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,
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


