International Journal of Applied Linguistics & English Literature
ISSN 2200-3592 (Print), ISSN 2200-3452 (Online)
Vol. 2 No. 1; January 2013

Can Automated Writing Evaluation Programs Help Students Improve Their English Writing?

Pei-ling Wang
Department of Applied Foreign Languages, National Kaohsiung University of Applied Sciences
415, Chien-Kung Road, Kaohsiung 807, Taiwan
Tel: 8867-3814526-3276 E-mail: peiling@cc.kuas.edu.tw

Received: 01-08- 2012 Accepted: 10-09- 2012 Published: 01-01- 2013
doi:10.7575/ijalel.v.2n.1p.6 URL: http://dx.doi.org/10.7575/ijalel.v.2n.1p.6

Abstract

This study explores the effect of the automated writing evaluation (AWE) on Taiwanese students writing, and whether student improvement and their perception of the program are related. Instruments included a questionnaire, 735 essays analyzed in *Criterion*, and a pre/post essay. Two classes of 53 college students participated in the study. Descriptive statistics, paired-samples t-tests, Pearson correlation, effect size, and regression were used to analyze the data. Results showed that students improved significantly in terms of the length of the essay and the scores awarded by the machine and the human raters. However, among the five essays, the first essay is the only one showing a significant level of consistency between student improvement and student attitude, and the correlation declined dramatically after the first essay. To conclude, this study may be of importance in confirming the usefulness of the AWE functions such as recursive revising and instant scoring, as well as in providing teachers with a better understanding of how student beliefs about the Criterion program might relate to their writing performance.

Keywords: AWE, Criterion, writing

1. Introduction

1.1 The Problem

While many studies have shown that students learn to write by writing (e.g. Brown, 2001; Elbow, 1973; Zamel, 1982), the National Commission on Writing (2003) pointed out, American students practice writing much less than they need. Similarly, Tsai (2010) claimed that the reason why many Taiwanese students' English writing skills are so poor lies in the fact that they seldom or never practice writing.

Actually, writing is not only a nightmare for students. Reading and correcting student's writing is also very time-consuming for teachers. Especially for Asian teachers who often have more than 50 students in one class, asking students to write more means teachers have to devote extended periods of time to assessing and giving comments on student work.

With the advent of the Internet, the topic of automated writing evaluation (AWE) has received considerable attention. Proponents of AWE maintained that the feature of immediate feedback of AWE can make learning more efficient and interesting (Frost, 2008; LinHuang, 2010; Moseley, 2006; Taylor, 1996); additionally, the AWE gives useful advice on organization and also objective feedback regarding the revision (Grimes, 2008; Phillips, 2007).

On the other hand, critics of AWE argued that the validity of AWE programs is doubtful. For example, McCurry's study (2010) showed that the machine did not grade the broad and open writing tasks as reliably as human raters. Other studies (Chen, 2006; LinHuang, 2010; Wang & Brown, 2007) also found that the machine tended to score higher than human graders.

Given the fact that AWE programs are usually very costly, it is necessary to know the effectiveness of the AWE program before schools purchase the license of a particular program. Unfortunately, there have been few studies on the use of the AWE programs, and the results of these studies are still conflicting and inconclusive.

1.2 Importance of the Problem

At present, *My Access* and *Criterion* are the two most popular AWE programs in Taiwan. However, there has been little research on the outcomes of these two programs in the Taiwan classroom setting. Studies examining Taiwanese student attitudes toward the program were scant, and most of them inspected *My Access* (Chen & Cheng, 2008; Yang, 2004; Yu & Yeh, 2003) instead of *Criterion*. Moreover, most of the previous studies (e.g. Frost, 2008; Moseley, 2006; Otoshi, 2005) only examined students' writing improvement in one genre of essay (e.g. persuasive writing), very little attention has been paid to other rhetorical modes such as process, cause/effect, and comparison/contrast essays. Furthermore,

previous studies seldom adopted a pre-essay and a post-essay design for human raters to compare student writing performance before and after the AWE treatment. In other words, the methods or designs of the previous studies could be considered inappropriate or inadequate.

In light of these concerns, this study has three purposes: (1) to discuss the effect of the *Criterion* program on Taiwanese students' writing improvement in three rhetorical modes of essays (i.e. process, cause/effect, and comparison/contrast essays); (2) to verify the students' writing improvement by comparing students' pre and post essays assessed by human raters; and (3) to analyze the relationship between students' writing improvement assessed by the program and students' attitudes toward the program.

In the following, the researcher reviews the relevant literature in the effect of the AWE programs and students perceptions of the programs, which inform the theoretical framework of this study.

1.3 Relevant Literature

Several studies (LinHuang, 2010; Flinn, 1986; Grimes & Warschauer, 2006; Warschauer & Grimes, 2008) have noted that the AWE programs made revision easier and also facilitated students writing skills. For example, LinHuang's study revealed that the program helped students to revise their writings, especially in the areas of grammar and spelling. Grimes and Warschauer also found that the AWE programs could motivate students to write more. Furthermore, some studies (Dikli, 2006; Wang, 2011; Yang, 2004) have confirmed that the tools provided by the AWE programs such as on-line dictionary and e-portfolio were very helpful to students. In Wang's study, for instance, participants generally believed that the tool of e-portfolio allowed them to examine their growth in writing, and also helped them value the process writing curriculum.

In contrast, some studies have shown that students were dissatisfied with the functions of AWE. In Cheng's study (2006), many students complained that the machine feedback was too vague to understand, and around half of the student felt *My Access* was slightly helpful to them. Likewise, both Yu and Yeh (2003) and Yang (2004) reported that most students believed the feedback from *My Access* was repetitive and similar, which might be useful for the primary revision.

The drawbacks of the AWE programs were also found in Chen, Chiu, and Liao's study (2009). In this study, the researchers examined the feedback messages provided by *My Access* and *Criterion* for 269 student essays. The results showed that although these two programs could identify about thirty types of grammatical errors, the feedback provided by these two programs was not entirely accurate. In fact, most of the machine feedback messages in *My Access* were false alarm; on the contrary, almost all of the feedback messages provided by *Criterion* had 70% accuracy.

According to the previous studies, AWE programs have both merits and drawbacks. However, since most of these studies were not conducted in an English as a foreign language (EFL) setting or did not investigate the effectiveness of *Criterion* by examining students' improvement in the different rhetorical modes of writing, this study tries to investigate this under-researched area.

1.4 Research Question

This study asks three research questions. First, are the students' essays for three different modes improved when the essays are scored by the *Criterion*? Second, do the students' English writing performances when assessed by human raters improve? Third, what is the relationship between students' writing improvement assessed by the *Criterion* and students' perceived effectiveness of the program?

2. Method

2.1 Participants

The participants of this study were two classes of sophomore students (58 students) majoring in English from a technical university in Southern Taiwan, who were taking the English writing course with the researcher. Therefore, this study uses a quasi-experimental and convenient sampling design (Creswell, 1994). Since this study examined students' writing samples of five different essays, only those who had submitted all assignments were counted as the valid participants. As a result, five students were excluded and 53 students (49 females and four males) became the participants of the study.

2.2 Instruments

The instruments include a questionnaire, student writing samples from the e-portfolio in *Criterion*, and a pre-essay and a post-essay. The questionnaire has 16 five-point Likert-scale type questions (student attitudes toward the effectiveness/functions of Criterion), 1 multiple-choice question (the effect of Criterion on student English ability improvements), and 12 open-ended questions (student evaluation of Criterion). The reliability of the questionnaire is estimated by Cronbach's α, which shows that the questionnaire is reliable ($\alpha = 0.76$).

2.3 Procedure for data collection

Data were collected in the fall semester of the 2010 school year. In the first week of the class, the teacher-researcher clearly explained how to use the program and also demonstrated various functions as well as the scoring mechanism of the program. Each student had one computer to practice on, and the teacher circulated in the computer lab to monitor student progress. After that, students wrote the pre-essay on paper.

During the semester, the teacher taught three modes of English writing (i.e. process, cause/effect, and comparison/contrast essays). After each mode was taught, the students had to write one to two topic-related essays in *Criterion*. In the end, student wrote one process essay, two cause/effect essays, and two comparison/contrast essays. Since students were requested to submit their drafts 3 times for each essay, a total of 735 writing samples were collected.

In order to make sure students understand the various functions in *Criterion*, the researcher gave each student ten to fifteen minutes tutoring in the computer lab. During the tutoring, the researcher and students reviewed the feedback provided by the *Criterion* together. Whenever students had difficulties in revising their drafts according to the machine feedback, the researcher offered some advice to assist students.

The researcher retrieved student essays from the electronic portfolios in *Criterion,* and recorded the total number of words and the scores for each student's first and last submissions. In the last week of the semester, students were asked to complete the questionnaire and wrote the post-essay on paper. Later, two experienced English writing teachers were invited to evaluate the subjects' pre- and post-essays following the rubric on the Joint Common Entrance English Writing Examination for Universities in Taiwan. The inter-rater reliability was quite high (r= 0.80, p=0.00).

3. Results

3.1 Student Writing Improvements Evaluated by the Machine and Human Raters

Regarding student writing samples in *Criterion*, Table 1 shows that the mean number of words in the student's essays increased and their scores also improved from the first essay to the fifth essay. For example, in the first paper, the mean number of words in the first version ranged from 136 to 444. In comparison with the first paper, the words in the first version in the fifth paper totaled between 254 and 840 words. Furthermore, in the first paper, student scores ranged from 3 to 5. When it came to the final version of the second paper, the minimum score was advanced to 4. The same tendency was found in the third paper. In the fourth and fifth papers, students had a performance score of at least 4 in their first attempt on the essays. Some students had achieved the highest score (6) in *Criterion* since their second paper submission; therefore, the maximum score did not change from the second paper to the fifth paper.

A paired-samples t-test was conducted to compare students' mean number of words in their first submission and their final submission in each essay. There was a significant difference in the words for the first submission (M=249.60, SD=73.90) and the final submission (M=268.43, SD=76.36) in the first paper, t(52)=-3.35, p=0.001; in the words for the first submission (M=308.02, SD=83.74) and the final submission (M=341.91, SD=74.85) in the second paper, t(52)=-6.09, p=0.000; in the words for the first submission (M=321.47, SD=98.51) and the final submission (M=355.02, SD=99.41) in the third paper, t(52)=-3.65, p=0.001; in the words for the first submission (M=391.79, SD=93.55) and the final submission (M=400.30, SD=92.07) in the fourth paper, t(52)=-3.49, p=0.001; and in the words for the first submission (M=398.64, SD=111.26) and the final submission (M=412.15, SD=101.98) in the fifth paper, t(52)=-3.21, p=0.002. These results suggest that students' improvements in the text length were statistically significant. Moreover, a paired-samples t-test was conducted to compare students' mean scores in their first submission and their final submission in each essay. There was a significant difference in the scores for the first submission (M=4.00, SD=0.62) and the final submission (M=4.30, SD=0.60) in the first paper, t(52)=-4.36, p=0.000; in the scores for the first submission (M=4.62, SD=0.59) and the final submission (M=5.09, SD=0.49) in the second paper, t(52)=-6.35, p=0.000; in the scores for the first submission (M=4.38, SD=0.59) and the final submission (M=4.74, SD=0.56) in the third paper, t(52)=-4.41, p=0.000; in the scores for the first submission (M=5.06, SD=0.53) and the final submission (M=5.34, SD=0.55) in the fourth paper, t(52)=-4.53, p=0.000; and in the scores for the first submission (M=4.85, SD=0.60) and the final submission (M=5.17, SD=0.58) in the fifth paper, t(52)=-4.95, p=0.000. These results suggest that students' improvements in the scores were statistically significant (See Table 2).

Although significant differences in the words were found between the first version and the final version of these five papers, the corresponding Cohen's d effect sizes, which ranged from 0.09 to 0.43, indicated no or small-to-medium levels of statistical power (See Table 2). More specifically, except the second paper (d= -0.43) and the third paper (d=-0.34), whose significant differences were close to medium in statistical power, the Cohen's d effect sizes of the other three papers were quite small. The effects of the significant difference of the fourth paper (d=-0.09) and the fifth paper (d=-0.13) were both below 0.2. These results seem to indicate that students had greater enthusiasm for writing in the beginning of the semester, but as the time passed, they became less diligent in expanding their papers. Possibly, it is because students had lots of assignments to accomplish as the semester was about to finish.

As for the significant difference in the scores between the first version and the final version of these five papers, the Cohen's d effect sizes underlying these significant levels ranged from -0.49 to -0.87 and demonstrated medium to large effects (See Table 3). These results showed that students writing scores in their first draft were better than their writing scores in their final drafts. The effect was especially evident in the second essay (d=0.87), which might be due to the fact that the instructor had tutored individual students in relation to their first draft of this essay in the computer lab before the students wrote their second drafts.

In order to justify whether students' writing skills improved at the end of the semester, in addition to those writing samples assessed by *Criterion*, students wrote a pre- and a post-essay on paper, which were evaluated by human raters. Similarly, the score differences between the pre-essays and post-essays achieved a significant level ($p<0.01$), which indicated that student writing skills improved. The Cohen's d effect size showed that the significant difference was

close to moderate level of statistical power (See Table 4).

Table 1. Words/Scores of First and Final submission

Essay	N.	Words of 1st submission		Words of final submission		Scores of 1st submission		Scores of final submission	
		Min	Max	Min	Max	Min	Max	Min	Max
1	53	136	444	143	444	3	5	3	5
2	53	127	528	206	532	3	6	4	6
3	53	104	565	104	567	3	6	4	6
4	53	214	700	254	700	4	6	4	6
5	53	254	840	277	763	4	6	4	6

Table 2. Improvements in the Mean Number of Words (N=53)

Essay	Words of 1st submission		Words of final submission		t-test		Cohen's d effect size
	Mean	SD	Mean	SD	t	p	d
1	249.60	73.90	268.43	76.36	-3.35	0.001**	-0.25
2	308.02	83.74	341.91	74.85	-6.09	0.000***	-0.43
3	321.47	98.51	355.02	99.41	-3.65	0.001**	-0.34
4	391.79	93.55	400.30	92.07	-3.49	0.001**	-0.09
5	398.64	111.26	412.15	101.98	-3.21	0.002**	-0.13

**=p<0.01, ***=p<0.001

Note: According to Cohen (1988), guideline for the d effect size, 0.2 a small effect, around 0.5 a medium effect, and 0.8 to infinity a large effect.

Table 3. Improvements in the Mean Scores (N=53)

Essay	Scores of 1st version		Scores of final version		t-test		Cohen's d effect size
	Mean	SD	Mean	SD	t	p	d
1	4.00	0.62	4.30	0.60	-4.36	0.000***	-0.49
2	4.62	0.59	5.09	0.49	-6.35	0.000***	-0.87
3	4.38	0.59	4.74	0.56	-4.41	0.000***	-0.63
4	5.06	0.53	5.34	0.55	-4.53	0.000***	-0.52
5	4.85	0.60	5.17	0.58	-4.95	0.000***	-0.54

***=p<0.001

Table 4. Student Writing Performance Evaluated by Human Raters

Pre-essay Score		Post-essay Score		t	p	Cohen's d effect size
Mean	SD.	Mean	SD.	3.081	0.003**	-0.39
10.87	2.26	11.82	2.59			

**=p<0.01

3.2 Relationship between Student Writing Improvement and Student Attitude

Among the 53 participating students, four students refused to answer the questionnaire. Therefore, only 49 students completed the attitude survey. Table 5 showed the descriptive statistics of students' attitudes and student improvement in their writing samples scored by the *Criterion.* Moreover, according to the result of the Pearson correlation test, there were positive correlations between student improvement in their writing samples scored by the *Criterion* and student attitude toward the usefulness of the program, but the relationships were not significant ($p<0.05$). In detail, the correlation tests showed student improvement in each essay's score and their attitudes were as follows: the first essay (r

=0.59, p=0.692), the second essay (r=0.004, p=0.981), the third essay (r=-0.031, p=0.832), the fourth essay (r=0.073, p=0.623), and the fifth essay (r=0.151, p=0.307). The results revealed that the first essay is the only one showing a strong correlation and the correlation decreased noticeably after the first essay. Perhaps the students did not feel the program was very helpful.

To examine to what extent writing score improvements explain the variation in student attitude toward the *Criterion* program, the researcher did a regression on these variables. The results showed that the scores only explain 3.9% of the variance in student attitude, R^2=0.039, F=0.339, p=0.886. That is, student writing score improvements do not significantly predict student attitude, β=0.06, t=0.43, p=0.66 (first essay); β=-0.06, t=-0.37, p=0.71 (second essay); β=-0.03, t=-0.22, p=0.82 (third essay); β=0.09, t=0.64, p=0.52 (fourth essay); and β=0.17, t=1.09, p=0.27 (fifth essay) (See Table 6).

Table 5. Student Writing Score Improvements and Student Attitude (N=49)

	Mean	SD
Improvements in the 1st essay	0.31	0.51
Improvements in the 2nd essay	0.49	0.55
Improvements in the 3rd essay	0.37	0.60
Improvements in the 4th essay	0.29	0.46
Improvements in the 5th essay	0.33	0.47
Attitude toward *Criterion*	3.34	0.36

Table 6. Regression for Writing Score Improvements and Student Attitude

	B	Standard Error of the Estimate	Beta (β)	t	Sig.
Latitude	3.286	0.101		32.67	0.000
Improvements in the 1st essay	0.050	0.114	0.069	0.437	0.665
Improvements in the 2nd essay	-0.040	0.107	-0.060	-0.372	0.712
Improvements in the 3rd essay	-0.021	0.093	-0.035	-0.222	0.826
Improvements in the 4th essay	0.079	0.122	0.099	0.644	0.523
Improvements in the 5th essay	0.135	0.123	0.177	1.099	0.278
R= .197 R^2= .039 Adjusted R^2= -.076		F=0.339 p=0.886			

Dependent variable: students' attitude

4. Discussion and Conclusions

This study discusses the issue of the effect of the use of the AWE program on Taiwanese college student writing skills. Student writing improvements were assessed by both the *Criterion* and human raters. In addition, the present study also tries to explore the relationship between student improvement assessed by the *Criterion* and student attitude toward the program.

The first finding of the study showed that students significantly wrote longer for each essay in the *Criterion*, and they also got higher machine scores. One possible explanation for this result is that editing and revising in *Criterion* is so convenient that students could make changes in their writing easily. If this is the case, then this study is in accord with the results of the previous studies (e.g. LinHuang, 2010; Warschauer & Grimes, 2008; Williamson & Pence, 1989) which confirmed that the feature of word processing was beneficial to writing. Another possible explanation is that writing for the machine might be like a computer game for some students. The recursive process of writing, submitting, getting instant scores and feedback, revising, resubmitting, and getting scores again is similar to a series of actions or instructions that the player must take or follow in order to find the hidden treasure in a computer game. Thus, students might be motivated to continually revise their drafts based on the machine advice in the hope that they could eventually get the prize - the highest score on a computer. The third possible explanation rests in the fact that the AWE programs rewarded essay length, which has been widely reported by the previous studies (e.g. Chen & Cheng, 2008; Grimes, 2008). After sending their drafts several times, students might have learned that if they wanted to get a higher machine score, they would have to write longer. Consequently, students might put more emphasis on the length of their essays in order to get a better score.

The second finding of the study indicated that students had significant improvement in their post-essays assessed by human raters, which confirmed that student writing skills were enhanced at the end of the semester. This finding is quite encouraging, but it should be interpreted cautiously. Possibly, the uses of the *Criterion* and the frequent writing practice may have contributed to students' improvement. In this study, each student was requested to submit 15 drafts and the *Criterion* evaluated a total of 735 submissions, which would be extremely difficult for a teacher to accomplish in one

semester. However, since the participants of this study were English majors, they were taking other English-related courses and probably had been working on other writing assignments while the study was conducted. Therefore, it would be arbitrary to declare that the improvement of student writing was exclusively due to the use of *Criterion*.

The third finding of this study was that there was no significant relationship between the machine score and student attitude toward the program. In fact, except the first essay showing positive correlation, the correlation among other essays decreased considerably, which might indicate the longer students tried the Criterion program, the more they felt the program was not useful. It is unclear, however, why students had such a negative evaluation.

The findings of this study lead to a number of implications. First, an AWE program is a good tool to motivate students to devote to the recursive process of drafting and revising. However, since the machine might not really understand the content of an essay, teachers had better randomly check student writing samples and provide consultation with individual student to clarify the vague or even incorrect machine messages (if there are any). Next, considering the machine might value a wordy but meaningless essay, teachers need to remind students about this drawback of the AWE program. Teachers should also encourage students to regard the quality of their essay more highly than the quantity of words or the machine scores. Furthermore, in the beginning of the class, teachers could show their own positive attitude toward the machine and patiently demonstrate various functions of the program to increase student confidence in the ability of the program, although it is also important to warn students not to blindly trust the machine scoring. Finally, according to Grimes (2008), "if AWE is used persistently and indiscriminately without a competent teacher or mentor and without authentic human audiences, then it is possible that students' beliefs about the social nature of writing may be distorted, as critics have feared (p.197)." Writing teachers who plan to incorporate an AWE program into the curriculum need to consider the importance of meaningful communication between the writer and the real reader. Future studies might display strategies for teachers to integrate the activity of peer feedback for revision with the application of AWE programs, which will help to mediate the limitations of the use of the AWE in the classroom setting.

Acknowledgements

Part of this paper was presented in the 2011 Symposium on Second Language Writing. National Taiwan Normal University. June 9-11, 2011. Taipei, Taiwan. The author would like to express her gratitude to the audience for their valuable advice.

References

Brown, H. D. (2001). *Teaching by Principles*. Addison Wesley Longman.

Chen, C. F. E., & Cheng, W. Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Teaching, 12* (2), 94-112.

Chen, H. J. (2006). Examining the scoring mechanism and feedback quality of My Access. *Proceedings of Tamkang University Conference on Second Language Writing.* Tamkang University, Taipei.

Chen, H. J., Chiu, T. L., & Liao, P. (2009). Analyzing the grammar feedback of two automated writing evaluation systems: My Access and Criterion. *English Teaching and Learning, 33* (2), 1-43.

Cheng, W. Y. (2006). The Use of a Web-based Writing Program in College English Writing Classes in Taiwan— A Case Study of MyAccess. Unpublished Master's thesis. National Kaohsiung First University of Science and Technology, Taiwan.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Creswell, J. W. (1994). *Research design.* Thousand Oaks, CA: SAGE Publications.

Dikli, S. (2006). An overview of automated scoring of essays. The Journal of Technology, Learning, and Assessment, 5 (1), 1-36. Retrieved from http://escholarship.bc.edu/ojs/index.php/jtla/article/view/1640

Elbow, P. (1973).*Writing without Teachers*. New York: Oxford University Press.

Flinn, J. (1986). *The role of instruction in revising with computers: Forming a construct for good writing* (ED 274963).

Frost, K. L. (2008). *The effects of automated essay scoring as a high school classroom intervention*. Unpublished Doctoral dissertation. University of Nevada, Las Vegas, USA.

Grimes, D. C. (2008). *Middle school use of automated writing evaluation: A multi-site case study*. Unpublished Doctoral dissertation. University of California, Irvine, USA.

Grimes, D. C., & Warschauer, M. (2006). Automated essay scoring in the classroom. American Educational Research Association (AERA) Annual Conference, San Francisco, CA, USA.

LinHuang, S. H. (2010).*The exploitation of e-writing in an EFL classroom: Potential and challenges*. Unpublished Master's thesis. I-Shou University, Taiwan.

McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing, 15*, 118-129. http://dx.doi.org/10.1016/j.asw.2010.04.002

Moseley, M. H. (2006).*Creating recursive writers in middle school: the effect of a writing program on student revision practices*. Unpublished Doctoral dissertation. Capella University, USA.

National Commission on Writing (2003). The Neglected "R": The Need for a Writing Revolution. New York, NY, College Entrance Examination Board.

Otoshi, J. (2005). An analysis of the use of Criterion in a writing classroom in Japan. *The JALT CALL Journal, 1*(1), 30-38.

Phillips, S. M. (2007). Automated essay scoring: A literature review. TASA Institute, Society for the Advancement of Excellent in Education, 1-70.

Taylor, J. (1996). *Computers: Tools of oppression, tools of liberation*. A paper presented at the annual meeting of the Conference on College Composition and Communication, Milwaukee (ED 434350).

Tsai, P. Y. (2010). Students' biggest writing problem- they never write! Retrieved from http://mag.udn.com/mag/campus/storypage.jsp?f_MAIN_ID=13&f_SUB_ID=1259&f_ART_ID=290730

Wang, Y. J. (2011). *Exploring the effect of using automated writing evaluation in Taiwanese EFL students' writing*. Unpublished Master's thesis. I-Shou University, Taiwan.

Wang, J., & Brown, M. S. (2007). Automated essay scoring versus human scoring: A comparative study. *The Journal of Technology, Learning, and Assessment, 6*(2), 1-28.

Warschauer, M., & Grimes, D. C. (2008). Automated writing assessment in the classroom. *Pedagogies, 3*, 22-36.

Williamson, M. M., & Pence, P. (1989). Word processing and student writers. In B. K. Britton & S. M. Glynn (Eds.) (1989). *Computer writing environments: Theory, research, and design.* (pp.93-127). Hillsdale, NJ: Lawrence Erlbaum.

Yang, N. D. (2004). Using My Access in EFL writing. *Proceedings of the 2004 International Conference and Workshop on TEFL & Applied Linguistics* (pp. 550-564). Taipei, Taiwan: Ming Chuan University.

Yu, Y. T., & Yeh, Y. L. (2003). Computerized feedback and bilingual concordance for EFL college students' writing. *Proceedings of the 2003 International Conference on English Teaching and Learning in the Republic of China* (pp. 35-48). Taipei, Taiwan: Crane.

Zamel, V. (1982).Writing: the process of discovering meaning. *TESOL Quarterly*, *16*, 195-209. http://dx.doi.org/10.2307/3586792