A comparison between Automated Writing Evaluation and Teachers' Feedbacks

----An Experimental Study based on Juku Correcting Net

W.Y. WU, S.R. ZHANG
Institute of Graphic Communication, Beijing, China

ABSTRACT: In order to find out whether Automated Writing Evaluation exemplified by Juku Correcting Net is reliable and valid and able to improve students' writing ability, the author compares it with teachers' feedbacks by analyzing students' samples and pre-test as well as post-test writing scores. The results show that the scores given by the teacher and Juku are consistent despite an insufficient degree of differentiation. Compared with feedbacks from teachers, the feedbacks from Juku are more general and just give students some hints about their usage of vocabulary and phrases in spite of offering some knowledge about how to distinguish some synonyms. After practising writing on Juku for a semester, students make some progress in it. It concludes that Juku is expected to be developed to a more intelligent degree and be combined with other means of feedbacks.

KEYWORD: Automated Writing Evaluation; Juku Correcting Net; Teachers' Feedbacks; Reliability; Validity

1 INTRODUCTION

Feedback is one of the most important aspects in English writing teaching process. Effective feedback can accurately express the readers' expectation of the writers and in turn encourage them to revise their writing based on it[1].

According to Longman Dictionary of Language Teaching and Applied Linguistics[2], feedback is defined as the comments or information the learners obtain from the teachers or other learners. In terms of writing, feedback refers to the input that comes back from readers to the writer [3]. Another expert [1] defined feedback as information that a reader offers to writers for further revision of composition, including comments, challenge, suggestions and etc.. Ur [4] further clarify the term by saying that feedback is the information that the writer receive about his or her performance of a writing task, always with the purpose of improving the performance. Based on the definitions above, correction and assessment composes the two distinguishable components of feedback. Correction refers to the detailed explanation and guidance that the reader provides in the expectation of helping the writer find out and then correct his or her shortcomings. Assessment means general and holistic comment or scoring that the reader gives the writer [5].

With the definitions above combined, in this paper, feedback refers to readers' responses to the writers' composition in forms of correction, comments, suggestions or grading in hope of facilitating the writers to improve their writing ability gradually.

In perspective of the origin of a feedback, it can be classified into 3 types: self feedback, peer feedback and teacher's feedback [1][6]. With the development of CALL (Computer-Assisted Language Learning), however, online writing and evaluation system emerged and has been widely adopted by a large number of teachers and students. Therefore, a fourth type of feedback appears, i.e., AWE (Automated Writing Evaluation). In this paper, teacher's feedback and AWE are focused.

Wiseman & Hunt [7] defined the teacher's feedback as information provided to students by teachers, either in oral or written form, which let them know about their writing performance in the process for the purpose of improvement; the more specific, clear and in depth the feedback, the better. AWE is defined as "the ability of computer technology to evaluate and score written essays"[8]. It was in the early 1960s that AWE system emerged in America [9]. Since then, a variety of AWE systems were developed in foreign countries. Among more than ten successful systems put into use, PEG(Project Essay Grade), IEA (Intelligent Essay Assessor) and E-rater(Electronic Essay Rater) are the

most popular ones. PEG bases scores and comments of a written essay mainly on the forms of language and IEA puts emphasis on the content while E-rater takes both of language form and content into consideration. These AWE systems, however, are mostly applied to the evaluation and feedback on essays by native English writers [10].

2 THE FEATURES AND APPLICATION OF JUKU CORRECTING NET

Juku Correcting Net (mentioned as Juku below) is an AWE system developed by Beijing Ciku Technology Co.ltd and is now widely used in thousands of colleges and universities in China. It is an online AWE system based on corpus and cloud computing in which score and comments are offered by measuring the distance between students' essay and the essays in the corpus. It not only stimulates students' interest in writing and revising their English essays but it also relieves the teachers' burden of correcting essays [11].

Juku Correcting Net enjoys the following features:

First, User-friendly;

Both teachers and students are supposed to get an account once they register on Juku. Then they can log onto it with the account. The teacher gets a number after assigning a writing task. When the students learns of the number from the teacher, they can log on, search the essay number and then finish it on the net and finally submit it.

Second, quick feedback;

Upon submission, student writers immediately get the revision suggestions and scores from Juku. The net automatically analyzes the essay, pointing out the mistakes of vocabulary, collocation, grammar and etc..

Third, sentence-by-sentence comment;

Juku not only points out language mistakes in terms of spelling, grammar, vocabulary, collocation and etc. in each sentence but it also analyzes the types of mistakes by offering student writers specific language points, suggested collocations and examples so as to guide students to improve their writing ability in linguistic aspects.

Fourth, Plagiarism detection;

Every piece of writing submitted is detected to find out whether it is partly or wholly identical to the writings from other classmates, the corpus or the internet for the teachers' reference.

Fifth, report on the common features of the whole class's essays;

Juku automatically summarizes the language features of all the essays of the same number submitted by the whole class, including the usage of vocabulary, frequency of vocabulary, the length of

sentence, different collocations, error distribution and etc..

Since 2012, Juku Correcting Net has been in use in our college. Both teachers and students think it does have the advantages mentioned above but it is not without some problems, for instance, the scores given by Juku are sometimes different from those given by teachers. Therefore, it is necessary to make a comparison between teacher's feedbacks and the feedbacks given by Juku so that it can be adopted in a more scientific way to improve students' writing capability.

3 A COMPARISON BETWEEN TEACHER'S FEEDBACKS AND AWE FEEDBACKS BASED ON JUKU CORRECTING NET

3.1 Research questions

In order to explore whether Juku Correcting Net can give the essays reliable and valid feedbacks in terms of scores, corrections and comments, this study aims at answering the following questions: (1) Are the scores given by the teacher and Juku consistent? (2) What are the differences between the comments given by the teacher and Juku? What are the disadvantages of Juku in the perspective of feedback? (3) Can Juku help students improve their writing ability after adopting it for one semester?

3.2 Subjects

For the convenience of the study, the participants are 46 sophomores of non-English majors taught by the author.

3.3 Research Period

This study lasts for one semester. All the students are required to sumbit one essay assigned by the teacher to Juku Correcting Net every 3 weeks and the net offers comments and scores immediately. Students revise their writing based on the comments given by Juku. In the 16-week semester, each student submits 5 essays.

3.4 Research Instruments

In view of the research questions above, the instruments of the study cover analysis of writing sample as well as pre-test and post-test.

For each writing task, 5 pieces of essays are picked out at random and are given feedbacks by both the teachers and the system. Their differences in feedback format, feedback emphasis, comments and scores are compared.

Pre-test and post-test writing tasks are designed to find out whether students make any progress in writing after one semester's adoption of Juku

Correcting net. Students are required to finish the tests in the classroom within 30 minutes, with the pre-test conducted in the first class of the semester while the post-test in the last class. Similar topics are selected to guarantee the pre-test and post-test writing tasks are of the same difficult level. In both tests, the outlines are given in Chinese so that students can understand the requirements clearly. Pre-test and post-test writing tasks are as follows:

Pre-test writing task:

You are allowed 30 minutes to write a composition on the topic of On Wechat based on the outline given below. You should write at least 120 words.

1. More and more people are using WeChat, including some elder people.

2. WeChat brings us a lot of conveniences, i.e. sending messages, writing diaries, reporting feelings, reading news and financing.

3. In my opinion,……

Post-test writing task:

You are allowed 30 minutes to write a composition on the topic of On Online Shopping based on the outline given in Chinese below. You should write at least 120 words.

1. Online shopping is rather popular nowadays and almost everyone have had such experience.

2. Online shopping brings us both conveniences and some bothers.

3. In my view,……

All quantitative analysis in this paper are conducted by SPSS 19.0.

3.5 Research Results

3.5.1 A Comparison of the scores given by Juku and teachers

As is mentioned above, for each writing task, 5 pieces of essays are picked out at random and are given feedbacks by both the teachers and Juku. Therefore, 25 pieces of essays are selected to be scored by the teachers altogether. To guarantee students' writings are scored and commented reliably, three teachers who have participated in marking CET-4 writing test are invited to give feedbacks to the same piece of essay in accordance to the same standard. Reliability analysis shows the three teachers' scores are consistent for Cronbach's Alpha is 947. It is assumed that the items are correlated when Cronbach's Alpha is more than .6.

The mean of the 3 teachers' scores of the same essay is finally regarded as the teacher's score, which is then compared with the score by Juku.

Reliability analysis demonstrates that the scores given by teachers and Juku are consistent (Cronbach's Alpha= .672>.6). Therefore, the scores by Juku are reliable.

Despite the consistency of the scores by the teacher and Juku, paired-samples t-test is conducted to find out whether there is statistically significant difference between the two groups of scores.

The result indicates that there is no significant difference between the scores given by teachers and Juku. That means the scores given by Juku are reliable and accurate to some extent.

However, when the 25 pieces of essays are divided into groups of high, middle and low scores according to the mean scores of the teachers, the analysis results are different.

In high-score group and low-score group, the scores given by teachers and Juku are significantly different since P is .002 and .032 respectively. In middle-score group, there is no significant difference between the scores from the two sources (P=.336) . In high-score group, the scores given by Juku are lower than those by teachers (Mean score by teachers=11.76; Mean score by Juku=8.64) while in low-score group the case is the opposite (Mean score by teachers=5.29; Means score by Juku=6.68). That means the degree of differentiation of the scores by Juku is insufficient.

3.5.2 Analysis of a writing sample

The following sample is selected from the 25 pieces of essays scored by both teachers and Juku.

University Students' Pursuit of Famous Brands

Nowadays, many university students like to wear the clothes and use the things of famous brands. They think their pursuit of latest release is important, If they don't do these, they will lose their face.

Dut I don't think so, One person wear clothes and eat food should suit their identity. We are university student, we shouldn't wear clothes and use things which is very expensive. Because we don't make money form woking yet. Even if we work, we also wear the decent and comfortable clothes, rather than famous brand.

It is important to use things compatibly. Compatible things are beautiful for us. Somrtime, famous brands don't suit you.

The following is a comparison of feedbacks by Juku and 3 teachers based on individual sentences of Paragraph 2 in this sample.

Sentence 2.1 *Dut I don't think so, One person wear clothes and eat food should suit their identity.*

Feedback by Juku:

[Sentence mistake] This sentence is grammatically wrong. Please check it.

[Low frequency warning] The phrase "suit...identity" cannot be found in the corpus，probably a Chinglish expression.

[Learning hint] Please distinguish the following words: clothes, clothing, coat, dress, garment, robe, gown, uniform, costume, suit.

[Correcting hint] "Clothes" means "clothing in general". Pay attention to its difference from "cloth".

[Correcting hint] Pay attention to the difference between "person" and "people".

Feedback by teacher 1: *1)"Dut": misspelling or misprint; correct it into "But"; 2) Comma should be corrected into period; 3) Double predicate in the sentence "one person……". Possible correction: One should wear clothes and use the things that suit their identity.*

Feedback by teacher 2: the same as Teacher 1.

Feedback by teacher 3: Teacher 3: *1) the same as Teacher 1 in the first two mistakes; 2) Possible correction for the sentence "one person……": What one person wears and eat should suit his or her identity.*

Sentence 2.2 *We are university student, we shouldn't wear clothes and use things which is very expensive.*

Feedback by Juku:

[Collocation counting] The phrase "use...thing" appears 436 times in teaching materials.

[Collocation counting] The phrase "wear...clothes" appears 2040 times in teaching materials.

[Learning hint] Pleaes distinguish the following words: clothes, clothing, coat, dress, garment, robe, gown, uniform, costume, suit.

[Correcting hint] The synonyms of "very" include extremely, exceedingly, remarkably and intensely。

[Correcting hint] "Clothes" means "clothing in general". Pay attention to its difference from "cloth".

Feedback by teacher 1: 1) "student" should be used as plural form "students"; 2) the two sentences should be connected by using a link word like "so", "therefore" and etc.; 3) "which is very……" should be corrected into "which are very……".

Feedback by teacher 2: the same as teacher 1.

Feedback by teacher 3: the same as teacher 1.

Sentence 2.3 *Because we don't make money form working yet.*

Feedback by Juku:

[Collocation counting] The phrase "make...money" appears 20192 times in teaching materials.

[Learning hint] Please distinguish the following words: mould, forge, form, shape.

Feedback by teacher 1: *1) "form": misspelling or misprint; correct it into "from"; 2) This sentence cannot be an independent sentence since it is an adverbial clause showing a reason. It should be dependent on the second half of 2.2. Possible correction for 2.1 & 2.2: We are university students, so we shouldn't wear clothes and use things which are very expensive, for we don't make money from work yet.*

Feedback by teacher 2: the same as teacher 1.

Feedback by teacher 3: *1) the same as teacher1; 2) Possible correction for 2.1 & 2.2: We are*

university students, so we shouldn't wear clothes and use things which are very expensive because we don't make money from work yet.

Sentence 2.4 *Even if we work, we also wear the decent and comfortable clothes, rather than famous brand.*

Feedback by Juku:

[Collocation counting] The phrase "wear...clothes" appears 2040 times in teaching materials.

[Learning hint] Both "brand" and "trademark" means "a formally registered symbol identifying the manufacturer or distributor of a product". Pay attention to their differences.

[Correction hint] "Clothes" means "clothing in general". Pay attention to its difference from "cloth".

[Correcting hint] Pay attention to the differences between decent and descent.

Feedback by Teacher 1: 1) Add "should / are supposed to" before the word also to make the mood of the sentence suitable; 2) "brand" should be used as plural form.

Feedback by Teacher 2: the same as teacher 1.

Feedback by Teacher 3: *1) the same as 1) and 2) in teacher 1; 2) delete the second comma and add "those of " before famous.*

To sum up, most of the feedbacks from Juku are based on vocabulary, phrases and collocations while those from the teachers cover all the mistakes, including mistakes in sentences. Besides, feedbacks by juku are nothing but hints, too general and not concise at all. In contrast, feedbacks by teachers are more clear and direct. They not only point out the mistakes but offer suggested corrections as well. However, teachers seldom offer students some knowledge about how to distinguish some synonyms while Juku does a really good job in this aspect, which does play a significant role in helping students grasp vocabulary more accurately by frequently polishing their essays in accordance with the suggestions by Juku [12].

As far as general comments on the sample are concerned, the differences between Juku and the teachers are as follows:

General comments by Juku: *Excellent spelling; skilled use of transitional words and link words; too many simple sentences; a few sentence mistakes; more use of academic words in the future.*

General comments by teacher 1: *The general structure is good and you have fairly good points of view. You need work further to improve the grammar such as connection between sentences. Pay attention to the proper use of punctuations. Type your piece more carefully. More reading will help you improve your writing.*

General comments by teacher 2: *The structure is well-organized and the view is clearly expressed. However, the mood of some sentences is not*

suitable. Pay attention to proper use of auxiliary verbs and modal verbs. More efforts need to be paid to solve your problems in punctuation and grammar.

General comments by teacher 3: *well-organized passage with well-illustrated viewpoints. Pay attention to your mistakes in punctuation and grammar, i.e. subject-verb agreement, broken sentences, singular and plural forms, double predicates and etc.*

Compared with teachers' comments, Juku doesn't give any feedbacks on the structure, contents, and logics of the passage and it just centers around the obvious literal mistakes. Moreover, the comment in Juku is too general and doesn't point out the specific weak points students need polish in the future. Therefore, like other automated Essay Scoring systems, Juku Correcting Net fails to give an accurate comment on the internal quality of an essay and it just generally comments on its language [13].

3.5.3   *pre-test and post-test*

Both of the pre-test and post-test are marked by Juku. All the 46 students' scores are analyzed by paired-samples t-test. The results shows that students have made fantastic progress in their writing ability after one semester's writing practice in Juku Correcting Net, for there is a statistically significant difference between pretest scores and post-test scores(P=.005).

To find out whether the progress is brought about by students' revision based on the suggestions given by Juku, the correlation ratio between submission times and post-test scores is analyzed, which indicates that students' post-test scores correlates with the times they submit their essays after revision because the Correlation Ratio=.281 and P= .029, which is within the acceptable range of less than level .05.

4   DISCUSSION

Major findings yielded from one-semester experimental research contain the following points:

First, generally speaking, the scores given by teachers and Juku are consistent. That means when Juku scores a high mark, the teacher's mark is also high and vise versa. Consequently, Juku is a reliable system in scoring essays. Nevertheless, the degree of differentiation of the scores by Juku is insufficient. Compared with teachers' scores, Juku tends to offer relatively lower scores to those excellent essays while giving higher scores to those poor essays.

Second, in the perspective of either sentence-by-sentence feedbacks or general comments, Juku mainly focuses on such language points as vocabulary, phrases and collocations while the teachers cover all kinds of aspects, including structure, contents, logics of the passage,

complexity, fluency and variety of sentences in addition to evident language points. Therefore, Teachers are more flexible in giving feedbacks. Besides, feedbacks by Juku are too general and not specific at all by just giving students some hints about their mistakes. On the contrary, feedbacks by teachers directly and clearly point out the mistakes and offer them suggested corrections. However, one valuable characteristic of Juku is that it offers students some knowledge about how to distinguish some synonyms. It is this outstanding advantage that helps students grasp vocabulary more accurately by frequently polishing their essays in accordance with the suggestions by Juku.

In view of the points mentioned above, it requires the efforts of both technicians and linguists to improve the reliability of Juku. A new automated essay scoring model is expected to be set up, which will contain structure, contents, logics of the passage as well as complexity, fluency and variety of sentences in addition to evident language points [11]. Moreover, in order to make up for its disadvantages, it's advisable to combine the teacher's feedback and peer feedback with the feedbacks by Juku.

Third, after practising writing on Juku Correcting net, students do make some progress in writing, which is a probable result of the circular process "submission, feedback and revision" that Juku allows. Based on this, Juku is a valid Automated Essay Scoring net and is capable of helping students improve their writing ability.

5   CONCLUSION

Real-time feedbacks by Juku Correcting Net help students improve their writing ability as a result of circular process "submission, feedback and revision". Regardless of its relatively reliable scoring and valid comments, however, Juku fails to give essays an all-around feedbacks due to its drawbacks in scoring model. Therefore, Juku is expected to be developed to a more intelligent degree and be combined with other means of feedbacks.

REFERENCES

[1] Keh, C.L. 1990. Feedback in the writing process: a model and methods for implementation. *ELT Journal*: 44.

[2] Richards, J.C. & Schmidt, R. 2003. *Longman Dictionary of Language Teaching and Applied Linguistics*. Beijing: Foreign Language Teaching and Research Press.

[3] Arndt, V. 1993. Response to writing: using feedback to inform the writing process. In M. Brock & L. Walters (Eds.), *Teaching Composition around the pacific Rim: Politics and Pedagogy*: 90-116. Cleveland: Multilingual Matters.

[4] Ur, P.A. 1998. *Course in language teaching: practice and theory*. Cambridge: Cambridge.

[5] Zhang Xuemei & Dai Weidong. 2001. Feedbacks on second language acquisition and language teaching. *Foreign Language World* (2): 2-8.

[6] Charles, M. 1990. Responding to problems in written English using a student self-monitoring technique. *English Language Teachers' Journal* (44): 193-286.

[7] Wiseman, D. G. & Hunt, G.H. 2001. *Best practice in motivation and management in the classroom*. Springfield: Charles C. Thomas Publisher LTD.

[8] Shermis, M. & Burstein, J. 2003. *Automated essay scoring: a cross disciplinary perspective*. Mahwah: Lawrence Erlbaum Associates.

[9] Page, E.B. 2003. Project essay grade: PEG. In M. D. Shermis & J. C. Burstein (eds), *Automated essay scoring: a cross-disciplinary perspective*. Mahwah: Lawrence Erlbaum Associates.

[10] Liang, Maocheng. & Wen, Qiufang. 2007. A critical review and implications of some automated essay scoring systems. *Media in Foreign Language Instruction* (5): 18-24.

[11] Shi Xiaoling. 2012. A tentative study on the validity of online automated essay scoring used in the teaching of EFL writing---exemplified by http://www.pigai.org. *Modern Educational Technology* (10): 67-71.

[12] He Xuliang. 2013. Reliability and validity of the assessment by the Pigaiwang on college students' writings. *Modern Educational Technology* (5): 64-67.

[13] Ge, Shili & Chen Xiaoxiao. 2009. The key problems and solutions in automated essay scoring for college English teaching in China. *Shandong Foreign Language Teaching Journal* (3): 21-26.