www.jrspelt.com

Journal for Research Scholars and Professionals of English Language Teaching

ISSN: 2456-8104
Vol. 8 | Issue 44 | July 2024

Impact Factor: 6.895 (SJIF)
www.jrspelt.com

AI Assistance in Language Education: AI-detection Accuracy and Students' Vocabulary Retention

Yuliyan Zhelyazkov (yulizhelyazkov@gmail.com), : https://orcid.org/0009-0002-4083-167X
Ph.D. Research Scholar, Universidad de Las Palmas de Gran Canaria, 35001 Las Palmas, Spain

Crossref/DOI: https://doi.org/10.54850/jrspelt.8.44.001

Abstract: *The emergence of Artificial Intelligence (AI) presents potential benefits and challenges in writing instruction. Thus, the adoption of AI-detection tools is necessary to preserve academic integrity. However, challenges persist regarding false positives and detecting AI-generated content combined with human-written text or paraphrasing. This study aims to evaluate the reliability and accuracy of Scribbr, an AI detection tool, and to assess the effectiveness of ChatGPT in facilitating language learning activities, specifically in the context of writing formal letters of application. To achieve these objectives, a research methodology comprising three main components was employed. The reliability and accuracy of Scribbr were assessed in differentiating between AI-generated and human-written texts. Students' writing outcomes were analyzed and compared, both qualitatively and quantitatively, to evaluate the impact of AI-generated content on writing proficiency. The results reveal the limitations of Scribbr due to false positive results, i.e. flagging human-written texts as AI-generated. The study also suggests that while some students successfully learned and applied the conventions of the formal application letter genre, others exhibited challenges in fully retaining the new vocabulary introduced through the AI-generated content. In conclusion, this research calls for a balanced approach to AI integration in education.*

Keywords: ChatGPT, Critical Thinking, EAP, GAI, Language Education, TEFL, Vocabulary, Writing Skills

Article History: **Received: 30 June -2024;** **Accepted: 15 July- 2024;** **Published/Available Online: 30 July- 2024**

1. Introduction

Throughout history, the evolution of education has been profoundly influenced by pivotal historical events and technological advancements. In the past, a few such milestones that come to mind, the invention of the printing press in the fifteenth century, the advent of the Industrial Revolution in the eighteenth and nineteenth centuries, and the invention of the computer and the Internet along with the World Wide Web in the last century, dramatically reshaped the educational landscape.

Today, as we face another technological revolution with the emergence of Generative Artificial Intelligence (GAI), there exists a profound opportunity to redefine the educational landscape once again. GAI, with its capacity for data analysis, pattern recognition, and machine learning, holds promise for revolutionising educational practices, from personalised learning experiences to enhanced assessment methodologies.

Scholars believe that AI can help students improve their writing skills (Wei, 2023; Al Mahmud, 2023; Arora et al., 2023; Nugroho et al., 2023; Bakai et al., 2023). Their studies discuss the enhancement of writing skills and language learning through learning using AI tools and integrated approaches to language instruction. For instance, participants in Wei's study attributed significant advancements in English language proficiency to AI-mediated instruction, citing improvements in writing, among other skills (8). Notably, they observed that these improvements correlated with higher grades and increased academic confidence (ibid.). Al Mahmud's study revealed that using Wordtune, an AI-powered digital writing tool, led to syntactic gains in students' writing, such as increased phrasal complexity and the use of conjoined and embedded clauses (1402). The same study also recommended the integration of AI-powered digital writing tools into writing classrooms to provide instant feedback, promote self-directed learning, and increase student engagement with writing (ibid.). Nevertheless, Al Mahmud warns of a potential uncritical reliance on digital writing tools that might lead students to become blind followers of writing technology, which could be counterproductive to their writing development (ibid.).

Indeed, while GAI holds the promise of revolutionising educational practices, concerns regarding its accuracy and implications for academic integrity have been raised. Al Mahmud is not the only one who has concerns about overreliance on AI. There is a risk that massive use of ChatGPT can lead to an overreliance on AI, which may hinder the development of independent learning skills (Kasneci et al., 2023). Another concern surrounding the integration of GAI in education revolves around its accuracy (Kohnke et al., 2023). GAI tools such as ChatGPT may produce verbose, repetitive, monotonous, and inaccurate responses, and sometimes lack appropriate sources and citations (Kohnke et al. 454).

The prime concern in academic discussion, however, has become the ethical misuse of AI, particularly resulting in plagiarism (Kasneci et al., 2023). This concern has prompted universities to reassess their Academic Conduct Policies, adapting them to the digital age. Scholars have examined the policies and perspectives of 100 top American universities regarding the use of AI writing tools, focusing on ChatGPT as a prominent example (Wang et al., 2023). Findings reveal three prevalent responses: 35.6% of universities lack clear guidance, 54.8% grant instructors autonomy in setting policies, and 9.6% permit conditional use with proper citation. No university has a policy that completely bans the use of GAI. Thus, as AI tool adoption grows, there is a need for educational institutions to use not only plagiarism detectors but as well as AI-detecting ones. Scholars underscore the importance of detecting AI-generated assignments and emphasise the ethical and professional necessity to prevent false classifications of human-written work (Cingillioglu 266). However, questions persist regarding the accuracy of such software in reliably detecting instances of plagiarism or unauthorised assistance.

Koe Driessen (2024) explores the efficacy of AI detectors in a recent study. Findings show that Scribbr's premium AI Detector demonstrated the highest accuracy at 84%, followed by QuillBot's and Scribbr's free AI Detectors, both achieving 78% accuracy. While Scribbr's premium version excels in precision, the free alternatives provide reliable performance without financial investment. However, challenges persist regarding false positives and detecting AI-generated content combined with human text or paraphrasing.

This research study investigates the efficacy of utilising AI-generated content in language learning contexts, with a specific focus on enhancing the writing skills of Spanish teenage students preparing for the Cambridge First Certificate in English (FCE) examination.

1.1. Aim

The research aims to determine the extent to which AI-generated content aids in improving students' writing skills and whether it can be effectively integrated into language learning curricula.

1.2. Objectives

In order to fulfil the set goals, the study is designed to meet the following objectives:

- To assess the reliability and accuracy of AI detection tools in distinguishing between AI-generated texts and those produced by students,
- To assess the effectiveness of AI-generated content, particularly ChatGPT, in facilitating language learning activities, specifically in the context of writing formal letters of application.

2. Materials and Methods

2.1. Participants

The participants in this study consist of ten Spanish teenage students enrolled in a first-year Cambridge FCE preparation course ranked at level B2 based on the Common European Framework of Reference for Languages (CEFR) at Idiomaster in Lucena.

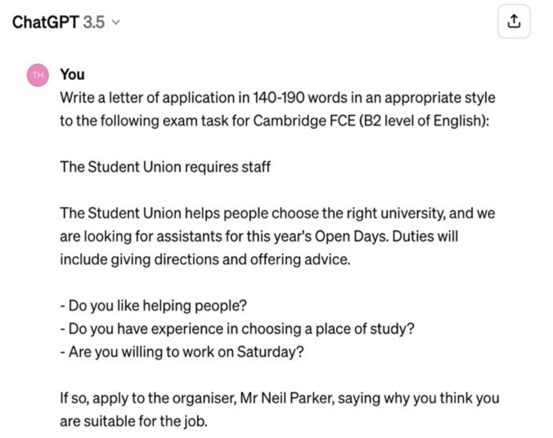
2.2. Procedure

Reliability and accuracy of Scribbr, a free AI detection tool

To conduct this research, I initially selected the Scribbr AI detector, a freely available tool known for its capability to identify AI-generated content. Subsequently, I evaluated the tool's accuracy through a series of testing texts, categorised into two distinct groups.

- ChatGPT-3.5 text, but humanised by 10 students
- Completely human-written texts by 10 students

In total, there were 20 texts across all categories, each ranging from approximately 100 to 300 words in length. The research procedure consists of the following steps:



ChatGPT 3.5

TH You

Write a letter of application in 140-190 words in an appropriate style to the following exam task for Cambridge FCE (B2 level of English):

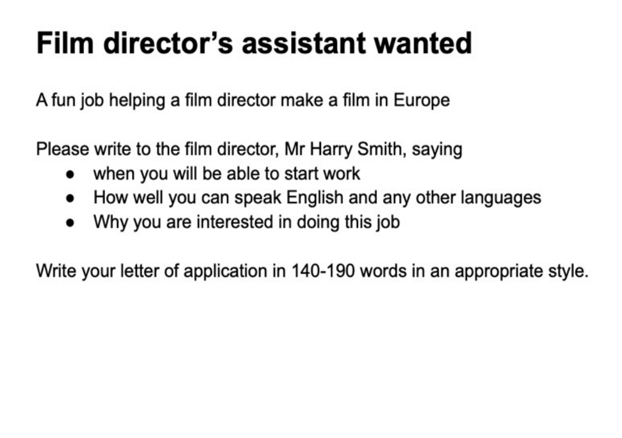
The Student Union requires staff

The Student Union helps people choose the right university, and we are looking for assistants for this year's Open Days. Duties will include giving directions and offering advice.

- Do you like helping people?
- Do you have experience in choosing a place of study?
- Are you willing to work on Saturday?

If so, apply to the organiser, Mr Neil Parker, saying why you think you are suitable for the job.

Figure 1: ChatGPT Prompt (Task 1)



Film director's assistant wanted

A fun job helping a film director make a film in Europe

Please write to the film director, Mr Harry Smith, saying

- when you will be able to start work
- How well you can speak English and any other languages
- Why you are interested in doing this job

Write your letter of application in 140-190 words in an appropriate style.

Figure 2: Controlled Writing Task (Task 2)

- AI-generated Letter Generation: ChatGPT 3.5 was prompted to generate a formal letter of application (Fig. 1).
- Authenticity Assessment: The AI-generated letter was processed through Scribbr to evaluate its authenticity.
- Student Rewrite Task: Spanish teenage students were provided with the AI-generated letter and instructed to rewrite it, using fictitious information. This task was completed as homework and students were advised to use their student books.
- Second Authenticity Assessment: The rewritten letters were subjected to Scribbr analysis to determine any changes in authenticity compared to the original AI-generated version.
- Controlled Writing Task: Students were asked to write original letters of application in response to a different exam task (Fig. 2). This task was conducted in a classroom setting, under exam conditions.
- Data Collection: The AI-generated letter and all student responses were collected and recorded for analysis.

Impact of AI-generated Content on Student's Ability to Produce a Formal Letter of Application

The procedure for examining the impact of AI-generated content on students' ability to produce formal writing, particularly in the context of crafting letters of application, is as follows:

- Using AntConc Software: AntConc is a software tool for text analysis and corpus linguistics research. It features word frequency analysis, providing insights into vocabulary patterns, and keyword analysis, identifying significant terms within a text corpus. The AntConc software was employed to facilitate vocabulary analysis and comparison of student writings.
- Creation of Ignore Lists: Three sets of ignore lists were compiled:
  - *Ignore List 1: Function words (pronouns, determiners, prepositions, and auxiliary words).*
  - *Ignore List 2: Words provided in the instructions for the first writing task.*
  - *Ignore List 3: Words provided in the instructions for the second writing task.*
- Generation of Control List: Using the AI-generated text as a reference, a control list comprising words not present in Ignore 1,2, or 3 was compiled. These words were considered new vocabulary introduced by the AI-generated content.
- Analysis of Student Writing with Control List: The control list is used to analyse both the students' first writings and their second writings. The presence or absence of words from the control list in each student's writing was noted down.
- Evaluation of Vocabulary Retention: The results were analysed to assess students' vocabulary retention from the AI-generated content.

2.3. Data Analysis

The collected data were analysed using both quantitative and qualitative methods. The quantitative analysis included comparing authenticity scores obtained from Scribbr for the AI-generated letter versus student-written letters. The qualitative analysis involved examining the content, structure, and language proficiency demonstrated in the student-written letters. Any identifiable personal information was redacted or anonymised to ensure participant confidentiality.

2.4. Ethical Considerations

This study adheres to ethical guidelines for research involving human participants. Informed consent was obtained from the parents or guardians of all student participants. Participants were assured of confidentiality, and their rights were protected throughout the research process. Additionally, steps were taken to minimise any potential risks of discomfort associated with participation.

3. Results

This section presents the study's results, starting with the first objective, i.e. determining the accuracy of the Scribbr AI-detection algorithm. Table 1displays the results of the Scribbr AI-detection tool from both Task 1 and Task 2.

Student	Scribbr Result Task 1 (%)	Used Vocabulary from the Control List	Scribbr Result Task 2 (%)	Retained Vocabulary from the Control List
		Head-words		Head-words
001	0	9	0	4
002	0	18	16	4
003	0	24	0	5
004	1	25	16	6
005	13	14	25	6

006	62	28	0	6
007	26	6	24	5
008	11	7	13	1
009	17	13	12	5
010	1	17	1	6

Table 1: Scribbr Results for Task 1 and Task 2

The second objective of the study was to determine the students' vocabulary retention. The results are displayed in two tables for Task 1 and Task 2 respectively.

Student Sample	Word count	Scribbr Result (%)	Control List		Ignore List Task 1		Ignore List Function words		Not in Lists		Total head words
			Head words	(%)	Head words	(%)	Head words	(%)	Head words	(%)	
001a	145	0	9	6.90	15	17.24	38	61.38	19	14.48	81
002a	212	0	18	9.86	15	10.80	40	55.87	44	23.47	117
003a	163	0	24	16.57	15	13.02	33	53.25	25	17.16	97
004a	152	1	25	16.88	15	13.64	37	58.44	16	11.04	93
005a	137	13	14	11.59	15	16.67	30	48.55	30	23.19	89
006a	144	62	28	22.97	17	14.86	29	48.65	18	13.51	92
007a	143	26	6	4.20	14	9.79	32	59.44	35	26.57	87
008a	115	11	7	7.76	12	11.21	29	56.03	23	25.00	71
009a	127	17	13	11.63	12	9.30	35	58.14	26	20.93	86
010a	221	1	17	8.97	16	10.76	51	59.19	37	21.08	121

Table 2: AntConc Data for Task 1

Student Sample	Word count	Scribbr Result (%)	Control List		Ignore List Task 1		Ignore List Function words		Not in Lists		Total head words
			Head words	(%)	Head words	(%)	Head words	(%)	Head words	(%)	
001b	182	0	4	2.75	10	9.34	27	62.64	38	25.27	79
002b	166	16	4	2.99	11	11.38	36	56.29	48	29.34	99
003b	169	0	5	2.87	13	12.07	34	64.94	29	20.11	81
004b	192	16	6	2.93	12	8.29	46	63.41	47	25.37	111
005b	179	25	6	3.31	13	13.26	36	58.01	40	25.41	95
006b	157	0	6	4.32	13	9.26	33	56.17	47	30.25	99
007b	283	24	5	2.11	12	8.10	50	59.51	71	30.28	138
008b	112	13	1	0.87	9	9.57	37	60.00	31	29.57	78
009b	158	12	5	3.70	15	12.35	38	57.41	38	26.54	96
010b	212	1	6	3.69	16	14.29	38	58.99	41	23.04	101

Table 3: AntConc Data for Task 2

4. Discussion

In Task 1, where students were tasked with humanising AI-generated content, half of the Scribbr results show a 0 or 1% likelihood of being flagged as AI-generated, as expected. This indicates that students successfully humanised the AI-generated text, resulting in texts that closely resemble human-written content. There are a few exceptions, such as texts 005 and 008, where Scribbr detected a small percentage of AI-like characteristics, potentially indicating areas where the humanisation process was less effective.

In Task 2, where all texts were human-written, the Scribbr results should ideally be 0% for all samples. However, there are seven instances, where Scribbr detected a non-zero percentage of AI-like characteristics. This suggests that Scribbr has produced false positives, erroneously flagging human-written texts as AI-generated. These false positives highlight potential limitations of inaccuracies in the Scribbr AI-detection algorithm.

Possible explanations for these false positives could include similarities between human-written texts and AI-generated content in terms of vocabulary, syntax, or writing style. Additionally, factors such as the complexity of the language used, or the presence of certain linguistic features may contribute to the misclassification of human-written texts.

The presence of false positives underscores the importance of critically evaluating the results of AI detection tools and considering them alongside other evidence or contextual factors. While AI detection tools can provide valuable insights, they are not infallible and may produce inaccurate results in certain cases.

The lack of correlation between students using vocabulary from the AI-generated letter and the Scribbr results can also be observed in Table 1. There is an inconsistency between the Scribbr results and the presence of AI-generated vocabulary in the humanised texts. There does not appear to be a clear correlation between Scribbr results and the presence of AI-generated vocabulary in the control list based on the provided data. Some texts with low Scribbr results still contain a significant amount of AI-generated vocabulary, while others with high Scribbr results have minimal AI-generated vocabulary. However, the lack of correlation suggests that Scribbr’s detection algorithm may not effectively identify AI-generated content solely based on vocabulary usage. Other factors, such as sentence structure, writing style, and syntactic patterns, may also influence Scribbr’s results.

Let us now see the results of the second objective: Have students truly acquired the skills necessary to compose a formal letter of application through the process of studying and humanizing an AI-generated exemplar? The results from Task 2 indicate otherwise. A quantitative analysis, as depicted in Table 3, underscores this assertion. The number and percentage of words from the control list varied across the text samples. While some samples contained a significant portion of the vocabulary from the previous writing task, others had fewer instances. This indicates an inconsistency in the students’ retention of newly acquired vocabulary through a non-traditional instructional approach.

Moreover, a notable proportion of words in each text sample were not present in the control list or the ignore lists, as demonstrated in Table 2 and Table 3. These potentially new words introduced by students or other sources may indicate creativity and diversity of language use in the writing samples. However, upon examining the nature of these new words to determine their relevance and impact on the overall text, it became clear that a portion of these newfound words is marred by misspellings, inaccuracies, or unsuitability within the context of the task at hand.

5. Conclusion

In conclusion, this study illuminates the multifaceted landscape surrounding the integration of AI detection tools in educational contexts, particularly in the realm of assessing student writing. Through a meticulous examination of Scribbr’s performance in identifying AI-generated content and its correlation with students’ writing outcomes, several key findings have emerged.

Firstly, while Scribbr demonstrates efficacy in discerning AI-generated content in some instances, the presence of false positives and inconsistencies highlights the limitations of relying solely on such tools for text analysis. Factors such as variation in humanisation efforts and the intricate nature of linguistic features may influence Scribbr’s accuracy and reliability.

Secondly, the study underscores the nuanced nature of language learning and writing proficiency. Despite students’ attempts to humanise AI-generated content and integrate newly acquired vocabulary, the variability in retention rates and the presence of misspellings or inaccuracies emphasise the need for human-guided instructional approaches and targeted feedback mechanisms. While AI tools can offer valuable resources for language learning, they should be complemented with targeted instruction that fosters critical thinking skills and self-study habits.

Overall, this research calls for a balanced approach to the integration of AI generation and detection tools in educational practices. Such a balanced approach will ensure that students not only acquire language proficiency but also develop the necessary skills to independently navigate and analyse various types of texts effectively. Moving forward, continued research and refinement of AI detection algorithms, coupled with pedagogical strategies that foster critical thinking and writing skills, will be essential in harnessing the full potential of AI in education.

5.1. Implications

The integration of AI into education is inevitable, and educators should embrace its potential rather than resist its presence. Instead of discouraging students from utilising AI tools, educators should empower them with the necessary skills to critically evaluate AI-generated texts. These entail teaching genre writing conventions, critical thinking and analytical skills, enabling students to discern the quality and reliability of AI-generated content.

However, professors, scientists, journal reviewers, and editors must maintain their role as gatekeepers in academia. They should continue to uphold standards of academic integrity by rigorously assessing the quality and authenticity of assignments, research, and publications. It is indeed through this balance between embracing AI technology and upholding academic standards that educators can effectively prepare students for the digital age while safeguarding the integrity of scholarly discourse.

5.2. Limitations of the study

This empirical study has several limitations worth noting. Firstly, it relies on a small sample size of Spanish teenage students. Thus, the findings may not be broadly applicable to all English learners using ChatGPT or similar AI-generated content.

Secondly, the study primarily examines the effectiveness of AI-generated content in improving writing skills, particularly in formal letter writing. While informative, it may overlook other potential impacts of AI technology in language education.

Additionally, the assessment of AI detection tools' reliability is based on a specific tool used in the study, potentially limiting generalisability. Future updates or variations in these tools' algorithms may affect the findings' validity.

Lastly, the absence of a control group limits direct comparisons and definitive conclusions about AI-generated content's effectiveness compared to traditional instruction methods.

5.3. Future Scope

This study sets the stage for future research examining ChatGPT's potential as a tool for teaching genre conventions in English for Academic Purposes (EAP) courses. Subsequent investigations could entail longitudinal studies to gauge its sustained effectiveness, exploration of its adaptability to diverse academic genres, and customisation for non-native English speakers. Moreover, integrating ChatGPT into EAP curricula alongside supplementary resource development and blending it with traditional instruction methods may optimise writing proficiency. Collaboration between ChatGPT, teacher guidance, and peer interaction warrants further exploration to enhance learning outcomes in EAP writing education. Through these avenues, future research can contribute significantly to advancing pedagogical strategies in EAP writing instruction.

References

Al Mahmud, Fawaz. ‘Investigating EFL Students’ Writing Skills Through Artificial Intelligence: Wordtune Application as a Tool’, *Journal of Language Teaching and Research*, vol. 14, no. 5, Sept. 2023, pp. 1395-1404, https://doi.org/10.17507/jltr.1405.28.

Arora, Ashima, et al. ‘Enhancing Writing Skills Through ChatGPT: An Experimental Study in the Context of Ergonomics’. *LimbajŞi Context*, vol. 1, no. XVI, Jan. 2024, pp. 93-104, https://doi.org/10.5281/zenodo.10452223.

Bakai, Yuliia, et al. ‘The Efficiency of Language Teaching through Integration in Future Philologists’ Foreign Language Competence Formation’. *Revista Amazonia Investiga*, vol. 12, no. 61, Feb. 2023, pp. 297–306, https://doi.org/10.34069/ai/2023.61.01.30.

Cingillioglu, Ilker. ‘Detecting AI-Generated Essays: The ChatGPT Challenge’. *International Journal of Information and Learning Technology*, vol. 40, no. 3, May 2023, pp. 259–68, https://doi.org/10.1108/IJILT-03-2023-0043.

Driessen, K. ‘Best AI Detector | Free & Premium Tools Compared’. *Scribbr*, 20 March 2024, https://www.scribbr.com/ai-tools/best-ai-detector/.

Kasneci, Enkelejda, et al. ‘Chatgpt for Good? on Opportunities and Challenges of Large Language Models for Education.’ *EdArXiv*, 30 Jan. 2023, pp. 1-13, https://doi.org/10.35542/osf.io/5er8f.

Kohne, Lucas, et al. ‘ChatGPT for Language Teaching and Learning’, *Technology Review*, vol. 54, no. 2, Aug. 2023, pp. 537-50, https://doi.org/10.1177/00336882231162868.

Nugroho, Arif, et al. ‘The Potentials of ChatGPT for Language Learning: Unpacking Its Benefits and Limitations’. *Register Journal*, vol. 16, no. 02, 31 Dec. 2023, pp. 224–47, https://doi.org/10.18326/register.v16i2.224-247.

Wang, Hui, et al. ‘Generative AI in Higher Education: Seeing ChatGPT Through Universities’, *Policies, Resources, and Guidelines*, Dec. 2023, http://arxiv.org/abs/2312.05235.

Wei, Ling. ‘Artificial Intelligence in Language Instruction: Impact on English Learning Achievement, L2 Motivation, and Self-Regulated Learning’. *Frontiers in Psychology*, vol. 14, 6 Nov. 2023, https://doi.org/10.3389/fpsyg.2023.1261955.