The Analysis of Online EFL Composition Rating System in China

Huijie Li
School of Foreign Languages
Harbin Institute of Technology
Harbin, China

Xiaohui Han
School of Foreign Languages
Harbin Institute of Technology
Harbin, China

Abstract—The major advantage of analytic rating is to give learners feedback and provide useful diagnostic information and about learners’ writing abilities. Automated Essay Scoring (AES) is at present popular for EFL assessment and rating. This paper makes a comparison of two major high-stake large-scale English tests in terms of writing subtests and argues that AES is a bridge connecting the needs of efficient rating and the minimum negative washback of the two tests. The paper takes Bingo online rating system as an example and analyses its reliability and validity.

Keywords—EFL composition; rating system; AES; Bingo

I. INTRODUCTION: METHODS OF WRITING RATING AND ONLINE RATING

There are three main ways of rating EFL writing in language learning and teaching field. They are mechanical methods, analytic method, and holistic method (Jiang, 2018; Yao, 2018)[1][2]. The mechanical method, also called error-count method, usually counts the number of errors and deducts marks according to the amount of the error. The error can be categorized into different groups, including grammatical, subjective, vocabulary, structural, and other errors. As to the second method, the analytic rating scheme separates and weights different features of the learners’ performance on a writing task and assigns separate scores to each feature. The major advantage of analytic rating is to give learners feedback and provide useful diagnostic information and about learners’ writing abilities (Brown, 2004)[3]. White (1984) argues for the advantages of holistic method over analytic scoring[4]. He holds that simpler rating scales generally take less time to score by and therefore generally cost less to use than do more complicated ones. Furthermore, such scores tend to simplify the rater training process, which can be of considerable benefit in situations where the rater pool sees frequent turnover or when there is a frequent amount of rater retraining.

Li & Tian (2018) summarize the revolution stages of Automated Essay Scoring (AES), which is defined as a computer technology for writing assessment and rating[5]. AES provides solutions to the writing scoring. AES develops in three phases: firstly it stemmed from 1960’s Project Essay Grade (PEG) developed by American scholars, which analysed an essay’s quality by certain features; Secondly, it grew during 1990’s out of ETS’ E-rater and Vantage Learning’s IntelliMetric, which compromised more factors like syntax, semantics, contextuality and structure (He, 2016); from the beginning of 21st century, AES matured based on Latent Semantic Analysis technique and accordingly many products came into being such as My Access!, Criterion, Writing Roadmap, etc. Bingo is one of the earliest online rating systems in China.

II. RATING OF EFL WRITING IN CHINA

A. 2 major high-stake large-scale English tests

Chinese students lag in writing compared to in reading (Yang, 2017)[6]. The consequence has part of its origin from the negative backwash effect of two large-scale high-stake English tests (Li & Kong, 2011)[7]. One is the National Matriculation Entrance Test (NMET) and the other is College English Test Band 4 (CET4). The former is for high-school leavers who have finished learning all the grammar rules and have a vocabulary range of 3000 to 3500 words. The latter is mainly for second-year College students, whose vocabulary have amounted to 5000 words. Nevertheless, the requirement for writing ability is much lower. NMET prescribes a composition of about 100 words, and CET asks for candidates to write a composition about 120 words. Both NMET and CET have a very specific limitations to the writing content as is shown in Table 1.

| Writing Section | Words limitation | Rubric words         |
|-----------------|-----------------|----------------------|
| NMET            | 100             | 110 Chinese characters |
| CET 4           | 120-180         | 37 Chinese characters The same rubric outline patter for over 10 years |

B. An example of NMET writing requirement

Suppose you are Li Hua. You learn from the Internet that an international organization is to hold a summer camp and invite students all over the world to join. Please write an email for application. Main contents include:

1. Self-introduction (including English ability)
2. Purposes of participation (introducing China to others; learning about other countries)
3. Wish for being allowed

This paper is sponsored by China’s Foreign Language Assessment Fund Project ZGWYCPJJ2018110B
C. An example of CET writing requirement

Excessive packaging

Outline
1. The phenomenon of excessive packaging is popular;
2. The causes for the phenomenon;
3. My view and suggestions on the phenomenon.

D. Dilemma Chinese tertiary students are facing to

From the rubrics listed above we can see that the requirements of NMET is no more than C-E translation; while CET 4 is obeying the same rubric pattern for over 10 years. The unchangeable pattern inevitably leads to Students’ using pre-prepared sentences patterns for the opening and closing sentences, as well as the composition structure.

Why this phenomenon happens? The predominant reason lies in the rating system. Since they are high-stake tests and they involve millions of candidates, the steadiness of rating process comes to the priority in designing a writing item. The more specific the item is, the steadier the rating process is, which is a good way to reduce the uncertainty in the large-scale test and shorten the time used for rating. However, the negative backwash of the writing test is gaining weight. The function of writing test is not fully developed. Student’s creativity in writing is not encouraged.

Thus, Chinese tertiary students are facing a dilemma: on the one hand, the good master of writing is highly demanded; on the other hand, there is no effective way to improve students’ writing ability. How to make a balance between student’s practice and teacher’s rating deserves thorough research. That is to say, there is a high-demand for the effective rating of EFL writing in China.

III. BINGO ONLINE COMPOSITION RATING SYSTEM IN CHINA

A. Introduction to Bingos

BingoEnglish online composition rating system is developed by Chinese technicians and scholars with the aim to provide timely feedback on students’ English writings. It is intended for Chinese EFL learners at tertiary level. It is a platform of individuality, intelligence and interactiveness (Jiang, 2018) [1]. It is a service product with the function of automatic rating on Students’ massive compositions.

The homepage of the website is friendly. Teachers can log onto the site with user name and pin number; while students have easy access to the designated composition by inputting the composition series number given by teachers in advance.

The procedure of using the system is as follows in Figure 1:

B. Advantages

The system has a couple of functions for teachers’ conveniences. Firstly, it saves a great amount of teacher’s time. The computer can rate the submitted compositions within seconds regardless of the massive articles and words. Secondly, it gives remarks relevant to the compositions, in terms of length, vocabulary use, and the complexity of sentence structures. Thirdly, it provides a word file which collects all the students’ compositions and remarks, which is convenient for teachers to download and edit the compositions. The system also provides an excel file with all the information, such as students’ number, scores, and remarks, which is very helpful for teachers to rank the scores in descending order or ascending order. By comparing the remark of high score student and that of low score, teachers have a clue of the improvement room of writing.

C. Reliability and validity

The revolutionary merit of the BingoEnglish online rating system lies in its swiftness of rating. But the reliability and validity actually are the core of the system (Wu, 2018) [8]. Its reliability is guaranteed by looking at the scores of repeatedly submitted compositions. In one term, each student submitted 4 compositions to the online rating system. The table shows the scores given to the repeatedly submitted articles. It can be seen that the given scores keep consistency to 99.5%.

Table 2 gives statistical evidence to improve that the scoring results of Bingo.
TABLE 2 THE RELIABILITY EVIDENCE OF BINGO ENGLISH ONLINE RATING SYSTEM

| Series number | Composition title | Student No. | Score | Correlation |
|---------------|-------------------|-------------|-------|-------------|
| 10509         | Which Is More Usef... | '1110310116 | 7.9/7.9 | Twice | 0.667* |
|               |                   | '1110410327 | 7.2/7.2 | Twice |          |
|               |                   | '1110420123 | 6.4/6.4 | Three times | 0.859** |
| 10507         | The Importance of... | '1110310115 | 8.1/8.1 | Twice | 0.878** |
|               |                   | '1110410302 | 7/7     | Twice | 0.925** |
|               |                   | '1110610424 | 7.3/7.1 | Twice |          |
| 10506         | The Part of the Ne... | '1110310108 | 5.6/5.9 | Twice | 0.837** |
|               |                   | '1110410328 | 6.4/6.4 | Twice |          |
|               |                   | '1110420122 | 5.7/5.7 | Three times |          |
|               |                   | '1110610426 | 6.8/7  | Twice |          |
|               |                   | '1110310113 | 8.5/8.6 | Twice |          |
| 10505         | How to Keep Psychol... | '1110420104 | 6.5/6.5 | Twice | 0.957% |
|               |                   | '1110420113 | 7.1/7.3 | Twice |          |
|               |                   | '1110420118 | 6.3/6.2 | Twice |          |
|               |                   | '1110420419 | 9.2/9.7 | Twice |          |
| 10476         | The Different Learni... | '1110610428 | 6.7/6.7 | Twice | 0.957% |
|               |                   | '1110610430 | 6.4/6.5 | Twice |          |

Table 3 shows the correlation of the average scores rated by the system and by the teachers.

It can be seen from the table that as to each composition, the correlation value is significant ranging from 0.667 to 0.925. It means that the rating score given by Bingo and by the teachers are significantly correlated, which gives strong proof of the validity of Bingo rating system.

D. Ss feedback based on a questionnaire

A questionnaire of 25 items was administered to 103 subjects. 98 valid answered questionnaires were collected. Of the 25 items, 21 items are constructed with multiple choices; 4 of them are open. The subjects are from 5 classes of 3 departments. 82 subjects are male, and 16 are female.

The following table shows the subjects’ attitude to online composition rating.

TABLE 4 STUDENTS’ ATTITUDE TO BINGO ENGLISH RATING SYSTEM

| Statement | Positive | Neutral | Negative |
|-----------|----------|---------|----------|
| Overall attitude to online submission and rating | 94.7% | 3.52% | 1.78% |
| If there is any improvement on writing competence | 92.23% | 5.11% | 2.66% |
| If it is a burden to submit online | 88.63% | 7.89% | 3.48% |
| If online submission is helpful for improving writing ability | 87.87% | 10.26% | 2.87% |
| If the rating given by the system is reasonable | 85.13% | 11.56% | 3.41% |
| If the given remark is reasonable | 81.47% | 12.61% | 5.82% |
| If the remark gives the direction for writing improvement | 77.65% | 14.34% | 8.01% |
| If it is necessary to submit compositions online | 88.57% | 10.39% | 1.04% |
| If you have the intention to use the system next term | 95.36% | 2.24% | 2.40% |

As to the open questions, the subjects gave detailed explanation. They listed various improvement areas in writing practice, including vocabulary buildup, sentence structure complexity, and the increasing length of the composition, etc. They summed up the merits of the online rating system in these aspects as objectivity, no bias, convenience for reviewing, making use of internet resources, and practicing typing, etc.
E. Concerns

While the online rating system has a number of merits, some flaws cannot be neglected. Firstly, cheating of many kinds, such as copying sentences from the Internet, cannot be prevented; secondly, some remarks are ambiguous or blurred and sometimes of little difference; thirdly, the spelling ability is degraded.

IV. CONCLUSION

Chinese tertiary students are facing a dilemma that they are eager to improve their EFL writing capability but they are given insufficient writing tasks. The reason is that it is impossible for teachers to read and rate all their compositions every week since each teacher usually has around 100 students at hand (Wang, 2017).

BingoEnglish online composition rating system, developed by Chinese technicians and scholars, aims to provide timely feedback on students’ English writings. It is intended for Chinese EFL learners at tertiary level. The use of the system to a large extent stimulates students’ interest in writing and the rating comments are expected to give direction for improvement. Statistics indicate that the reliability and validity of the rating can be guaranteed. Students’ feedback on the system reveals that most students have positive attitude towards the software. The main concern of the system is how to avoid cheating of any form.

REFERENCES

[1] J. Jiang, “Research on Feedback of AESS in Senior High School English Teaching – based on Bingo System”[D]. Harbin Normal University, 2018.
[2] J. Yao, “Automated Essay Scoring in the Formative Assessment of Chinese EFL College Students” [D]. SHISU, 2018.
[3] H. Brown. “Language Assessment - Principles and Classroom Practice”[M]. Upper Saddle River, Pearson Education (US).2004.
[4] White, J., & Lightbown, P.M. (1984). Asking and answering in ESL classes. Canadian Modern Language Review. 40, 228-244.
[5] Y. Li, X. Tian, TIAN Xia-chun, “An Empirical Research into the Reliability of iWrite 2.0” [J]. Modern Educational Technology. 2018,28(02):75-80. In Chinese.
[6] Q. Yang, “Analysis of Writing Marking Schemes of CET 4 & 6” [J]. Journal of Higher Education. 2017(16):181-182+185. In Chinese.
[7] Q. Li, W. K, “Review on EFL Writing Assessment Scoring”[J]. Foreign Language Testing and Teaching. 2011(04):18-26+41.
[8] X. Wu, “Rating Scales for Summary Writing Assessment: construction and validation”[D]. SHISU, 2018.
[9] K. He, “The Key Technology Research on Automated Scoring” [D]. Hubei University, 2016. In Chinese.
[10] F. Wang, “An Empirical Study on the Effectiveness of Automated Essay Scoring in EFL Writing” [J]. College Foreign Language Teaching & Research. 2017:107-119+167-168. In Chinese.