Research on the Evaluation of Graduate English Writing Course based on iWrite Assessment System

Zhouting Ma[1*] and Xiaoying Song[2]

[1] Shenyang Aerospace University, Shenyang Liaoning 110134, China
[2] Shenyang Aerospace University, Shenyang Liaoning 110134, China
50817217@qq.com

Abstract. The iWrite writing evaluation system assesses students' writing from four dimensions: vocabulary, grammar, sentence structure, and content. This study selected 30 valid samples from the experimental class and compared the differences between computer-assisted evaluation and manual evaluation via data analysis to examine the effectiveness and credibility of computer-assisted evaluation. The research has found that there are slight differences in the focus and scoring criteria between iWrite assisted evaluation and manual evaluation — the former focuses more on vocabulary, grammar, and structure, while the latter focuses more on viewpoints and logic. Apart from improving the validity and objectivity of writing evaluation, the iWrite evaluation system can greatly enhance the evaluation effect and efficiency, thereby effectively polishing students' writing output ability. This synthesis of computer assisted evaluation and manual evaluation will contribute to the improvement of the quality and effectiveness of college English writing teaching and learning.

Keywords: iWrite; computer-aided assessment; graduate writing course

1 Introduction

Automated Essay Scoring (AES) is a modern technology based on computer network platforms for evaluating and grading essays [1]. In 2015, iWrite English writing teaching and evaluation system 2.0 was jointly designed and developed by Professor Maocheng Liang and his research team from Beijing Foreign Studies University and Foreign Language Teaching and Research Press. This system utilizes the operation of multiple intelligent systems to achieve multidimensional, high-speed, and intelligent evaluation of English compositions [2]. The iWrite2.0 system has been studied and used by numerous scholars and teachers in China, and its reliability and validity are constantly being tested. Scholar Zhouchun He [3] found that the system has the advantages of easy operation and comprehensive feedback. Yanling Li [4] obtained empirical data on the reliability of the scoring system by comparing the consistency rate between iWrite and manual assessment of over 600 essays. This study will compare and analyze the different criteria and focus of iWrite and manual assessment to check the credibility and objectivity of iWrite evaluation system.

1.1   Comparison of manual assessment and computer-aided assessment

This study selected 30 valid samples and compared iWrite evaluation with teacher evaluation. These 30 samples cover various score ranges from low to high, and the specific differences are shown in figure 1:
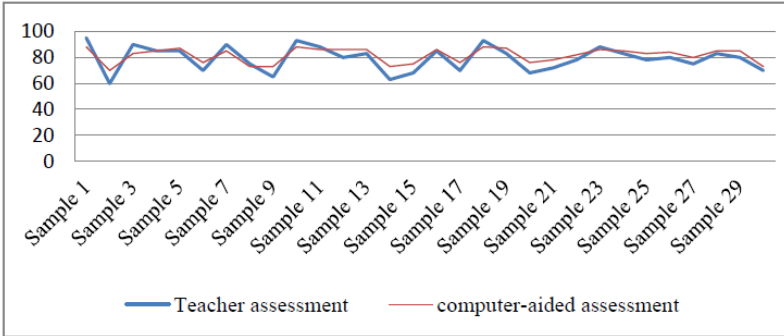

Fig. 1. Comparison of teacher assessment and iWrite assessment

Through comparison, it can be seen that iWrite evaluation scores are slightly higher than teacher evaluation scores in the low to medium score range, slightly lower than teacher evaluation scores in the high score range, and similar in the middle score range. Overall, the difference in scores given by teacher evaluation is greater, while the score span of iWrite is smaller.

1.2   Analysis of manual assessment and iWrite assessment

In addition to the above comparison, this study conducted comparative analysis on 7 dimensions and the specific details, which are shown in table 1 below:

Table 1. Analysis of manual assessment and iWrite assessment

	Manual Assessment	iWrite Assessment
Samples	30	30
Average score	79.5	81.6
Standard deviation	9.481	5.661
Median	80	84.5
Minimum	60	73
Maximum	95	88
Skewness	-0.259	-0.657
Kurtosis	-0.817	-1.092

Among the 30 valid samples with an average distribution of low, medium, and high scores, the average score of teacher evaluation is lower (average score=79.5), and the average score of iWrite evaluation is slightly higher (average score=81.6).

The mean square deviation of teacher evaluation is larger (Standard deviation=9.481), and the mean square deviation of iWrite evaluation is smaller (Standard deviation=5.661). This indicates that the score span of manual evaluation is much larger, which can also be seen from the highest and lowest scores, The D-value between the highest and lowest scores in manual evaluation is 35 points (max=95, min=60), while the D-value between the highest and lowest scores in iWrite evaluation is 15 points (max=88, min=73). The Kurtosis of iWrite is smaller than that of manual evaluation, indicating that iWrite evaluation has more samples below the average score.

1.3   Improvement of compositions after revised by iWrite assessment system

This study records students' scores after being revised according to the suggestions offered by iWrite, and compares the original scores with the scores of the revised versions. The improvement of compositions after revised by i-Write assessment system is shown in figure 2; and the comparison of the three versions of the samples is shown in table 2.
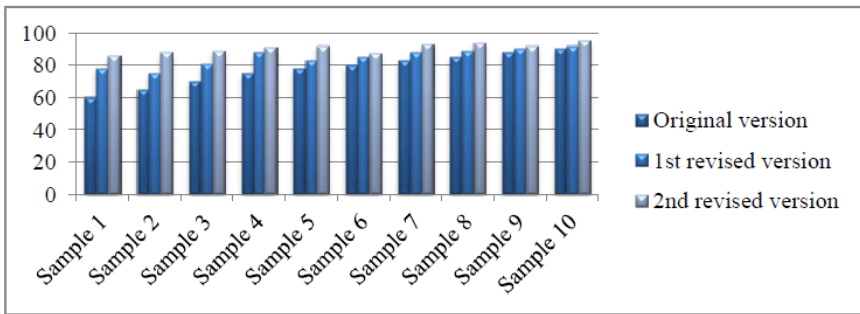

Fig. 2. Improvement of compositions after revised by iWrite assessment system

Table 2. Comparison of the three versions of the samples

	Original Score	Improvement after 1st revision	Improvement after 2nd revision
Sample 1	60	30%	10.26%
Sample 2	65	15.38%	17.33%
Sample 3	70	15.71%	9.88%
Sample 4	75	17.33%	3.41%
Sample 5	78	6.41%	10.84%
Sample 6	80	6.25%	2.35%
Sample 7	83	6.02%	5.68%
Sample 8	85	4.71%	5.62%
Sample 9	88	2.27%	2.22%
Sample 10	90	2.22%	3.26%

It can be seen from the above table that the lower the grade of the compositions (60-75), the greater the increase (15.38%-30%) after the revision according to the revision comments given by iWrite. Part of the reason is that iWrite's score will increase significantly after students correct some spelling and syntax errors and improve sentence structure; For samples with scores higher than 78, the improvement after modification was significantly declined (2.22% -6.41%), because high segmented compositions have fewer spelling errors and grammar problems, and most of the modifications are made to integrate sentences. The second modification generally improves less than the first modification, also because the space for improvement is reduced after the first correction of spelling and syntax error.

1.4   Criteria in manual assessment and iWrite assessment

The grading differences shown in the above research are actually due to the different grading criteria for teacher evaluation and computer-aided evaluation, as is shown in figure 3:
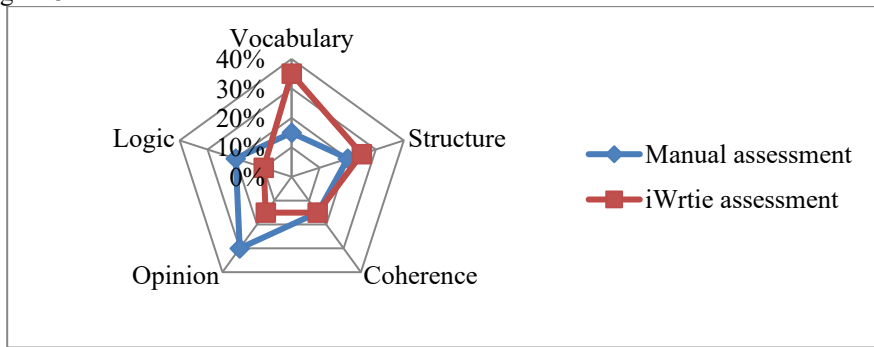

Fig. 3. Criteria in manual assessment and iWrite assessment

The focus of teacher assessment lies more on viewpoints, followed by logic and structure; while the computer-assisted assessment spots and gives corresponding suggestions on the spelling of words and syntax errors, paying less attention to logic and opinions. In brief, teacher evaluation, compared with computer-aided assessment, focuses more on the content of the composition rather than its form; while computer-assisted assessment evaluates grammar, words and other forms more than content and logic. This finding corresponds with other scholar's research on the analysis of the differences and comparison of the above two assessments [5].

2   Implication to college English writing course

2.1   To integrate computer-aided assessment with manual assessment

The iWrite automatic evaluation system evaluates students' grammar, vocabulary and syntax [6]. Teachers can make necessary supplements and propose modification sug-

gestions based on computer-aided evaluation feedback. This evaluation method will be more fair and objective [7]. Teacher evaluation compensates for the insufficient evaluation of article content and viewpoints, while computer-aided evaluation compensates for the low efficiency and heavy burden of teacher evaluation. The two assessments can complement each other and fully leverage the auxiliary role of the iWrite evaluation system.

2.2   To track the students learning curve

Teachers can stimulate students' interest, allowing them to modify and enhance vocabulary, grammar, and structure based on the iWrite evaluation system's ratings and modification suggestions, gradually improving the accuracy of language expression [8]. The system records each student's work and forms a learning archive, including information such as scores on each essay task, class ranking, etc., to facilitate the students to have a clearer understanding of their learning trajectory.

2.3   To utilize the iWrite resources

The iWrite system provides a rich library of writing questions, and teachers can encourage students to choose topics of interest for practice [9]. The types of question banks include various English exams, which cover a wide range of topics, including ethics, politics and economy, energy and environment, education and culture, technological development, daily life, and other topics [10]. The genres of question banks include argumentative papers, expository texts, memorandums, and various types of letters. These question banks will effectively facilitate the students to hone their writing skills.

3   Conclusion

Overall, iWrite has a high reliability in automatic scoring and can parallel manual assessment. This research result is consistent with other domestic research findings guide the students to better exploit the computer-aided assessment. Computer aided evaluation excels in evaluating and correcting spelling and syntax errors, while teacher evaluation focuses more on the content and logic. The manual assessment can leverage the auxiliary role of the iWrite evaluation, which can greatly boost the objectivity and efficiency of the evaluation of college English writing courses.

Acknowledgement

form Research Project — Approval General Project *Research and Practice on Customized Teaching Mode of College English Curriculum in Engineering Colleges Based on BOPPPS Teaching Model in the Context of New Liberal Arts*.

Reference

1. SHERMIS M D, BURSTEIN J. Automated essay scoring: across disciplinary perspective [C] Mahwah, NJ: Lawrence Erlbaum Associates, 2003：7-16.
2. Lin Shi. Journal of Mudanjiang University. "iWrite+" Integrated Practice of Teaching, Learning and Evaluation of English Writing in Large Classes. DOI: 10.15907/j.cnki.23-1450.2021.02.017. Vol.30 No.2. Feb. 2021.109-114.
3. Zhouchun He, Yanzhi Gong: A Case Study on the Application of iWrite English Writing Teaching and Evaluation System 2.0 [J] Journal of Chengdu Airlines Vocational and Technical College, 2017,33 (3): 29-32.
4. Yanling Li, Xiachun Tian. Research on the Reliability of Writing 2.0 Online English Composition Scoring [J] Modern Educational Technology, 2018,28 (2): 75-80.
5. Lan Luan1，Lianzhong Dong2*. Journal of Higher Education. The Reliability of the Automatic Composition Scoring System and Its Implications for College English Writing Teaching: A Comparative Analysis of iWrite System Scoring and Manual Scoring.[J].2022. 33.94-97.
6. Nannan Zhao. (2022) Construction of a Formative Evaluation System for College English Blended Teaching Supported by the Rain Classroom Learning Platform. [J] Science & Technology Vision. (8): 106-109.
7. Lailin Hu, Genping Liu. (2014) Application of Modern Educational Technology [M]. Beijing Higher Education Press, Beijing.
8. Shermis M D, Burstein J. (2003) Automated essay scoring: A cross-disciplinary perspective [C]. Mahwah, NJ: Lawrence Erlbaum associates. 7-16.
9. Shili Ge, Xiaoxiao Chen. (2009) Problems and Countermeasures in the Study of Automatic Scoring of College English Composition [J]. Shandong Foreign Language Education (3): 21-26.
10. Pan Huafang. (2021) A Study on the Influence of Formative Evaluation on College English Autonomous Learning Ability [J] Overseas English, (15): 142-143.