Tapping into the Pedagogical Potential of infinigoChatIC: Evidence from iWrite Scoring and Comments and Lu & Ai’s Linguistic Complexity Analyzer

Danping Dong
School of Foreign Languages, Yunnan University
2 Green Lake Road, 650000 Kunming, PRC
Email: 775330038@qq.com

Received:11/03/2023          Accepted:03/08/2024   Published:04/24/2024

Abstract
ChatGPT has drawn rapid interest among linguists and language teachers, with academics and industries rushing to tap into its potential. Yet the pedagogical potential of ChatGPT for L2 English writing remains to be further explored academically. This study aims to examine the effects of ChatGPT, or one of its Chinese versions called infinigoChatIC, in improving the quality of L2 writing and in facilitating the development of L2 proficiency among some postgraduate students learning an English reading and writing course in a prestigious university in Southwest China. Specifically, the main research questions focus on how infinigoChatIC-polished L2 English writings perform in comparison to L2 learner writings, and how the improvement of English writing is reflected in the changes of lexical and syntactic complexity, as evidenced by measured data and its pedagogical significance. A set of 24 English writings are collected as part of regular homework assignments, then this set of English writing is polished by the teacher researcher in the infinigoChatIC dialogue box with the prompt “please correct grammatical errors in the following English text so that it is grammatically accurate.” These two sets of writings are scored in iWrite automatic scoring system and then analyzed in Lu & Ai’s web-interfaced linguistic complexity analyzer. The study finds that infinigoChatIC can effectively improve the quality of L2 English writing, with lexical complexity vastly improved and yet surprisingly overall syntactic complexity slightly reduced. Pedagogical implications are discussed.

Keywords: iWrite scoring and comments, Lu & Ai’s linguistic complexity analyzer, L2 English writing, pedagogical potential of infinigoChatIC

Cite as: Dong, D. (2024). Tapping into the Pedagogical Potential of infinigoChatIC: Evidence from iWrite Scoring and Comments and Lu & Ai’s Linguistic Complexity Analyzer Arab World English Journal (AWEJ) Special Issue on ChatGPT, April 2024: 124-137.
DOI: https://dx.doi.org/10.24093/awej/ChatGPT.8
Introduction

ChatGPT has drawn rapid interest among linguists and language teachers, with industries, including language service, rushing to tap into its instrumental potential. ChatGPT, or Generative Pre-trained Transformer (GPT) models, especially GPT 3, based on the Transformer Model created by Google in 2017, has been widely used in natural language processing, including language translation (Eke, 2023). ChatGPT can produce human-like responses that are coherent and relevant to a variety of prompts, like questions and requests (Gilson et al., 2023; Pavlik, 2023). Compared to traditional Machine Translation tools, it is widely assumed that ChatGPT can produce language translations that are more fluent, grammatically accurate, and contextually appropriate. However, despite buzzes of tests and practices and seminars and workshops, few researches have been conducted to assess statistically and qualitatively whether and how ChatGPT-polished texts are better than students’ self-written and MT-produced essays, particularly regarding linguistic complexity. This study aims to provide quantitative and qualitative evidence for the effective use of ChatGPT, in this case, infinigoChatIC (https://www.infinigo.com/chatIC), which is a freely accessible and available version of ChatGPT in China, in the instruction of L2 English writing. This study will shed light on the approaches that L2 English writing teachers and students may take to improve both the writing quality and proficiency of students concerning ChatGPT and similar AI tools. The study will try to answer the following three questions:

a. How do infinigoChatIC-polished L2 English writing perform in comparison to learner L2 writings as evidenced by measured data?

b. How is the improvement of English writing reflected in the changes of lexical and syntactic complexity, again as evidenced by measured data?

c. What pedagogical insights could be generated in this process?

Literature Review

Linguistic complexity is a statistical construct associated with L2 writing proficiency and the quality of L2 writing, often perceived as the indicators, diagnostics, or proxies of the latter (Wolfe-Quintero et al., 1998; Ortega, 2003; Ortega, 2012). Linguistic complexity involves two dimensions: lexical and syntactic. Lexical complexity is a multi-dimensional construct (Xu & Casal, 2023; Lu, 2010; Lu, 2011; Ai & Lu, 2013; Lu & Ai, 2015; Kyle & Crossley, 2015, Kyle et al., 2018) involving significant dimensions, such as lexical density, lexical diversity, and dozens of fine-grained measures, such as verb variation, corrected verb sophistication, etc. Syntactic complexity is also a multi-dimensional concept with each dimension requiring one or more different measures appropriate for that dimension (Bulté & Housen 2012; Lu, 2011; Norris & Ortega, 2009; Lu & Ai 2015), such as length of production unit (MLC, MLS, MLT), amount of subordination (C/T, CT/T, DC/C, DC/T), amount of coordination (CP/C, CP/T, T/S), degree of phrasal sophistication (CN/C, CN/T, VP/T), overall sentence complexity (C/S).

It is often hypothesized that both lexical and syntactic complexity can strongly predict L2 proficiency and the quality of L2 writing, with lexical sophistication being an exceptionally robust predictor (e.g., Kim & Crossley, 2018). In a study where syntactic complexity and lexical sophistication were analyzed in association with grade levels of responses to picture description tasks, it was found that MLS and MLC, two indices of syntactic complexity, AWL or academic word list, an index of lexical sophistication, are strong predictors of the writing grade levels (Barkaoui, 2024).
However, looking at native English writings’ baseline data, after a certain point, a decrease in syntactic complexity might be interpreted as a sign of higher proficiency (Pallotti, 2009). This inconsistency creates an exciting space for research. To add to the complexity of this issue, it is discovered that syntactic complexity differs among NS and NNS with different L1 backgrounds (Lu & Ai, 2015). Besides, it was found in a study (Chen, 2024) that compared to native English writing, EFL and translated English writings in China exhibit both simplification tendencies as reflected in shorter sentences and lower usage of subordination structures, and improved complexities as reflected in the use of phrasal coordination and complex nominals. It is thus reasonable to argue that more evidence is required to further the understanding between linguistic complexity and L2 English writing quality and proficiency.

This study sets out to examine the effective use of ChatGPT (in this study, infinigoChatIC) in polishing L2 English writing by students, either self-written or MT-assisted, and thus through comparative analyses of such writings in terms of its lexical and syntactic complexity, pointing to the specific language knowledge for students to learn, and also offering a huge opportunity for the students highly driven to improve their writing by accessing ChatGPT anytime and anywhere as they wish.

Method

This is a cross-sectional study that is conducted in the natural learning process instead of under experimental conditions, as the latter may not be able to reflect the actual situation and hence lack ecological validity (Van Lier, 2000; Van Lier, 2004). The 24 L2 English writing samples are produced under non-experimental conditions as part of the regular coursework.

Participants

The participants are 24 postgraduate students pursuing a Master’s degree in a high-ranking university in Southwest China in the spring semester of 2023. These participants chose to submit their writings as an optional homework assignment. This is a small number: 24 out of 121 students in three classes that the researcher taught. Therefore, it is a convenience sampling. To improve their L2 writing skills, the teacher researcher assigns a writing task each week with a prompt differing from one week to the other. The study chose one week’s essay writing as the research object, polished the 24 writings with ChatGPT, and then had them scored in iWrite to compare the differences in scores concerning different dimensions such as language, content, text structure, and technicalities, and then have them analyzed via Prof. Lu Xiaofei’s syntactic and lexical analyzers to get the statistics, again to compare the differences between the two sets of texts. Then, based on the statistical findings, the study will explore how the teacher can use the results of comparative analysis for pedagogical purposes.

It should be noted that the writing assignment is not to write an essay in the typical sense but to post a writing as a form of discussion about the same topic. Specifically, the students are expected to express their view and then argue for it. Therefore, the writing is informal, with no requirement about the number of words used in the writing or strict requirements concerning text structure. Students are encouraged to use MT tools they are familiar with to assist their writing and also to read parallel English texts for information.
**Research Instruments**

ChatGPT has a lot of models and iterations. The technology used in this study is infinigoChatIC, one of the numerous Chinese versions of ChatGPT created by Chinese tech companies, which is the only one found freely available to the public. InfinggoChatIC(https://www.infinigo.com/chatIC) is an AI-powered chatbot platform driven by advanced NLP (Natural Language Processing) algorithms. It can generate human-like responses in both Chinese and English, equipped with grammar-checking and language correction tools that can identify and correct errors in English text. The researcher has used infinigoChatIC to polish the 24 L2 English writings by the postgraduate students in her university with the same prompt: Please revise the following English paragraph and ensure that it is grammatically correct. The results created by infinigoChatIC are then copied and pasted into the iWrite system for scoring and evaluation to compare them with students' self-written or MT-assisted essays.

iWrite is an automatic English teaching and scoring system(http://iwrite.unipus.cn) designed by Professor Liang Maocheng and his team for intelligent diagnosis of relevance and coherence of an English essay written by L2 English learners and offers immediate feedback on grammar and usage. iWrite serves as the official scoring system for the FLTRP-ETIC Cup English Writing Contest. It has won the championship in 2020 CGED. iWrite serves over 600,000 users and 1700 educational institutions in over 300 cities of more than 30 provinces. iWrite offers not only a global score in a hundred-mark system but also specific scoring concerning the language, content, text structure, and technical specification in a five-star system, as well as a score in terms of coherence and relevance in a hundred-mark system. It also diagnoses the text grammatically. The detailed data of both the sets of the English essay and their ChatGPT-polished versions are collected and analyzed to offer an insight into the potential of infinigoChatIC in L2 writing instruction.

To provide more statistical evidence as to the potential role of infinigoChatIC in L2 writing instruction, Web-based Lexical Complexity Analyzer and Web-based L2 Syntactic Complexity Analyzer(https://aihaiyang.com/software/) are used to analyze the lexical and syntactic complexity of the two sets of texts under research in this study.

**Results**

**Statistics from iWrite Scoring and Related Case Studies**

Table one indicates that ChatGPT-polished texts use slightly fewer words than students' self-written or MT-assisted writing, in sharp contrast to the rise of mean iWrite global scoring from 65.375 to 79.158, by 17.4%. The number of words used might indicate the increase of lexical density via the process of nominalization, which in turn leads to the increase of the iWrite global scoring. This hypothesis is to be tested by more data in the text.

<table>
<thead>
<tr>
<th></th>
<th>Number essays</th>
<th>Total word count</th>
<th>Mean of essay length</th>
<th>Mean of iWrite global scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-written or MT-assisted</td>
<td>24</td>
<td>2558</td>
<td>106.58</td>
<td>65.375</td>
</tr>
<tr>
<td>ChatGPT-polished</td>
<td>24</td>
<td>2446</td>
<td>101.92</td>
<td>79.158</td>
</tr>
<tr>
<td></td>
<td>↓ 4.38%</td>
<td>↓ 4.37%</td>
<td>↑ 17.4%</td>
<td></td>
</tr>
</tbody>
</table>

Table two shows that the InfinigoChatIC-polished texts have risen in the iWrite scoring of all dimensions. The enhancement in language is most prominent(31.1%), with that in technicalities coming second(11.3%), with the enhancement of coherence least prominent(up by 1.2%).

---

[127]

Arab World English Journal
www.awej.org
ISSN: 2229-9327
Table 2. *iWrite* scoring data

<table>
<thead>
<tr>
<th></th>
<th>Language (mean)</th>
<th>Content (mean)</th>
<th>Text Structure (mean)</th>
<th>Technicalities (mean)</th>
<th>Coherence (mean)</th>
<th>Relevance (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-written or MT-assisted</td>
<td>3.354</td>
<td>4.021</td>
<td>2.396</td>
<td>4.042</td>
<td>95.625</td>
<td>89.782</td>
</tr>
<tr>
<td>ChatGPT-polished</td>
<td>4.396</td>
<td>4.417</td>
<td>2.583</td>
<td>4.5</td>
<td>96.75</td>
<td>93.167</td>
</tr>
<tr>
<td></td>
<td>↑31.1%</td>
<td>↑9.8%</td>
<td>↑7.8%</td>
<td>↑11.3%</td>
<td>↑1.2%</td>
<td>↑3.8%</td>
</tr>
</tbody>
</table>

It is found that infinigoChatIC excels in correcting language errors. Here is a typical example of students’ writing enhanced by infinigoChatIC.

I think genius is decided by gene and learning the day after tomorrow. Genius is innate, which is absolutely reasonable. There are some people are outstanding that they feel unreal in the world, for example, Picasso’s achievements in painting art, Jordan’s achievement in basketball and Michael’ Jackson’s achievements in pop music. Their talents are definitely closely related to genetics, and they have inherited the top talents. However, the viewpoint that genius is acquired also makes sense. There are so many people who have achieved great success through their unremitting efforts. So we must acknowledge that some people rely on postnatal efforts to become more outstanding.(a student’s writing, with punctuation errors retained.)

The above text showcases a typical student L2 writing, which the student writer is completely disengaged from, a case that students write in the Chinese and have the Chinese text translated into English by MT, as an error could only be made by MT tools rather than humans, as shown in the following:

“I think genius is decided by gene and learning the day after tomorrow”.

There is also a sentence that adopts the Chinese syntactic rule so that the readability is quite low:

“There are some people are outstanding that they feel unreal in the world, for example…”

Fortunately, the infinigoChatIC-polished writing has eliminated the above errors; furthermore, variations of words are used for the same meaning, as shown in the use of “pioneering work, unrivaled talent, and innovation” to substitute the single word “achievements.” Here is the infinigo-polished version:

“My belief is that genius is determined by both genetics and learning. The innate talent is undoubtedly reasonable. There are individuals that appear otherworldly, such as Picasso's pioneering work in painting, Jordan's unrivaled talent in basketball, and Michael Jackson's innovation in pop music.”

Besides the improvement of language, it is interesting that although the enhancement of coherence shows least prominently in statistics(1.2%), it is very significant in a case analysis. The following is a text written by a student, Cathy(Pseudonym):

I believe that genius is formed through nurture rather than nature.

Firstly, human intelligence is the result of the interaction between genes and the environment, but the difference in innate intelligence between individuals is not significant. This means that postnatal effort is crucial. Secondly, the term "genius" greatly diminishes the importance of people's postnatal effort and dedication, while exaggerating the significance of innate conditions, which can lead to misunderstandings. Finally, since our innate conditions are similar, I believe that through postnatal effort, practice, exploration,
curiosity, and critical thinking, we can still shine and become geniuses in certain fields. (The underlines are added by the researcher.)

In contrast, the infinigo-ChatIC-polished text reads as follows:

I hold the belief that genius is mostly nurtured from postnatal care rather than innate qualities. For one, human intelligence is determined by complex interactions between genetics and the environment, but the inheritance of native intelligence amongst individuals is not vastly different. Therefore, it is the efforts made after birth that shape our minds significantly. Furthermore, I believe that the term "genius" underestimates the significance of postnatal effort and hard work while exaggerating the relevance of innate conditions, potentially causing misconceptions. Ultimately, since our innate conditions are alike, I am of the opinion that with sufficient postnatal dedication, practice, curiosity, exploration, and critical thinking, we can all excel and become geniuses in our respective domains. (The underlines are added by the researcher.)

The replacement of “firstly...secondly...finally” with “for one...furthermore...ultimately” is meaningful for EFL teachers to instruct students to learn advanced coherent and linking devices.

In another case study, the student’s self-written text is effectively improved in cohesion and coherence. The student’s original text reads:

Genius, in my opinion, it means that someone shows strong learning and application ability in a certain field or multiple fields, and I think that these abilities require innate genetic inheritance. Generally speaking, the IQ above 140 is considered a genius. If the IQ score is too low or insufficient, it will not reach the level of learning and understanding about higher order knowledge. But if a genius doesn't learn new things, he can't do anything. (The underlines are added by the researcher.)

The infinigo-enhanced text, in contrast, has also enhanced the cohesion and coherence of the text although the prompt is just to revise the English text to conform to the English grammar rules. In my opinion, genius refers to someone who demonstrates exceptional learning and application abilities in one or more fields. I believe that these abilities are often innate and inherited genetically. Typically, an IQ score above 140 is considered indicative of genius-level intelligence. Conversely, if an individual's IQ score is too low or insufficient, they may struggle to comprehend and learn higher-order knowledge. However, even if someone is a genius, they must continue to learn and acquire new knowledge in order to apply their abilities effectively. (The underlines are added by the researcher.)

The comparison of the above two texts shows that linking devices such as “typically...conversely...however” are instrumental for EFL teachers to highlight adequate and better linking devices in L2 writing instruction.

**Qualitative Analysis of iWrite Comments on the Two Sets of Writings**

Besides scoring, the iWrite platform provides comments on the quality of student writing. The iWrite comments on the texts are coded artificially into six themes in table three.

<table>
<thead>
<tr>
<th>Table 3. Six themes of the iWrite comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelling</td>
</tr>
<tr>
<td>Sentences</td>
</tr>
<tr>
<td>Grammar and variety and richness of grammatical structures</td>
</tr>
<tr>
<td>Diversity and richness of vocabulary</td>
</tr>
<tr>
<td>Word count</td>
</tr>
<tr>
<td>Cohesive devices</td>
</tr>
</tbody>
</table>

---

Arab World English Journal  
www.awej.org  
ISSN: 2229-9327
The researcher has gathered all 48 comments on the two sets of writings with 24 in each set and made a qualitative analysis. In terms of spelling, it is clear that all of the 48 texts, both students’ self-written or MT-assisted writing and the ChatGPT-polished texts, are free of any errors, as all of them are commented as “the spelling is wonderful and free of any errors”. Thanks to technological assistance, spelling is no longer a problem.

Regarding sentences, the comments show that with MT-assisted L2 English writing, fluency is easily attained, thus playing a lesser role in the global scoring of the texts. Specifically, “sentences are relatively smooth” (10 times, separately or with other comments) concerning texts with a complete range of global scores, including 91.3, 77.6, 64.7, 53.4, 78.3, 63.4, 43, 60.6, 67.7, 78.4, 76.1, 85.5. “Sentences are very fluent” (11 times, separately or with comments on phrasing) concerning texts with a vast range of global scoring, like 86.4, 84.4, 83.4, 80.2, 78.7, 79.4, 72.3, 67.9, 66.2. Alternatively, the comments contain the words “flowing” or “fluid”, like “sentences are relatively flowing”, or “sentences are very flowing”. The comment “sentences are relatively flowing” is used 13 times to describe texts with a whole range of global scoring from 40.1 to 94.3. “Sentences are very flowing” is used 12 times to describe texts whose scores range from 53.1 to 88.1. This shows that with the assistance of MT tools, students’ L2 writings show enhanced and even perfect proficiency on the sentence level, to the extent that well-constructed sentences may play a role in the global scoring of the writings, but the role would be pretty slight. Nonetheless, when an infinigoChatIC-polished text is scored 80.9, in contrast with 64.7 of the student’s original text, the comment does change from “sentences are relatively smooth” to “sentences are very flowing.” In another case, when the ChatIC-polished text scores 94.3, from 84.9 of the student’s text, the comment changes from “sentences are relatively flowing” to “sentences are relatively flowing, with varied and rich syntactic structures.” This change of comments applies to many of the texts in the study.

With regards to grammatical accuracy and variety and richness of grammatical structures, the comments include mostly “There are relatively few grammar errors, demonstrating a good understanding of grammar structures,” which is used to describe the grammar of 22 texts; “Grammatical structures are accurate, and there are no grammatical errors in the essay,” used to describe 13 texts; “there are no grammatical errors in the essay, demonstrating a good understanding of grammar structures,” used to describe six texts. It is shown that the assistance of MT has eliminated most of the grammatical errors in the students’ writing, thus being unable to differentiate between texts with different scores. However, in one case where a rise of 21.8 points, the highest rise of scores among the 24 texts through infinigoChatIC polishing, from 53.1 to 74.9, the comment on grammar is changed from “there are some grammatical errors, and revisions are recommended” to “there are no grammatical errors in the essay, demonstrating a good understanding of grammar structures.” This points to the role that grammar plays in the global scoring of an English essay.

However, it is evident through the comments that with the assistance of infinigoChatIC, one of the most significant changes in scores occurs due to variations and richness of vocabulary. 18 out of the 24 texts written by students with or without MT assistance are commented as follows: “Variations of vocabulary should be enhanced”; “Vocabulary should be more varied”; “The vocabulary is limited, and use of more advanced words and phrases is recommended.” “The text does not show a good grasp of advanced vocabulary.” “The text shows a rather limited vocabulary.” And with infinigoChatIC-polishing, all 18 texts are commented as follows: “The vocabulary is
varied” or the “the vocabulary is rich.” It shows that infinigoChatIC improves the register of the words and phrases used when instructed to correct grammatical errors.

Here is a typical case of enhanced vocabulary complexity, mainly in the form of nominalization:

Version 1: I think genius is the result of the interaction between genes and environment. Essentially, the brain's capacity for intelligence depends on three factors: the number of nerve cells, how well they function, and the complexity of the connections between them. The number of nerve cells in each human brain is basically the same, which is the material basis of intelligence.

However, whether the function of nerve cells is normal and active, and whether the connections between nerve cells are rich or not depends on whether the brain is healthy, whether the nutrition of the brain is sufficient, whether the training is in place and so on. (a discussion draft written by a student with the help of MT)

With the polishing of InfinigoChatIC, the above text reads as follows:

Version 2: My opinion is that the emergence of genius is the outcome of the interplay between genetics and environmental factors. Fundamentally, the brain's potential for intelligence relies on three factors: the quantity of nerve cells present, their effectiveness, and the complexity of their interlinking. The number of nerve cells in each person's brain is fundamentally similar, being the foundation of cognitive ability. Nonetheless, the functionality and activeness of nerve cells and the richness of their interconnections depend on the brain's overall health, sufficient nutrients, proper training, among other things.

The two sets of vocabulary at different complexity levels are listed in Table 4 as follows:

Table 4. Comparing the vocabulary complexity of the above two texts (V1 & V2)

<table>
<thead>
<tr>
<th>Vocabulary complexity</th>
<th>Version 1</th>
<th>Version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think</td>
<td>Genius</td>
<td>My opinion is the emergence of genius</td>
</tr>
<tr>
<td>Genus and environment</td>
<td>The number of</td>
<td>Genetics and environmental factors</td>
</tr>
<tr>
<td>How well they function</td>
<td>Their effectiveness</td>
<td></td>
</tr>
<tr>
<td>However</td>
<td>Nonetheless</td>
<td></td>
</tr>
<tr>
<td>whether the function of nerve cells is normal and active</td>
<td>the functionality and activeness of nerve cells</td>
<td></td>
</tr>
<tr>
<td>whether the connections between nerve cells are rich or not</td>
<td>the richness of their interconnections</td>
<td></td>
</tr>
<tr>
<td>whether the brain is healthy</td>
<td>the brain's overall health</td>
<td></td>
</tr>
<tr>
<td>whether the nutrition of the brain is sufficient</td>
<td>sufficient nutrients</td>
<td></td>
</tr>
<tr>
<td>whether the training is in place</td>
<td>proper training</td>
<td></td>
</tr>
<tr>
<td>And so on</td>
<td></td>
<td>among other things</td>
</tr>
</tbody>
</table>

It is very intuitive for EFL teachers to instruct students to use this table to acquire in a very detailed way about the complexity of vocabulary.

Based on a thorough study of the iWrite comments on the word count or length of the texts, it is found that infinigoChatIC, under the prompt of revising the English text grammatically so that it is grammatically accurate, does not affect the length of the texts in any vital way, mostly retaining the same length (20 out of 24 texts), in some cases slightly increasing and reducing the word count (4 out of 24 texts). Despite the rise of scores due to enhanced lexical and syntactic complexity.
and the elimination of grammatical errors and technical irregularities, insufficient word count or text length leads to a low score. A text that contains 39 words is scored 48, although it is almost perfect linguistically.

With regards to coherence, 18 out of the 24 students’ self-written or TM-assisted writings and their infinigoChatIC-polished versions are commented as “the use of transitional words is appropriate”, or “the writing shows adept use of transitional techniques” or “the use of transitional words is effective”. In four cases, the student’s self-written or TM-assisted writings are commented as “transitional words should be used to enhance the structure of the writing” or “linking words should be used between sentences to improve the flow of the writing.” In contrast, their infinigoChatIC’s polished versions are commented as “the use of transitional words is effective.” This proves that infinigoChatIC can enhance coherence of the writing. Nonetheless, in two cases, both of the students’ writing and their polished versions lack linking devices.

**Lexical Complexity Statistics from Professor Lu’s Autonomous Analyzer**

Then, to derive statistical evidence for the enhancement of vocabulary complexity in the infinigoChatIC-polished texts, the researcher analyzed the lexical complexity of a total of 24 texts written by students and their counterparts polished by infinigoChatIC via Professor Lu Xiaofei’s web-interfaced lexical complexity analyzer (Lu 2010; Lu 2011; Ai & Lu 2013; Lu & Ai 2015). Table Five displays the results of the comparison.

Table 5. *Comparison of lexical complexity of the texts by students and polished by InfinigoChatIC*

<table>
<thead>
<tr>
<th>LD</th>
<th>LS1</th>
<th>LS2</th>
<th>VS1</th>
<th>VS2</th>
<th>CVS1</th>
<th>NDW</th>
<th>NDWZ</th>
<th>NDWERZ</th>
<th>NDWESZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.52</td>
<td>0.32</td>
<td>0.36</td>
<td>0.11</td>
<td>3.80</td>
<td>1.38</td>
<td>658</td>
<td>32</td>
<td>39.80</td>
</tr>
<tr>
<td>2</td>
<td>0.55</td>
<td>0.40</td>
<td>0.42</td>
<td>0.16</td>
<td>8.17</td>
<td>2.02</td>
<td>766</td>
<td>37</td>
<td>40.10</td>
</tr>
<tr>
<td>↑5.8%</td>
<td>↑25%</td>
<td>↑17%</td>
<td>↑45%</td>
<td>↑115%</td>
<td>↑46.3%</td>
<td>↑16.4%</td>
<td>↑15.6%</td>
<td>↑70.75%</td>
<td>↑12.8%</td>
</tr>
<tr>
<td>TTR</td>
<td>MSTTR</td>
<td>CTTR</td>
<td>RTTR</td>
<td>LOGTTR</td>
<td>UBER</td>
<td>VV1</td>
<td>SVV1</td>
<td>CVV1</td>
<td>LV</td>
</tr>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.75</td>
<td>9.09</td>
<td>12.85</td>
<td>0.82</td>
<td>19.47</td>
<td>0.44</td>
<td>58.19</td>
<td>0.42</td>
</tr>
<tr>
<td>2</td>
<td>0.27</td>
<td>0.80</td>
<td>10.20</td>
<td>14.43</td>
<td>0.84</td>
<td>21.04</td>
<td>0.49</td>
<td>79.29</td>
<td>6.30</td>
</tr>
<tr>
<td>↑8%</td>
<td>↑76.7%</td>
<td>↑12.2%</td>
<td>↑12.3%</td>
<td>↑12.43%</td>
<td>↑8.06%</td>
<td>↑11.4%</td>
<td>↑36.3%</td>
<td>↑16.9%</td>
<td>↑4.76%</td>
</tr>
<tr>
<td>VV2</td>
<td>NV</td>
<td>ADJV</td>
<td>ADVV</td>
<td>MODV</td>
<td>UBER</td>
<td>VV1</td>
<td>SVV1</td>
<td>CVV1</td>
<td>LV</td>
</tr>
<tr>
<td>1</td>
<td>0.10</td>
<td>0.41</td>
<td>0.08</td>
<td>0.05</td>
<td>0.05</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>0.40</td>
<td>0.10</td>
<td>0.05</td>
<td>0.15</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 shows the analysis results of the two sets of the first seven texts.
Table 6. Comparison of the complexity level of syntactic structures of texts written by students or polished by InfinigoChatIC (the first seven texts)

<table>
<thead>
<tr>
<th>W</th>
<th>S</th>
<th>VP</th>
<th>C</th>
<th>T</th>
<th>DC</th>
<th>CT</th>
<th>CP</th>
<th>CN</th>
<th>MLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>844</td>
<td>47</td>
<td>115</td>
<td>102</td>
<td>60</td>
<td>42</td>
<td>31</td>
<td>22</td>
<td>111</td>
</tr>
<tr>
<td>2</td>
<td>773</td>
<td>49</td>
<td>92</td>
<td>75</td>
<td>48</td>
<td>26</td>
<td>22</td>
<td>22</td>
<td>111</td>
</tr>
</tbody>
</table>

Table 7. Comparison of the complexity level of syntactic structures of texts written by students or polished by InfinigoChatIC (the second nine texts)

<table>
<thead>
<tr>
<th>W</th>
<th>S</th>
<th>VP</th>
<th>C</th>
<th>T</th>
<th>DC</th>
<th>CT</th>
<th>CP</th>
<th>CN</th>
<th>MLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>992</td>
<td>52</td>
<td>145</td>
<td>109</td>
<td>56</td>
<td>48</td>
<td>29</td>
<td>30</td>
<td>115</td>
</tr>
<tr>
<td>2</td>
<td>978</td>
<td>54</td>
<td>139</td>
<td>105</td>
<td>54</td>
<td>47</td>
<td>34</td>
<td>34</td>
<td>131</td>
</tr>
</tbody>
</table>

Table 8. Comparison of the complexity level of syntactic structures of texts written by students or polished by InfinigoChatIC (the last eight texts)

<table>
<thead>
<tr>
<th>W</th>
<th>S</th>
<th>VP</th>
<th>C</th>
<th>T</th>
<th>DC</th>
<th>CT</th>
<th>CP</th>
<th>CN</th>
<th>MLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>785</td>
<td>52</td>
<td>113</td>
<td>94</td>
<td>48</td>
<td>44</td>
<td>29</td>
<td>20</td>
<td>97</td>
</tr>
<tr>
<td>2</td>
<td>767</td>
<td>46</td>
<td>85</td>
<td>72</td>
<td>44</td>
<td>28</td>
<td>20</td>
<td>22</td>
<td>98</td>
</tr>
</tbody>
</table>

The analysis results of the two sets of the following nine texts are shown in Table 7, with the differences highlighted in different colors.

Table 8 shows the results of the analysis of the two sets of the last 8 texts, with the differences highlighted in different colors.
The figures of some indices fall across all three installments, including word count (W), verb phrases (VP), clauses (C), T-unit (T), dependent clause (DC), clause per sentence (C/S), clause per T-unit (C/T), and those of some indices fall across two installments, including complex T-unit ratio, T-unit per sentence (T/S), dependent clause per T-unit (DC/T), dependent clause per clause (DC/C), verb-phrase per T-unit, and mean length of sentence (MLS) and also complex T-unit (CT), totaling 14 indices.

In contrast, the data of some indices rise across all three installments, including complex nominal per clause (CN/C), complex nominal per T-unit (CN/T), coordinate phrase per clause (CP/C), coordinate phrase per T-unit (CP/T), mean length of clause (MLC), mean length of T-unit (MLT), and the data of some indices rise across two installments, including complex nominal (CN) and coordinate phrase (CP) and sentences, a total of 9 indices.

Discussion
Generally speaking, the syntactic complexity falls with the polishing of infinigoChatIC, with mainly the number of clauses and verb phrases and T-unit and mean length of sentence falling, in contrast to the increase of nominals, coordinate phrases, mean length of clauses and number of sentences.

The result is, in a student’s words, “The polished article has significantly improved its grammar by modifying long and complex sentences into more concise ones, splitting and recombining sentences to condense their meaning, and avoiding fragmented and cluttered sentences. Using simplified sentences makes the article more accessible and easier to understand.” The above data collected from Professor Lu’s syntactic complexity autonomous analyzer provides statistical evidence for students’ intuitive feelings.

This pattern of infinigoChatIC polishing of English texts, specifically, reducing clauses, verb phrases, and T-units and, shortening sentences and adding nominals, and coordinate phrases and, extending clauses provides new evidence to the research finding by Lu & Ai (2015) that the relationship of syntactic complexity in L2 writing to L2 proficiency might not be so straightforward as previously considered. One of the reasons is that there might be an L1 transfer effect in the L2 writing in terms of syntactic complexity (Lu & Ai 2015). For example, despite the previous claim that greater use of coordinate sentences indicates beginning or lower levels of L2 proficiency (Bardovi-Harlig, 1989; Norris and Ortega, 2009), the L2 English writing by Chinese students scored as the lowest level of proficiency among the 7 NNS groups in a study conducted by Lu & Ai in 2015, was found to contain significantly less syntactic coordination than the NS group. This may be explained by an L1 transfer effect, i.e. that in the Chinese language, independent clauses are linked through punctuation instead of coordinating conjunction (Lu & Ai 2015). As a result, when infinigoChatIC polishes the text, it tends to add coordinate phrases.

The use of infinigoChatIC yields great insights into the language knowledge that the L2 student writers need to acquire to improve their proficiency and accordingly the quality of their L2 writing, which is very helpful for EFL teachers. Furthermore, the use of ChatGPT can be encouraged among students to improve the quality of their writing and in this process enhance their L2 writing proficiency. In the study, the prompt is limited to having grammatical errors corrected. However, it is just a step away for EFL teachers and for students to give more varied prompts to ChatGPT, whatever versions mostly commonly seen in the world, to provide ideal language samples and to enhance students’ writing, and to give instructions on how to improve L2 English writing. The teacher researcher has assigned a homework requiring students to combine
Pigai, iWrite, and ChatGPT to correct grammatical errors, and to enhance their writing and then present a study report on the improvement of the writing as shown by Pigai or iWrite automatic scoring system, and what they have learned in this process. The reports submitted by students show that the combined use of pigai, iWrite and ChatGPT has dramatically improved the quality of their writings. In the best case, a student has successfully improved the score of his writing from 69 to 91.5. In addition, by comparing their original writing with the polished version, the students learn in a very intuitive and vivid way what a good English essay looks like and where they can improve.

**Conclusion**

The study aims to examine the pedagogical potential of ChatGPT, or one of its Chinese versions infinigoChatIC, in helping improve the quality of L2 student English writings. The objectives are to answer how infinigoChatIC-polished student English writings perform in comparison to students’ self-written or MT-assisted writings, as evidenced by the results of iWrite scoring and comments and by the results of linguistic complexity analysis by Lu & Ai’s web-interfaced linguistic complexity analyzer, and explore how AI tools like infinigoChatIC facilitate the development of L2 English writing proficiency.

In conclusion, the three research questions have been answered.

First, ChatGPT-polished students’ L2 English writing score significantly higher than their self-written and MT-produced texts in the automated scoring systems such as iWrite.

Second, ChatGPT-polished students' L2 English writing score significantly higher than their self-written and MT-produced texts in terms of lexical complexity and in some measures of syntactic complexity, specifically nominals, coordinate phrases, mean length of clauses, and number of sentences.

Third, the use of ChatGPT can facilitate L2 English writing instruction by providing insights into the language knowledge and essay writing skills that the students may need to improve their L2 writing, both for EFL teachers and for the students themselves.

The current study is significant as it may point to a new gap in the extant literature concerning the L1 transfer effects on L2 writing, explicitly shown in the use of verb phrases (VP), clauses (C), T-unit (T), dependent clause (DC) as well as nominals and coordination as more research is needed with the advent of ChatGPT and other new technologies.

**Funding**

This research is not funded.

**Acknowledgments**

Not applicable.

**Conflicts of Interest**

The authors declare no conflict of interest.

** Authenticity**

This manuscript is an original work.

**Artificial Intelligence Statement:**

AI and AI-assisted technologies were not used.
About the author

Danping Dong, Master of English Language and Literature, is a lecturer at the School of Foreign Languages at Yunnan University, PRC, where she teaches English reading and writing courses, English listening and speaking courses, and English for International Academic Conferences to students pursuing a master’s or a doctor’s degree.

ORCID: https://orcid.org/0009-0009-2724-0921

References


Gilson, A. et al. (2023). How does CHATGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. JMIR Medical Education, 9 (1) (2023), Article e45312, 10.2196/45312


