

An Empirical Study on the Impact of an Automated Writing Assessment on Chinese College Students' English Writing Proficiency

Shuwen Wang^{*}, Ran Li

School of Foreign Languages, Southwest Petroleum University, Chengdu, China

Email address:

wangshuwenswpu@163.com (Shuwen Wang)

^{*}Corresponding author

To cite this article:

Shuwen Wang, Ran Li. An Empirical Study on the Impact of an Automated Writing Assessment on Chinese College Students' English Writing Proficiency. *International Journal of Language and Linguistics*. Special Issue: *Linguistic and Pedagogical Language Issues in the 21st Century Classrooms*. Vol. 7, No. 5, 2019, pp. 218-229. doi: 10.11648/j.ijll.20190705.16

Received: June 29, 2019; **Accepted:** August 27, 2019; **Published:** September 11, 2019

Abstract: Automated writing evaluation (AWE) is an online essay scoring system which can provide feedback and revising advice to teachers and students. In this paper, an empirical study was carried out to explore the impact of Writing Roadmap2.0 (WRM2.0)-an automated writing assessment system on the English writing proficiency, which is reflected in three dimensions-the language form, the contextual structure and the writing quality of non-English major freshmen in China. In this study, 100 participants were divided into the experimental class (EC) and the controlled one (CC) at random, with 50 ones in each class. Both qualitative method and quantitative method were adopted for data collection and analysis, including the pre- and post-tests on WRM2.0, teacher-assessed writing task and interviews. The results revealed that while there was no significant difference in the writing proficiency in pre-test on WRM2.0 between EC and CC, the former outperformed the latter in both post-test on WRM2.0 and teacher-assessed writing task in the final exam in two dimensions: the language form and the writing quality. With regard to the aspect of contextual structure, EC benefited a little on WRM2.0. Generally speaking, this empirical study observed positive impact of WRM2.0 on writing proficiency of L2 students in China. It is expected that the findings will provide references for the further integration of AWE with writing teaching and learning in the EFL classroom.

Keywords: Automated Writing Evaluation, Writing Roadmap 2.0, Writing Proficiency, Language Form, Contextual Structure, Writing Quality

1. Introduction

Writing is not a series of discrete chronologically ordered tasks, but rather a complex of recursive, embedded activities [1], which means that writing should not be regarded as a final product but a recursive process, covering “writing-feedback-reflection-revision-rewriting-refeedback-reflection-re-revision...” In this recursive process, writing proficiency will be improved more or less. However, for the college students in China, English writing teaching and learning are a big challenge due to the big class size, the heavy workload, the delayed feedback and sometimes too general comments, which makes the recursive process infeasible. Automated writing evaluation (AWE), an online essay scoring system, is designed to provide instant feedback, revising

advice and enormous chances of revision to the users, which will make the recursive writing feasible. It is argued that the use of AWE in English writing can be taken as a means to close the gap between actual and desired performance levels through the provision of timely feedback, to improve the writing quality, and to improve the cost-effectiveness of education [2, 3]. Besides, the timely corrective feedback, to some degree, will help learners consciously notice their problems in their writings and then take actions to modify or even recast them. In other words, in the foreign language acquisition, efficient and sufficient feedback could trigger the learners' notice of their errors and such notice could, in turn, help to improve their learning proficiency [4-7]. This paper, by employing qualitative and quantitative analyses, is designed to investigate the impact of an automated writing

assessment on Chinese college students' English writing proficiency.

2. Literature Review

In the 1960s, the first AWE was created in the United States and then was adopted in teaching writing. Later, with the rapid development of computer and Internet, more AWE systems, such as MY Access!, Project Essay Grade, Intelligent Essay Assessor, Electronic Essay Rater, Writing Roadmap 2.0 (WRM2.0) and *pigaiwang*, were developed and widely used in English writing practice as well as in large-scale English tests, like TOEFL and the GRE. These systems are designed to provide a range of feedback from individualized reports on grammatical errors, such as word choice, fluency, conventions [8, 9], to holistic evaluations concerning the content, voice, organizational and mechanical aspects of essay writing for both first and second language writers [10-14].

Generally speaking, research on AWE mainly falls into three groups [15].

The first group discusses the reliability and the validity of AWEs, with mixed findings. Some researchers doubt the reliability of e-assessment [16-18], while more studies yield positive findings [19-24]. For example, Attali et al. reported that human-computer agreement (.72 to .79) was lower than human-human one (.78 to .79) for the GRE [16]. McCurry and Tsai Min-hsiu also reported pessimistic results of reliability of AEE system, with lower reliability of machine scoring than that of human scoring [17, 18]. However, Rich et al. reported high interrater reliability of WRM2.0 and human raters, indicating that such e-feedback was reliable and reasonable [19]. Shermis and Hamner examined nine AES platforms and ultimately concluded that the kappas of scoring engines ranged from .60 to .84, close to the kappas of human raters—.61 to .85, indicating high agreement [22]. He conducted the study of 30 writing samples at random in Chinese Learner English Corpus (CLEC) at *pigaiwang* and concluded that *pigaiwang* was reliable except that the scores on this system were significantly higher than those by human raters [23]. Li and Tian studied 645 essays collected from English Test for International Communication (ETIC) held in China at iWrite English Writing Assessment and Evaluation System 2.0 (iWrite 2.0), which is developed by Foreign Language Teaching and Research Press and Beijing Foreign Studies University, and found that the reliability of iWrite 2.0 was 0.721, indicating a high consistency. They also reported that iWrite 2.0 highly agreed with human raters [24]. But in terms of the validity of AWEs, a majority of studies agreed that AWEs could provide useful feedback and sometimes advice on spelling, word choice, grammatical structures and punctuation, but failed in contents and ideas [22, 25, 26]. For example, Herrington and Moran showed that Write Placer Plus and Intelligent Essay Assessor could only provide information on structural features but not on clichés and plagiarism [25]. He reported that *pigaiwang* could evaluate sentences and provided sufficient information on words and grammars, but could not explicitly evaluate organizations,

styles, coherence and cohesion [23].

The second group focuses on the empirical studies on the efficiency of AWEs on students' writing performance, including the English native learners and the non-English native ones from primary school, middle school, college and graduate school [19, 27-37]. The findings are mixed. Some research observe unsatisfactory results. For example, Shermis, Burstein and Bliss studied 1072 high school students at random on Criterion and reported that no significant difference in writing performance was found between the experiment group and the control group in the statewide examination [27]. Wilson and Andrada also reported that e-assessment insufficient to improve the overall writing quality [28]. Saricaoglu observed mixed findings in Turkey. He explored the impact of AWE on the improvement of English as a second language learners' causal explanations within two cause-and-effect essays across pre- and post-tests. Results revealed statistically significant changes in learners' causal explanation within only one cause-and-effect essay while no significant improvement was observed across pre- and post-test [29]. However, a majority of research report positive impact of AWEs on students' overall writing quality. For example, Vantage Learning delivered a series of reports on *MY Access!*, stating that the learners who used this system in their writing practice performed better in the statewide examination, but no pre-tests were reported which reduced the credibility [30]. Warschauer & Grimes and Grimes collected the data from the statewide writing assessments of the students in Grades 6-12 in California and found that the learners who used *MY Access!* and Criterion to do their writing tasks performed better than those who didn't [31, 32]. Rich, et al. studied the impact of WRM2.0 on students in Grades 6-12 in West Virginia for two years and reported the learners who used this system significantly outperformed [19]. White et al. also reported such positive impact of WRM2.0 on WESTEST 2 online writing assessment scores [33]. However, these two studies ignored the participants' data of pre-exams. Wilson et al. examined data of grade 4-8 students who used PEG to participate in statewide computer-based benchmark writing assessment and showed positive impact on writing quality across revisions. But they also admitted the unvaried feedback and lack of feedback on writing strategies [34]. Palermo and Thomson examined 829 middle school students' argumentative writing performance combined with NC Write and self-regulated strategy development instruction or traditional writing instruction by quantitative method and qualitative method. Results suggested that incorporating AWE into writing instruction supported improvements in students' writing quality [35]. Most Chinese studies also agree that AWEs have positive effect on English writing proficiency of Chinese EFL (English as a foreign language) learners, especially in the mechanical aspects and writing quality [36-42]. For example, in her longitudinal study on the impacts of WRM2.0 on Chinese college EFL learners' writing proficiency, Wang demonstrated that RWM2.0 had enduring positive impact on the improvement of the participants' writing proficiency. But she also pointed out that statistically

significant improvements in language form and overall writing quality were observed while non-significant improvements were found in textual structure, ideas and contents [41]. However, only qualitative method is adopted in her study. Li, through textual analysis and interview, explored the impact of the integrated feedback of *pigaiwang*, the peers and the teacher on students' writing revision in the aspect of the number, type and effect. The results showed that the integrated feedback exerted a positive influence on students' writing revision [42]. However, no control group and pre-test are designed and only three writing tasks are observed in her study. Therefore, it is less convinced to claim the positive effects of AWEs on students' writing achievements because some studies are lack of a control group or pre-test.

The third type of research concentrates on the survey of teachers' and students' attitudes toward AWEs and indicates slightly positive feelings [3, 32, 43-47] mixed with few negative ones [48]. For example, Chen & Cheng investigated EFL undergraduates' perspectives on *MY Access!* in Taiwan and found that students' general attitude was not positive because they criticized that this system provided insufficient and vague feedback on coherence and content development [48]. Warschauer & Grimes and Grimes & Warschauer conducted studies to examine teachers' and students attitudes toward *MY Access!* in primary schools and middle schools in the USA. They found that although both teachers and students acknowledged benefits from AWE, such as time-saving, reducing grading and timely feedback, they placed some unwarranted trust in the system [44, 45]. Wang, et al. analyzed that their participants had positive attitude toward e-feedback since they could find their writing weakness and took due revisions immediately and continuously, and thus their autonomy and motivation could be heightened [46]. Rolim & Isaias investigated teachers' and students' opinions on e-assessment in higher education in Portugal and other countries, reporting that both teachers and students believed that AWE was advantageous [47]. In China, Wang and Yang also reported mixed feelings of WRM2.0 among teachers and students [49, 50].

From the above discussion, we could conclude that till now, studies on the AWEs have achieved a lot. However, few studies discuss whether the students who receive e-assessment will get similar achievement when teacher-assessed. This empirical study, taking WRM2.0 as an automated writing assessment and 100 EFL freshmen majoring in Mechanical Engineering as participants, is designed to investigate the impact of WRM2.0 on Chinese college students' English writing proficiency at three dimensions—language form, contextual structure and writing quality.

In the following sections, we introduce the research design, including WRM2.0, research context, participants, research questions and research methods, and provide research procedure and data collection. A discussion of our findings then follows. Finally, we finish with concluding remarks.

3. Research Design

In this section, a brief introduction of automated writing

evaluation tool-WRM2.0, research context, participants, research questions and research methods are presented respectively.

3.1. Automated Writing Evaluation Tool-WRM2.0

In this study, WRM2.0, developed by CTB/McGraw-Hill, was taken as a writing tool. This system is an online essay scoring tool that provides unlimited writing practice and automatic feedback with score, instructional guidance and robust report. In general, it provides timely form-focused feedback, including Word choice, Fluency and Conventions, context-focused feedback concerning Organization, Voice, and Ideas and Content, an overall quality-focused feedback provided as a holistic score, with the scoring on 6-point rubric scale (from 0 to 5) and the scoring analysis of each writing dimension (that is subjective comments). Besides, on the easy-to-use interface, some tutoring tools, such as Hint (providing tips for improving essays), Tutor (tutorial on grammar and syntax), Thesaurus (providing synonym) and Tree (pointing out mistakes in sentence structure and providing correct uses) are available for students to adopt proper strategies while self-editing their drafts until they are satisfied with their work. What's more, an e-portfolio is produced for each student as well as the whole class, and is also available for teachers to collect enough information covering the individual student's writing strengths and areas for improvement and the general writing proficiency of the whole class. It is evident that WRM2.0 is designed to afford diagnostic assessment, formative assessment, and summative assessment for students to obtain more feedback about their writing work and undertake self-directed and self-regulated writing skills, and for teachers to get general evaluation about their students' writings at the levels of language form (i.e. form-focused feedback), contextual structure (i.e. context-focused feedback) and writing quality (i.e. holistic), and then make curricular adjustments purposely.

3.2. Research Context

This study lasted about 16 weeks in an integrated course which covered reading, writing, listening and speaking. There were five English class periods per week, including three ones for reading and writing, and two for listening and speaking, with 45 minutes in each class period. The participants were taught by the same teacher, a young woman with 9 years of teaching experience, and were required to take the same tasks of proficiency examination, pre- and post-tests, and final examination. They also followed the same syllabus, used the same textbooks and finished five writing topics with identical directions in these 16 weeks.

3.3. Participants

The participants in this study were 100 freshmen majoring in mechanical engineering in the same university and were separated into two classes—one as experimental class (EC) and the other the controlled one (CC) at random, with 50 students respectively. Students in each class were

approximately balanced in terms of male/female mix. In EC, participants were predominantly male ($n=37$) with female ($n=13$), which was similar to those in CC—35 males and 15 females. And the participants were aged between 17 and 20 years old, with a median age of 18. Besides, based on their teacher-assessed English scores (with the full marks as 100) in the English proficiency examination (concerning only reading and writing) held in the university before the students were divided into individual classes, the participants' English proficiency could be described as mid-advanced (65% of all the freshmen at university). Table 1 presents the details of the proficiency-based scores in EC and CC.

Table 1. Proficiency test of EC and CC.

Item	Descriptive statistics	Proficiency test		<i>t</i>	<i>p</i>
		EC	CC		
Score	Means	74.081	74.450	435	.664
	S. D.	3692	8503		

Table 1 shows that CC performed a little better than EC in this proficiency test, but no significant difference was found ($p=0.664 > 0.05$), which indicates that CC and EC had equivalent English proficiency prior to the empirical work, and thus the results at the end of this work would be explainable, reasonable and reliable.

Participants' other characteristics, such as motivation, cognitive styles and learning strategies, were not taken into consideration in this study due to the interference of too many variables. In general, EC and CC were matched in the dimensions of major, gender, age, teacher, primary English proficiency and syllabus.

3.4. Research Questions

This study, by using WRM2.0 as the automated evaluation system, intends to add some insights to EFL with the investigation of the following three research questions:

- 1) What is the impact of WRM2.0 on the language form of students' writings?
- 2) What is the impact of WRM2.0 on the contextual structure of students' writings?
- 3) What is the impact of WRM2.0 on the writing quality of the students' writings?

3.5. Research Methods

In order to answer the above three questions, quantitative method and qualitative method are adopted, including pre-test and post-test on WRM2.0, teacher-assessed writing task in the final examination, and interview with an English teacher and six participants in EC.

Prior to this empirical work, a pre-test was conducted on WRM2.0 to examine participants' levels of writing proficiency. All 100 participants (50 in the EC and 50 in the CC respectively) were trained 40 minutes about the use of WRM2.0 and then were required to finish a composition titled *Enjoyable Work* with less than 120 words in 30 minutes in the computer classrooms at the same time period, with the tutoring tools in WRM2.0 shielded. At the end of the 16th

week, each participant was required to submit a post-test writing titled *My Hobby* with less than 120 words following the same procedure as that in the pre-test. During the 16 weeks of this empirical work, EC and CC were required to finish the same writing tasks, including the titles, the length, the requirements and the time limitation. The only differences lied in the writing environment and assessment: EC finished on WRM2.0 and received e-assessment while CC wrote on the paper and obtained teacher assessment. In addition, all 100 participants took the same final examination held in the classrooms at the same time at the end of the 17th week organized for all of the freshmen by the university. The writing part in the final examination, titled *Can We Live Without Mobile Phones?*, was scored (scores ranging from 0 to 15) according to the CET-4 (College English Test Band 4, which is a national English proficiency test for college students with high reliability and validity in China) scoring criteria by another English teacher who had rich experience in CET-4 writing assessment. After the experiment, the English teacher and six participants in EC were invited to take part in the 10-minute semi-structured interview individually to share their opinions on WRM2.0. Among the six students, three were boys and three girls; two at advanced level (Student A and Student B, with average e-assessed scores between 4 and 5), two at mid-advanced level (Student C and Student D, with average e-assessed scores between 2 and 3), and two at low level (Student E and Student F, with average e-assessed scores between 0 and 1). In order to collect adequate information, the interview was conducted in Chinese and then translated into English by the researchers.

4. Data Collection and Analysis

The data were gathered from three aspects, that is, scores of pre- and post-tests on WRM2.0, scores of teacher-assessed writing task in the final examination, and interviews with an English teacher and six participants in EC. As introduced in Section 3.1, three aspects provided by WRM2.0 were considered as language form, that is, Word choice, Fluency and Conventions, and another three aspects—Ideas and Contents, Organization and Voice—were regarded as contextual structure. As to the overall writing quality, the holistic score was taken. In addition, only holistic score was offered in teacher assessment, because the overall writing quality is popularly reported by the holistic score in EFL classrooms in China. Chandler also used holistic ratings as a measure for students' overall writing quality [51]. For comparison of the results of the language form, contextual structure and overall writing quality provided by WRM2.0 between pre- and post-tests, a paired samples *t*-test was carried out by SPSS 19.0 to find out the effect of the treatment. For comparison of the results between CC and EC, an independent samples *t*-test was done by SPSS 19.0 to check whether there was significant difference in the above mentioned three dimensions provided by WRM2.0. Besides, the 100 teacher-assessed writings in the final examination were considered as overall writing quality. The data were also

keyed into SPSS 19.0 and then an independent samples *t*-test was computed to see what EC behaved when teacher-assessed. What's more, the semi-structured interviews of the English teacher and her six students from EC were audio-taped with their permission and then transcribed verbatim in the 17th week.

5. Findings and Discussion

In this section, research findings, including language form,

Table 2. Results of the Language Form (Independent *t*-test).

Item	Descriptive Statistics	Pre-test				Post-test			
		CC	EC	<i>t</i>	<i>p</i>	CC	EC	<i>t</i>	<i>p</i>
Word choice	Means	2.190	2.098	0.955	.342	2.300	2.674	-2.955	.004
	S. D.	4573	5049			6506	6144		
Fluency	Means	2.456	2.472	-164	.870	2.506	2.878	-2.634	.010
	S. D.	4572	5139			7385	6753		
Conventions	Means	2.504	2.454	0.480	.633	2.542	2.898	-2.629	.010
	S. D.	5307	5116			6752	6790		

Table 2 shows that in the pre-test, there were no significant differences in these three aspects respectively between EC and CC in the language form, including Word choice ($p=0.342 > 0.05$), Fluency ($p=0.870 > 0.05$) and Conventions ($p=0.633 > 0.05$). But after 16 weeks of treatment, EC improved more significantly than CC in Word choice ($p=0.004 < 0.05$), Fluency ($p=0.010 < 0.05$) and Conventions ($p=0.010 < 0.05$), indicating that WRM2.0 did have positive impact on the

contextual structure and writing quality, are presented and analyzed respectively.

5.1. Positive Impact of WRM2.0 on Language Form

As mentioned above, language form is examined from three aspects, concerning Word choice, Fluency and Conventions. The results on WRM2.0 are presented in Table 2 and Table 3 respectively.

language form in EC due to the timely and detailed assessments, and revision functions on WRM2.0, whereas in CC, even if the students could get some feedback on spellings or sentence structures, they seldom received specific information about how to correct those mistakes or polish the language.

The pre-and post-tests results of CC and EC are listed in descriptive statistics and paired samples *t*-test in Table 3.

Table 3. Results of the Language Form (Paired *t*-test).

Item	Descriptive Statistics	CC				EC			
		Pre-test	Post-test	<i>t</i>	<i>p</i>	Pre-test	Post-test	<i>t</i>	<i>p</i>
Word choice	Means	2.190	2.300	-1.042	.302	2.098	2.674	-5.290	.000
	S. D.	4573	7462			5049	6144		
Fluency	Means	2.456	2.506	-0.425	.673	2.472	2.878	-3.241	.002
	S. D.	4572	7385			5139	6753		
Conventions	Means	2.504	2.542	-0.342	.734	2.454	2.898	-3.642	.001
	S. D.	5307	6752			5116	6790		

As can be seen in Table 3, in CC, although there appeared a little improvement in Word choice, Fluency and Conventions respectively, the differences failed to reach statistical significance ($p=0.302 > 0.05$, $p=0.673 > 0.05$, $p=0.734 > 0.05$ respectively). In contrast, Table 3 shows that students in EC made statistically significant improvements in the surface features, that is, Word choice ($t=-5.290$, $p=0.000 < 0.05$), Fluency ($t=-3.241$, $p=0.002 < 0.05$) and Conventions ($t=-3.642$, $p=0.001 < 0.05$), suggesting that to some degree, WRM2.0 could help learners to correct word-level errors and sentence-level errors efficiently by providing timely and corrective feedback, which agrees with the previous studies [36, 39-41].

In addition, all of the students interviewed mentioned that they used the e-feedbacks very often to identify errors in the language form and clarify some confusion they had in writing. For example,

"... It is great to see my scores immediately. WRM2.0 tells me what the errors are. And the most amazing thing is that it also tells me how to correct them. Later on, I realize my

problems and manage to solve them, such as the word spelling, punctuation and some other grammatical rules. You know, I often misspell "government" but now I think that I will never make that mistake. I also begin to pay more attention to punctuation because sometimes WRM2.0 tells me that my mistakes are caused by the wrong use of comma...." (Student A)

"...WRM2.0 tells me there are some mistakes in the sentences. Sometimes I can correct them but sometimes I can't. I think English sentences are too difficult..." (Student E)

We can conclude that the interviewees were satisfied with the computer-generated feedback on the language form and agreed that the tutoring tools, providing tips, synonyms and partial solutions, could guide their revisions. However, due to the complexity of the English grammar, the advanced students could respond more actively and efficiently to the e-feedback on the language form than their low-level peers.

The teacher stated:

"I think the most amazing point about WRM2.0 is that it

cannot only point out the spelling errors but also offer some corrections and more words to choose from. I think it's the students' job to make sure their spellings are correct. But the students in CC often repeat the same mistakes again and again although every time I point them out and require them to make correction. This drives me crazy.... However, when it comes to Fluency, I don't think WRM2.0 works efficiently. English grammars are very complicated. Some students can only recognize the mistakes in tense, plural form, speeches of the word, but fail in the complicated sentences...."

Her response indicates that WRM2.0 was advantageous in partially playing the teacher's role in identifying and correcting the mechanic problems contained in students' writings, thus reducing her workload in locating the surface-level problems again and again. However, she also agreed students with lower level of language proficiency could not react effectively to the feedback on WRM2.0, supporting Student E's statements.

In general, EC outperformed CC in the language form and had positive attitude toward WRM2.0 in its surface-level feedback. We believe that the reasons beyond such improvements may be as follows. 1) Given more practice in English writing, students can gradually be aware of the nature of English grammar and make due development in the language form. WRM2.0 is designed to provide numerous writing and revising frequency, which indicates that the users could get extra feedbacks every time they submit their products. Thus, in this recursive process of "writing-revision-rewriting-re-revision", they not only got more writing practice, but also recognized their weaknesses while experiencing success. On the contrary, it was not feasible for students in CC to get further feedbacks for their revised products due to the big class size, indicating that CC presented a linear writing process. According to the Noticing Hypothesis, noticing is the necessary and sufficient condition for converting input to intake [4]. In other words, only when the learners notice their linguistic deficiency, can they deal with the problems. It seems that the sufficient and frequent e-feedback helps EC to bring their attention to their deficiency and then try to bridge the gap. But it is a little difficult for the low-level learners to turn "e-input" into intake successfully even if they receive frequent feedback. 2) Detailed feedback on students' mistakes are of some help to reduce the possibility of making the same mistakes [52, 53]. WRM2.0 is designed as an assessment tool as well as a learning tool. As mentioned above, during the process of

each writing practice, students in EC turned to the tutoring tools, such as Hint, Tutor, Thesaurus and Tree to get detailed prompts of mistakes, diagnosis and revision advice. Gradually, they were clearly aware of their weaknesses and managed to avoid repeating those errors in the later writing tasks. It seems that frequent and continual e-feedbacks do have potential positive effect on the long-term improvement in the language form. However, in CC, the teacher usually gave a holistic score on the basis of the overall evaluation of the writing. Even if sometimes the teacher pointed out the surface-level problems, it was a big challenge for her to offer suggestions or further feedback on revisions due to the big class size and heavy workload (At this university, an English teacher usually gave lectures to 4 or 5 classes with more than 300 students a term). 3) The positive impact comes from the timely feedback on mistakes instead of the mistakes themselves [54]. Different from the delayed teacher-assessment, WRM2.0 is able to provide timely and inexhaustible feedback to students' products. Once EC submitted their products, they received assessment and tips. On the contrary, in the traditional teacher-assessed writing, students usually got their feedback two or more weeks after submission. Some of them even forgot what they did in the former writing task. As a consequence, it is likely that some students might lose enthusiasm in revision and be left behind in the aspects of diversity, correctness, norm and fluency of vocabulary, and sentence structure.

To sum up, our findings support the claim made by Fathman & Walley and Kepner, that detailed error feedback leads to decrease in the number of errors [52, 53]. And they are also consistent with Ross' study [55],

"EFL writers can assimilate only a small proportion of corrective feedback into their current grammatical system, especially when the corrections are not detailed enough to be applied to the more complex and problematic aspects of word order and syntax."

5.2. Some Positive Impact of WRM2.0 on Contextual Structure

The second research question is to find out the impact of WRM2.0 on the contextual structure. The results of pre- and post-tests between the EC and the CC in this aspect, including Organization, Voice, and Ideas and Contents, are listed respectively in both independent samples *t*-test statistics (Table 4) and paired samples *t*-test statistics (Table 5) below.

Table 4. Results of the contextual structure (independent t-test).

Item	Descriptive Statistics	Pre-test				Post-test			
		CC	EC	<i>t</i>	<i>p</i>	CC	EC	<i>t</i>	<i>p</i>
Ideas and Contents	Means	2.310	2.288	0.173	.863	2.380	2.690	-1.882	.063
	S. D.	5884	6820			7418	8979		
Organization	Means	2.492	2.412	0.178	.439	2.570	2.796	-1.613	.110
	S. D.	5642	4592			7014	6996		
Voice	Means	2.584	2.482	0.909	.366	2.524	2.848	-2.210	.029
	S. D.	5723	5502			7096	7560		

As can be seen in Table 4, in the pre-test, there were no significant differences between CC and EC in the contextual structure, including Ideas and Contents ($p=0.863 > 0.05$), Organization ($p=0.439 > 0.05$) and Voice ($p=0.366 > 0.05$), which indicates that EC and CC had equivalent writing level in this dimension before the experiment began. In contrast, in the post-test, EC seemed to significantly outperformed CC in Voice ($p=0.029 < 0.05$), while in Organization and Ideas and Contents, although EC ($M= 2.796$ and $M= 2.690$ respectively)

did better than CC ($M=2.570$ and $M=2.380$ respectively), there were no significant differences ($p=0.110 > 0.05$; $p=0.063 < 0.05$). It can be concluded that WRM2.0 had some, but insignificant impact on EC's improvement in the contextual structure.

The pre- and post-tests results of the contextual structure between EC and CC are listed in descriptive statistics and paired samples *t*-test in Table 5.

Table 5. Results of the contextual structure (paired t-test).

Item	Descriptive Statistics	CC		<i>t</i>	<i>p</i>	EC		<i>t</i>	<i>p</i>
		Pre-test	Post-test			Pre-test	Post-test		
Ideas and Contents	Means	2.310	2.380	-0.520	.605	2.288	2.690	-2.603	.012
	S. D.	5884	9513			6820	8979		
Organization	Means	2.492	2.570	-0.708	.482	2.412	2.796	-3.507	.001
	S. D.	5642	7014			4592	6996		
Voice	Means	2.584	2.524	0.493	.624	2.482	2.848	-2.864	.006
	S. D.	5723	7096			5502	7560		

Table 5 demonstrates that CC showed no differences statistically in Ideas and contents ($p=0.605 > 0.05$) and Organization ($p=0.482 > 0.05$) although they made some improvements in these two aspects. Besides, there was even a decrease in Voice, with means from 2.584 in pre-test to 2.524 in post-test. In general, after spending 16 weeks of further English learning, CC seemed to show a dissatisfactory performance in the contextual structure in their English writings. The results ascertain some previous studies [38-41]. On the contrary, EC made some improvements in this dimension, especially in Organization ($p=0.001 < 0.05$) and Voice ($p=0.002 < 0.05$), although no significant difference was found in Ideas and Contents ($p=0.012 > 0.05$). This can be explained as EC could get timely context-focused feedback on WRM2.0, which inspired them to realize the significance of the contextual structure in the writing.

Interviews showed that students held mixed opinions of the contextual structure provided by WRM2.0. Four of the six students interviewed (all of them are at advanced level and mid-advanced level) believed that they began to benefit from the context-focused feedback while stating the weaknesses with WRM2.0. For example,

"...in the past, I often corrected the errors in word spelling or grammar, but now I find I was wrong. Ideas, content and organization are also important because if I want to get higher scores in these parts I have to pay more attention to them...but WRM2.0 only tells me that I need to better my organization, but it doesn't tell me how. Sometimes I try again and again, but the result is still not good or even worse. I'm a little disappointed." (Student C)
"...in the class, the teacher shifts to the analysis of the text structure, the transitional words and phrases. She also tells us how to enrich the ideas. I think all of these are very useful when writing..." (Student D)

"WRM2.0 tells me that organization, ideas and voice are very important factors in English writing, but it doesn't tell me how to better my work. Fortunately, in the intensive reading class, the teacher instructs us to analyze the

structure of the reading materials and often asks us to find out the transitional words. I think that is very useful. The most important thing is that she instructs us to analyze how the writer develops the passage." (Student A)

Their response reveals that WRM2.0 was a little weak in providing explicit feedback and suggestions on the contextual structure, but it aroused some participants' awareness of the components of a good writing.

However, the other two low-level interviewees complained that they were confused of the content-focused feedback on WRM2.0, for example,

"I don't know what the comments on my writing mean—'Ideas are extremely limited or simply unclear' or 'The writing lacks coherence'. How to clarify the ideas or achieve coherence is a big problem. In addition, I revised several times, but received the same comments. I feel very frustrated." (Student F)

It is not surprising that some participants focused on surface-level revision instead of the contents due to their limited cognition in those areas. But they acknowledged that they got some directions on transitional words from their English teacher.

The teacher also pointed out:

"WRM2.0 is useful in some aspects. I don't spend a lot of time responding to the mechanics of student writing again and again. Instead, I can use e-feedback to find out the weaknesses in my students' writings and devote myself to designing the class. By reading the e-feedbacks, I begin to pay closer attention to the other important aspects of an English writing, such as ideas, organization and genre. In teaching reading materials, I shift to the structure and encourage students to broaden their ideas. But this is quite different from my former teaching experience. I used to concentrate on the words and phrases in the passage. Now I realize the importance of ideas and organization. But I am still a fresh in this area. And I need some professional guidance in reading and writing teaching. I believe that I will do much better in the future."

This statement aligns with the previous studies that in ESL class, teachers respond most frequently to mechanical errors as the most important criterion for responding to student writing but ignore the context-focused feedback, such as contents, textual rules, rhetoric devices and genre [56-59]. If the learners took insufficient notice concerning the contextual structure, they would not “notice” their problems or gaps consciously, let alone intake conversion. This may be the major reason for the non-significant improvements in Ideas and Contents in both EC and CC.

In addition, by reading the context-focused analytic scores, the teacher was also aware that she should respond to the construction of meaning, such as idea, textual rules, genre and logic. And this awareness would, of course in turn, instruct her class teaching, and then inspire her students' conscious notice. But for those low-level learners, their notice of negative points in writings could not contribute to contextual improvement in their writings unless it was fully understood. This indicates that they need more practice and human instruction to get cognitively engaged.

On the other hand, it must be pointed out that the narrative feedback in the context offered by WRM2.0 is too general, too vague and sometimes even stereotyped. For example, a student, whose writing was scored 4.3 in Ideas and Contents, received the narrative e-feedback as “Your score means your writing is on its way to being engaging to readers.”, which is the same as the other two students whose scores were 3.5 and

4.0 respectively in the same area. This indicates that WRM2.0 can successfully inspire learners' and teachers' “notice” in context, but it fails to provide meaningful and instructional guidance in this dimension. This is also coincided with the corpus established by WRM2.0 that the revision frequency in the contextual dimension is much lower than that in the language form.

Given these findings, it can be inferred that some positive impacts of WRM2.0 on the contextual structure were observed, but the achievements varied among the learners with different learning proficiency.

5.3. Positive Impact of WRM2.0 on Writing Quality

The third research question asks whether WRM2.0 helped to improve students' writing quality over time. Here, it should be noted that writing quality on WRM2.0 is decided by the comprehensive evaluation on all of the dimensions instead of an average score of those dimensions. And in order to answer this question more comprehensively, we will observe from two aspects: e-assessment on WRM2.0 in post-test and teacher-assessment in the final examination.

5.3.1. E-assessment on WRM2.0

The results of pre-and post-tests between EC and CC in writing quality are listed in both descriptive statistics and paired samples *t*-test statistics in Table 6 and Table 7.

Table 6. Results of the writing quality (independent *t*-test).

Item	Descriptive Statistics	Pre-test				Post-test			
		CC	EC	<i>t</i>	<i>p</i>	CC	EC	<i>t</i>	<i>p</i>
Holistic	Means	2.258	2.204	0.466	.642	2.328	2.672	-2.240	.027
	S. D.	5922	5668			7062	7754		

Table 6 shows that before the experiment, there was no significant difference ($p=0.642 > 0.05$) between CC and EC in writing quality, but after 16 weeks, EC outperformed significantly CC ($p=0.027 < 0.05$), suggesting that EC

benefited from WRM2.0.

The pre- and post-tests results of CC and EC in writing quality are listed in descriptive statistics and paired samples *t*-test in Table 7.

Table 7. Results of the writing quality (paired *t*-test).

Item	Descriptive Statistics	CC				EC			
		Pre-test	Post-test	<i>t</i>	<i>p</i>	Pre-test	Post-test	<i>t</i>	<i>p</i>
Holistic	Means	2.258	2.328	-0.572	.570	2.204	2.672	-3.427	.001
	S. D.	5922	8650			5668	7754		

Table 7 demonstrates that there appears a general increasing trend in writing quality in both CC and EC, which indicates that continual English learning would of course contribute to better writing quality. However, no significant improvement in CC was found ($p=0.570 > 0.05$) compared to statistical significance in EC ($p=0.001 < 0.05$), suggesting that WRM2.0 worked. This agrees qualitatively with the studies carried out in the previous studies [36, 41, 60].

Interviews also revealed that the teacher and a majority of the students used formative e-assessments to reinforce their learning and gain a deeper insight into writing. For example, “...every time when I am writing, I want to see the higher and higher scores. I'm just trying to get what I can on

WRM2.0, learn more about the topic because sometimes it expands on the question even if you do get it right it gives you a little bit more information about why that is the case. I think this is very useful.” (Student B)

“... I usually write down what I've got wrong, so I can figure out where I've gone wrong... I enrich my vocabulary and realize that I should always keep the readers in my mind while writing.” (Student E)

It can be concluded that on the one hand, the system was helpful to student writers in noticing and solving some language problems, such as spelling, vocabulary, basic grammar, etc. We know that such corrections on surface features did contribute partially to the improvement of the

quality. On the other hand, WRM2.0 failed to arouse users' "deep" notice of contents, especially the low-level learners.

The teacher showed a mixed feeling,

"I think the timely e-portfolio provided on WRM2.0 is very useful for me to shape the class teaching. By reading the e-feedbacks on the individual student and the whole class, I can find out the areas of weakness and then target teaching to students' needs. In addition, I begin to shift my focus to the context and leave the language form to the system. But to tell the truth, this process is really

challenging because I need to improve myself in writing coherence and cohesion." (Teacher)

It seems that the teacher began to develop a better understanding of the English writing and took some active responses to e-feedback in contents.

5.3.2. Teacher-Assessment in the Final Examination

In order to achieve a comparable result, we switched the teacher-assessed scores (from 0 to 15) to e-assessed score range (from 0 to 5). The findings are presented in Table 8.

Table 8. Writing quality in the final examination.

Item	Descriptive statistics	CC	EC	t	p
Holistic	Means	2.783	2.937	-2.004	.044
	S. D.	5671	5116		

Table 8 presents that EC also significantly outperformed CC in writing quality when teacher assessed in the final examination ($p=0.044 < 0.05$). This further validates the opinion that when students receive both holistic assessment and dimensional assessment, they would get improvement in both contents and language form in the process of revision [13, 41, 52, 61, 62]. However, the holistic score is decided by all of those six dimensions on WRM2.0 rather than an average score of those dimensions, suggesting that the increase or decrease in any dimension will be responsible for the ups-and-downs of writing quality.

This study shows that WRM2.0, which provides continuous writing practice, automatic scoring, instructional guidance and descriptive reporting holistically and dimensionally, helped the learners to consciously notice the components of a good writing and pro-actively identify strength and weakness in their products with a view to directing their revision, even though there are few meaningful instructional guidance in the contextual structure. Besides, a lot of e-feedbacks are available to the teacher about areas for improvement, which provides clear guidance about what kinds of instructional activities to be designed in the class.

Generally speaking, in terms of learning benefits, e-assessments can be seen as a means to close the gap between actual and desired performance levels through the provision of timely feedback [3]. It has been proved that L2 student writers eliminate more language errors, improve the language accuracy and fluency over time with effective feedback [52, 63, 64]. And this study found that students were capable of using e-assessments to improve or reinforce their essays by identifying errors or misconceptions, and thus to enhance their writing accuracy (i.e. language form), fluency (i.e. contextual structure) and overall quality to some degree. In the meanwhile, the teacher could turn to the timely and sufficient e-feedbacks to shape the in-class activities with certain targets.

6. Conclusion

This study addresses three issues of e-assessment: What is the impact of WRM2.0 on the language form of students'

writings? What is the impact of WRM2.0 on the contextual structure of students' writings? What is the impact of WRM2.0 on the writing quality of the students' writings? With qualitative method and quantitative method, we observe that the Chinese college English learners who received e-assessment on WRM2.0 outperformed their peers in CC in these three aspects. Besides, a better performance of EC in writing quality was also observed even when teacher assessed than their peers in CC. Therefore, it can be concluded that e-assessment plays a positive role in the process of writing just as some researchers ascertain that e-assessment not only provides ESL teachers and students with individualized reports concerning grammatical errors, ideas, organization and skills, but also offers them an opportunity to monitor the writing process [8, 9, 13, 14, 65-68]. Thus, e-assessment can be seen as a means to help students notice the gap and reinforce their understanding of core concepts [69], close the gap between actual and desired performance levels through the provision of timely feedback [3] and provide sufficient information for the teacher to shape "just-in-time-teaching" [70].

While the results of this research indicate that e-assessment does have the potential to enhance the effectiveness of student writing in the language form, the contextual structure and the writing quality, the sample size is still relatively small, the participants are all drawn from the same major of a single university, and the experiment period may not be long enough. The issues, such as whether the benefits of e-assessment would hold on in the long run, how to integrate e-assessment into the ESL writing classroom, etc. require further study.

Acknowledgements

This study is supported by Project the Ministry of Education of China (Project number 19XJA740008), Southwest Petroleum University (Project number X2018JGZDI035) and Sichuan Center for Education Development Research (Project number CJF19023). We would like to extend our appreciation to the generous contribution of CTB/McGraw-Hill.

References

- [1] Flower, L. and Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *CCC*, 32, pp. 365-387.
- [2] Alexander, S. (2001). E-learning developments and experiences. *Education & Training*, 43, 4/5, pp. 240-248.
- [3] Nicol, D. (2007). Laying a foundation for lifelong learning: Case studies of e-assessment in large 1st-year classes. *British Journal of Educational Technology* 38 (4), pp. 668-678.
- [4] Schmidt, R. (1990). The role of consciousness in second language learning, *Applied Linguistics*, 11, pp 129-158.
- [5] Swain, M. (1995). Three functions of output in second language learning [A]. In Cook, G & Seidlhover, B. (Eds.), *Principles and Practice in the Study of Language* [M]. Oxford: Oxford University Press.
- [6] Van Patten, S. (1996). *Input Processing and Grammar Instruction in Second Language Acquisition* [M]. Norwood, NJ: Ablex.
- [7] Qi, S. & Lapkin, S. (2001). Exploring the role of noticing in a three stage second language writing task [J]. *Journal of Second Language Writing*, 10, pp. 277-303.
- [8] Liou, H.-C. (1994). Practical considerations for multimedia courseware development: An EFL IVD experience. *CALICO Journal*, 11 (3), pp. 47-74.
- [9] Warden, C. & Chen, J. (1995). Improving feedback while decreasing teacher burden in R. O. C. ESL business English writing classes. In P. Bruthiaux, T. Boswood & B. Du-Babcock (Eds.), *Explorations in English for professional communications* (pp. 125-137), Hong Kong: City University of Hong Kong.
- [10] Brock, M. (1990). Customizing a computerized text analyzer for ESL writers: Cost versus gain. *CALICO Journal*, 8 (2), pp. 51-60.
- [11] Brock, M. (1993). Three disk-based text analyzers and the ESL writer. *Journal of Second Language Writing*, 2 (1), pp. 19-40.
- [12] Burston, J. (2001). Computer-mediated feedback in composition correction. *CALICO Journal*, 19 (1), pp. 37-50.
- [13] Ferris, D. R. (1993). The design of an automatic analysis program for L2 text research: Necessity and feasibility. *Journal of Second Language Writing*, 2 (2), pp. 119-129.
- [14] Leacock, C. (2004). Scoring free-responses automatically: A case study of a large-scale assessment. *Examens*, 1 (3), pp. 102-117.
- [15] Tang, J. L., Rich, C. S. & Wang, Y. H. (2012). Technology-enhanced English language writing assessment in the classroom. *Chinese Journal of Applied linguistics*, 35 (4), pp. 385-399.
- [16] Attali, Y., Bridgeman, B., & Trapani, C. (2010). Performance of a generic approach in automated essay scoring. *Journal of Technology, Learning, and Assessment*, 10 (3), pp. 1-21
- [17] McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing*, 15 (2), pp. 118-129.
- [18] Tsai, Min-hsiu. (2012). The consistency between human raters and an automated essay scoring system in grading high school students' English writing. *Action In Teacher Education*, 34 (4), pp. 328-335.
- [19] Rich, C. S., Harrington, H., Kim, J. & West, B. (2008). Automated essay scoring in state formative and summative writing assessment. Paper presented at the annual meeting of the American Educational Research Association, March 2008, NYC.
- [20] Gibbs, G. (2006). Why assessment is changing. In C. Bryan & K. Clegg (Eds). *Innovative assessment in higher education* (pp. 11-22): London: Routledge.
- [21] Wang, J. & Brown, M. S. (2007). Automated essay scoring versus human scoring: A comparative study. *Journal of Technology, Learning, and Assessment*, 6 (2), pp. 4-28.
- [22] Shermis, M. D. & Hamner, B. (2013). Contrasting state-of-the-art automated scoring of essays. *Handbook of Automated Essay Evaluation: Current Applications and New Directions*, pp. 313-346.
- [23] He, X. L. (2013). Reliability and validity of the assessment by the Pigaiwang on college students' writings. *Modern Educational Technology*, 23 (5), pp. 64-67.
- [24] Li, Y. L. & Tian, X. C. (2018). An empirical research into the reliability of iWrite2.0. *Modern Educational Technology*, (2), pp. 5-80.
- [25] Herrington, A. & Moran, C. (2001). What happens when machines read our students' writing? *College English*, 63 (4), pp. 480-499.
- [26] Matzen Jr., R. N., & Sorensen, C. (2006). E-write as a means for placement into three composition courses: A pilot study. In P. F. Ericsson & R. H. Haswell (Eds.), *Machine Scoring of Student Essays: Truth and Consequences* (pp. 130-147). Logan, UT: Utah State University Press.
- [27] Shermis, M. D., Burstein, J. C., & Bliss, L. (2004). The impact of automated essay scoring on high stakes writing assessments. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA.
- [28] Wilson, J. & Andrada, G. (2013). Examining patterns of writing performance of struggling writers on a statewide classroom benchmark writing assessment: The utility of dynamic assessment. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- [29] Saricaoglu, A. (2018). The impact of automated feedback on 12 learners' written causal explanations. *ReCall*, 31 (2), pp. 189-203.
- [30] Vantage Learning, (2007). *MY Access! Efficacy Report*. Newtown, PA: Vantage Learning. <http://www.vantagelearning.com/school/research/myaccess.html>
- [31] Warschauer, M. & Ware, P. (2006). Automated essay scoring, defining classroom research agenda. *Language Teaching Research*, 10 (2), pp. 1-24.
- [32] Grimes, D. (2008). Assessing automated assessment: Essay evaluation software in the classroom. Retrieved December 12th 2016 from <ftp://ftp.ics.uci.edu/pub/grimesd/AutoAssessment.pdf>

- [33] White, L., Hixson, N., D'Brot, J., Perdue, J., Foster, S. & Rhudy, V. (2010). Research brief, impact of Writing Roadmap 2.0 on WESTEST 2 online writing assessment scores. <http://wvde.state.wv.us/oa/pdf/research/Research%20Brief%20-%20WRM2.0%20Impact%20FINAL%2001.27.10.pdf>
- [34] Wilson, J., Olinghouse, N. G. & Andrada, G. N. (2014). Does automated feedback improve writing quality? *Learning Disabilities: A Contemporary Journal*, 12 (1), pp. 93-118.
- [35] Palermo, C. & Thomson, M. M. (2018). Teacher implementation of self-regulated strategy development with an automated writing evaluation system: Effects on the argumentative writing performance of middle school students. *Contemporary Educational Psychology*, (7), pp. 255-270.
- [36] Jiang, X. Q., Cai, J. & Tang, J. L. (2011). Impacts of An Automated Essay Scoring Tool on the Development of Writing Proficiency of Chinese College EFL Learners, *Shandong Foreign Language Teaching Journal*, 145 (6), pp. 36-43.
- [37] Rich, C. S. (2012). The impact of online automated writing evaluation: A case study from Dalian. *Chinese Journal of Applied Linguistics*, 35 (1), pp. 83-99.
- [38] Wang, S. W. & Xian, Y. C. (2012). An empirical study on the efficacy of error correction practice by using an automated writing evaluation system, In *English Teaching Reform in the Digital Age—The Application of Educational Assessment Technology in Writing Instruction*, Tang, J. L., et al. (eds.), pp 254-264.
- [39] Gao, M. X. & Luan, X. H. (2013). The impact of online automated essay scoring system Writing Roadmap 2.0 on writing quality of college students, *Journal of Jilin Institute of Chemical Technology*, 20 (4), pp. 88-90.
- [40] Zhou, L. (2015). The effect of online English writing platforms on college students' syntactic ability, *TEFLE*, 165, pp. 26-29.
- [41] Wang, S. W. (2017). A longitudinal study on the impacts of an automated writing assessment system on the English writing proficiency of Chinese college EFL learners, *Shandong Foreign Language Teaching*, 177 (2), pp. 51-61.
- [42] Li, G. F. (2019). Impact of the integrated feedback based on AWE on students' writing revision. *Foreign Language Education*, 40 (4), pp. 72-76.
- [43] Hoon, T. (2006). Online automated essay assessment: Potentials for writing development. Retrieved December 12th, 2016 from <http://ausweb.scu.edu.au/aw06/papers/refereed/tan3/paper.html>
- [44] Warschauer, M. & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3 (1), pp. 22-36.
- [45] Grimes, D. & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment*, 8 (6), pp. 1-44.
- [46] Wang, Y., Shang, H. & Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26 (3), pp. 234-257.
- [47] Rolim, C. & Isaias, P. (2018). Examining the use of e-assessment in higher education: Teachers and students' viewpoints. *British Journal of Educational Technology*, 4 (50), 1785-1800.
- [48] Chen, C. F. & Cheng, W. Y. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, (2), pp. 94-112.
- [49] Wang, S. W. (2011) On On-line English Writing Feedback with Writing Roadmap 2.0 Automated Evaluation System. *Modern Educational Technology*, 21 (3), pp. 76-81.
- [50] Yang, L. (2013). On the application of AWE system in high-level students' EFL writing learning. *Modern Educational Technology*, 23 (5), pp. 73-77.
- [51] Chandler, J. (2000). The efficacy of error correction for improvement in the accuracy of L2 student writing. Paper presented at the AAAL Conference, Vancouver, BC.
- [52] Fathman, A. & Walley, E. (1990). Teacher response to student writing: Focus on form versus content, in B. Kroll (ed.), *Second Language Writing: Research Insights for the Classroom*, Cambridge: Cambridge University Press.
- [53] Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *Modern Language Journal*, 75, pp. 305-313.
- [54] Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler, *Journal of Second Language Writing*, 13 (4), pp 337-343.
- [55] Ross, S. (1982). The effects of heuristic feedback on EFL composition. *JALT Journal*, 4, pp. 97-108.
- [56] Applebee, A. N. (1981). *Writing in the secondary school (NCTE Research Rep. No. 21)*. Urbana, IL: National Council of Teachers of English.
- [57] Cai, J. G. (2002). Influence of CET writing requirements and scoring criteria on Chinese students' compositions, *Journal of PLA University of Foreign Languages*, 25 (5), pp. 49-53.
- [58] Vann, R. J., Meyer, D. E., & Lorenz, F. O. (1984). Error gravity: A study of faculty opinion of ESL errors. *TESOL Quarterly*, 18, pp. 427-440.
- [59] Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19, pp. 79-101.
- [60] Tang, J. L. (2014). How to integrate an automated writing assessment tool in the EFL classroom, *FLLTP*, 1, pp. 49-57.
- [61] Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, 9, pp. 227-257.
- [62] Feng, L. & Gao, S. F. (2012). A study on effects of rating methods in college English writing formative assessment: A comparative study of holistic and mixed rating methods, *Journal of Beijing Jiaotong University (Social Sciences Edition)*, 11 (3), pp. 126-131.
- [63] Lizotte R. (2001). Quantifying progress in an ESL writing class. *MATSOL Currents*, 27 (1), pp. 7-17.
- [64] Robb T. Ross S, Shorted I. (1986). Salience of feedback on error and its effect of EFL writing quality. *TESOL Quarterly*, 20, pp. 83-93.

- [65] Bolt, P. (1992). An evaluation of grammar-checking programs as self-help learning aids for learners of English as a foreign language. *Computer Assisted Language Learning*, 5 (1-2), pp. 49-91.
- [66] Dalgish, G. (1991). Computer-assisted error analysis and courseware design: Applications for ESL in the Swedish context. *CALICO Journal*, 9 (2), pp. 39-56.
- [67] Healey, D. (1992). Where's the beef? Grammar practice with computers. *CAELL Journal*, 3 (1), pp. 10-16.
- [68] Nutta, J. (1998). Is computer-based grammar instruction as effective as teacher-directed grammar instruction for teaching L2 structures?. *CALICO Journal*, 16 (1), pp. 49-62.
- [69] McCausland, W. D. (2003). *Extended case study: Computer aided assessment and independent learning in macroeconomics*. Bristol: University of Bristol.
- [70] Novak, G. M., Patterson, E. T., Gavrín, A. D. & Christian, W. (1999). *Just-in-time-teaching: blending active learning with web technology*. New Jersey: Prentice Hall.