

A Study of Automated L2 Writing Evaluation by Japanese College Students: Eva Text Analysis

HIGUCHI Akihiko*, HIGUCHI Takako**

(Received 23 October, 2012)

Abstract

This study is an on-going study dealing with one of the automated writing evaluation (AWE) systems, Eva Text Analysis of L2 writing by Japanese EFL college students (see note 1). Through the medium of Eva Text Analysis, this present study aims to identify the advantages, disadvantages, as well as changes, in the students' attitudes and reactions to their evaluation of L2 writing. At the same time, this study evaluates possible backwash effects after introducing Eva Text Analysis to L2 writing classes, and the possibility for establishing autonomous learners. In order to obtain relevant data, a longitudinal study spanning almost four months was conducted. The quantitative and qualitative aspects of students' attitudes and reactions were assessed.

As a result, we found that the use of Eva Text Analysis was successful not only for the assessment of L2 writings but also for improving L2 writings by the students. At the same time, Eva Text Analysis was also helpful for enhancing autonomous learners and for motivation of the students' L2 studying. In other words, Eva Text Analysis was helpful significantly for students' autonomy in L2 writings and also for students' motivation in studying L2 writing. They were important findings in this study. By introducing AWE, instructors can save a lot of time when revising students' writing at a surface level. Students can evaluate their written work anywhere as many times as they want, thus promoting the establishment of autonomous learners. Encouraging students to become autonomous learners is crucial.

However, there were some limitations to be considered in this study. First, the number of students in this study was limited (only 28 in total). If this study collected much more than 28, the result of this study would have been different from the present data results. For example, the readability grades would have been different, and the number of words/sentences and T.T.R. (Type Token Ratio) values they would also have been different from the present study results. Pedagogical implications were also discussed in this study.

Key Words: automated writing evaluation (AWE), Eva Text Analysis, L2 writing, motivation, autonomy
T.T.R (Type Token Ratio)

Introduction and Rationale

Revision procedures in L2 writing settings (English writing in this case) have always been recognized as one of the main processes (Sommers, 1988, Attali, 2004). However, it is problematic for both language teachers and students. For teachers, it is time consuming to revise the students' written work given the usual class size (Warschauer & Grimes, 2008; Warschauer & Ware, 2006). To address this problem, many types of feedback have been employed in L2 writing classes worldwide for the past two decades. They are teach-

* Professor of Kagoshima University, Faculty of Education

** Part time lecturer of Kagoshima University, Center of Education

er oral/written feedback, peer feedback, self-evaluation, computer-mediated feedback, and so forth (Hyland, 2003; Hyland & Hyland, 2006). Various researchers have examined these approaches, revealing both positive and negative aspects.

In addition to the feedback methods enumerated above, AWE systems (such as Criterion e-rater by Educational Testing Services (ETS), and MY Access! by Vantage Learning) have been developed recently. In Japan, an AWE system, the “Eva Text Analysis” implemented by Kyoto Notre Dame University, is available for anyone free of charge via the Internet. Such systems received some attention because they might not only help reduce the instructors’ burden, they also allow students to evaluate their own written performance at home (Hyland & Hyland, 2006; Warschauer & Ware, 2006). Nevertheless, despite their potential, AWE systems remain among the least explored instructional methods. Due to their novelty it is still unclear whether AWE systems provide effective feedback on L2 writing (Hearst, 2000; Warschauer & Ware 2006).

This study investigates the students’ progress when employing the Eva Text Analysis to evaluate their L2 writing performance, as well as the their overall reactions to the system and possible backwash effects.

The Present Study

Subject

The subjects in this study were 28 first year college students where the authors have been teaching English as a foreign language. They were the students belonged to the Faculty of Education.

Method

The 28 students in the course were given L2 writing compositions for three times from April to July for four months. The title of the composition was “Why do many people attend college or university? Put your opinion in English (80-100words)”, and the same title was given to them in May, June, and July. Then the students input their writings to PC (Eva Text Analysis) and recorded their text analysis in terms of: (1) factor analysis by using questionnaire survey; and (2) text analysis by using ANOVA (Analysis of Variance)

Factor analysis

First, factor analysis was conducted in which questionnaire survey were used for obtaining the data. Eight questions were asked to the students. This factor analysis was conducted with four scales, which consisted of three main groups of factors. These three factors from the eight questions were extracted by MLE (Maximum Likelihood Estimation) and factor analysis by varimax rotation. The first factor group was named as “the use of Eva Text Analysis”, and the second factor group was “effects by group learning” and the third factor was “the use of process approach”.

Text Analysis (ANOVA)

For the text analysis, Eva Text Analysis was conducted and six items were recorded: (1) readability grades, (2) token, (3) Type Token Ratio (T.T.R), (4) the number of sentences, (5) average number of words/sentences, and (6) nominalization. They were analyzed by ANOVA (Analysis of Variance).

Hypotheses

This study hypothesized the following points both from the factor analysis by using questionnaire survey and from the text analysis by using ANOVA in which six items were analyzed. The following three points were suggested as hypotheses in this study.

- (1) From the factor analysis, students in the course will show the strong interest in using Eva Text Analysis because they had almost no chance to learn this PC-based text analysis in high school.
- (2) From Eva Text Analysis, students in the course will show some improvements in their L2 writings although these improvements will not be significant level because L2 writing proficiency will not be improved rapidly within a limited time.
- (3) From Eva Text Analysis, students in the course will show almost the same level of readability grades as other students in the course. This is because the class members are assigned according to their English test results of the entrance examination (National Center for University Entrance Examination).

Results and Discussions

Results from the Factor analysis

The factor analysis was conducted by the data obtained from the 28 students in the Faculty of Education (see figure 1). Here Eva Text Analysis is abbreviated as ETA.

Table 1

Name of factors, questions	1	2	3
The first factor items. The use of ETA.			
Is ETA useful for L2 writing assessment?	.979		
Do you want to use ETA outside of the class?	.648		
Interested in ETA for L2 writing assessment?	.585		
The second factor items. Effects by group learning			
Is brain storming useful in L2 writings?		.887	
Is writing 1st draft in groups useful?		.719	
Is peer feedback in L2 writings useful?		.620	
The third factor items. The use of Process Approach			
Did you know process approach in L2 writings?			.999
Accumulation rate	26.82	53.17	69.37

This factor analysis was conducted with four scales. Statistically if factor items have more than 0.4 in factor loadings, they are considered as items that consist of a factor group. As a factor analysis of this study, there are three main groups of factors. These three factors from the eight questions were extracted by MLE and factor analysis by varimax rotation.

As a result, the first factor group was based on “the use of Eva Text Analysis”, and the second factor group was “effects by group learning” and the third factor “the use of process approach”. In the accumulation rate in figure 1, the total number of the first factor items was 26.82 ($.979 + .648 + .585 = 26.82$). The total number of the first factor items and the second factor items were 53.17 ($.979 + .648 + .585 + .887 + .719 + .620 = 53.17$). The accumulation rate all together, if added from the first factor items to the third factor items, were 69.37. In other words, these three factors showed almost 70% all together.

When looking closely at the three factor items from the first factor to the third factor items, they can be categorized into three factor groups. They can be categorized depending on kinds of methods, and instruments can be used when studying L2 writing. This means that this study considers that the 28 students in the course were expecting their improvements in their L2 writing proficiency. At the same time, they were also thinking about how to teach L2 writings to their learners. They are students in the faculty of education and they will be engaged in teaching jobs in the future. Therefore, the three factor items were categorized in the factor analysis in the questionnaire data results.

As for the hypothesis (1), from the data results in figure 1, the 28 students in the course had a strong interest in using Eva Text Analysis for their L2 writing assessment.

Results from Eva Text Analysis (ANOVA)

Six items were analyzed in Eva Text Analysis. They were readability grades, token, T.T.R (Type Token Ratio), the number of sentences, average number of words/sentences, and nominalization. They were recorded from their L2 writings for three times, May, June, and July.

Table 2. Readability Grades

	1	2	3		
M	10.10	9.13	9.49		
SD	6.38	2.57	2.85		
N	28	28	28		
	SS	df	MS	F	p
Readability Grade	13.73	1.036	13.255	.738	.402
Error Range	502.434	27.973	7.962		

As for readability grades, ANOVA was conducted but significant effects were not found ($F=.738$, $df=1.036/27.973$, $p<.402$).

Although there were slight differences among the Mean score (10.10, 9.13, 9.49), values in SD (Standard Deviation) changed smaller in the second and third writings than in the first writings (6.38, 2.57, 2.85). This means that 28 students learned L2 writing effectively step by step in a group and they improved their L2 writings. This is because SD values in the second and third writings became smaller than in the first writing in terms of the readability grades. In other words, the variance in the readability grades by the students became smaller and getting close to the Mean score. If the students keep on their L2 writings more than four, five, and six times, and collect the data results, this study will be able to obtain more accurate data and more accurate tendency. This was a very interesting finding in this study.

Table 3. Token

	1	2	3		
M	97.18	98.21	98.5		
SD	64.42	69.31	68.32		
N	28	28	28		
	SS	df	MS	F	p
Token	27.071	2	13.536	.609	.547
Error Range	1199.595	54	22.215		

As for token, ANOVA was conducted but significant effects were not found ($F=.609$, $df=2/54$, $p<.547$). There were slight differences and almost no changes among Mean score in token. This was also the same as in SD values in token.

There were almost no changes among average number of words/sentences, and readability grades in the students' L2 writings for three times, this study claims that 28 students seemed to have almost the same English language proficiency. In fact, this is true because the students in the course were assigned to the class depending on their English test score in the National University Center Entrance Examination.

Table 4. TTR

	1	2	3		
M	.96	.66	.67		
SD	1.88	.30	.28		
N	28	28	28		
	SS	df	MS	F	p
TTR	1.66	1.002	1.659	.668	.421
Error Range	67.174	27.043	2.484		

As for TTR, ANOVA was conducted but significant effects were not found ($F=.668$, $df=1.002/27.043$, $p<.421$). Although there were no significant effects found in TTR values, the Mean score and SD values in the second and third writings became much smaller than in the first L2 writings. Since there were almost no changes in the values of token (see figure 3), this study assumes that the number of types in L2 writings decreased.

Besides, SD value in TTR in the second and the third L2 writings became smaller than the one in the first L2 writing. This means that there is a possibility that the 28 students learned L2 writing effectively step by step in a group and they improved their L2 writings. This is because, as in the case of readability grades in figure 2, SD values in TTR in the second and third writings became smaller than the one in the first writing.

Table 5. The number of sentences

	1	2	3		
M	6.56	6.68	6.79		
SD	2.06	1.49	1.71		
N	28	28	28		
	SS	df	MS	F	p
The number of sentences	.643	1.358	.474	.634	.477
Error Range	27.357	36.656	.746		

As for the number of sentences, ANOVA was conducted but significant effects were not found ($F=.634$, $df=1.358/36.656$, $p<.477$). There were no significant changes in the Mean scores, in average number of words/sentences, and in token. Besides, there were almost no changes in the scores of token and readability grades for three times, 28 students in the class seemed to have almost the same English language proficiency.

Table 6. Nominalization

	1	2	3		
M	.82	.75	.82		
SD	1.16	1.00	1.16		
N	28	28	28		
	SS	df	MS	F	p
Nominalization	.009	1.518	.006	.194	.763
Error Range	13.239	40.989	0.323		

As for nominalization, ANOVA was conducted but significant effects were not found ($F=.194$, $df=1.518/40.989$, $p<.763$). There were almost no changes in SD values and Mean scores for three times. Therefore, the 28 students in the class did not aware of the use of nominalization in their L2 writings for three times. In fact, the author (teacher) did not refer to the use of nominalization during the classes.

Hypotheses Testing

Three hypotheses were suggested in this study. Followings are the results in the bracket.

(1) From the factor analysis, students in the course will show the strong interest in using Eva Text Analysis

because they had almost no chance to learn this PC-based text analysis in high school.

(Judging from the factor analysis, questionnaire in this study, 28 students in the course showed strong interests in using Eva Text Analysis. And none of them had used it before. Therefore, this hypothesis was rightly proved.

(2) From Eva Text Analysis, students in the course will show some improvements in their L2 writings although these improvements will not be significant level because L2 writing proficiency will not be improved rapidly within a limited time.

(Judging from Eva Text Analysis by using ANOVA, readability grades by the students were slightly improved although their values were not so significant as we had hypothesized. The SD values in the readability grades were 6.38, 2.57, and 2.85 respectively. Therefore, this hypothesis was rightly proved.

(3) From Eva Text Analysis, students in the course will show almost the same level of readability grades as other students in the course. This is because the class members are assigned according to their English test results of the entrance examination, National Center for University Entrance Examination.

(Judging from Eva Text Analysis by using ANOVA, the 28 students in the class learned L2 writings effectively step by step in a group and they improved their L2 writings. This was because SD values in the second and third writings became smaller than in the first writing in terms of the readability grades. The variance in the readability grades by the students became smaller and getting close to the Mean score. However, there were individual differences in terms of the readability grades for three times. Some improved slightly within a limited of time, but others did not. Therefore, this hypothesis was not successfully improved and there were individual differences judging from their data.)

Other findings

Some other findings were also obtained from the questionnaire in which the students put their personal opinions on the use of AWE, Eva Text Analysis. They can be summarized as follows:

- (1) It was fun for many students in the course to use Eva Text Analysis because of its objectivity, easy access, free of charge, but by all means, autonomous learning. The students in the course were able to learn L2 writings individually particularly its evaluation by using Eva Text Analysis.
- (2) They were able to study individual L2 writings outside the class and obtain objective data results of their individual L2 writing assessments.
- (3) The use of Eva Text analysis was also helpful in terms of motivation for the students. Even the students who dislike studying English became interested in studying L2 writings, trying to improve some of the grammatical items in the text analysis such as nominalization, the average number of words/sentences,

and TTR.

These findings were something new this study had not expected at the beginning of this study. In particular, autonomous learning was recognized significantly as a successful finding in this study.

Pedagogical Implications

Factor analysis and text analysis were conducted in this study, examining college students' written work, using AWE. Factor analysis showed that students were interested in Eva Text Analysis as a means for improving their own writing skills. It also revealed that students were interested in finding more about the available tools when they learn. This implied that students in the Faculty of the Education already had an instructors' point of views on their own learning environment. Text analysis showed that the students learned L2 writings effectively in a group. Furthermore, readability grades and T.T.R. indicated that their L2 writing skills improved. These findings suggested that the use of Eva Text Analysis can lead to the betterment of students' writing.

By introducing AWE, instructors can save a lot of time when revising students' writing at a surface level. Students can evaluate their written work anywhere as many times as they want, thus promoting the establishment of autonomous learners. Encouraging students to become autonomous learners is crucial. Researchers have reported that it raises learners' motivation and leads to more effective learning (Cotteral & Crabbe, 1992; Jiao, 2005; Little, 1996a, 1996b).

Conclusion and Limitations

The use of Eva Text Analysis was successful not only for the assessment of L2 writings but also for improving L2 writings by the students. At the same time, as we have seen in this study, Eva Text Analysis was also helpful for enhancing autonomous learners and for motivation of the students' L2 studying. Eva Text Analysis was helpful significantly for students' autonomy in L2 writings and also for students' motivation in studying L2 writing. They were important findings in this study. However, there were some limitations to be considered in this study.

First, the number of students in this study was limited (only 28 in total). If this study collected much more than 28, the result of this study would have been different from the present data results. For example, the readability grades would have been different, and the number of words/sentences and T.T.R. (Type Token Ratio) values they would also have been different from the present study results.

Second, the number of words in L2 writings was limited (about 80-100). Therefore, in the subsequent study, the number of words should be changed to 130-150 words. In so doing, the different data will be obtained, and the data results should be examined. How the data results will be changed according to the number of words in L2 writings? They should also be examined in the next step of L2 writings. Here, we

should be careful to the change in the value of T.T.R. according to the total number of words in the written texts.

Third, this study spent for three to four months but this was not enough. If we keep on studying the same topic by using the same subjects more than six months, some new findings will be obtained from the study. Therefore, the duration of testing L2 writing is also important in the subsequent study.

Finally, Eva Text Analysis can only assess surface-level corrections (such as the number of sentences and words). Needless to say, the basic building blocks of writing should include not only these grammatical/structural aspects but also meaning. Therefore, it is important to research what kinds of instruction/interventions may help improve meaning-level revision when employing an ETA system.

They were limitations to be considered in this study, and they should be studied in the subsequent studies with further elaborations.

Note

1. Eva Text Analysis is a web-based automated writing evaluation system available on the internet <http://poets.notredame.ac.jp/cgi-bin/evatext> established by Kyoto Notre Dame University.

References

- Attali, Y. (2004). Exploring the feedback and revision features of Criterion. Paper presented at the National Council on Measurement in Education conference, April 2004, San Diego, CA.
- Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly*, 28, 181-188.
- Cotteral, S., & Crabbe, D. (1992). Fostering Autonomy in the Language Classroom: Implications for Teacher Education. *Guidelines*, 14, 11-22.
- Hearst, M. (2000). The debate on automated essay grading. *Intelligence Systems* 15, 22-37.
- Hyland, K. (2003). *Second Language Writing*. Cambridge: Cambridge University Press.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39, 83-101.
- Jiao, L. (2005). Promoting EFL learner autonomy. *Sino-US English Teaching*, 17, 27-30.
- Little, D. (1996a). Learner Autonomy: Some Steps in the Evolution of Theory and Practice. *TEANGA: The Irish Yearbook Of Applied Linguistics*, 16, 1-13.
- Little, D. (1996b). The politics of learner autonomy. Paper presented at the Fifth Nordic Workshop on Developing Autonomous Learning. Copenhagen, Denmark.
- Sommers, N. (1980). Revision strategies of students writers and experienced adult writers. *College Composition and Communication*, 31, 378-388.
- Warschauer, M. & Grimes, D. (2008). Automated Writing Assessment in the Classroom. *Pedagogies: An International Journal*, 3, 22-36.
- Warschauer, M. & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research* 10, 1-24.