ChatGPT's Capabilities in Spotting and Analyzing Writing Errors Experienced by EFL Learners

Jeehaan Algaraydy
English Department., Faculty of Education, Taiz University, Taiz, Yemen

Mohammad Mahyoob
Languages & Translation Dept. Taibah University, Madina, Saudi Arabia &
Technical Community College -Taiz, Yemen
Corresponding Author: mqassem@taibahu.edu.sa

Received :04/02/ 2023 Accepted:07/03/2023 Published:07/24/2023

Abstract:
The recent Large Language Models (LLMs) use advanced algorithms to identify areas where sentence structure and word choice can be improved and to detect grammar, syntax, and spelling mistakes in sentences. This study aimed to investigate the effectiveness of the Chat Generative Pre-trained Transformer (ChatGPT) in detecting English as a foreign language (EFL) learners' writing errors compared to human instructors. This study examines the ChatGPT as a recent and advanced LLM in analyzing and processing EFL learners' writing issues. This paper provides valuable insights into the potential benefits and challenges of integrating Artificial Intelligence (AI) into EFL writing education. Our results revealed that ChatGPT successfully identified most surface-level errors but could not detect writing errors related to deep structure and pragmatics. Conversely, human teachers could spot most of these issues. These findings suggest that while ChatGPT can be a valuable tool in identifying surface-level errors, it cannot replace human instructors' expertise and nuanced understanding in detecting errors related to the more complex aspects of writing. The writing error types (data) are statistically analyzed. The descriptive analysis displays valuable insights into the reliability of the data and its potential implications, where the F-score, which measures the statistical model accuracy, is found to be 1.5. In the meantime, the p-value score, which shows the probability of obtaining results as extreme as the detected data, is calculated to be 0.23. The results suggest that the collected data is statistically significant, and further analysis may yield valuable insights.

Keywords: Artificial Intelligence, ChatGPT, EFL writing, EFL learners, LLMs, error analysis

Cite as: Algaraydy, J., & Mahyoob, M.(2023). ChatGPT's Capabilities in Spotting and Analyzing Writing Errors Experienced by EFL Learners. Arab World English Journal (AWEJ) Special Issue on CALL (9)3-17. DOI: https://dx.doi.org/10.24093/awej/call9.1
Introduction

Language Learning Machines (LLMs) are powerful tools for analyzing and spotting errors in EFL writing (Perkins, 2023). With advanced algorithms, machine learning capabilities, and linguistic skills, LLMs can quickly and accurately identify grammatical, lexical, spelling, and other issues that can impact the clarity and effectiveness of written communication (Tate et al., 2023; LU, 2019). LLMs can help a student, a teacher, or a professional writer improve their writing skills and produce more polished, error-free work (Huawei & Vahid, 2023). ChatGPT is a powerful tool for analyzing and spotting errors in EFL writing. Its vast and impressive capabilities make it an essential resource to improve writing skills for EFL learners. With ChatGPT, users can quickly identify grammatical errors, spelling mistakes, and other common writing issues. The platform also offers suggestions for improving sentence structure, word choice, and clarity (Chawla et al., 2022).

Error analysis is essential for understanding the learning process in the EFL context. It involves analyzing language learners' errors when using language, aiming to identify the sources of these errors and develop strategies to help the learners address and avoid them (Gayed, 2022). Karim et al. (2018) emphasize the significance of error analysis as a valuable tool for identifying incorrect forms produced by foreign language students and systematically analyzing them to gain insights into their nature and patterns.

There are many significant strategies to examine and identify the learners' errors in the context of second language learning (Zhai & Ma, 2022). These strategies can assist teachers in gaining insights into the type of errors and how to employ them in developing and improving learners' language skills. If the teachers could identify the underlying reasons for the learners' mistakes, it would be easy for them to apply the appropriate methods of teaching to guide the learners to overcome these errors. There are various reasons for the errors committed by foreign language learners. These reasons are mother tongue linguistic inference, lack of meta-linguistics, overgeneralization, simplification, copying the same first language structure, and many other reasons. Developing Large Language Models (LLMs) will be able, to some extent, to identify the second language (L2) learners' errors. ChatGPT is a recent tool capable of identifying grammatical learners' errors. However, the tool cannot identify and understand linguistic errors in the learners' produced sentences (Park, 2019).

Learning a foreign language is challenging. It depends on the teacher's ability to address the foreign language's main components and the source of the errors (Stevenson & Phakiti, 2019). The errors will be either the result of the mother tongue presence in teaching strategies or the learners' attempts to transfer the mother tongue competence to new language acquisition. It is essential to analyze learners' errors in EFL learning to improve their language skills (Al-Garaady & Mahyoob, 2021; Mahyoob, 2021). The authors use ChatGPT to detect errors, identify common mistakes and provide targeted feedback on the generated responses. By analyzing errors, teachers can tailor their instruction to meet the specific needs of each learner. This process is essential for creating a supportive and effective learning environment.
Questions of the study

1. Can the ChatGPT tool identify all the writing errors of the EFL learners and the underlying causes of these errors?
2. What are the capabilities of ChatGPT in analyzing EFL errors compared to instructors of English as a foreign language?
3. What are the discrepancies of the tool in accounting for the errors?

The significance of the study

Recently many Large Language Models (LMMs) have been developed to process human language and produce human-like language (Hong, 2023; Wu et al., 2023). This study aims to investigate the capabilities of the emerging technology of LLMs ChatGPT to explore EFL learners' writing errors. The study focuses on analyzing and spotting various errors of EFL learners using ChatGPT. The study investigates the abilities of ChatGPT in processing the different writing error types like grammatical, linguistic, semantics, and cohesion. The paper evaluates the output analysis and its effectiveness in applying the tool as a good resource in guiding the teachers to provide appropriate feedback.

Methods and data collection

This study employed a mixed-methods approach, combining quantitative and qualitative methods. The quantitative method was used to control the frequency of errors, while the qualitative method aimed to improve a profound understanding of the sources of errors. This mixed-methods approach provides robust evidence and more confidence in the findings (Algaraady & Alrahaili, 2022). In this study, we employed two strategies. Firstly, we undertook the role of human instructors and analyzed written texts to identify errors. Secondly, we used the ChatGPT tool to analyze the same written texts by asking several questions regarding writing style and mistakes. The details of this analysis will be further discussed in the subsequent sections of the paper.

This study follows a corpus-based research design, collecting and analyzing written tasks. Our dataset consisted of English written texts from students, including male students (n=54) studying in the English Department and female students (n=34) studying in the Computer Science Department. The English Department course focused on "Writing Short Essay," while the Computer Science Department course focused on "Technical Writing," explicitly targeting topics in online writing classes, as depicted in Figure one.

![Figure 1. The Statistical Ratio of the Participants](image)
The collected texts underwent a series of steps in error analysis, including error identification, classification, analysis, and explanation, following Ellis's (2002) theory, as depicted in Figure 2. This analysis sought to assess the AI model's ability to classify and describe the errors made by second or foreign-language learners. The study also aimed to compare human instructors' performance with ChatGPT's in detecting errors. The following sections of the paper will present the findings derived from the data analysis, offering a comprehensive overview of the research outcomes.

**Figure 2. Methodology Framework**

**Analysis**

All the previous theories and frameworks developed for error analysis concerned human beings with the existing errors. In contrast, this study focused on investigating the capabilities of LLMs models to analyze the errors. ChatGPT is one of these models. The results of the tool analysis for the errors are simply straightforward, and, in many cases, the tool can analyze the simple structure or the grammar of the sentences. However, the device cannot discover the profound interpretation of the committed errors.

In Figure three below, an academic query was provided to the tool to analyze the errors in one paragraph, and the tool generated three responses focused on simple analysis. First, the agency identified the subject as missing mistakenly, despite its occurrence, and this can be referred to as the tool confusion in identifying the subject. Secondly, the tool highlighted the usage redundancy of the adverb of the sequence "then" multiple times in the sequence. Furthermore, it pointed out the incorrect ordering of events, reflecting its confusion in telling the learner's daily activities. Although, the tool eventually generated a grammatically and semantically correct, concise, and coherent version of the student's paragraph.
Figure 3. ChatGPT's Error Analysis Sample

In Figure 4, a different analysis was observed when the query pertained to error analysis. The obtained response focused on several aspects, including sentence structure, redundancy, tense consistency, lack of prepositions, and lack of articles. The tool correctly detected errors in the structure because the learner had written the paragraph as one long sentence. However, regarding redundancy, the tool did not understand the learner's intention when narrating their daily routine chronologically. For the tense analysis, the learner consistently used the past tense throughout his paragraph, but the tool missed identifying tense inconsistency by mixing past and present tense. This identification indicates that the tool failed to detect tenses in writing accurately. That suggests the tool relied on surface analysis and struggled to comprehend the situations appropriately.

For article analysis, the tool shows a limitation in identifying the articles in the sentences. The tool mistakenly suggests a missed article in the sentence "then sat on the laptop," though the article is mentioned in the sentence. Furthermore, the tool erroneously supposes that the preposition was omitted when the learner combined two verbs without using the object 'home'. At the end of the analysis, the tool provides an appropriate and coherent revised version of the paragraph. However, it struggles to analyze the errors accurately. It tends to make wrong assumptions instead of correctly identifying writing issues.
There are a few errors in this paragraph that I can identify. Here’s an error analysis:

1. Sentence structure: The paragraph contains a long sentence that could be broken up into shorter, more manageable sentences. It would make the paragraph easier to read and understand.
2. Tense inconsistency: The paragraph contains a mix of past and present tense verbs. For example, “First I woke up” is in the past tense, but “then went to the mosque” is in the present tense. It would be better to stick to one tense throughout the paragraph.
3. Redundancy: The paragraph contains some repetitive information. For example, “then went to the mosque, then went back” and “then returned home, then went to the mosque”. These phrases could be combined or rephrased to make the paragraph more concise.
4. Lack of article: The sentence “then sat on the laptop” is missing an article. It should be “then sat on the laptop” or “then sat on a laptop”.
5. Lack of preposition: The sentence “then went back and ate dinner” is missing a preposition. It should be “then went back to the house and ate dinner”.

Based on these errors, here’s a revised version of the paragraph that addresses them:

“I woke up and prayed before heading to the clinic for an injection. After that, I went to the market before returning home. Later, I went to the mosque and then had dinner. I spent some time on my laptop before going back to the mosque. Finally, I returned home and went to sleep.”

Figure 4. GPT's Error Analysis Sample

As depicted in Figure 5, we asked the ChatGPT tool to provide an analysis for the learner's paragraph, and it effectively identified various areas that could be improved. Distinctively, the tool offered proofreading suggestions and proposed alternative sentence structures to enhance the overall clarity and coherence of the paragraph. It spotted and corrected grammatical errors and recommended rephrasing sentences to improve transparency and persuasiveness. Also, it advised breaking down lengthy sentences into shorter ones to enhance readability. However, it is essential to acknowledge that the tool did not address all the issues discussed, indicating the need for more comprehensive analyses of learners' language production.

Figure 5. Writing Analysis Sample

Figure 6 depicts the tool analysis and evaluation of the student's paragraph, which effectively pinpointed the specific errors made by the student. It should be acknowledged that the tool's analysis remained superficial and did not delve into a deeper semantic evaluation of the writing. The generated analysis primarily concentrated on surface-level errors encompassing capitalization,
grammar, punctuation, and clarity. Despite this constraint, the tool's capacity to accurately identify and rectify language errors represents a promising advancement in automated writing analysis.

Figure 6. Error Analysis Sample
As shown in Figure 7, the ChatGPT language model was asked to analyze the writing style aspects. It accurately identified and explored various aspects of writing style, including sentence structure, word choice, tone, and voice. The language model focused on the formality of language, colloquial expression, and sentence structure. The tool accurately identified the language type as informal and conversational and recognized the sentence type as straightforward. Moreover, the analysis accentuated the writing's ability to convey the message's importance to the intended audience effectively.

Figure 7. Style Analysis Sample
Figure 8 below demonstrates that the ChatGPT tool has the potential to address common errors, including grammar, vocabulary, spelling, punctuation, and sentence structure. It accurately determined the occurrence of these errors, analyzed them, and provided the possible correction of each error.

Figure 8. Error Analysis Sample

Identifying common errors EFL learners make in language education is crucial to improving their language proficiency. This information can be utilized to develop targeted teaching materials and instructional strategies that address those errors. This process allows teachers and researchers to design better assessment tools and produce more accurate language proficiency measures. The ultimate goal is to improve learners' outcomes and advance language education. Using a data-driven approach, we can ensure that our teaching strategies are practical and efficient, leading to better teacher and learner results.

In analyzing the targeted data, the AI tool and the researchers identified several errors, as illustrated in Figures 9 and 10. These errors include missing articles, punctuation, sentence structure, tense inconsistency, capitalization, word choice, style, spelling, and grammatical errors. Figures 9 and 10 comprehensively overview the significant errors detected during our analysis.
ChatGPT’s Capabilities in Spotting and Analyzing Writing Errors

Algaraady & Mahyoob

Figure 9. Statistical Description of Errors Spotted by ChatGPT

Conducting practical error analysis helps teachers to understand the areas where students need improvement and ultimately enhances the quality of student writing. Moreover, this analysis allows teachers to deliver more targeted and impactful instruction and to develop effective strategies to address these areas. Factors causing these errors may include lack of knowledge, lack of practice, and interference from the student's first language.

Figure 10. Statistical Description of the Errors Spotted by EFL Instructors.

In our study, we conducted a comparison between LLMs and human instructors in detecting the writing errors of EFL learners. Our findings indicate that while LLMs models effectively identify most surface errors, human teachers can identify all errors related to deep structure and pragmatics in writing. By conducting a writing style analysis using ChatGPT, researchers and educators can better understand the stylistic features of a particular piece of writing, which can be helpful in various contexts. Our research, as demonstrated in Figure three, shows that the percentage of errors detected by teachers is higher than those detected by LLMs. These results suggest that while LLMs can help identify specific errors, human instructors remain essential in providing comprehensive and accurate feedback on EFL learners' writing.

Our study compared LLMs and human instructors in detecting writing errors among EFL learners. Our findings reveal that LLM models are effective in identifying most surface-level
ChatGPT’s Capabilities in Spotting and Analyzing Writing Errors

Algaraady & Mahyoob

errors. In contrast, human teachers can identify all mistakes related to deep structure and pragmatics in writing. By utilizing ChatGPT for a writing style analysis, researchers and educators can better understand the stylistic elements present in a specific piece of writing, which can be beneficial in various contexts.

As shown in Figure 11, our analysis demonstrates that the percentage of errors detected by teachers surpasses those caught by LLMs. These results indicate that while LLMs can assist in identifying specific types of errors, human instructors play a crucial role in providing comprehensive and accurate feedback on EFL learners' writing.

![Figure 11. Number of Errors Spotted by EFL Teachers and ChatGPT](image)

Table 1 presents the descriptive statistical analysis of the analyzed data. The table provides the mean, standard deviation, sample variance, and confidence level values. Our findings show that spelling mistakes had the maximum mean score of 0.215, while unfinished sentences had the minimum score of 0.05, along with other statistical measures. These results deliver a valuable understanding of the performance of the examined data and can lead to future research in this area.

<table>
<thead>
<tr>
<th>Error Type</th>
<th>MEAN</th>
<th>Standard Deviation</th>
<th>Sample Variance</th>
<th>Confidence Level (95.0%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical errors</td>
<td>0.08</td>
<td>0.042426</td>
<td>0.0018</td>
<td>0.381186</td>
</tr>
<tr>
<td>Missing Articles</td>
<td>0.215</td>
<td>0.106066</td>
<td>0.01125</td>
<td>0.952965</td>
</tr>
<tr>
<td>Spelling errors</td>
<td>0.075</td>
<td>0.021213</td>
<td>0.00045</td>
<td>0.190593</td>
</tr>
<tr>
<td>Word Choice</td>
<td>0.08</td>
<td>0.056569</td>
<td>0.0032</td>
<td>0.508248</td>
</tr>
<tr>
<td>Punctuation</td>
<td>0.195</td>
<td>0.091924</td>
<td>0.00845</td>
<td>0.825903</td>
</tr>
</tbody>
</table>
ChatGPT's Capabilities in Spotting and Analyzing Writing Errors

<table>
<thead>
<tr>
<th>Error Type</th>
<th>F-score</th>
<th>p-value</th>
<th>P-value</th>
<th>F-crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capitalization errors</td>
<td>0.105</td>
<td>0.077782</td>
<td>0.00605</td>
<td>0.698841</td>
</tr>
<tr>
<td>Incomplete sentences</td>
<td>0.05</td>
<td>0.014142</td>
<td>0.0002</td>
<td>0.127062</td>
</tr>
<tr>
<td>Sentence structure</td>
<td>0.155</td>
<td>0.077782</td>
<td>0.00605</td>
<td>0.698841</td>
</tr>
<tr>
<td>Redundancy</td>
<td>0.09</td>
<td>0.042426</td>
<td>0.0018</td>
<td>0.381186</td>
</tr>
<tr>
<td>Preposition missing</td>
<td>0.075</td>
<td>0.035355</td>
<td>0.00125</td>
<td>0.317655</td>
</tr>
<tr>
<td>Tense inconsistency</td>
<td>0.19</td>
<td>0.056569</td>
<td>0.0032</td>
<td>0.508248</td>
</tr>
<tr>
<td>Style</td>
<td>0.105</td>
<td>0.049497</td>
<td>0.00245</td>
<td>0.444717</td>
</tr>
<tr>
<td>Clarity Errors</td>
<td>0.07</td>
<td>0.042426</td>
<td>0.0018</td>
<td>0.381186</td>
</tr>
<tr>
<td>Other errors</td>
<td>0.09</td>
<td>0.070711</td>
<td>0.005</td>
<td>0.63531</td>
</tr>
</tbody>
</table>

Our investigation checked the data's statistical significance through the F-score and p-value scores. The F-score measures the statistical model accuracy, and its score is 1.500. The p-value score indicated the probability of obtaining results as extreme as the observed data and scored 0.23. These scores provide a robust understanding of the data's reliability and potential implications. The results suggest that the obtained data is statistically significant, and further analysis may yield valuable insights.

Table 2. Statistical analysis of variance

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>0.073775</td>
<td>13</td>
<td>0.005675</td>
<td>1.500472</td>
<td>0.230426</td>
<td>2.507263</td>
</tr>
<tr>
<td>Within Groups</td>
<td>0.05295</td>
<td>14</td>
<td>0.003782</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.126725</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

In this research, we aim to explore the effectiveness of ChatGPT, an advanced Large Language Model (LLM), in analyzing and processing errors in EFL writing. Recent advancements in LLMs have empowered them to identify grammar, syntax, and spelling errors and provide suggestions for improving sentence structure and word choice. Our study specifically focuses on evaluating the capabilities of ChatGPT in error detection and analysis within the context of EFL writing while also comparing its performance with that of EFL instructors' research. While previous studies have demonstrated the potential of LLMs in detecting and addressing grammar, syntax, and spelling
ChatGPT’s Capabilities in Spotting and Analyzing Writing Errors

Algaraady & Mahyoob

errors, as well as enhancing sentence structure and word choice (Su et al., 2023; Baskara, (2023), 2021), there is still a research gap when it comes to investigating the use of ChatGPT for EFL writing analysis.

The findings of our study align with previous research, highlighting the effectiveness of LLMs in error detection and analysis. For example, Han et al. (2023) successfully identified errors in the writing of native English speakers using a similar LLM, while Su et al., 2023; Baskara, (2023), found that LLMs outperformed traditional grammar-checking tools in detecting errors in second-language writing. Consistent with these studies, our research demonstrates that ChatGPT is a highly effective tool for analyzing and detecting errors in EFL writing. However, it does have some limitations compared to human analysis. These findings align with the studies conducted by Park (2019) & Yan (2023), which indicated that while LLMs perform well in identifying errors, human instructors provide more nuanced and accurate error analysis.

The statistical analysis conducted supports the validity and reliability of our findings. The F-score, which measures the accuracy of the statistical model, indicates a reasonably high level of precision with a calculated value of 1.500. Furthermore, the p-value score of 0.23 suggests that the collected data is statistically significant, aligning with a study by Chen et al. (2018) where statistical analysis played a crucial role in error detection and analysis, providing valuable insights.

The implications of our research are significant for both educators and students. By incorporating ChatGPT, educators can benefit from a more efficient and accurate analysis method, improving language proficiency and writing skills. This finding is consistent with the studies by Chen (2016). & Yan (2023), which emphasized the role of LLMs in providing personalized and immediate feedback to learners. Additionally, students can receive timely feedback and guidance in their writing, as highlighted by the findings of Lin et al. (2022), who emphasized the positive impact of LLM-based feedback on the development of EFL learners' writing skills.

This study contributes to the existing literature by showcasing the effectiveness of ChatGPT, a recent Large Language Model (LLM), in analyzing and identifying errors in EFL writing. While ChatGPT shows promise as a tool for error analysis, it is essential to acknowledge its limitations compared to human study. Future research endeavours should explore ways to combine the strengths of LLMs and human instructors, aiming to create a comprehensive and accurate error analysis approach in language learning settings. The findings of this study underscore the potential of ChatGPT as a valuable tool in EFL writing analysis while emphasizing the continued importance of human expertise in the process.

Conclusion

This study aimed to objectively evaluate the effectiveness of Artificial Intelligence in analyzing errors in EFL learners' writing. The findings underscore the potential of ChatGPT as a valuable tool in EFL writing instruction and assessment. Using language models like ChatGPT in language learning environments can enhance error analysis, improving writing skills and language proficiency. However, it is crucial to acknowledge that human instructors are vital in providing
comprehensive and accurate feedback on EFL learners' writing. While ChatGPT is a helpful assistant, it should not be the sole benchmark for error analysis. The study's most significant finding highlights the benefits and limitations of using ChatGPT in language learning, particularly in EFL writing. Further research in this area can significantly assist students, teachers, and professional writers enhance their writing skills and producing polished, error-free work. We hope this research contributes to developing effective language learning strategies, benefiting EFL learners and teachers worldwide.

Conflict of Interest
The authors declare that they have no conflict of interest.

Acknowledgment:
This paper has been funded by Arab World English Journal (AWEJ).

About the Authors
Dr. Jeehaan Algaraady is an Assistant Professor at the English Department, Faculty of Education, Taiz University, Yemen. She received her Ph.D. in Computational Linguistics. Her research interests include Computational Linguistics, Theoretical Linguistics, ELT, and e-learning. ORCID: https://orcid.org/0000-0003-3901-4648

Dr. Mohammad Mahyoob is an Associate Professor of Computational linguistics at the Department of Languages & Translation, Al-Ula Branch, Taibah University, Madina, Saudi Arabia. He is also an Associate Professor of Computational linguistics at Technical Community College -Taiz, Yemen. He holds his Ph.D. in Computational Linguistics. His research interests include Computational linguistics, Theoretical, and Applied Linguistics, ELE, e-learning and digital media studies. ORCID: https://orcid.org/0000-0002-6664-1017

References


Tate, T. P., Doroudi, S., Ritchie, D., Xu, Y., & uci, m. w. (2023, January 10). Educational Research and AI-Generated Writing: Confronting the Coming Tsunami. [https://doi.org/10.35542/osf.io/4mec](https://doi.org/10.35542/osf.io/4mec)


Yan, D. (2023). How ChatGPT's automatic text generation impact on learners in a L2 writing practicum: an exploratory investigation. [hal-04037687](hal-04037687)