ISSN 2663-8169 Vol. 6, Issue 10: 184-190, DOI: 10.25236/IJNDE.2024.061027

A Comparative Study on the Effectiveness of Different Automated Writing Evaluation for College English Writings: Taking Pigai and Grammarly as Examples

Li Weiyang

Nanning Normal Univeristy, Guangxi, Nanning, China

Abstract: *This article compares the effectiveness of two automated English writing evaluation tools, the domestic Pigai and the foreign Grammarly, in college English writing teaching. Based on the TEM-8 composition samples of English majors from a university in Nanning, the study evaluates the differences in scoring accuracy, feedback content and form, and student acceptance between these two tools and compares them with the TEM-8 scoring standards. The results show that while both tools have advantages in providing immediate feedback and increasing learning convenience, they have limitations in assessing creativity and complexity. Therefore, it is suggested that these tools should be used as auxiliary tools for teaching and learning, rather than as the sole standard to replace professional human judgment. At the same time, suggestions are made on how teachers should integrate technology and humanities in teaching, and how students should use technological tools to improve their writing skills.*

Keywords: *Automated Writing Evaluation (AWE), English Writing, Pigai, Grammarly, Integration of Technology and Humanities, Creative Thinking*

1. Introduction

Firstly, against the backdrop of globalization, world cultures are evolving towards diversity, English, as a widely used global language, plays a significant rolein the dissemination of culture. We are currently in an unprecedented era of major change, making it crucial to strengthen our cultural confidence and effectively tellChinese stories to the world. English major students in China can promote culturalexchange and dissemination between China and other countries through English writing, acting as a bridge. By showcasing a country's cultural achievements andvalues, they can enhance the national image and soft power. However, writing ahigh-quality English composition has always been a challenge for English educators. In English learning, the cultivation of English writing skills is of paramount importance[1]. Improving English writing skills requires students to work on aspectssuch as vocabulary, grammar, syntax, discourse structure, language expression, andlogical thinking abilities. In other words, the continuous improvement of English writing proficiency to some extent also represents the continuous improvement of overall English proficiency. Therefore, to master English well, it is imperative tostrengthen the cultivation of English writing skills.

Secondly, with the continuous advancement of artificial intelligence technology especially breakthroughs in the fields of Natural Language Processing (NLP) andmachine learning, we have witnessed the development of online essay gradingsystems. These systems utilize advanced algorithms and models to analyze andevaluate students' English writing, providing immediate feedback and suggestions for improvement. Online essay grading systems break the limitations of time and space, allowing students to practice writing and study at any time and place, greatly enhancing the flexibility and convenience of learning; these systems also serve as auxiliary tools for English teachers, helping them to assess students' work more efficiently, thus freeing up more time for personalized teaching.

In summary, the author has selected two essay grading software or websites from many that provide this service, namely the domestie Pigai and the foreign Grammarly. By placing college students' writings into these two grading systems for correction, the aim is to explore the differences in grading accuracy, feedback content and form, and student acceptance between difierent automated grading

ISSN 2663-8169 Vol. 6, Issue 10: 184-190, DOI: 10.25236/IJNDE.2024.061027

systems, providing areference for the effective use of automatie evaluation systems in college English writing teaching.

2. Literature Review

2.1 Automatic Writing Evaluation Overview

Automatic Writing Evaluation (AWE) is a system that uses technology to assess written texts, based on an automatic scoring engine developed from computational. linguistics. This type of assessment can range from the lexical, syntactic, andgrammatical levels to more advanced levels of semantics and discourse. Regarding its origins, it can be traced back to the 1960s when it was known as the Page EssayGrade (PEG). This program applied multiple regression analysis to measure thequality of writing using previously hand-scored essays as reference texts (Shermis.Mzumara, Olson, and Harrington, 200l). In the 1990s, Educational Testing ServiceETS) collaborated with Vantage Learning to develop e-rater and Intellimetric. Thesetwo scoring programs are considered more comprehensive in assessing the vocabulary, syntax, and discourse of writing (Elliot and Mikulas, 2004). Pearson Knowledge Technologies also acquired Intelligent Essay Assessor to score essays. Whatdistinguishes this program from other AWE programs is that it uses latent semanticanalysis and compares the semantic meaning of the text with a corpus under similartopics. Nowadays, various AWE tools can be used to score and evaluate differentforms of writing, such as Criterion and My Access (Li, Link, and Hegelheimer, 2015).

AWE consists of two components. First, AWE tools generate automated scores.They are designed to provide a quick and comprehensive analysis of the text's assessment results so that the evaluation process is not time-consuming. Teachers canuse the scores as a source of feedback to determine future actions. Second, AWE tools provide automatic written corrective feedback (Ranalli, 2018). Students can use the feedback to independently revise their papers, but the role of the teacher remains crucial. As Iswanti et al. (2019) found, this teacher role implies the benefits of implementing teacher-student conferences[2].

2.2 Status Quo of Pigai and Grammarly

2.2.1 Status Quo of Pigai

"Pigai" is an educational platform that utilizes cutting-edge natural language processing technology and a rich corpus of linguistic resources to provide intelligent assessment services for English writing. The platform can quickly perform multidimensional analysis of compositions within l to 2 seconds and provide comprehensive overall scoring. Unlike traditional scoring methods, the scoring system of "Pigai" not only evaluates the overall quality of the composition but also goes into detail with each sentence, identifying errors through 192 detailed dimensions, helping students to clearly identify deficiencies in their writing and thus make more targeted improvements.

The scoring mechanism of "Pigai" relies on cloud computing technology comparing students' compositions with a standard corpus to determine the score. This scoring method is not only eficient but also accurate, providing students with anobjective assessment of their writing abilities. The scoring process is divided intothree parts: scores, overall comments, and sentence comments. The score section shows the total score of the composition, the overall comment section provides an evaluation of the overall quality of the composition, and the sentence comment section specifically points out the errors and improvement points of each sentence. When using "Pigai", teachers can evaluate compositions from multiple dimensions, including grammatical correctness, error distribution, vocabulary richness, and frequency of use. The system also deeply analyzes the use of vocabulary in the composition, including the vocabulary quantity, average vocabulary, and the use of high-frequency and ultra-high-frequency vocabulary. The uniqueness of "Pigai" lies in the fact that it not only points out errors but also provides suggestions for improvement. These suggestions are based on the corpus of native English speakers, helping students to avoid common mistakes and enhance the naturalness and authenticity of their language. In addition, the platform can also identify and correct "Chinglish", helping students to understand and use English more accurately[3].

In summary,"Pigai"is a comprehensive, easy-to-use, and accurate English writing teaching platform. It not only helps students improve their writing skills but also facilitates the teaching work of teachers, providing an efficient and intelligent solution for English teaching.

ISSN 2663-8169 Vol. 6, Issue 10: 184-190, DOI: 10.25236/IJNDE.2024.061027

2.2.2 Status Quo of Grammarly

With the continuous advancement of technology, Natural Language Processing(NLP), Artificial Intelligence (Al), and infonnation retrieval technologies have permeated every aspect of our daily lives, especially in the fields of education andwriting. The emergence of online writing assessment tools such as e-rater and Intelligent Essay Assessor (lEA) marks an important milestone in edueational technology. These tools can not only analyze the content and structure of the text but also deeply understand the ideas presented in the article, and score accordingly, greatly improving the efficieney and accuracy of writing assessment.

The establishment of Grammarly and the launch of its writing assistance platform have pushed the application of this technology to new heights. Since 2009 Grammarly has used advanced Al technologies, including machine learning, deep learning, and natural language processing, to analyze millions of sentences, thereby learning and mastering complex grammatical rules and text structures. Through cloud technology, Grammarly can update its database in real time to ensure the accuracy andtimeliness of the assessment. Grammarly's service versions are designed to meet theneeds of different users. The free version provides basie writing assessment funetions, ineluding grammar and spelling checks. The premium and business versions offer more comprehensive services, including plagiarism detection. This feature helps users identify and avoid potential academic misconduct by comparing with the vast resources in the ProQuest database, ensuring that their writing meets strict academic standards.

In addition, the premium and business versions also provide context-sensitive writing suggestions, allowing users to adjust the wording, tone, and format of the article according to specific contexts, thereby enhancing the professionalism and persuasiveness of the article. This personalized service makes Grammarly not only a writing assistant but also a partner that can help users improve their writing skills. Grammarly's scoring system integrates multiple dimensions, including correctness, clarity, engagement, and delivery, providing users with a comprehensive writing assessment. This scoring method not only helps users identify and correct errors but also encourages users to pursue higher writing quality. Although the last two scoring standards require a premium subseription, they provide valuable feedback for users who pursue excellent writing.

Finally, Grammarly marks sentences that may contain errors with color coding and provides suggestions for modification, which not only improves the user's editing efficiency but also makes the entire writing and revision process more intuitive and friendly. This innovative user interface design further reflects Grammarly's unremitting efforts to enhance the user experience[4].

3. Research Subject and Method

In this study, the composition part of TEM-8 completed by a class of junior English majors in a university in Nanning is taken as the research subject. Two kindsof automated writing evaluations, Pigai and Grammarly, are used to score respectively and then the dimensions and scores of the composition scoring are compared. In order to ensure the validity of the score, the author also compares the two with the scoring standards of the composition part of TEM-8 respectively, and makes these dimensions and data into Excel tables for convenient comparison.

4. Research results

4.1 Comparison between Grammarly and TEM-8 Writing Scoring Standards

4.1.1 Analysis of Data of Grammarly

The author first inputted the composition part of the 2021 specialized English test (TEM-8) for the students of the class into Grammarly in sequence, obtaining scores on a hundred-point scale. Since the composition part of the TEM-8 accounts for 20 points out of the total test score, the author adjusted the hundred-point scale scores from Grammarly to a 20-point scale for comparison. It should be noted that the scores provided by Grammarly represent the percentage of the article quality above the articles already inputted into the platform. The author compiled the data of the entire class and finds from the score distribution that the students' scores show a clear dispersion, with the highest score being 13.8 points and the lowest 5.8 points. This indicates a significant difference in writing abilities among

ISSN 2663-8169 Vol. 6, Issue 10: 184-190, DOI: 10.25236/IJNDE.2024.061027

students. The full score is 20 points, but most students did not reach half of the score, indicating that the overall writing level needs improvement.

In terms of scoring thinking dimensions, it can be concluded that students face varying degrees of challenges in the four key dimensions of writing. In the “Correctness” aspect, there is a large fluctuation in the number of errors, ranging from 6 to 44, indicating that students need more practice and guidance in the use of grammar, spelling, and punctuation. The scores in the “Clarity” dimension show that some students’ articles may not be clear enough in structure and logic, and they needto improve their organizational skills and expressiveness. The low scores in the “Engagement” dimension reflect that many students’ compositions may lack engagingelements, and improvements are needed in the opening of the article, the construction of arguments, and the selection of examples to enhance the appeal of the article. Finally, in the “Delivery” aspect, students also have some issues with the appropriate use of tone and voice, which may affect the persuasiveness and reading experience of the article. Overall, students’ performance in these dimensions shows a certain degree of imbalance, and targeted guidance and practice are needed to improve their writingskills. From the overall trend, students’ problems in the “Correctness” dimension arethe most prominent, which may be an area that teachers need to pay special attention to in teaching. At the same time, “Clarity” and “Engagement” are also areas that need to be strengthened to improve the quality and appeal of students’ writing.

4.1.2 Comparison with TEM-8 Writing Scoring Standards

Comparing with the TEM-8 writing scoring standards, it is evident that Grammarly’s scoring dimensions assess various aspects of language use, including correctness (grammar, spelling, punctuation), clarity (ease of understanding), engagement (the appeal of content and reader involvement), and delivery (the appropriateness of using active or passive voice). These dimensions help students identify and improve specific issues. The TEM-8 scoring standards, on the other hand, categorize grades based on the overall quality of writing and the effectiveness of communication, emphasizing the organization of the article, clarity of the theme, sufciency of supporting details, and appropriateness of language use. It focuses more on the overall impact of the article on the reader, Grammarly’s scoring criteria are automated, focusing on the formal accuracy of language and do not fully cover the depth, creativity, or complexity of writing content. The TEM-8 grading standards are assessed by humans, examining not only the accuracy of language but also the relevance, logic, consistency, and impact on reader understanding. This type of scoring is more comprehensive but also more subjective, relying on the judgment of the graders.

In terms of score distribution, Grammarly’s scores are a continuous range from (to 20, with each student receiving a specific score based on the number of language issues in their composition. This score can very precisely reflect the student’s technical performance in language use. The TEM8 grading standards are interval-based, with each interval corresponding to a proficiency level, from 5 to 1each with a detailed description that outlines the characteristics of compositions at that level and the difficulties readers may encounter. This method is convenient for quickly identifying the general level of students but is not as detailed as a continuous score. From the perspective of student performance, teachers and students can understand the students’ performance and existing problems in specific dimensions such as language correctness, clarity, engagement, and delivery through Grammarly’s scoring. Combined with the TEM-8 grading standards, further analysis can be madeon how these issues affect the overall writing quality and communication effectiveness of students. For example, if a student has a low score in clarity onGrammarly, it may mean that there are problems with the logical organizational structure of their composition, which corresponds to the requirements of “logical organizational structure” and “clarity of theme” mentioned in the TEM-8 grading standards.

4.2 Comparison between Pigai and TEM-8 Writing Scoring Standards

4.2.1 Analysis of Data of Pigai

Analyzing the students’ writingscores,we cansee that thescores are concentrated between 14 and 17 points, indicating that most students are close to but have not reached the full mark. This distribution may suggest that students have acertain foundation in writing but still have room for improvement. The gap between the highest and lowest scores reveals a significant difference in writing abilities among students. Dimension values may represent the proportion of scores or error rates of students on each dimension. A value close to l indicates that the student has hardly any errors in that dimension, while a lower value means there are more issues. These values can be converted into percentages to more intuitively display the performance of students on each dimension. By comparing the values of students

ISSN 2663-8169 Vol. 6, Issue 10: 184-190, DOI: 10.25236/IJNDE.2024.061027

on different dimensions, we can assess the consistency of students in various aspects of writing. For example, if a student scores high on “Text Structure” and “Content Relevance” but low on “Words” and “Sentences”, it may indicate that the student does well in terms of text organization and content relevance but needs to strengthen vocabulary use and sentence construction.

From an overall trend perspective, if most students generally score low on aspecific dimension, it may indicate an area where students generally need to improve. For instance, if most students have a low value on the “Sentences” dimension, it may suggest that more teaching and practice are needed on sentence structure and complexity. By comparing the total writing score with the values of each dimension.we can analyze the relationship between the overall quality of the composition and the performance on each dimension. Generally, students with higher total composition scores also perform better on each dimension, indicating that all aspects of the composition are interrelated. For example, a student who scores high on “Content Relevance” may provide more coherent and targeted arguments overall.

4.2.2 Comparison with TEM-8 Writing Scoring Standards

From the perspective of scoring standards, the scoring system of Pigai assesses the technical aspects of compositions in a quantified manner, such as the correctness of vocabulary, the grammatical structure of sentences, the organizational structure ofthe text, and the relevance of content. The values of these dimensions may represent the proportion of errors or issues, with higher values indicating fewer technical issues. The TEM-8 grading standards focus more on the overall completion of the writing task and the effectiveness of communication, with grading levels decreasing from 5 to1, each with detailed descriptions, including organizational structure, clarity of theme, sufficiency of supporting details, accuracy, and appropriateness of language. This scoring method emphasizes the reader’s reading experience and the difficulty of understanding. Looking at the grading levels, the continuous numerical scoring of Pigai may require some conversion work to correspond with the TEM-8’s interval-based grading system. For example, a high score on Pigai (close to 20 points) may correspond to Band 5 in the TEM-8 grading standards, indicating that the composition has achieved a high level in both technical and communicative effectiveness. Students with lower scores on Pigai may need to improve on multiple dimensions to reach a higher level in the TEM-8 grading standards.

From a technical perspective, technical accuracy is the foundation of overall writing performance but is not the only determining factor. A student may do well in vocabulary and sentence structure, but if the text lacks clear logic and organization, or if the content is not sufficiently relevant to the topic, the overall communicative effect may still be negatively affected. The TEM-8 grading standards emphasize the importance of the text’s organizational structure, clarity of theme, relevance of content, and supporting details, in addition to the technical level. These factors together determine the overall quality of writing and the reader’s level of understanding Automated Scoring Systems like Pigai may have advantages in consistency and objectivity as they are based on preset algorithms to assess compositions. However, such scoring may lack the ability to recognize subtleties in writing, such as creativeuse of language or depth of argument. The TEM-8 grading standards are usually assessed by humans, which may make the scoring more flexible and sensitive, capable of capturing the complexity and nuances of writing. But this also means that scoringmay be influenced by the subjective judgment of the graders, and different gradersmay have different scores for the same composition. Therefore, students and teachers should consider the differences between the scoring systems when using them, and if possible, combine the advantages of both scoring systems to provide a more comprehensive assessment.

4.3 Comparison between Grammarly and Pigai

From a grading perspective, Grammarly’s scoring dimensions focus on the accuracy of language and the effectiveness of expression, including correctness clarity, engagement, and delivery. These dimensions help students identify and improve issues with grammar, spelling, punctuation, and vocabulary usage. Pigai’s scoring dimensions, on the other hand, place more emphasis on the structure and content of writing, including the richness and accuracy of vocabulary, the complexity and fluency of sentences, the logic and coherence of text structure, and the relevance and persuasiveness of content. Grammarly uses algorithms to analyze texts, detecting specific errors and deficiencies in language use, such as grammatical mistakes, spelling errors, and improper punctuation, and then provides a score reflecting linguistic accuracy. Pigai’s scoring method may also employ automated technology but focuses more on assessing the overall structure and content quality of compositions, such as whether the text is organized reasonably and whether the content revolves around the central theme.

ISSN 2663-8169 Vol. 6, Issue 10: 184-190, DOI: 10.25236/IJNDE.2024.061027

Grammarly's scoring results can serve as atool for students to self-check and correct language errors, helping them improve the linguistic quality of their writing and making their expressions more accurate and natural. Pigai's scoring results can guide students to pay more attention to the arrangement of text structure and the logic of content when writing, helping them organize more persuasive and attractive arguments.

From a technical standpoint, while technical scoring cannot fully represent writing ability, it is an essential component of overall writing proficiency. For instance, a student might score high on Grammarly for accurate language use, but this does not mean that their text structure and content development are equally excellent. Conversely, a student might score high on Pigai for text structure and content relevance, but this does not guarantee they are free from language errors. Therefore, by integrating the scoring results from both platforms, a more comprehensive assessment and enhancement of students' writing abilities can be achieved. A single scoring dimension or method cannot fully reflect a student's writing ability. Combining the scoring results firom Grammarly and Pigai provides a more holistic view, helping students understand their performance in various aspects of writing. For example, a student may need to improve linguistic accuracy on Grammarly while enhancing text organization and content depth on Pigai. Through comprehensive evaluation, students can more effectively target improvements in their writing skills and teachers can gain a more thorough understanding of students' learning needs and room for progress.

5. Suggestions

5.1 Suggestions for Teachers

Teachers should promote the integration of technology and humanities. The worship of scientific knowledge and technical means not only reduces educational activities to mere technical issues but also alienates the practitioners and subjects of educational actions[5]. As teachers in the new era, while utilizing automated scoring tools, they should emphasize the importance of manual scoring. This means that in teaching, attention should not only be paid to the language forms and structures that algorithms can recognize but also to students' creative thinking and personalized expression. Teachers should guide students on how to be creative while following language rules, such as by guiding students to analyze the characteristics of different writing styles, encouraging them to try different writing styles and techniques, and discussing language expression and rhetorical methods in literary works.

Teachers should maintain continuous skill development. Encourage students to view writing as a skill that is continuously developed, rather than just part of course assignments. This includes encouraging students to participate in extracurricular writing activities, such as writing clubs, blog writing, and entering writing competitions. Teachers can provide writing exercises on different topics and types to help students apply and develop their writing skills in a variety of contexts. In addition, teachers can also teach students how to improve their writing skills through reading, such as analyzing excellent writing samples and learning how to construct arguments and use thetorical strategies.

Teachers should use data to guide their teaching. With the development of educational informatization, educational big data has been increasingly applied in the management and evaluation of college teachers. By collecting, storing, mining, and analyzing relevant teacher information, data support for teacher evaluation can be achieved[6]. Teachers should regularly review and analyze the data reports provided by the automatic scoring system to gain a macro understanding of the writing abilities of the entire class. These data can help teachers identify the common strengths and weaknesses of the class in specific writing dimensions. For example, if the data show that most students score low on the "Text Structure" dimension, teachers can arrange special courses or workshops to focus on how to build clear paragraphs and logical flow. In addition, teachers can also use this data to adjust teaching methods and materials to ensure they meet the specific needs of students and promote the progress of each student in writing.

5.2 Suggestions for Students

Students should utilize technological tools. In the era of the surge in artificial narrow intelligence (ANI), it has become indispensable for students to use new intelligent tools to assist in their learning. From the perspective of writing, students should regard automatic scoring systems as one of the auxiliary tools for improving their writing skills. These systems can quickly identify language errors and offer suggestions for improvement, but they cannot replace in-depth manual review. Students

ISSN 2663-8169 Vol. 6, Issue 10: 184-190, DOI: 10.25236/IJNDE.2024.061027

should learn to critically accept the feedback from automatic scoring systems understand the limitations of these tools, and integrate feedback from teachers and peers to comprehensively assess their writing. At the same time, students can use these tools to practice identifying and correcting common grammatical, spelling, and punctuation errors, thereby improving the quality of their first drafts.

Students should possess creative thinking. In writing, students should been couraged to go beyond basic language accuracy and develop creative thinking. The creative thinking in writing, which enhances both focused and divergent thinking, often unfolds following a "diverge—converge—diverge again—converge again" approach[7]. This means adopting novel perspectives when constructing arguments andusing creative language and rhetorical devices in expression. Students can stimulate their creativity by reading literary works of different styles and genres, learning how to enhance the expressiveness of their writing through rhetorical means such as metaphor, simile, and personification. Additionally, participating in writing workshops, discussions, and creative writing exercises can help students develop new ways of thinking about problems and apply these methods to their own writing, thereby creating articles that are more engaging and persuasive.

6. Conclusion

Firstly, the two systems offer complementary scoring dimensions, with Grammarly focusing on linguistic details, including correctness, clarity, engagement, and delivery, while Pigai emphasizes text structure and content relevance. This complementarity highlights the multidimensionality of writing skills, indicating that we should consider various elements at different levels when assessing writing. By combining the evaluations of both systems, a comprehensive analysis of writing skills can be provided for students, helping them identify strengths and weaknesses in language use and text organization.

Secondly, although AWE offers the convenience of rapid feedback for teachers and students, it also has its limitations. These systems may not fully replicate the nuanced assessments made by human graders, especially in evaluating creativity and complexity, Therefore, AWE should serve as an auxiliary tool for teaching and learning, rather than the sole standard replacing human professional judgment. Teachers can use these tools to guide their teaching plans, while students can use them for self-assessment and timely adjustment of writing strategies.

Lastly, the use of AWE should promote the development of personalized learning. By analyzing the detailed scoring reports provided by the systems, students can devise targeted learning plans to improve their writing skills purpose fully. At the same time, as technology continues to advance, these systems are also being continuously optimized. Teachers and students should keep abreast of the latest developments and adaptively integrate technology into teaching and learning to achieve the best teaching outcomes.

References

[1] Ding Wangdao. A Manual of English Writing [M]. Beijing: Foreign Language Teaching and Research Press, 2003.
[2] Miranty,D., & Widiati, U. Automated Writing Evaluation (AWE)in HigherEducation: Indonesian EFL Students' Perceptions about Grammarly Use across Student Cohorts [J]. Pegem Journal of Education and Instruction, 2021, 11(4), 126-137.
[3] Dong Yue. Research on the Application of Intelligent Correction in Junior High School English Writing Teaching [D]. Supervisor: Xiao Kairong. Southwest University, 2022.
[4] Wang Chuankai. Research on the Application of Automatic Correction Systems in College English Writing Teaching: A Case Study of Grammarly [J]. Overseas English, 2023, (09): 100-102.
[5] Weng Yucen, Li Lun. The Humanistic Dilemma and Critique of Technologism in Education [J]. Research on Educational Technology, 2024, 45 (07): 19-24+31.
[6] Wei Haiying. Research on the Reform of University Teachers' Management and Evaluation under the Perspective of Big Data [J]. Forum on Education and Teaching. 2024,(26): 181-184.
[7] Qin Xiaofei. Cultivation of Creative Thinking in Junior High School Chinese Topic Composition [D]. Supervisors: Chen Yurong; Jin Rungen. Jiangxi Science and Technology Normal University, 2022.