Research Journal in Advanced Humanities

RJAH

https://doi.org/10.58256/tf8dyv05

Open Access

Check for updates

RESEARCH ARTICLE

**Section:** *Literature, Linguistics & Criticism*

# Leveraging artificial intelligence applications to develop English language skills and foster sustainable development awareness among Saudi university students: A step towards future economies

**Ali Lamouchi[1], Mustafa Mohammed Sadek Elsayed[2], Mohamed Sayed Abdellatif[3] & Shaban Aly[4]**

[1]Department of English Language and Literature, Prince Sattam bin Abdulaziz University, Saudi Arabia

[2]Unit of Scientific Research, Qassim University, Buraydah, Saudi Arabia

[3]Department of Psychology, Prince Sattam bin Abdulaziz University, Saudi Arabia

[4]Department of Mathematics, Al-Azhar University, Assiut, Egypt

**Correspondence:** a.lamouchi@psau.edu.sa

## ABSTRACT

This research aims to examine the effectiveness of utilizing artificial intelligence (AI) applications in enhancing Saudi undergraduate students' English language skills while simultaneously promoting their attitudes toward sustainable development, aligning with the Kingdom's vision for building future-ready economies. A total of 23 first-year students from the Applied College at Prince Sattam bin Abdulaziz University participated in the study, with 12 students assigned to the experimental group and 11 to the control group. The study employed an AI-based training program, an English language proficiency test, and an attitude scale toward sustainable development. Using a quasi-experimental design and statistical analyses, including Mann-Whitney and Wilcoxon tests, the findings revealed statistically significant differences favoring the experimental group in both English language skills and attitudes toward sustainability. Moreover, the results confirmed that the positive effects of the AI-based training persisted in follow-up measures, indicating sustainable impact. By demonstrating the role of AI tools in strengthening students' communication skills and sustainable mindset, this study highlights the importance of integrating advanced technologies in higher education to prepare human capital for knowledge-based economies. Recommendations and directions for future research are presented to further support the strategic national priority of transitioning toward diversified, innovation-driven economic growth.

**Keywords:** Artificial Intelligence, English language skills, sustainable development attitudes, future economies, digital transformation, Saudi university students

**Research Journal in Advanced Humanities**

Volume 6, Issue 3, 2025

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

## ARTICLE HISTORY

Submitted: **18 June 2025**

Accepted: **09 August 2025**

Published: **18 August 2025**

## HOW TO CITE

Lamouchi, A., Sadek Elsayed, M., Abdellatif, M., & Aly , S. (2025). Leveraging artificial intelligence applications to develop English language skills and foster sustainable development awareness among Saudi university students: A step towards future economies. *Research Journal in Advanced Humanities, 6*(3). https://doi.org/10.58256/tf8dyv05

**Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited**

## Introduction

Artificial intelligence (AI) advancements are having a profound impact on numerous academic, scientific, industrial, business fields as well as our daily lives. According to Chiu et al. (2023), artificial intelligence (AI) is the ability of digital systems to carry out tasks that would typically need human intelligence. Related technologies include machine learning, image analysis, voice recognition, and language comprehension. AI applications have significantly improved computer performance and advanced the field of teaching foreign languages. For AI technology to have a significant impact on educational practice, it must be used in language instruction efficiently (Chen, 2020). Four functions of AI in education were identified by Hwang et al. (2020): policy-making adviser, intelligent learning tool or partner, intelligent tutor, and intelligent tutee. This demonstrated how AI plays a variety of functions in education and its significance in enhancing learning outcomes and assisting with policymaking.

Numerous indicators suggest that scholarly interest in using AI in education has been increasing over the past few years (Jain & Raghuram, 2023). AI's applications in foreign language education have significantly expanded due to the quick development of large data management and natural language processing technologies (Li, 2020). According to Moulieswaran et al. (2023), the application of AI in language education has not only demonstrated definite benefits but also encouraged curriculum development and teaching strategy optimization. Modern technologies create realistic virtual environments and intelligently support collaborative learning. These developments give students access to the proper help when they need it, empowering them to deal with problems and challenges that arise in real life (Luckin et al., 2016).

On the one hand, the relationship between AI and language education has been the subject of numerous studies. Pokrivcakova (2019) noted that teachers and students of foreign languages could use a variety of AI-powered resources to streamline their teaching and learning procedures. AI integration in foreign language education offers students quick, and individualized assistance. On the other hand, while maintaining respect for human dignity and security, international communities are quickly putting themselves in a position to guarantee the application and role of AI in creating inclusive knowledge societies and helping nations achieve the 17 Sustainable Development Goals (SDGs) of the UN (Bachmann et al., 2022). Given that AI offers both significant opportunities and challenges, it is necessary to focus on and gather more empirical cases and discussions from various development sector stakeholders. These discussions should highlight the growing relationship between AI and the development agenda, position the roles of AI, identify risks and limiting factors that need to be addressed, and put policies in place to fully realize the potential of societal impact (Holmes & Miao, 2023).

Accordingly, the focus of this research is to investigate the effectiveness of using Artificial Intelligence applications in developing Saudi undergraduate students' English language skills and enhancing their attitudes toward sustainable development. It seeks to provide insights into the current state of AI in foreign language education in the Saudi higher education context. Researchers, educators, and decision-makers can use the findings from this study to enhance teaching methods and raise the standard of English language education. Additionally, it could help sustainable development professionals in organizing and carrying out workable plans and procedures while taking possible hazards into account.

## Literature Review

English is more than a language; it is a doorway to opportunities, a cross-border means of communication, and a key skill required by employers globally (Adriansen et al., 2022). Because English is the lingua franca of business, science, diplomacy, and the internet, being proficient in it can lead to opportunities for education, employment, and international cooperation Al-Smadi et al., 2020, Lamouchi, A., Mohamed Sayed Abdellatif, & Hafez, W. E. (2024). Future developments in AI's role in English language learning promise even more creative solutions. The potential of AI to improve language learning's effectiveness and accessibility is becoming more apparent as technology develops (Wei, 2023). These advancements bring in a time when language competency is more relevant and achievable than ever before, and they also fit in perfectly with the larger objectives of education and sustainable development (Kamalov et al., 2023).

Artificial Intelligence and language education have been the subject of numerous studies. According to Pokrivcakova (2019), teachers and students of foreign languages can use a variety of AI-powered applications to simplify their teaching and learning process. Learners can receive rapid, individualized support when AI is

included into foreign language instruction. At a Turkish university, the application of ChatGPT in L2 education was investigated in the study of Karataş et al. (2024). The results demonstrated that ChatGPT improved students' motivation and engagement while having a favorable effect on writing, grammar, and vocabulary learning. This study offered teachers recommendations and strategies for using ChatGPT efficiently. According to Ribeiro (2021), artificial intelligence (AI) is the field of computer science that focuses on creating intelligent systems capable of performing writing tasks.

Academic writing in English has greatly benefited from the use of AI-based technology applications. Wei et al. (2023) investigated how Automated Writing Evaluation (AWE) training affected the writing skills of EFL students. The study showed that AWE may be useful in enhancing writing skills, underscoring the need of implementing AWE in instructional strategies. According to a study by Lu (2019), English language learners who utilize AI in their writing can get help with grammar and language correction, which helps them find and fix mistakes. By the same token, a study by Rahman et al. (2022), showed that students demonstrated a positive impact when they used AI-based technology to edit and offer feedback during the writing process. According to Chen (2023), artificial intelligence (AI) gained popularity by offering a variety of technological tools for academic writing, such as QuillBot, DeepL, Google Translate, ChatGPT, and others. In a classroom setting for academic writing, English teachers and students could utilize appropriate Artificial Intelligence applications. Thus, AI can help students write academically in English.

According to the researchers' experience teaching, some Saudi undergraduate students struggle with proper grammar, coming up with ideas, and using the right language when writing academically in English. The use of artificial intelligence applications might be helpful in overcoming these challenges. Mali (2022), asserted that technology can serve as an effective tool to support students in overcoming challenges during the English academic writing process. According to Setyani et al. (2023), students may be encouraged to express their thoughts in academic writing in English by using AI-based technology tools. According to a meta-analysis by Tajik (2025), students who used AI-supported platforms proved faster improvement in their writing skills than those in traditional classrooms. Thus, technology and artificial intelligence have the potential to help students with their academic writing in English.

Even though there is a growing corpus of research on the use of artificial intelligence to improve English language skills worldwide, there are still significant gaps in the literature that pertain to using Artificial Intelligence in English language education in Saudi higher education context. In order to fill these gaps, the current study investigates the effectiveness of using artificial intelligence applications in developing Saudi undergraduate students' English language skills and enhancing their attitudes toward sustainable development. The study will also add to the current debate about the best practices for integrating AI tools in English language instruction, with a special emphasis on the distinctive cultural and educational context of Saudi Arabia.

**Theoretical Framework**

The Technology Acceptance Model (TAM) serves as the theoretical framework for this study, which explores the effectiveness of using Artificial Intelligence applications in developing Saudi undergraduate students' English language skills and enhancing their attitudes toward sustainable development. The Technology Acceptance Model (TAM), first put forth by Davis (1989) and then expanded upon by Venkatesh and Davis (2000), is a well-known framework for comprehending users' acceptance and adoption of technology, especially in educational settings.

The two main components of the Technology Acceptance Model (TAM) are perceived usefulness (PU) and perceived ease of use (PEOU), which have a significant impact on people's adoption of technology. PEOU is the degree to which consumers believe that a certain technology is simple and easy to use. Conversely, PU represents users' opinions about how well technology can help them perform and accomplish their duties.

PEOU and PU hold great significance in the context of English language education. Learners' willingness to adopt AI in their English language classrooms is greatly influenced by how easily English language learners perceive AI-integrated language education tools to be incorporated into their learning practices (PEOU) and how beneficial they believe these tools to be in improving their language learning experiences (PU).

This study is theoretically based on the two main components of the Technology Acceptance Model discussed above to examine the effectiveness of using artificial intelligence applications in developing Saudi

undergraduate students' English language skills and enhancing their attitudes toward sustainable development. The study anticipates a positive effect of using artificial intelligence applications in developing Saudi undergraduate students' English language skills and enhancing their attitudes toward sustainable development.

**Research Questions**

The research primary question is: What is the effectiveness of using artificial intelligence applications in developing Saudi undergraduate students' English language skills and enhancing their attitudes toward sustainable development?

From the research primary question, the following questions arise:

- Are there statistically significant differences between the mean scores of the experimental group and the control group on the four components of the English language skills test and their total score after implementing the program?
- Are there statistically significant differences between the mean scores of the experimental group on the four components of the English language skills test and their total score in the pre- and post-tests?
- Are there statistically significant differences between the mean scores of the experimental group on the four components of the English language skills test and their total score in the post- and follow-up tests?
- Are there statistically significant differences between the mean scores of the experimental group and the control group on the components of the Attitude Toward Sustainable Development Scale and their total score after implementing the program?
- Are there statistically significant differences between the mean scores of the experimental group on the components of the Attitude Toward Sustainable Development Scale and their total score in the pre- and post-tests?
- Are there statistically significant differences between the mean scores of the experimental group on the components of the Attitude Toward Sustainable Development Scale and their total score in the post- and follow-up tests?

**Methodology**

**Research Design**

In their effort to collect data, the researchers used the quasi-experimental method to identify the effectiveness of using an Artificial Intelligence applications-based training program (independent variable) on developing Saudi undergraduate students' English language skills and enhancing their attitudes toward sustainable development (dependent variables) under conditions in which the researchers controlled the intervening factors in the two groups (experimental and control) that could affect the dependent variables.

**Sample of the Study**

During the second semester of the 2024/2025 academic year, the first-year students at the Applied College, Prince Sattam bin Abdulaziz University in Wadi Al-Dawasir Governorate, Kingdom of Saudi Arabia, participated in this study. As a first step, the sample was purposefully selected by the researchers. The number of participants in the pilot study included 120 students at the Applied College, with an average age of (19) years and a standard deviation of (1.25); this was to ensure the psychometric properties of the study instruments.

The second step was selecting the main study participants in which the researchers chose 23 students, based on their academic achievement in English classes, from the Applied College in order to test the research hypotheses. They included 12 students majoring in computer science and programming (experimental group) and 11 students majoring in information security (control group). In terms of ethical considerations, the participants were briefed on the purpose of the study prior to its commencement, and their informed consents were obtained.

The third step involved controlling intervening variables that could potentially affect how the independent variable (the training program) affects the dependent variables (English language skills and the attitudes toward sustainable development). The study was restricted to male students solely in order to eliminate the influence

of gender, one of the intervening variables that the researchers believe might have an impact on the dependent variables. An additional intervening variable is the trainer, a researcher who provided training to the experimental group on how to use artificial intelligence applications. Prior to the study, it was also confirmed that the two groups were equivalent in terms of age, English language skills, and attitudes toward sustainable development. The results showed that there were no statistically significant differences between the average scores of the experimental and control groups in the pre-test for each of (age, English language skills and its components, and the attitudes toward sustainable development and its components), which indicates the equivalence of the two groups in the pre-test.

**Instruments**

1- **The Training Program**

prepared by the researchers addressed the following topics:

*Basics of the Training Program:* The program's content was modified to consider the characteristics, attitudes, and general issues of the participants that impact on their mental well-being and thought patterns. Open content, such as scenarios, exercises, and real-world issues, was incorporated into the training program to encourage students to look outside the confines of university courses. The training program was developed by consulting prior relevant studies and psychological frameworks on artificial intelligence applications.

The program has both a general aim that it seeks to achieve, and procedural behavioral objectives for each session. These procedural objectives must be integrated to achieve the overall aim of the program.

There is a tight relationship between every component of each program session, including the session objectives, subject matter, skills employed, duration, exercises, and evaluation instruments.

Promoting positivity among participants by actively involving every student in the program, offering stimuli, activities, and tasks that generate excitement and energy, preventing boredom by creating session content on PowerPoint slides, teaching students through cooperative learning techniques, holding contests, presenting puzzles, and having participants act out scenarios in a theatrical presentation.

To guarantee stability and improvement, each component undergoes multiple training sessions.

By employing a range of strategies, the sessions motivate students to actively participate, reduce boredom, and accomplish the objectives of each session and, ultimately, the overall aim of the program as a whole.

To make sure students understand the program's activities, formative assessment is used throughout training sessions. Summative assessment is used at the conclusion of the program using an evaluation form tailored to the objectives of each session.

Creating special sessions for each Artificial Intelligence application to teach participants about each mechanism and provide them with an overview of it at the beginning of the program.

The program consists of three sections: knowledge, training, and practical application. Information about artificial intelligence applications is given, followed by sessions in which participants receive training on these applications and, finally, practice using these skills.

**Program Preparing Steps:**

The following steps were taken in order to prepare this program for the current study:

Identifying the program's objectives: The researchers used the definition of artificial intelligence applications to identify the program's objectives. The content, approaches, strategies, and evaluation procedures were selected in accordance with these objectives. After deciding on the program's overarching aim, the sub-objectives, followed by procedural objectives for each session were defined.

The general aim of the program: The program includes a general aim and specific procedural objectives for each session. The general aim of the program is "to train university students on artificial intelligence applications in order to develop their English language skills and promote their attitudes toward sustainable development." The program's objectives focus on digital citizenship skills.

- Program Content: with an emphasis on sustainable development, the program covers topics and skills essential for university students with low English language skills.
- The program's training techniques include lecture, dialogue and discussion, cooperative learning,

brainstorming, direct and symbolic modeling, self-assessment of ideas and behavior, role-playing, conscious work, virtual thinking, gradual reinforcement, positive self-talk, alternative methods, individual and group exercises through offering questions like: What would you do in the following situations?

– Program duration and number of sessions: The program was distributed over ten weeks, during which 20 training sessions (each lasting 60 minutes) were held, two sessions per week, gradually ranging from easy to difficult, in addition to one session for each of the pre-test, post-test, and follow-up tests. The following table illustrates the content of the training program.

**Summary of Training Program Sessions**

1. Introduction Session: Orientation, rapport building, and clarifying the role of AI in English learning and sustainability.
2. Reading with ChatGPT: Enhance reading comprehension on sustainability through interactive dialogues.
3. Writing with ChatGPT: Develop structured paragraph writing on sustainability topics.
4. Basic Error Correction with Grammarly: Improve writing accuracy by identifying and fixing grammar and spelling mistakes.
5. Style Improvement with Grammarly: Refine writing style for coherence and clarity in academic texts.
6. Pronunciation with ELSA Speak: Correct English pronunciation and identify individual articulation challenges.
7. Recording Conversations with ELSA Speak: Build fluency through recording and self-assessment of spoken English.
8. Listening with YouGlish: Develop listening skills using real-life English examples from authentic videos.
9. Listening Comprehension & Vocabulary with YouGlish: Expand vocabulary and understand usage contexts.
10. Virtual Conversation with ChatGPT: Practice spontaneous speaking on sustainability to build confidence.
11. Debate with ChatGPT: Develop argumentation and persuasion skills on environmental topics.
12. Writing with Write & Improve: Write paragraphs and receive feedback to enhance structure.
13. Writing Improvement Cycles: Engage in iterative writing-review cycles for measurable progress.
14. Summarizing & Paraphrasing with ChatGPT: Practice critical reading, summarizing, and rephrasing texts.
15. Integration of Grammarly & Write & Improve: Combine tools for advanced writing correction and refinement.
16. Fluency Practice with ELSA Speak & YouGlish: Integrate pronunciation and listening for natural intonation.
17. Analyzing Texts with ChatGPT: Develop analytical reading of academic texts on sustainability.
18. Pronunciation Shadowing with YouGlish: Improve pronunciation by mimicking authentic speech.
19. Integrated Project: Use all tools to create a comprehensive presentation or article on sustainability.
20. Closing Session: Conduct post-tests, gather feedback, and evaluate the overall training impact.

**Program Evaluation**: The program was evaluated in light of the foundations and philosophy on which it is based. The program evaluation procedures include a number of steps, which are as follows:

Pre-test: A pre-test was conducted on the experimental group members to measure the skills of using artificial intelligence applications before applying the program.

Formative assessment: Using a session objectives assessment form that was created by the researchers and given out at the conclusion of each program session, the researchers tracked program progress during the sessions. In order to assess the session's level of benefit, the program's progress, and the necessity of avoiding and resolving program flaws, the researchers also depended on talking about homework assignments.

Summative assessment: after the completion of the program, the summative assessment of the program sessions was conducted through:

– Conducting a post-test measurement of the skills in using artificial intelligence applications on the members of the experimental group.
– Homework: Students answered the questions at home.

- Applying experimental procedures' effectiveness scale.
- Conducting a follow-up test after one month on the use of AI application skills among the experimental group to verify the retention of learning impact after training.

**Program validity:** The program was presented to a panel of five distinguished referees, who were professors of educational psychology and educational technology, in order to assess the training program's procedures, goals, and content, as well as to get their feedback and determine whether the assessment was appropriate for the session's goals. Changes were done appropriately based on the observations and recommendations gathered. Once the referees' judgments were reviewed, it was evident that the program was suitable for its intended use. Their views validated that the arbitration components were appropriate for the program, and the referees' agreement on the arbitration components ranged from 80% to 100%. Additionally, the researchers conducted three program sessions with five students enrolled in exploratory study to make sure that the aims, activities, approaches, and content of the program were appropriate for the participants and that the session time was suitable.

2- **English Language Skills Test**

To assess the targeted English language skills among undergraduate students, a standardized English language skills test was designed and validated following rigorous scientific procedures.

Test Construction: The test focused on the four core language skills: listening, speaking, reading, and writing. Its items were developed based on the approved content of university-level English language courses and aligned with the Common European Framework of Reference for Languages (CEFR).

Test Items: The final version included twenty items with a balanced representation of language skills: five items measured reading comprehension, five assessed listening skills (accompanied by audio recordings), five evaluated grammar and vocabulary, and five addressed writing and speaking abilities through task-based prompts.

Expert Review: A panel of five experts in English curricula and language assessment reviewed the test to ensure linguistic accuracy, relevance, and alignment with educational objectives and students' proficiency levels. Pilot Testing and Psychometric Analysis: A pilot administration was conducted with a sample of 30 students to examine the psychometric properties. The results demonstrated robust validity and reliability indicators:

- Face validity: The expert panel agreed by 94% on item appropriateness.
- Construct validity: Exploratory factor analysis confirmed four distinct factors representing the language skills, with factor loadings ranging from 0.54 to 0.81.
- Reliability: Cronbach's alpha coefficient for the full test was 0.87. Subscale reliabilities were: reading (0.84), listening (0.81), writing (0.85), and speaking (0.79).
- Difficulty and Discrimination: Difficulty indices ranged from 0.31 to 0.74; items outside the acceptable range were revised or omitted. Discrimination indices ranged between 0.36 and 0.71, within acceptable standards.
- Test Duration: The average completion time was 50 minutes; therefore, 60 minutes was allocated to ensure sufficient time for open-ended tasks.

The statistical findings confirmed that the test possesses robust psychometric qualities and can be reliably used to measure university students' English language proficiency across the four essential skills.

3- **Scale of Attitudes Toward Sustainable Development**

A dedicated scale was developed to measure students' attitudes toward sustainable development, drawing on established theoretical models and prior empirical studies.

Scale Design: The scale encompassed four key domains: environmental, social, economic, and educational. It consisted of 20 positively worded items, evenly distributed (five items per domain). Responses were recorded on a five-point Likert scale ranging from "strongly agree" (5) to "strongly disagree" (1). Higher scores reflected stronger positive attitudes toward sustainable development. Total possible scores ranged from 20 to 100 points.

Content Validity: Content validity was confirmed through review by five faculty experts specializing in educational psychology. The reviewers' agreement rates ranged from 80% to 100%, indicating high relevance and clarity of the items.

Internal Consistency: Item-total and domain-total correlations were calculated using Pearson's correlation coefficient for a sample of 120 students. All coefficients were statistically significant at the 0.01 level, with values ranging from 0.511 to 0.890, demonstrating strong internal consistency.

Scale Reliability: Reliability was further ensured using both Cronbach's alpha coefficient and the Guttman split-half method. The alpha coefficient for the entire scale was 0.765, while the split-half reliability coefficient was 0.818. Both exceeded the acceptable threshold of 0.70, confirming the scale's reliability for measuring attitudes toward sustainable development among university students.

**Results**

**Result of the first hypothesis:** The first hypothesis states that "there are statistically significant differences between the mean ranks of the scores of the experimental and control group members in the components of the English language skills test and the total score after applying the program in favor of the experimental group." To test this hypothesis, the researchers used the Mann-Whitney test for significance of differences, and the following table shows the results.

Table 1. Differences between Control and Experimental Group in Post-measurement of English language skills (Mann-Whitney Test)

| English language skills | Group | N | Mean of ranks | Sum of ranks | U | Z | Sig. | Effect Size «r»[1*] |
|---|---|---|---|---|---|---|---|---|
| Listening | Control | 11 | 6.00 | 66.00 | 0 | 4.11 | <0.001 | 0.858 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |
| Speaking | Control | 11 | 6.00 | 66.00 | 0 | 4.01 | <0.001 | 0.839 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |
| Reading | Control | 11 | 6.00 | 66.00 | 0 | 4.02 | <0.001 | 0.839 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |
| Writing | Control | 11 | 6.00 | 66.00 | 0 | 4.20 | <0.001 | 0.876 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |
| Total Score | Control | 11 | 6.00 | 66.00 | 0 | 4.09 | <0.001 | 0.853 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |

It is evident from the previous table that there are statistically significant differences at the 0.01 significance level between the mean ranks of the experimental and control groups in the components of the English language skills test and the overall score after the program was applied, in favor of the experimental group. This indicates the validity of the first hypothesis, as the mean ranks of the experimental group were higher than those of the control group, confirming a statistically significant higher level of English language skills in the experimental group compared to the control group.

To ensure the effectiveness of the training program used in developing English language skills and its components, the researcher calculated the effect size of the program using the R effect size formula for the Mann–Whitney U test: $r=Z/\sqrt{(n_1+n_2)}$, Where: Z is the test statistic value (Mann-Whitney), n_1 is the number of individuals in the first group, n_2 is the number of individuals in the second group. The effect size is considered large if it is greater than or equal to (0.5), while if the value is between (0.3, 0.5), the effect size is "medium." If the value is less than or equal to (0.1), the effect size is "weak" (Pallant, 2016, 229). Looking at the effect size values in the previous table, it is clear that they are large. This means that the members of the experimental group showed an increase in their English language skills and components after the program

1 * Mann–Whitney U test: , Where r is the ES estimate for the Mann–Whitney U test, Z the Z score produced by SPSS, the number of observations for first group, and the number of observations for second group on which Z is based. The interpretation of ES relied on the benchmarks proposed by Pallant (2016, 229). small, r = 0.1; medium, r = 0.3; large, r = 0.5.

was implemented compared to the control group, which confirms the effectiveness of the training program in enhancing English language skills.

**Result of the second hypothesis:** The second hypothesis states that "there are statistically significant differences between the average ranks of the scores of the individuals in the experimental and control groups in the components of the scale of attitudes towards sustainable development and the total score after implementing the program in favor of the experimental group." To test this hypothesis, the researcher used the "Mann-Whitney" test to determine the significance of the differences. The following table shows the results.

Table 2. Differences between Control and Experimental Groups in Post-measurement of attitudes toward sustainable development (Mann-Whitney Test)

| Attitudes toward sustainable development | Group | N | Mean of ranks | Sum of ranks | U | Z | Sig. | Effect Size «r»[2*] |
|---|---|---|---|---|---|---|---|---|
| Environmental field | Control | 11 | 6.00 | 66.00 | 0 | 4.20 | <0.001 | 0.876 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |
| social field | Control | 11 | 6.00 | 66.00 | 0 | 4.01 | <0.001 | 0.837 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |
| Economic field | Control | 11 | 6.00 | 66.00 | 0 | 4.13 | <0.001 | 0.862 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |
| Education Field | Control | 11 | 6.00 | 66.00 | 0 | 4.21 | <0.001 | 0.878 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |
| Total Score | Control | 11 | 6.00 | 66.00 | 0 | 4.08 | <0.001 | 0.851 |
| | Experimental | 12 | 17.50 | 210.00 | | | | |

The previous table shows that there are statistically significant differences at a significance level of (0.01) between the average ranks of the scores of the members of the experimental and control groups in the components of the scale of attitudes towards sustainable development and the total score after applying the program in favor of the experimental group. This means that the second hypothesis is validated; as the average ranks of the experimental group were higher than the average ranks of the control group, which confirms the higher level of attitudes towards sustainable development in the experimental group than in the control group by a statistically significant difference.

To verify the effectiveness of the training program used in developing attitudes towards sustainable development and its components, the researchers calculated the effect size of the program using the R equation for the effect size mentioned above. Looking at the effect size values in the previous table, it is clear that they are high. This means that the members of the experimental group had a higher level of attitude towards sustainable development and its components after implementing the program compared to the control group. This confirms the effectiveness of the training program in developing attitudes towards sustainable development.

**Result of the third hypothesis:** The third hypothesis states that "there are statistically significant differences between the average ranks of the scores of the experimental group members in the components of the English language skills test and the total score in the pre- and post-tests in favor of the post-test." To test this hypothesis, the researchers used the Wilcoxon test to determine the significance of the differences, and the following table shows the results.

2

Table 3. Differences between the post and follow-up measurements of experimental group in English language skills (Wilcoxon Signed Ranks Test)

| English language skills | Ranks | *N* | Mean of ranks | Sum of ranks | *Z* value | Sig. | Effect Size «r»[3*] |
|---|---|---|---|---|---|---|---|
| Listening | Negative ranks | 0 | .00 | .00 | 3.09 | <0.001 | 0.658 |
| | Positive ranks | 12 | 6.50 | 78.00 | | | |
| | Ties | 0 | | | | | |
| Speaking | Negative ranks | 0 | .00 | .00 | 3.08 | <0.001 | 0.656 |
| | Positive ranks | 12 | 6.50 | 78.00 | | | |
| | Ties | 0 | | | | | |
| Reading | Negative ranks | 0 | .00 | .00 | 3.07 | <0.001 | 0.654 |
| | Positive ranks | 12 | 6.50 | 78.00 | | | |
| | Ties | 0 | | | | | |
| Writing | Negative ranks | 0 | .00 | .00 | 3.10 | <0.001 | 0.660 |
| | Positive ranks | 12 | 6.50 | 78.00 | | | |
| | Ties | 0 | | | | | |
| Total Score | Negative ranks | 0 | .00 | .00 | 3.07 | <0.001 | 0.654 |
| | Positive ranks | 12 | 6.50 | 78.00 | | | |
| | Ties | 0 | | | | | |

It is clear from the previous table that all (z) values are statistically significant at the 0.01 level, which means there are statistically significant differences between the mean ranks of the experimental group members' scores on the English language skills test and its components in the pre-test and post-test, in favor of the post-test; as the mean of the positive ranks (post-test) is greater than the mean of the negative ranks (pre-test), which confirms the validity of the third hypothesis. This is an indication of the effectiveness of the training program used in developing English language skills and its components among the experimental group members.

To ensure the effectiveness of the training program used in developing English language skills and its components, the researchers calculated the effect size of the program using the Wilcoxon signed-rank test effect size formula: $r=Z/\sqrt{(2\times N)}$, where Z is the test statistic value (Wilcoxon), and N is the number of participants (Pallant, 2016, 229). Looking at the effect size values in the previous table, it is clear that they are large, which means that the members of the experimental group showed an increase in their level of English language skills and components after the implementation of the program compared to the pre-implementation, confirming the effectiveness of the training program in developing English language skills and components.

**Result of the fourth hypothesis:** The fourth hypothesis states that "there are statistically significant differences between the mean ranks of the experimental group members' scores in the components of the scale of attitudes towards sustainable development and the total score in the pre-test and post-test, in favor of the post-test." To test this hypothesis, the researchers used the Wilcoxon test for significance of differences, and the following table illustrates the results.

3

Table 4. Differences between the post and follow-up measurements of experimental group in attitudes toward sustainable development (Wilcoxon Signed Ranks Test)

| Attitudes toward sustainable development | Ranks | *N* | Mean of ranks | Sum of ranks | *Z* value | Sig. | Effect Size «r»[4*] |
|---|---|---|---|---|---|---|---|
| Environmental field | Negative ranks | 0 | .00 | .00 | 3.09 | <0.001 | 0.658 |
| | Positive ranks | 12 | 6.50 | 78.00 | | | |
| | Ties | 0 | | | | | |
| social field | Negative ranks | 0 | 0 | 0 | 3.10 | <0.001 | 0.660 |
| | Positive ranks | 12 | 7.00 | 91.00 | | | |
| | Ties | 0 | | | | | |
| Economic field | Negative ranks | 0 | 0 | 0 | 3.07 | <0.001 | 0.654 |
| | Positive ranks | 12 | 7.00 | 91.00 | | | |
| | Ties | 0 | | | | | |
| Education Field | Negative ranks | 0 | 0 | 0 | 3.10 | <0.001 | 0.660 |
| | Positive ranks | 12 | 7.00 | 91.00 | | | |
| | Ties | 0 | | | | | |
| Total Score | Negative ranks | 0 | 0 | 0 | 3.06 | <0.001 | 0.652 |
| | Positive ranks | 12 | 7.00 | 91.00 | | | |
| | Ties | 0 | | | | | |

It is clear from the previous table that all (z) values are statistically significant at the 0.01 level, which means there are statistically significant differences between the mean ranks of the experimental group members' scores on the scale of attitudes towards sustainable development and its components in the pre-test and post-test, in favor of the post-test; as the mean ranks of the positive ranks (post-test) are greater than the mean ranks of the negative ranks (pre-test). This confirms the validity of the fourth hypothesis. This indicates the effectiveness of the training program used in developing the attitudes towards sustainable development and its components among the experimental group members.

To ensure the effectiveness of the training program used in developing the attitude towards sustainable development and its components, the researcher calculated the effect size of the program using the previous R equation for effect size. Looking at the effect size values in the previous table, it is clear that they are large. This means that the experimental group members showed an increased level of attitude towards sustainable development and its components after the implementation of the program compared to the pre-implementation. This confirms the effectiveness of the training program in developing the attitude towards sustainable development and its components.

**Result of the fifth hypothesis:** The fifth hypothesis states that "there are no statistically significant differences between the mean ranks of the scores of the experimental group members in the components of the English language skills test and the overall score in the post-test and follow-up test (after one month)." To test this hypothesis, the researchers used the Wilcoxon test for significance of differences. The following table illustrates the results.

Table 5. Differences between Post and Follow-up Measurements of Experimental group in English language skills (Wilcoxon Signed Ranks Test)

| English language skills | Ranks | *N* | Mean of ranks | Sum of ranks | *Z* value | Sig. |
|---|---|---|---|---|---|---|

4

| | | | | | | |
|---|---|---|---|---|---|---|
| Listening | Negative ranks | 1 | 1.00 | 1.00 | 1.000 | 0.317 |
| | Positive ranks | 0 | .00 | .00 | | |
| | Ties | 11 | | | | |
| Speaking | Negative ranks | 0 | .00 | .00 | 0.000 | 1.000 |
| | Positive ranks | 0 | .00 | .00 | | |
| | Ties | 12 | | | | |
| Reading | Negative ranks | 1 | 1.00 | 1.00 | 1.000 | 0.317 |
| | Positive ranks | 0 | .00 | .00 | | |
| | Ties | 11 | | | | |
| Writing | Negative ranks | 0 | .00 | .00 | 1.000 | 0.317 |
| | Positive ranks | 1 | 1.00 | 1.00 | | |
| | Ties | 11 | | | | |
| Total Score | Negative ranks | 2 | 2.00 | 4.00 | 0.577 | 0.564 |
| | Positive ranks | 1 | 2.00 | 2.00 | | |
| | Ties | 9 | | | | |

It is clear from the previous table that there are no statistically significant differences between the mean ranks of the experimental group members' scores on the English language skills test and its components in the post-test and follow-up test (after one month). This means that the fifth hypothesis is valid, as all values of (z) were not statistically significant, as they were all greater than (0.05).

**Result of the sixth hypothesis:** The sixth hypothesis states that "there are no statistically significant differences between the mean ranks of the experimental group members' scores in the components of the scale of the attitudes towards the sustainable development and the overall score in the post-test and follow-up test (after one month)." To test this hypothesis, the researchers used the Wilcoxon test for significance of differences. The following table illustrates the results.

Table 6. Differences between Post and Follow-up Measurements of Experimental group in Attitudes toward sustainable development (Wilcoxon Signed Ranks Test)

| Attitudes toward sustainable development scale | Ranks | *N* | Mean of ranks | Sum of ranks | *Z* value | Sig. |
|---|---|---|---|---|---|---|
| Environmental field | Negative ranks | 1 | 1.00 | 1.00 | 1.000 | 0.317 |
| | Positive ranks | 0 | .00 | .00 | | |
| | Ties | 11 | | | | |
| social field | Negative ranks | 1 | 1.00 | 1.00 | 1.000 | 0.317 |
| | Positive ranks | 0 | .00 | .00 | | |
| | Ties | 11 | | | | |
| Economic field | Negative ranks | 1 | 1.00 | 1.00 | 1.000 | 0.317 |
| | Positive ranks | 0 | .00 | .00 | | |
| | Ties | 11 | | | | |
| Education Field | Negative ranks | 0 | .00 | .00 | 1.000 | 0.317 |
| | Positive ranks | 1 | 1.00 | 1.00 | | |
| | Ties | 11 | | | | |
| Total Score | Negative ranks | 2 | 2.25 | 4.50 | 0.816 | 0.414 |
| | Positive ranks | 1 | 1.50 | 1.50 | | |
| | Ties | 9 | | | | |

It is clear from the previous table that there are no statistically significant differences between the mean ranks of the experimental group members' scores on the sustainable development attitude scale and its components in the post-test and follow-up test (after one month). This means that the sixth hypothesis is valid, as all values of (z) were not statistically significant, as they were all greater than (0.05).

**Discussion**

This research investigated the effectiveness of using artificial intelligence applications in developing Saudi undergraduate students' English language skills and enhancing their attitudes toward sustainable development. By verification of the validity of the first hypothesis, it was noted that the members of the experimental group showed an increase in their English language skills and components after the program was implemented compared to the control group, which confirms the effectiveness of the training program in enhancing English language skills. This result aligns with earlier findings of prior studies, like Bilad et al., (2021), who confirmed that the impact of AI on learning outcomes is evident in the enhanced proficiency of students across various language skills. In the same vein, Biletska et al., (2021) asserted that through interactive exercises, instant feedback, and adaptive content, AI empowers students to improve their speaking, listening, reading, and writing abilities.

The second hypothesis was verified, and the results showed that the members of the experimental group had a higher level of attitude towards sustainable development and its components after implementing the program compared to the control group. This finding aligns with Bachmann et al., (2022), who emphasized that while maintaining respect for human dignity and security, international communities are quickly putting themselves in a position to guarantee the application and role of AI in creating inclusive knowledge societies and helping nations achieve the 17 Sustainable Development Goals.

The third hypothesis' verification revealed that there were statistically significant differences between the average ranks of the scores of the experimental group members in the components of the English language skills test and the total score in the pre- and post-tests in favor of the post-test. Interestingly, this hypothesis is supported by this result which is consistent with a similar trend in a meta-analysis by Tajik (2025), who assured that students who used AI-supported platforms proved faster improvement in their writing skills than those in traditional classrooms.

The result of the verification of the fourth hypothesis indicated that there were statistically significant differences between the mean ranks of the experimental group members' scores in the components of the scale of attitudes towards sustainable development and the total score in the pre-test and post-test, in favor of the post-test. This aligns with both Wei, (2023); and Kamalov et al., (2023), who found that one of the most profound shifts facilitated by AI is the move toward student-centered learning. By empowering learners to set their pace, explore their attitudes, and take ownership of their learning journey, AI aligns with the principles of learner agency and autonomy. This approach resonates with the educational ideals of SDG 4, which calls for a learner-centric approach to education.

The result of the verification of the fifth hypothesis showed that there were no statistically significant differences between the mean ranks of the scores of the experimental group members in the components of the English language skills test and the overall score in the post-test and follow-up test (after one month). This aligns with Pokrivcakova (2019) who noted that both instructors and learners of foreign languages could use a diversity of AI-powered applications to simplify their teaching and learning processes. She asserted that AI integration in foreign language education offers students quick, and individualized assistance.

The result of the verification of the sixth hypothesis stated that there were no statistically significant differences between the mean ranks of the experimental group members' scores in the components of the scale of the attitudes towards sustainable development and the overall score in the post-test and follow-up test (after one month)." This aligns with the findings of the Bachmann et al., (2022) above-mentioned study.

**Conclusion**

This study set out to investigate the effectiveness of using artificial intelligence applications in developing Saudi undergraduate students' English language skills and enhancing their attitudes toward sustainable development. The study involved 23 first-year students from Prince Sattam bin Abdulaziz University's Applied College,

including 12 in the experimental group and 11 in the control group. An AI-based training program, a test of English language proficiency, and a scale measuring attitudes toward sustainable development were among the tools the researchers employed. The study results showed that there were statistically significant differences between the average ranks of the scores of students in the experimental and control groups in both the English language skills test and the scale of attitudes toward sustainable development after implementing the training program, favoring the experimental group. This was revealed by employing a quasi-experimental approach and appropriate statistical treatments, such as the Mann-Whitney and Wilcoxon tests.

The results of the study showed that, following the implementation of the training program, the experimental group's average ranks of scores on the scale of attitudes toward sustainable development and the English language skills test differed statistically significantly from those of the control group. The results also demonstrated that the experimental group's pre-test and post-test measurements differed statistically significantly in favor of the post-test, and that there were no significant differences between the post-test and follow-up tests one month later, suggesting that the program's impact persisted. The usefulness of deploying AI apps to improve Saudi undergraduate students' English language skills and attitudes toward sustainable development was thoroughly examined by the researchers. Lastly, they offered a number of suggestions and recommendations for future research.

**Recommendations**

The results of this study may have broad implications and recommendations for practice and research in a number of situations and for a large number of stakeholders. Researchers and practitioners in higher education institutions must be interested in the possible factors that motivate students to use AI applications for academic purposes, given the surge in the use of these tools by college students. As AI applications continue to permeate every part of our lives, educational institutions must now more than ever develop policies and procedures that maximize the benefits of compulsive AI use while reducing its drawbacks. The results of this study provide a foundation for future research on the efficacy of utilising AI apps to enhance students' academic performance in different nations and educational settings. The results of this study showed that individual motivation for utilizing AI applications for educational purposes is more significant than demographic characteristics like age and gender, which has various practical ramifications for AI practitioners and other stakeholders. Teachers should talk about the expectations for using AI applications in the classroom and how they can be used academically as part of their open communication with students.

**Limitations**

The quasi-experimental data collection method used to determine the impact of using artificial intelligence applications on developing the English language skills of Saudi undergraduate students and enhancing their attitudes toward sustainable development is one of the study's limitations. Using other direct measures and qualitative research methods such as observations, interviews, and experimental methods to evaluate the efficacy of using AI applications in higher education institutions could help overcome these limitations and test the findings presented in this paper in other higher education contexts. Utilizing the triangulation approach, students' semi-structured and focus group interviews may provide more detailed information that can be utilized to validate the study's findings. Furthermore, the sample size was small in relation to the overall number of AI application users in other Saudi universities, even if the data was gathered from a reasonable population of AI application users at just one Saudi university. It is possible that the sample does not accurately reflect the total number of Saudi Arabian students. Furthermore, future research should aim for a more balanced participant gender distribution because of the study sample's overrepresentation of male participants (00.0%). Finally, neither individual differences nor learning styles were used to categorize students in this study. Future research should clarify how different learning styles and AI application-based teaching methods relate to one another. When implementing the findings in new higher education contexts, these limitations need to be considered.

**Funding**

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/02/33067)

**References**


Adriansen, H. K., Juul-Wiese, T., Madsen, L. M., Saarinen, T., Spangler, V., & Waters, J. L. (2022). Emplacing English as lingua franca in international higher education: A spatial perspective on linguistic diversity. *Population, Space and Place*, *29*(2), e2619. https://doi.org/10.1002/psp.2619

Al-Smadi, O. A., Rashid, R.A., & Altamimi, D.H. (2020). A linguistic ethnography of teacher talk in an English for Medical Purposes Classroom. *International Journal of Arabic-English Studies*, *20*(1), 51-66. DOI:10.6007/ijarped/v6-i3/3169

Bachmann, N., Tripathi, S., Brunner, M., & Jodlbauer, H. (2022). The contribution of data-driven technologies in achieving the sustainable development goals. *Sustainability, 14*(5), 2497. https://doi.org/10.3390/su14052497

Bilad, M. R., Yaqin, L. N., & Zubaidah, S. (2023). Recent Progress in the Use of Artificial Intelligence Tools in Education. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika*, *7*(3), 279–315. https://doi.org/10.36312/esaintika.v7i3.1377

Biletska, I. O., Paladieva, A. F., Avchinnikova, H. D., & Kazak, Y. Y. (2021). The use of modern technologies by foreign language teachers: developing digital skills. *Linguistics and Culture Review*, *5*(S2), 16-27. https://doi.org/10.21744/lingcure.v5nS2.1327

Chen, J. (2020). On the Application of Artificial Intelligence Technology in Foreign Language Teaching. *Journal of Beijing International Studies University*, *42*(2), 14. https://doi.org/10.12002/j.bisu.272

Chen, Tzeng-Ji. 2023. "ChatGPT and Other Artificial Intelligence Applications Speed up Scientific Writing." *Journal of the Chinese Medical Association* Publish Ah:351–53. doi: 10.1097/jcma.0000000000000900.

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Davis, F.D. (1989). Perceived Usefulness, Perceived Ease Of Use, And User Acceptance. MIS Quarterly, 13(3), 319-340.

Holmes, W., & Miao, F. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing.

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jiang, R. (2022). How does artificial intelligence empower EFL teaching and learning nowadays? A review on artificial intelligence in the EFL context. *Frontiers in Psychology*, *13*, 1049401. https://doi.org/10.3389/fpsyg.2022.1049401

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

Karataş, F., Abedi, F.Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12574-6

Lamouchi, A., Mohamed Sayed Abdellatif, & Hafez, W. E. (2024). Linguistic accommodation strategies in Non-Arabic speaking doctors' interactions with patients in Saudi hospitals and its role in enhancing healthcare for them. *Research Journal in Advanced Humanities*, *5*(4). https://doi.org/10.58256/j2whs893

Li, R. (2020). Using Artificial Intelligence in Learning English as a Foreign Language: An Examination of IELTS LIULISHUO as an Online Platform. *Journal of Higher Education Research*, *1*(2), 85-89. https://doi.org/10.32629/jher.v1i2.178

Lu, Xiaoxia. 2019. "An Empirical Study on the Artificial Intelligence Writing Evaluation System in China CET." *Big Data* 7(2):121–29. doi:10.1089/big.2018.0151.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Mali, Yustinus Calvin Gai. 2022. "Teaching Reflection of Using Technology in Two Indonesian EFL Classrooms: An Autoethnography." *Journal on English as a Foreign Language* 12(1):1–21. https://doi.10.23971/jefl.

v12i1.2860 .
Moulieswaran, N., & NS, P. K. (2023). Investigating ESL Learners' Perception and Problem towards Artificial Intelligence (AI)-Assisted English Language Learning and Teaching. *World Journal of English Language*, *13*(5), 290-290. https://doi.org/10.5430/wjel.v13n5p290
Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge. *ng Psychology, 54*(3),295-305.
Pokrivcakova, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education*, *7*(3), 135-153. http://dx.doi.org/10.2478/jolace-2019-0025
Rahman, Nurul Ajleaa Abdul, Luqmanul Hakim Zulkornain, and Nurul Huda Hamzah. 2022. "Exploring Artificial Intelligence Using Automated Writing Evaluation for Writing Skills." *Environment-Behaviour Proceedings Journal* 9(SI9):547–53. doi: https://doi.org/10.21834/ebpj.v7iSI9.4304.
Ribeiro, Jair. 2021. *A.I. in 2020: A Year Writing about Artificial Intelligence*. Independently Published.
Setyani, Erin Dwi, Eusabinus Bunau, and Yanti Sri Rezeki. 2023. "The Influence of Grammarly towards Indonesian EFL Students' First-Degree Thesis Writing Confidence." *Elsya : Journal of English Language Studies* 5(1):54–67. https://doi.10.31849/elsya.v5i1.6773.
Tajik, A. (2025). Exploring the role of AI-driven dynamic writing platforms in improving EFL learners' writing skills and fostering their motivation. https://doi.org/10.21203/rs.3.rs-5788599/v1
Venkatesh, V. & Davis, F.D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), 186-204.
Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese