

Experimental Study on the Validity of AES Systems in the College EFL Classroom

Li Li

School of Foreign Languages
Shanghai University of Political Science and Law
Shanghai, China
lily2211@126.com

Abstract—Online Automated Essay Scoring (AES) is used to promote the teaching of EFL writing partly, but it can't work very effectively for some reasons, mainly including misuse and the shortcomings of AES itself. The study tries to investigate the validity of AES by analyzing a questionnaire survey by means of a statistical tool SPSS and discusses the problems existing in the use of AES and the shortcomings in AES itself. Then accordingly some suggestions are presented on how to use AES more effectively so as to enhance students' stimulus and motivation in writing and improve their writing ability and on how to improve AES.

Keywords—automated essay scoring; EFL writing teaching; validity analysis

I. INTRODUCTION

English writing teaching is the important component of college English teaching in China, but it has been widely acknowledged that English writing teaching is a rather tough and challenging task. To a certain degree, it is considered as a burden for teachers to respond student essays. Every teacher of English has a large number of students. If they assign frequent assignments, it might be rather time-consuming to offer individual feedback to student essays. Consequently, parts of teachers are unwilling to assign more essays, and though some do, they only assess part of student essays, namely half or a third of all the students. What's more, while assessing these essays, most of teachers merely underline the mistakes made in the essays and then grade or score them, which demotivates students' interest in writing and minimizes the effective feedback to student essays. In effect, feedback is an essential aspect of the writing process. Students need to receive feedback from the teacher in order to increase their writing quality. As far as the students are concerned, some of them can't put stress on the feedback that their teacher provides. Though part of them revises the essays accordingly, they seldom put in a lot of time and energy into the content of essays. Instead, they emphasize the form more, which can't be helpful for them to improve their writing quality and increase their writing ability. In the process of English writing, a number of students can't complete their writing actively. After writing their essays, they don't read any more and ignore the very important step of revision. So the only reader and evaluator is their teacher. Teachers and students can't interact with one another. Nor can't students themselves. Therefore, students don't have much knowledge of how to improve the writing quality by revision. Students' initiative can't be displayed in the English writing.

As the growing use of both computers and the Internet in English language teaching, such online communication methods as online discussion, online writing, QQ, forums, SKYPEE, are employed as an instructional strategy and put into practical teaching by part of teachers in China, which contributes to the reform carried out in English language teaching. In particular, a variety of AES systems are used to help teachers to assess student essays and students to improve their writing proficiency to a great extent.

The paper mainly discusses the validity of AES systems in enhancing students' interest in writing and improving their writing proficiency. The study is a naturalistic classroom-based inquiry that was conducted in five EFL college writing classroom contexts in a university in China. Data acquired mainly include four aspects: students' self-evaluation of their writing ability, the effectiveness of AES system, the practical use of feedback of AES, and online peer-assessment. In order to analyze these data, the tool of SPSS is employed with the purpose of studying the validity of AES. Accordingly, the paper puts forward some effective suggestions on how to make better use of AES to motivate students' interest in writing and improving their writing proficiency.

Meanwhile, recommendations are offered to better AES systems and improve the effectiveness of the AES systems.

The study aims at bring the AES systems into full play in the writing teaching in order to promote the transformation of writing teaching from traditionally result-oriented approach to process-oriented approach and call special attention to cooperative learning of students so as to enhance students' interest in writing and improving their writing proficiency.

II. LITERATURE

AES is defined as the computer technology that evaluates and scores the written prose [1][9]. AES can be traced back to the 1960s in America. With the growing development of computer technology, AES systems have been improved a lot and are being improved. In order to make the large-scale essay scoring process more practical and effective, Project Essay Grader (PEG) was developed by Ellis Page in 1966 upon the request of College Board [2]. It utilizes proxy measures to assess the quality of essays. But it has been criticized for ignoring the semantic aspect of essays and focusing more on the surface structures[3][4]. In the 1990s, with the advance of computer technology, more AES systems, such as Intelligent Essay Assessor (IEA), the Electronic Essay Rater (E-Rater) were developed to meet the

requirements. IEA analyzes and scores an essay using a semantic text analysis method called Latent Semantic Analysis (LSA). It is claimed that unlike other AES systems, IEA's main focus is more on the content related features rather than form related ones. However, this doesn't mean that IEA offers no feedback on formal aspects, i.e., grammar and punctuation, in an essay. However, the system doesn't evaluate the creativity and reflective thinking. E-Rater was developed by the Educational Testing Service (ETS) to evaluate the quality of an essay by identifying linguistic features in the text [6]. E-Rater uses natural language processing (NLP) techniques, which identify specific lexical and syntactic cues in a text, to analyze essays [4]. Later, artificial Intelligence (AI) was introduced to the development of AES systems. IntelliMetric, developed by Vantage learning, is known as the first essay-scoring tool that was based on AI. Like, E-Rater, IntelliMetric relies on NLP, which determines "the meaning of a text by parsing the text in known ways according to known rules conforming to the rules of English language" [12]. Another AES system, named My Access, is known as the instructional application of IntelliMetric. My Access is a web-based writing assessment tool that relies on Vantage Learning's IntelliMetric automated essay scoring system. The main purpose of the program is to offer students a writing environment that provides immediate scoring and diagnostic feedback; that allows them to revise their essays accordingly; and that motivates them to continue writing on the topic to improve their writing proficiency. ETS' Criterion, a web-based instructional writing tool, uses the E-Rater engine to provide both scores and targeted feedback. It allows students to improve their writing skills while working independently with immediate, detailed feedback on grammar, spelling, mechanics, usage, and organization and development.

Writing Roadmap utilizes validated items that contain short prompts as well as reading passages for students to write about. Teachers can select writing assignments across five genres from CTB's assignment library or they can create and use their own assignments. Writing Roadmap employs CTB's patented Mosaic™ automated essay-scoring (AES) high-stakes technology to drive a newly developed generic scoring engine that is calibrated to provide grade specific scores. Reporting is immediate, provides guidelines for improving writing to students and provides insights to educators on their students' writing proficiency to inform instruction. Scoring by Writing Roadmap matches human scoring.

However, the study on AES systems in china is comparatively later than western countries. What's more, early study focused mainly on the introduction of major AES systems developed in the USA and other countries [5][8][10][13]. Nevertheless, with the increasing development of Chinese science and technology, more and more researchers set about researching and developing their own AES systems which are more suitable to Chinese learners. At present, some major AES systems are being put into use, including the Automated Essay scoring developed by Liang Maocheng in 2005, Bingoenglish developed by School of International studies of Zhejiang University and Zenghui Internet

company, Pigai based on the corpus and cloud computing, and Good Point developed by the Hong Kong Polytechnic University. All of these AES systems are at the initial stage, so there is still a long way to go. Chinese Scholars and researchers have made some meaningful explorations in the application of AES systems in the English language teaching[5][7][14], but more empirical study from practical teaching still need carrying out.

The experiment employs Pigai as the platform to study the validity of the AES system in the College EFL classroom. Pigai is a web-based instructional writing tool based on the corpus and cloud computing. It returns targeted instruction and grammar feedback within seconds by comparing student essays with those in the corpus. Besides, the system provides an online platform for peer evaluation.

III. EXPERIMENTS

A. Tasks

The study mainly discusses the validity of AES systems in enhancing students' interest in writing and improving their writing proficiency. Data acquired mainly include four aspects: (i) students' self-evaluation of their writing proficiency; (ii) the effectiveness of AES system; (iii) the practical use of feedback of AES system, and (iv) online peer-assessment.

B. Participants

The study is a naturalistic classroom-based inquiry that was conducted in five EFL college writing classroom contexts in a university in China. Totally, 230 non-English majors are involved in the experiment, who are from different departments except English Language Department.

C. Materials and Methods

Questionnaire, considering the four above-mentioned aspects, is employed as a main method in the present study. Totally twenty-three multiple choice questions are presented, including six multiple answers and seventeen single answers. In order to ensure its reliability and validity, the questionnaire is designed and modified after consulting relevant EFL teachers and interviewing students.

D. Data collection and analysis

In order to make sure of the objectivity of participants' evaluation, participants in the study are from different classes whose EFL teachers are also different correspondently. The questionnaires are completed by these participants in class, who cooperate very well. 230 questionnaires are distributed and recalled, among which 9 are not completed. As a result, 221 questionnaires are valid and its efficiency is 96.1%.

Data collected in the present study mainly include rough drafts and final ones of participants, their writing's online assessment and feedback, their questionnaires and records about their interviews.

In order to analyze these data, the software of SPSS is employed with the purpose of studying the validity of AES. Exploratory Factor analysis (EFA) is adopted to analyze

construct validity of single choices in the questionnaire (in Table I).

TABLE I. KMO AND BARTLETT TEST

Test		Value
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.724
Bartlett's Test of Sphericity	Approx. Chi-Square	646.173
	Df	136
	Sig.	0.000

It is shown in Table I that Kaiser-Meyer-Olkin Measure of Sampling Adequacy is 0.724, which is bigger than 0.5. Besides, P is 0.000, which is smaller than 0.05. All of these indicate that the data collected can be used in factor analysis and are statistically significant. Factor analysis shows that most of the factor loading is bigger than 0.5 (generally > 0.3). Variables reflect the differences of more than half of online writing (60%). These analysis show that the questionnaire is well-constructed.

IV. DATA ANALYSIS

By means of descriptive study on the basis of the SPSS software, multiple-choice questions with a single answer are calculated and its mean, standard deviation, variance and percentage are worked out accordingly; to those with multiple answers, their percentage and case percentage are figured out respectively.

A. Data analysis of Participants' writing

In the 221 questionnaires, 87.3% thought of their writing ability as just-so-so or bad. As to the difficulty in writing in English, 39.8% thought it difficult; 52.5% thought it not very difficult; and only 7.7% thought it very difficult. Table II shows its mean is 2.45 and its SD is 0.403. Generally, it is not easy for participants to write in English. When asked their frequency of writing in English, 71.6% acknowledged that they only wrote when assigned by their teachers. Only two participants (0.9%) wrote frequently. As to the use of writing resources on the Internet or in the books, few (5.4%) often used them, 30.6% used them when coming across new topics and 21.3% never used respectively.

It is found that difficulty in writing vertically existed though participants have studied English for many years, which challenges writing teaching and indicates it will be an important and urgent task to improve their writing proficiency. Data also show that though they are not good at writing in English, participants are lack of autonomy and self-reliance. They depend much on their teachers, because they seldom write in English after class actively and make use of rich writing resources, which is somewhat worrying. Therefore, EFL teachers should take the responsibility of helping students enhance their stimulus and improve their ability of autonomy so as to improve their writing proficiency.

TABLE II. PARTICIPANTS' WRITING

Questions	N	Mean	SD	ANOVA
Self-evaluation	221	3.22	0.681	0.464
Difficulty in writing	221	2.45	0.635	0.403
Writing frequency	221	2.76	0.599	0.358
Use of writing resources	221	2.64	0.876	0.767

B. Data analysis of help given by AES

When asked whether their interest was enhanced by AES, 31.7% chose "great" and 45.2% chose "ordinary". As to whether AES contributes to improving their writing proficiency, "great" and "ordinary" take up 36.7% and 43.4% respectively, and only 2.3% thought it "helpless". (shown in Table III)

TABLE III. HELP GIVEN BY AES IN IMPROVING PARTICIPANTS' WRITING

Questions	N	Mean	SD	ANOVA
Improving their interest	221	2.73	0.956	0.915
Improving their ability	221	2.68	0.843	0.710

Nowadays, it is not tough for college students to operate computers, so most of them will be not very anxious to use the computer and surf the Internet, which does good to practice their writing on the basis of AES. Data show that most of the participants take an active attitude towards online writing and approve of the assistance given by AES. It is found that comparatively speaking, the greatest help lies in vocabulary and grammar, because among the participants, case percentages of both amount to 61.8% and 54.8% respectively. In addition, case percentages of organization and coherence are also 39.6% and 41.0% respectively. (shown in Table IV)

TABLE IV. HELP PROVIDED BY AES

Options of the Question	N	percentage	Case percentage
content	19	3.7%	8.8%
organization	86	16.8%	39.6%
coherence	89	17.4%	41.0%
vocabulary	134	26.2%	61.8%
grammar	119	23.3%	54.8%
syntax	64	12.5%	29.5%
total	511	100.0%	235.5%

C. Data analysis of attitude towards AES feedback

It is found that 34.5% and 43.3% thought of AES feedback as "helpful" and "ordinary" respectively and only 1.4% didn't think it helpful. The overwhelming majority of the participants (85%) thought of AES feedback as acceptable. 21.7% took AES feedback seriously and revised their first draft very carefully and 58.4% also revised theirs carefully according to AES feedback. 20.8% would consult

some resources while revising their writing (shown in Table V).

TABLE V. PARTICIPANTS' ATTITUDE TOWARDS AES FEEDBACK

Questions	N	Mean	SD
Impact of AES feedback	221	2.68	0.843
Acceptance of AES feedback	221	2.47	0.834
Revision based on AES feedback	221	2.50	0.766
Consulting materials while revising	221	2.04	0.768

Compared with teacher feedback, the biggest benefit of AES is that AES can give students an instant feedback. In fact, the interview with students showed that more students preferred teacher feedback, but it was really annoying that teachers could not correct all of the writings or give a feedback in time. AES could compensate for this shortcoming. It must be pointed out that though most of the students accepted AES feedback, there existed many shortcomings. For example, students remarked that comments made by AES were too general, not specific, and somewhat "cold". What's more, a "bad-faith" essay could fool AES into giving a high score. All of these need to be improved.

However, most of the students are willing to revise their drafts carefully according to AES feedback. It is acknowledged that the process of revising, correcting and modifying one's own essay is essential in the writing training. In the past, students could not find their errors and mistakes made in the writing and didn't know how to revise and correct them. With the help of AES feedback, students can correct some errors and mistakes especially about vocabulary, grammar and collocation. They also can obtain more information by using the corpus provided by AES. Besides, more students can consult relevant materials actively, which is undoubtedly a big step forward. To a certain extent, students' autonomous ability can be developed and improved.

D. Data analysis of attitude towards online peer evaluation

Most of the participants (77.6%) were willing to evaluate peer's writing online. When asked whether online peer evaluation is helpful to his/her writing, 93.7% answered "yes" (shown in Table VI).

TABLE VI. ATTITUDE TOWARDS ONLINE PEER EVALUATION

Questions	N	Mean	SD	ANOVA
Willingness of online peer evaluation	221	1.83	0.771	0.595
Whether online peer evaluation is helpful	221	2.89	0.939	0.883

81% thought they could learn other students' strong points and 74.2% agreed that the biggest benefit of online peer evaluation was that they learned how to avoid some errors and mistakes made by peers. At the same time, they learned to read essays critically. Of course, the shortcomings of online peer evaluation could not be avoidable. 71.4% concerned that peer evaluation would be formalistic(shown in Table VII).

TABLE VII. BENEFITS AND DRAWBACKS OF ONLINE PEER EVALUATION

Questions	Options	N	Percent age	Case percent age
Benefits of online peer evaluation	Improving writing interest	60	9.5%	27.1%
	Reading articles critically	137	21.7%	62.0%
	Learning others' strong points	179	28.4%	81.0%
	Learn to avoid mistakes made by peers	164	26.0%	74.2%
	Giving the fullest play to one's strength	90	14.3%	40.7%
	total	630	100.0%	285.1%
Draw-backs of online peer evaluation	Considering peer's face	124	24.9%	56.4%
	Anxiety of one's own ability	100	20.1%	45.5%
	Only focusing on form	117	23.5%	53.2%
	Being formalistic	157	31.5%	71.4%
	total	291	498	100.0%

Peer evaluation has been generally supported in the literature as a "potentially valuable aid for its social, cognitive, affective, and methodological benefits" [11]. It is a very effective way to improve students' writing proficiency and promote their cooperation. AES provides such a platform of peer evaluation on which students can evaluate other students' writing anonymously, which can avoid the matter of face and stimulate their initiative to write, because students pay more attention to their image in his/her peers' eyes and would like to display his best to peers. Therefore, they will be much more careful to revise their drafts and be more confident. In the process of evaluation, they become more critical readers and revisers through readings other's writings critically. Gradually, students become more sensitive to some errors and mistakes and learn to analyze from different perspectives, which is beneficial to develop the habit of autonomy.

V. DISCUSSION

Accordingly, the paper puts forward some effective suggestions on how to make better use of AES to motivate students' interest in writing and improving their writing proficiency.

First of all, AES in China still has a long way to go. Failing to detect the content related features of an essay (organization, style, cohesion, etc.), AES systems don't provide instructional feedback for the students. Sometimes, AES systems don't actually read and understand essays as humans do. Human raters may directly evaluate various intrinsic variables of interest, such as diction, fluency and grammar, in order to produce a score. AES systems use approximations or possible correlates of these intrinsic variables. In real assessment, AES system also makes some mechanical mistakes. In the interview with the students, it was mentioned again and again by the students that some expressions, especially those new and authentic expressions

that they have just learned from the textbook and put into writing, were assessed as mistakes or errors and were not accepted by AES system. Obviously, AES based on the corpus has much room to improve. It is suggested that at least two functions should be added to the AES system. One is to keep the originals of the students so that they can make a comparison after revision and have much knowledge of their progress in writing. The other is to strengthen the function of essay copy detection. Suggestion is put forward that technology of information retrieval should apply to the analysis of the content of the essay so as to prevent students from copying or downloading directly.

Secondly, teachers still play an irreplaceable role in the writing training. In order to indeed improve students' writing proficiency, AES systems should be combined with human assessment. Teachers' careful feedback is also crucial for students' development. It seems that AES systems can reduce the teachers' workload. In fact, their load cannot be reduced at all, if teachers do offer students help. Instead, it just puts forward new requirements. As a teacher, he/she must take a right attitude towards AES systems and gives students reasonable guidance. Teachers' instruction and involvement are indispensable to the development of students' writing proficiency. To make full use of online peer evaluation, teachers must instruct the students how to make an objective judgment and also list reasonable and specific evaluation criteria so that they can evaluate their own essay and others' effectively and decrease the blindness and arbitrariness of evaluation. After assigned homework online, students will complete their essay within the specified time, revise their own essays based on AES feedback and submit the revised draft. Then, teachers click the menu of peer evaluation. After self-evaluation and peer evaluation, it's teacher's turn to evaluate the essays that have been revised several times. Generally speaking, these essays are much better than the first drafts. At this time, teacher feedback is very useful and meaningful to the accuracy of their writing and the improvement of writing skills. Teacher feedback compensates for the drawbacks of AES. For instance, AES system can't present its advantages in evaluating organization, discourse coherence, cohesion etc. when teachers read students' essays, more attention can be paid to these points, which are weak points in the students' writing. The process of "writing-correcting-revising" not only realizes the transfer of writing knowledge but also enhances students' interest and writing proficiency. Also, it contributes to developing students' critical thinking, enforcing the consciousness of autonomic learning and forming the habit of independent learning.

Finally. Students' self-consciousness and initiative are the foundation of bringing AES systems into the fullest play. Online writing relies on network environment and provides students with abundant resources. AES feedback presents valuable guidance to the students in revising their essays.

However, if the students couldn't perform the project as designed by their teachers or attach importance to AES feedback, AES wouldn't perform as expected. Therefore, students must be goal-oriented, diligent, self-regulated and self-disciplined.

Undoubtedly, it is apparent that more effective input will promote students' writing proficiency with the application of AES into college English writing teaching. Students get more writing practice without adding to instructor workload, and instructors can concentrate on the content and style of students' work and teach higher level writing skills. Though not perfect, as a new means and approach, AES system can, to some extent, solve the problem in writing teaching caused by lack of writing classes. It has been a real and viable alternative and complement to human scoring.

REFERENCES

- [1] D. Grimes, "Assessing automated assessment: Essay evaluation software in the classroom," [ftp://ftp.ics.uci.edu/pub/grimesd/AutoAssessment.pdf](http://ftp.ics.uci.edu/pub/grimesd/AutoAssessment.pdf), accessed 06/20/2010.
- [2] E. B. Page, "Project essay grade: PEG," *Automated essay scoring: A cross-disciplinary perspective*, pp.43-54, 2003.
- [3] G. K.Chung, and H. F. O'Neil, *Methodological approaches to online scoring of essays*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles,1997.
- [4] K. Kukich, "Beyond automated essay scoring," *IEEE intelligent systems*, vol.15, no.5, pp.22-27,2000.
- [5] J. L. Tang, and Y. A. Wu, "Using automated writing evaluation in the classroom: a critical review," *Journal of Foreign Language Teaching and Research*, vol. 43, no.2, pp.115-124, 2011.
- [6] J. C. Burstein, , & Marcu, D. , Andreyev, S, & Chodorow, M. *Towards automatic classification of discourse elements in essays*[A]. In *Proceedings of the 39th annual meeting of the Association for Computational Linguistics [C]* , France, 2001: 90 - 92
- [7] L. Gui, "Application of Good Point online system in English writing teaching," *Modern Educational Technology*, no.4, pp.79-82, 2011.
- [8] M. C. Liang, and Q. F. Wen, "A critical review anf implications of some automated essay scoring systems," *Computer-Assisted Foreign Language Education*. no.117, pp.18-24, 2007.
- [9] M. D. Shermis, and J. C. Burstein, "Automated essay scoring: A cross-disciplinary perspective", Routledge, 2003.
- [10] N. Han, "An overview of several automated essay scoring systems," *China Examinations*,no.3, pp.38-44, 2009.
- [11] P. Rollinson, "Using peer feedback in the ESL writing class," *ELT Journal*, vol.59, no .1, pp.23-30, 2005.
- [12] S. Elliot, "IntelliMetric: from here to validity. Automated essay scoring: A cross-disciplinary perspective," *Automated essay scoring: A cross-disciplinary perspective*, pp. 71-86, 2003.
- [13] X. X. Chen, and S. L. Ge, "A review of automated essay scoring," *Journal of PLA University of Foreign Languages*. vol. 31, no. 5, pp.78-83, 2008.
- [14] X. L. Shi, "A tentative study on the validity of online automated eassay scoring used in the teaching of EFL writing," *Modern Educational Technology*, vol.22, no. 10, pp. 67-71,2012.