the flexibility of AI-generated tasks, some learners—particularly those with lower confidence or limited experience with digital tools—required additional scaffolding. These students tended to rely more heavily on instructor support, suggesting that AI-enhanced environments may privilege learners who already possess higher levels of academic autonomy.

4.3 Differential Effects across Learner Profiles

The findings indicate that AI tools do not impact all learners equally. Students with stronger linguistic foundations and greater familiarity with exam formats appeared to benefit more from AI-supported activities, particularly in higher-level tasks requiring inference, paraphrasing, or stylistic transformation.

Conversely, learners who struggled with foundational skills or experienced higher levels of test anxiety showed more modest gains. In these cases, traditional methods—such as explicit teacher explanation and guided practice—remained essential. This suggests that AI tools are most effective when integrated into a blended pedagogical model rather than used as a standalone solution.

4.4 Instructional Efficiency and Teacher Practices

From the instructors' perspective, AI tools significantly reduced the time required for task creation and material adaptation. Compared with traditional approaches, which often involve modifying publisher-provided resources, AI-supported design enabled rapid generation of level-appropriate materials.

Nevertheless, this efficiency was partially offset by the need for careful review and editing of AI-generated content to ensure accuracy, clarity, and pedagogical appropriateness. This highlights that AI does not eliminate teacher involvement but reshapes it, shifting the instructor's role from material creator to pedagogical curator.

5. Discussion

The purpose of this study was not merely to demonstrate the feasibility of using AI tools in EFL exam preparation, but to examine how such tools influence learning processes, instructional practices, and pedagogical equity. The findings provide insight into both the potential and the limitations of AI-supported language education.

5.1 Interpreting Learner Engagement through the Action-Oriented and Constructivist Frameworks

The increased engagement observed during AI-supported review sessions can be interpreted through the action-oriented and constructivist frameworks that underpin CEFR-aligned pedagogy. From an action-oriented perspective, learners are viewed as social agents who use language to complete purposeful tasks. The AI-generated activities in this study supported this principle by allowing instructors to design exam-oriented tasks that simulated authentic communicative demands rather than isolated linguistic drills.

Recent empirical research supports this interpretation. A scoping review by Xiao et al. (2025) demonstrates that generative AI tools such as ChatGPT are most effective when used to support task-based and meaning-focused instruction in EFL contexts. Similarly, Ferreiro-Santamaria (2024) found that students perceived AI-supported activities as more engaging when tasks were clearly aligned with communicative goals and assessment criteria, rather than used for generic language practice.

From a constructivist standpoint, the adaptability of AI-generated materials reduced cognitive overload by aligning task difficulty with learners' proficiency levels. This allowed students to actively construct meaning rather than struggle with tasks that exceeded their current linguistic resources, which is consistent with CEFR-informed scaffolding principles.

5.2 Learner Autonomy, Self-Regulation, and Unequal Benefits

The findings also align with theories of learner autonomy and self-regulated learning, both of which emphasize learners' ability to monitor progress, identify weaknesses, and adjust strategies accordingly. In this study, AI-supported review sessions enabled students to progress at individual paces and focus on specific areas of difficulty, fostering greater autonomy compared with traditional teacher-led review sessions.

Empirical evidence from Vo and Nguyen (2024) confirms that EFL learners generally report positive perceptions of ChatGPT as a tool for independent learning, particularly for exam preparation and language practice outside the classroom. However, the same study highlights disparities in benefit, with students possessing higher digital literacy and academic confidence deriving greater advantage from AI tools.

This pattern mirrors the present findings, where not all learners benefited equally from AI-supported tasks. Students with lower proficiency levels or higher test anxiety continued to rely on explicit teacher guidance, suggesting that AI tools alone cannot compensate for gaps in foundational skills or affective barriers. These results reinforce the need for blended pedagogical models that combine AI-supported autonomy with structured instructor mediation.

5.3 Comparison with Traditional Exam-Preparation Methods

The contrast between AI-supported and traditional exam-preparation methods reveals important pedagogical trade-offs. Traditional approaches—characterized by standardized textbooks, fixed mock exams, and teacher-led correction—offer reliability and transparency but limited adaptability to heterogeneous learner needs.

AI-supported review sessions addressed some of these limitations by enabling rapid differentiation and customized task generation. An action research study by Dorasamy et al. (2024) similarly found that ChatGPT-supported activities enhanced learner engagement and perceived relevance in foreign language learning contexts, particularly when integrated into existing instructional routines rather than used as standalone tools.

However, traditional methods retained value for learners who required explicit explanation and guided practice. This supports the argument made by Nguyen and Hoang (2025) that generative AI should be positioned as a pedagogical supplement rather than a replacement for established instructional practices in EFL education.

5.4 Academic Integrity and the Risk of Over-Reliance on AI

Consistent with recent literature, the findings raise concerns related to academic integrity and learner over-reliance on generative AI. Although AI tools in this study were restricted to review and practice activities, their ability to generate complete responses presents a risk that learners may prioritize AI-produced output over their own language development.

A recent literature review by Nguyen and Hoang (2025) emphasizes that while generative AI can enhance learning efficiency, it also blurs boundaries between support and substitution. Without explicit ethical guidance, students may come to view AI as an authoritative source rather than a learning aid. This underscores the importance of integrating metacognitive and ethical instruction alongside AI-supported activities.

5.5 Critical and Ontoepistemological Perspectives on AI in Language Education

From a critical perspective, recent reviews caution that generative AI tools are not pedagogically neutral. Nguyen and Hoang (2025) emphasize that while AI can enhance efficiency and personalization in EFL learning, it may also reinforce dominant linguistic norms and encourage over-reliance if not carefully mediated. Similarly, Liu et al. (2025) note that generative AI systems reflect epistemological assumptions embedded in their training data, which may limit their responsiveness to learners' sociocultural and multilingual realities.

In multilingual and transcultural educational contexts, this limitation is particularly salient. AI-generated materials may inadequately capture learners' identities, affective dimensions, and contextualized language use. Therefore, AI should be positioned not as an epistemic authority, but as a pedagogical tool whose outputs require critical evaluation and adaptation by instructors.

6. Conclusions

This study examined the implementation of AI-supported review sessions using ChatGPT and MagicSchool in CEFR-aligned English exam preparation at Istinye University. The findings indicate that AI tools can enhance task personalization, student engagement, and instructional efficiency when compared with traditional methods such as textbook-based practice, static mock exams, and exclusively teacher-led material design.

At the same time, the results demonstrate that AI integration does not benefit all learners equally. Differences in learner autonomy, digital literacy, and affective factors influence how students engage with AI-supported activities. Furthermore, issues related to academic integrity, AI bias, over-reliance on automated language, and data security highlight the need for responsible and critically informed implementation.

Rather than replacing traditional approaches, AI tools should be understood as complementary resources that extend instructors' pedagogical repertoires. When combined with human mediation, reflective teaching practices, and critical awareness of sociocultural contexts, AI can contribute meaningfully to language education.

Future research should incorporate empirical measures of learning outcomes and draw on critical and ontoepistemological perspectives in applied linguistics to better understand how AI-mediated instruction interacts with learners' identities, multilingual repertoires, and educational trajectories.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The author acknowledges the support of the Erasmus+ Programme, which enabled the academic visit to Istinye University during which part of this research was conducted. Institutional support from Istinye University and Universidad Europea del Atlántico is also gratefully acknowledged. No special agreements concerning authorship apply to this work. The author thanks colleagues for their assistance and informal feedback during the preparation and revision of the manuscript.

References

Al-Zoubi, A. (2024). Designing IoT introductory course for undergraduate students using ChatGPT. In M. E. Auer, U. R. Cukierman, E. Vendrell Vidal, & E. Tovar Caro (Eds.), Towards a hybrid, flexible and socially engaged higher education (pp. 383–394). Springer. https://doi.org/10.1007/978-3-031-51979-6_40

British Council. (2025). Aptis ESOL General. https://www.britishcouncil.es/examenes/aptis/general

Cambridge University Press & Assessment. (2025). *English language assessment*. https://www.cambridgeenglish.org/es/

Consejo de Europa. (2020). Marco común europeo de referencia para las lenguas: Aprendizaje, enseñanza y evaluación. *Volumen complementario*. Servicio de publicaciones del Consejo de Europa. https://www.coe.int/lang-cefr

Dorasamy, M., Jun, L. Y., & Kaliannan, M. (2024). Generative AI and ChatGPT for foreign language learning: An action research study. *South Asian Journal of Higher Education, 31*(6), 94-199. https://10.62206/sajm.31.6.2024.94-119

ETS TOEFL. (2025). *Contenidos del examen*. https://toefltest.mx/about/

Ferreiro-Santamaria, G. (2024). Exploring the role of ChatGPT in English teaching within higher education settings. *International Journal of Trends and Developments in Education, 4*(1), 44-58. http://10.5281/zenodo.11204781

Financial Times. (2014, February 23). British middle class splits in two [Video]. YouTube. https://www.youtube.com/watch?v=C5lUZ9OXG3s

IELTS. (2025). IELTS Academic. https://ielts.org/take-a-test/test-types/ielts-academic-test

Liu, J., Sihes, A.J.B. & Lu, Y. How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learn. Environ, 12*(48). https://doi.org/10.1186/s40561-025-00406-0

National Geographic. (2019, September 5). Science fiction inspires the future of science [Video]. YouTube. https://www.youtube.com/watch?v=IAZAIdLwhGw&t=3s

Oxford University Press. (2025). Oxford Test of English. https://elt.oup.com/feature/global/oxford-test-of-english/oxford-test-of-english?cc=global&selLanguage=en

Nguyen, T. H., & Hoang, T. N. H. (2025). Investigating the Promises and Perils of Generative AI in EFL Learning in Higher Education: A Literature Review. *AsiaCALL Online Journal, 16*(1), 1–15. https://doi.org/10.54855/acoj.251611

Ruiz Mendoza, K. K. (2023). El uso de ChatGPT 4.0 para la elaboración de exámenes: Crear el prompt adecuado: The use of ChatGPT 4.0 for test development: Creating the right prompt. *LATAM Revista Latinoamericana De Ciencias Sociales Y Humanidades, 4*(2), 6142–6157. https://doi.org/10.56712/latam.v4i2.1040

Ruksana, T. P. (2025). AI-driven pedagogy: Unveiling ChatGPT's influence in education. *Journal of Applied Science, Engineering, Technology and Management, 2*(1), 9–14. https://doi.org/10.61779/jasetm.v2i1.2

Tilak, S., Lincoln, J., Miner, T., Christensen, T., Jankowski, J., & Setnar, J. (2024, June 24–29). MagicSchool AI as a cybernetic tool for course design: A framework to facilitate machine-teacher partnerships [Conference presentation]. 18th International Conference on Sociocybernetics Artificial Intelligence: New perspectives for social theory and practice, Krakow, Poland.

Ulloa Espinoza, C. M., Ojeda Morán, M., Bedoya Gutiérrez, A., López Proaño, A., Palacios González, J., & Angulo Quiñónez, O. (2024). ChatGPT: Generación de música infantil en contextos educativos: ChatGPT: Generation of children's music in educational contexts. *LATAM Revista Latinoamericana De Ciencias Sociales Y Humanidades, 5*(2), 565–577. https://doi.org/10.56712/latam.v5i2.1894

Vo, A., & Nguyen, H. (2024). Generative artificial intelligence and ChatGPT in language learning: EFL students' perceptions of technology acceptance. *Journal of University Teaching and Learning Practice, 21*(4). https://doi.org/10.53761/fr1rkj58

Xiao, F., Zhu, S., & Xin, W. (2025). Exploring the landscape of generative AI-powered writing instruction in EFL education: A scoping review. *ECNU Review of Education*, 1-19.