Criterion in a Writing Class: Help or Hindrance

Olesya Shatunova

ライティングクラスにおけるクライテリオン（*Criterion*）： そのメリットデメリット

オレシア シャトノバ

Abstract

　　With technological innovations steadily entering education Automated Writing Evaluation（AWE）software has been applied to the evaluation and assessment of English writing performance and support revision. An automated evaluation has been a topic of hot debate and has been perceived as "both a boon and a bane in the struggle to improve writing instruction" （Grimes and Warschauer, 2010）. Criterion is a one of such Web-based learning tools that aims to support writing instruction across many different levels and several genres. Once students submit their essays, Criterion provides a performance summary that includes holistic scores and the corresponding feedback on each error. Although Criterion has been used in writing classes for the past decade in Japan, the actual benefits of Criterion has not been fully investigated. This paper presents a descriptive study of how and whether Criterion feedback can improve academic writing skills of students at a tertiary level by analyzing how students used feedback provided by the software.

Key words: Automated Writing Evaluation（AWE）, feedback, academic writing

　　自動ライティング評価ソフトウェア（AWE）が英語のライティングの向上とその修正支援サポートとして適用されている。本論文は、どのようなフィードバックが大学生のライティングスキルを向上させるのか、そしてその日本の大学生はどのようにしてこのフィードバックをアカデミックライティングのスキルアップに反映させているのか、について述べている。

Introduction

The expansion of globalization has led to an increased significance of the role of written English, as it has been a major source of information for a multitude of fields, both popular and academic. Writing is an indispensable part in the mastery of language, as well as it allows measurements of many other skills. Learning how to write in a second language is one of the most challenging aspects of second language learning. "Even for those who speak English as a first language, the ability to write effectively is something that requires extensive and specialized instruction"（Hyland, 1996）.

The Role of Feedback

Feedback has long been regarded essential for the development of second language（L2）writing skills: it provides information to students on how successful their writing is and what still they need to improve in their writing; it also helps them understand what good writing is（Hyland and Hyland, 2006; Leki, Cumming, and Silva, 2008）, as well as feedback teaches both the conventions of writing and L2 grammatical forms（Hedgcock and Lefkowitz, 1994; Paulus, 1999）. In process-based, learner-centered classrooms feedback is as an important developmental tool that helps learners to move through multiple drafts.

Up until the 1970s, written feedback was largely concerned with linguistic accuracy, and most attention was paid to error corrections. It was traditionally provided by teachers at the end of the writing process. However, in the 1970s and 1980s emphasis was placed not on a final product but on the writing process itself, as writing was regarded as "non-linear, exploratory, and generative process whereby writers discover and reformulate their ideas as they attempt to approximate meaning"（Zamel, 1985）. This process-based writing pedagogy focused not on isolated parts of texts or grammatical features, but rather on discovering ideas, drafting, and revising. The shift in writing practice also had a significant influence on feedback practices. Teachers sought to support writers through multiple drafts by providing feedback and suggesting revisions during the writing process rather than at the end of it. In addition, feedback practices have

transformed over the past 20 years, and while teacher written feedback remains the main feedback, it is now often combined or supplemented with peer feedback, oral conferences, and even computer-delivered feedback (Hyland and Hyland, 2006).

Traditionally, the authority to provide feedback to students has been in the hands of teachers. However, it is time-consuming for teachers to correct all student errors and to give individualized feedback. As a result, the feedback teachers provide to students might be delayed. When it takes a week or two to get feedback, the flow of the learning process breaks, so students tend to lose interest in the assignment and motivation to improve their work. The revision cycle needs to happen as quickly as possible, so students still recall what they did and thought at the time they wrote their assignment.

With technological advancement and the expansion of the Internet, computerized feedback provided by automated writing evaluation (AWE) software, has exerted and increased influence on writing instruction (Warschauer and Ware, 2006). Unlike traditional feedback approaches such as feedback provided by teachers and peers, AWE can be operated independently, and it can give students near-instant feedback.

Automated Writing Evaluation (AWE)

Recently automated writing evaluation (AWE) software has been applied to the evaluation and assessment of English writing performance and support revision in both L1 and L2 settings. Automated writing evaluation has been a topic of hot debate and has been perceived as "both a boon and a bane in the struggle to improve writing instruction" (Grimes and Warschauer, 2010). Automated evaluation (AWE) systems have been under development since the 1960s, when a national network of US universities, known as the College Board, supported the development of project Essay Grade to help score thousands of high school student essays (Page, 2003 cited in Warschauer and Ware, 2006). However, the results were far from expectation, and it was not until the 1980s, when microcomputers were introduced, the interest in the Project Essay Grade was once again renewed,

and Writer's Workbench was created (Warschauer and Ware, 2006). This system instead of scoring essays provided feedback to writers, although the feedback was quite limited, such as flagging misspelled words and identifying long and short sentence; however, the precedent was made – providing the feedback.

In the 1990s, ETS developed e-rater, Vantage Learning created Intellimetric, and Intelligent Essay Assessor scoring machine was developed by Pearson Knowledge Technologies. Similar to Page Essay Grade, e-rater and Intellimetric use regression model based on a corpus of human-graded essays; however, these two scoring engines could analyze broader range of lexical, syntactic, and discourse elements (Attali and Burshtein, 2006). Intelligent Essay Assessor uses latent semantic analysis technique to evaluate essays by comparing semantic meaning of a created text with a broader corpus of textual information on a similar topic.

Criterion

Criterion is a Web-based learning tool that aims to support writing instruction across many different levels (from Grade 4 to GRE level) and several genres. The feedback provided by the software can be orientated toward English language learners, e. g., practicing for writing TOEFL essays, and therefore has been recently marketed more specifically as an English language tool (Lim and Kahng, 2012).

Teachers can design a writing assignment either from selecting a category (from grade 4 to GRE level), a topic mode (e.g. persuasive, informative), and an essay topic, or teachers can create an original essay prompt. Additionally, a teacher can set a writing time limit. A teacher can see the results of the Criterion scoring, as well as the whole-class and individual student feedback. Summary tables and charts can give insights of overall class's performance and common patterns of writing errors. "Criterion also lists individual students' performances, including their holistic scores and the analytic feedback regarding the five categories: grammar, usage, mechanics, style, and organization" (Lim and Kahng, 2012). Also, a teacher and a student can have online dialogues to discuss the essay and

feedback. Students have the opportunity to create a plan before writing their essays using one of the eight graphic organizers. The essays may be typed, or copied and pasted, into the text box at the bottom of the page. According to Anderson (2013), "Students may format the essay with bold text, italics, underlining, bullets, and numbering. They may also use the Spell Check and Thesaurus features before submitting the essays." Once an essay is submitted, a performance summary is generated that presents a holistic score and the number of errors and the corresponding feedback on each error. The bar graph below shows the feedback in Usage category.
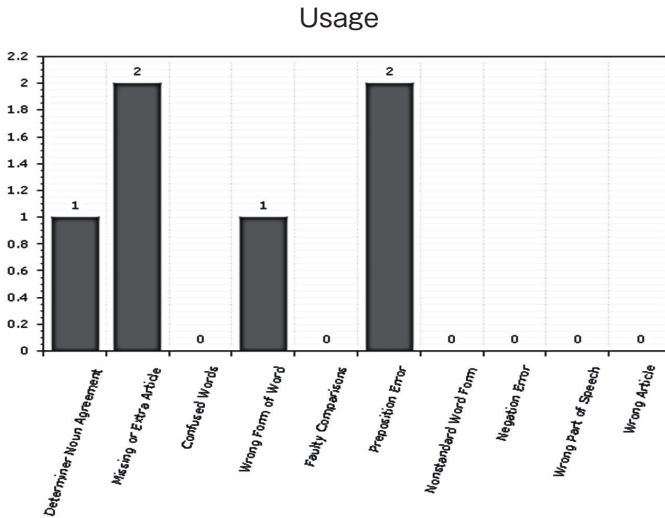
Usage

Criterion has two applications that are based on natural language processing (NLP) methods. *Critique* is an application that evaluates and provides feedback for errors in grammar, usage, and mechanics, identifies the essay's discourse structure, and recognizes potentially undesirable stylistic features. The companion scoring application, *e-rater* version 2.0, extracts linguistically-based features from an essay and uses a statistical model of how these features are related to that used in top-scoring essays on the same prompt and assigns a holistic score to the essay (Burstein,

Chodorow, and Leacock, 2004).

The Study Overview

Research Question: how and whether Criterion feedback can improve academic writing skills of learners at a tertiary level by comparing human raters' feedback with feedback provided by the software in the following areas of feedback: Grammar, Usage, Mechanics, and Style.

Procedure:

This study was conducted using 30 Japanese first-year students' essays submitted to Criterion as the first draft; the students receive comments from the instructor on their second draft.

Data analysis:

In this study error detection between Criterion and human instructors was compared, however, the main focus of the study was the analysis of students' revision.

Results and Discussion

Table 1
Grammar: Feedback and Error Correction

Trait	Error detection by *Criterion*			Error Correction by Students		
	Total number of errors	Incorrectly detected errors	Undetected Errors	Correct Revision of errors detected by *Criterion*	Incorrect Revision of errors detected by *Criterion*	Correct Revision of errors that were not detected by *Criterion*
Fragments	33	30	3	4		1
Run-on Sentences	2					
Garbled Sentences	10	4	4	5		
Subject-Verb Agreement	49	2	25	31	3	1
Ill-formed Verbs	6		3	3	1	
Pronoun Errors	1				1	

Table 1 shows the error detection by Criterion and error correction by students. As for feedback on grammar, many fragment errors were incorrectly flagged by Criterion because of bad text formatting; also titles and headings were detected as fragments. As for the garbled sentences, students were able to review half of such sentences, while the garbled sentences that were no flagged by Criterion were not revised by students. Criterion could detect 49 Subject Verb (SV) agreement errors out of 75, and students were able to correct 62% of SV errors detected by Criterion. However, students were not able to identify SV errors that were not flagged by Criterion. In addition, there was no correction of possessive errors or missing/wrong words provided by Criterion.

Table 2
Usage: Feedback and Error Correction

Total number of errors	Error Correction by Students			Error Correction by Students		
	Total number of errors	Incorrectly detected errors	Undetected Errors	Correct Revision of errors detected by *Criterion*	Incorrect Revision of errors detected by *Criterion*	Correct Revision of errors that were not detected by *Criterion*
Determiner Noun Agreement	19		2	9		
Missing or Extra Article	172		51	101		
Preposition Error	8		13	1		1

As Table 2 shows, the most useful feedback by Criterion on Usage was provided on Article usage: students were able to correct 59% of Article errors detected by Criterion. The least useful feedback from Criterion was on Preposition Usage: Criterion could detect only 8 errors out of 21, with only one successful correction made by a student.

Also, errors such as confused words, wrong form of word, faulty comparison, nonstandard word form, negation error, wrong part of speech and wrong article were not corrected by *Criterion*.

Table 3
Mechanics: Feedback and Error Correction

Trait	Error detection by *Criterion*			Error Correction by Students		
	Total number of errors	Incorrectly detected errors	Undetected Errors	Correct Revision of errors detected by *Criterion*	Incorrect Revision of errors detected by *Criterion*	Correct Revision of errors that were not detected by *Criterion*
Spelling	40	20		23		
Missing Comma	32	6	16	17	5	
Extra Comma	11	3	4	7		

As Table 3 shows, the most effective feedback was provided on Comma usage; students could revise successfully 70% of comma errors. Most spelling errors were incorrectly detected by Criterion: many words that related to Japanese culture and computer terms were flagged as errors. Numerous errors were not detected, such as missing initial capital letter, missing question mark, missing final punctuation, missing apostrophe, hyphen error, fused words, compound words, and duplicates.

Table 4
Style: Feedback and Error Correction

Trait	Error detection by Criterion			Error Correction by Students		
	Total number of errors	Incorrectly detected errors	Undetected Errors	Correct Revision of errors detected by *Criterion*	Incorrect Revision of errors detected by *Criterion*	Correct Revision of errors that were not detected by *Criterion*
Repetition of Words	475			11		
Short Sentence	52	36		10	1	
Passive Voice	10	9	1	1	2	

As Table 4 shows, the most useful feedback by Criterion was provided on Word Repetition. However, students failed to revise their essays, as only 11 words out of 475 words were replaced with appropriate synonyms or

pronouns.

Out of 52 Short Sentence errors 36 errors were incorrectly detected because of bad text formatting. However, out of 16 Short Sentence errors, students could successfully revise 10 Short Sentence errors. Criterion flagged ten Passive Voice errors; although, nine such errors were mistakenly flagged.

Problems with Feedback provided by Criterion

One of the main problems with *Criterion* feedback was its inconsistency. For example, in the following extract from a student's essay:
Some people stay up late and lack[1] of sleep reduce[1] their energy. Also, lack of sleep cause the confusion of the life rhythm.
Criterion Feedback:

1 　主語と動詞が呼応していない可能性があります。文を読み直して、主語と動詞（数：単数か複数）が一致するよう修正してください。
This subject and verb may not agree. Proofread the sentence to make sure the subject agrees with the verb.

Criterion marked the Subject-Verb agreement mistake in one sentence; however, in the following sentence the same error type was not flagged.

Also, some error correction by *Criterion* was unclear and difficult to understand; consequently, no correction was attempted by the students.
For example,

In 2011.3.11, that accident was[1] happened[1] in Fukushima and the influence continues now.

1 　綴りの誤りなどが存在するため、文章のこの部分の内容がわかりにくくなっています。
This part of the sentence contains an error or misspelled word that makes your meaning unclear.

Because of vagueness of the Criterion feedback, there was no correction in a subsequent draft of the student's essay.

Additionally, students rely entirely on the feedback from Criterion. Thus, while revising their writing, they do not pay attention to the errors

that were not flagged by *Criterion*, and as the results of this study show there was almost no revision of errors that were not detected by *Criterion*.

Conclusion

Criterion cannot correct all mistakes in students' writing, it is not consistent in error correction, and it cannot replace instructor's feedback. However, understanding limitations and strengths of *Criterion*, knowing how to make best use of the features in the software for their specific classroom situation, as well as giving clear guidelines to students how to use the software, writing teachers can use it as one of the feedback tools that assists students in their writing development. As Grimes and Warschauer (2010, p. 34) state,

> Mindful use of AWE can help motivate students to write and revise, increase writing practice, and allow teachers to focus on higher level concerns instead of writing mechanics. However, those benefits require sensible teachers who integrate AWE into a broader writing program emphasizing authentic communication, and who can help students recognize and compensate for the limitations of software that appears more intelligent at first than on deeper inspection.

Although providing feedback (computerised, teacher-based, or peer-based) is essential, the most important issue concerns how a teacher can encourage students to think critically about their writing, reflect on the feedback they receive, and use it efficiently.

References
Anderson, C. (2013). Student user manual. Criterion. ETC
　　　Retrieved October 15, 2017 from http://criterion.etc.org

Attali, Y. & Burshtein, J. (2006). Automated essay scoring with e-rater v 2.0 *Journal of Technology, Learning and Assessment, 4-3*, pp.1-30.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated Essay Evaluation: The

Criterion Online Writing Service. *American Association for Artificial Intelligence*, 25, (3). pp.27-36.

Grimes, D. & Warschauer, M. (2010). Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation. *The Journal of Technology, Learning, and Assessment, 8,* (6), pp.4-43.

Hedgcock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity in second language writing. *Journal of Second Language Writing*, 3, pp.141-163.

Hyland, K. (1996). *Second language writing*. Cambridge: Cambridge University Press.

Hyland, K. & Hyland, F. (Eds.) (2006). *Feedback in Second Language Writing: Contexts and Issues*. New York: Cambridge University Press.

Leki, I., Cumming, A., & Silva, T. (2008). *A Synthesis of Research on Second Language Writing in English*. New York, NY: Routledge.

Lim, H., & Kahng, J. (2012). A Review of Criterion: Software for automated essay scoring. *Language Learning and Technology, 16* (2), pp.38-45.

Paulus, T. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8, pp.265-289.

Warschauer, M. & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research 10, 2*, pp.157-180.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19 (1), pp.79-101.