S. L. Wong et al. (Eds.) (2010). Proceedings of the 18th International Conference on Computers in Education. Putrajaya, Malaysia: Asia-Pacific Society for Computers in Education.

Analyzing the Effects of Peer Review Activities in the EFL Writings

Sayuri YOSHIZAWA[a*,b], Takao TERANO[b] & Atsushi YOSHIKAWA[b,c]
[a]*Hoshi University, Japan*
[b]*Tokyo Institute of Technology, Japan*
[c]*The Japan Institute for Educational Measurement, Inc., Japan*
*s-yoshizawa@hoshi.ac.jp

Abstract: In this paper, we propose a new method to analyze the effect of peer review activities in the English-as-a-Foreign-Language (EFL) writing classes. Although a number of studies on peer review activities have been conducted in the last two decades, no method that can objectively analyze the effects of the approach has been proposed thus far. We surmised the degree of difference of the essays by focusing on the estimation of similarity of the documents using the normalized compression distance (NCD), a compression-based classification algorithm. We have applied the NCD measures to the student reports in the college EFL writing courses. The proposed method was found to be beneficial in evaluating the writings of the pre and post peer review activities.

Keywords: peer review activity, collaborative learning, normalized compression distance

Introduction

Peer review activities are one of the collaborative learning approaches, which will help develop logical thinking skills and insight into their own writings. The method using peer review activities, also known as peer response or peer feedback, was originally proposed and introduced by P. Elbow in the 1970s as a promising process approach for writing without teachers in the first language compositions [1]. It was then applied to the second language education and recognized as a beneficial way to polish the writings. In the late 1980s the researches on the peer review activities in the English as a second language (ESL) writings attracted considerable attention in the United States and then in some Asian countries such as China, Singapore, and Taiwan. Numerous studies on the implementation of peer review have been conducted and verified its effects [2]~[9]. On the contrary, some studies suggest that students with certain cultural backgrounds tend to feel reluctant to comment on peers' writings and prefer teacher comments, because they either have a lack of confidence in their language ability or wish to stay in harmony with their peers [10][11]. Although an increasing number of teachers of the Japanese language employ peer review activities, the approach has not been widely used among the teachers of English in Japan.
This paper presents a new method based on the Kolmogorov complexity to objectively analyze the effects of peer review activities on the English as a foreign language (EFL) writings. The proposed method estimates the degree of similarity of the documents by use of the normalized compression distance (NCD), a compression-based classification algorithm to cluster any similar objects, which was first formulated by Cilibrasi and Vitányi [12] on the basis of the normalized information distance (NID) proposed by Li et al [13]. In order to lighten a burden of correcting students' essays, which is a time-consuming task for any language teachers, we have applied the NCD measures to the student essays to surmise the degree of difference between the first and the revised drafts by focusing on the estimation of similarity of the documents.

The remainder of this paper is organized as follows. Section 1 shows our method of the experiment, the results using the holistic scores and the NCD measure are presented in Section 2, and the whole experiment is concluded in Section 3.

1. Method of the Experiment

1.1 Subjects

The present experiment was conducted at a pharmaceutical university in Tokyo, Japan. A total of 35 fourth-year Japanese students, who are enrolled in the six-year course which is compulsory to become a pharmacist, participated in the experiment. They had been educated and received at least nine years of instruction in English. The number of male and female subjects was 10 and 25 respectively. The original data included 37 students, but 2 of them were eliminated from this analysis because they did not attend the anonymous peer review activity. The participants have experienced at least one of the three identifiable peer review activities prior to this experiment.

1.2 Procedures

1.2.1 Essay Writing

On the very class of the experiment, the participants were provided with a written assignment and asked to write an opinionative essay about *kampo* medicines and the health insurance systems. The participants were instructed to write a logical essay with approximately two hundred words. The provided time for the essay writing was forty-five minutes. They were permitted to use dictionaries. After the completion of the first draft, all essays were collected and typed in a word form to make them unidentifiable.

1.2.2 Anonymous Peer Review Activities

After one week of the essay writing, the participants underwent the peer review activities. The anonymous essays were distributed to the participants and twenty-five minutes were provided for the activity in Japanese. The reviewers were told to underline any ambiguous sentences in the peer's essay while reading and were allowed to make any changes or suggestions they could make.

The following written instructions were given before the activity: point out any spelling, grammatical, and syntax errors; discover any omission of background information and inconsistency of logical structure; make any suggestions to improve the peer's essay; and state overall impressions. More specific instructions were also provided orally prior to the activity in order to help students understand how to make comments (e.g. pay attention to the third person singular present form, the sequence of tenses, and examine the adequacy of the reasons to support the writer's final statement).

Upon receiving the comments from the peer, all essays with the reviewer's comments were collected again and returned to each author by the instructor. Then, another twenty minutes was provided to revise the essays.

1.2.3 Analysis Using Holistic Scoring

The first and the revised essays were assessed by the instructors based on the adequacy and organization of the content, the cohesion and logicality of the scripts, and the overall quality. The holistic scoring used for the evaluation is shown in Table 1.

Table 1 A Rubric for Holistic Scoring

Score	Description
5	Sufficient background information and reasons for the claimed statement; good presentation of coherent logical development
4	Some but insufficient background information and reasons for the claimed statement; good presentation of coherent logical development
3	Sufficient background information and reasons for the claimed statement; either incoherent or illogical development
2	Insufficient background information or reasons for the claimed statements; either incoherent or illogical development; less than 80% of the designated length
1	*(Not acceptable)* Little or no background information and reasons for the claimed statement; incoherent and illogical development

1.2.4 The Normalized Compression Distance (NCD)

The normalized compression distance (NCD) is an outcome of the mathematical theoretical development based on the incomputable notion of algorithmic complexity developed by Kolmogorov in the late 1960s. In 2004 the formula called the Normalized Information Distance (NID) was proposed by Li et al. [11], and the next year Cilibrasi and Vitányi [10] designed the NCD on the basis of the NID. The NCD is applicable to clustering objects of any kind, such as music, texts, or gene sequences.

If x and y are the two objects concerned, and c(x) is the compressed size of x using compressor c, then the

$$\mathrm{NCD}(x,y) = \frac{c(x \cdot y) - \min\{c(x), c(y)\}}{\max\{c(x), c(y)\}}$$

where c(y) is the compressed size of y using compressor c and c(x·y) is the compressed size of concatenated objects x and y. The NCD is supposed to be between 0 and 1.

The closer the NCD(x,y) is to zero, the more similar x and y are. In this study, we set c(x) is the compressed size of the first draft, c(y) is the compressed size of the revised draft, and c(x·y) is the compressed size of the concatenation of the first and the revised essays.

2. Results

2.1 Holistic Scores of the Essays and Surface-Level Errors

Seventeen out of the thirty-five essays (48.6%) received higher scores, fifteen (42.9%) received the same scores, and three (8.5%) received lower scores than the first drafts, as shown in Table 3. The shadowed sections are the scores of the essays whose contents were scarcely changed even after the revision. Four reviewers (11.4%) could not point out any spelling, grammatical, or syntax errors even though some obvious mistakes were found in the texts. These four reviewers were below the English proficiency level and their essays were indeed indicated five to fifteen such errors. The scores with an asterisk were the essays with many surface-level errors (more than ten grammatical and/or spelling errors).

Table 2 Evaluations of the Essays

Essay#	Score of the first draft	Score of the revised draft	Essay#	Score of the first draft	Score of the revised draft
1	3*	4	20	4	5
2	3*	5	21	4	4
3	2	2	22	4	4
4	2	1	23	2	3
5	1*	2	24	2	3
6	2	2	25	3	3
7	4	5	26	4	4
8	3	3	27	3	1
9	4	4	28	3	4
10	3	3	29	3*	4
11	3	5	30	5	5
12	3*	4	31	3	4
13	3	4	32	3*	4
14	3	3	34	4	5
15	4*	4*	35	4	2
16	1*	2	36	4	5
17	2	4	37	4	4
18	5	5			

Regrettably, the scores of the three revised essays (essay #4, 27, and 35) were lower than the first drafts. It was revealed from the comments submitted along with the revised essays that those writers were uncertain of the way to incorporate the peer comments into their writings.

2.2 The NCD Results of the Essays

The normalized compression distance ranged from .11 to .44 as shown in Table 3.

Table 3 Compressed Size and the NCD Results of the Essays

Essay#	c(x)	c(y)	c(x·y)	NCD	Essay#	c(x)	c(y)	c(x·y)	NCD
17	417	558	662	.44	24	452	493	573	.25
27	519	361	640	.34	20	585	621	736	.24
35	620	498	778	.32	29	539	556	671	.24
34	619	676	832	.32	12	550	523	673	.24
5	407	424	534	.30	14	521	558	651	.23
23	536	554	701	.30	9	487	549	614	.23
11	553	560	719	.30	22	564	582	695	.23
2	649	671	847	.30	26	566	559	686	.21
13	505	540	664	.29	6	430	448	525	.21
32	605	653	797	.29	3	382	395	465	.21
16	506	469	642	.29	30	561	573	674	.20
7	553	615	731	.29	18	654	661	776	.18
1	485	500	620	.27	21	609	548	709	.18
4	430	419	541	.26	25	591	591	697	.18
10	449	445	566	.26	8	560	574	661	.18
36	552	549	692	.26	15	687	681	788	.15
31	485	500	620	.25	37	546	549	605	.11
28	660	645	821	.25					

The shadowed sections were the same as in Table 2 except the essay #10, whose content in the revised draft was scarcely changed from the first draft. The revised version of this essay contained some rephrased sentences, which resulted in a higher NCD value.

2.3 Holistic Scores and the NCD Results of the Essays

The holistic scores of the essays with the NCD value below .23 were unchanged. The changes made in those essays were the only minor surface-level ones that were pointed out by the peer. The contents of the essays were unchanged as the NCD measure indicated the similarities of the documents pre and post peer review activities. From the experiment, the NCD measure was found useful in properly evaluating some of the essays.

The results of the evaluation using the holistic scores are easily comprehensible for the English language teachers. It is, however, a time-consuming task to score while reading each essay. On the other hand, the NCD measure is an effective means for finding the essays that contain noticeable changes in the content, although the quality of the changes made cannot be determined by the NCD results. In the present case, the revised essays with the NCD of .23 or lower contained few changes in the content as well as in the structure.

3. Conclusions

In this paper we presented a new method based on the Kolmogorov complexity to objectively analyze the effects of peer review activities on the English as a foreign language writings. The proposed method estimated the degree of similarity of the documents by use of the normalized compression distance. Within the present experiment, it was found quite beneficial in preestimating and evaluating the degree of difference in the contents of the writings before and after the peer review activities. This study provided the ground that supports the use of the new method as a potentially valuable aid for the language teachers in reducing the time-consuming task. However, further analyses will be necessary to establish the objective measurement and beneficial prior instructions for more effective peer review approach. For further research, we plan to make qualitative analyses of the students' production, determine what factor to be utilized as an indicator, and establish an objective measurement that would reduce the burden of the language teachers.

References

[1] Elbow, P. (1998). *Writing Without Teachers, 2nd ed.* New York, NY: Oxford University Press.
[2] Chaudron, C. (1984). The effects of feedback on students' composition revisions. *RELC Journal 15*(2): 1-14.
[3] Mittan, R. (1989). The peer review process: Harnessing students' communicative power." In D. Johnson and D. Roen (eds.), *Richness in Writing: Empowering ESL Students:* 207- 19, New York: Longman.
[4] Nelson, G. & Murphy, J. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? *TESOL Quarterly 27*(1): 135-142.
[5] Connor, U. & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing 3* (3): 257-276.
[6] Mendonça, C. & Johnson, K. E. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly 28* (4): 745-769.
[7] Tsui, A. B. M. & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing 9*(2): 147-170.
[8] Liu, J. & Hansen, J. G. (2002). *Peer Response in Second Language Writing Classrooms*. Ann Arbor, MI: The University of Michigan Press.
[9] Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal 59*(1): 23-30.
[10] Zhang, S. (1995). Re-examining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing 4*(3): 209-222.
[11] Carlson, J.G. & Nelson, G.L. (1996). Chinese students' perceptions of ESL peer response group interaction. *Journal of Second Language Writing 5*(1): 1-19.
[12] Cilibrasi, R. & Vitányi, P. (2005). Clustering by compression. *IEEE Transactions on Information Theory 51*(4): 1523-1545.
[13] Li, M., Chen, X., Li, X., Ma, B. & Vitányi, P. (2004). The similarity metric. *IEEE Transactions on Information Theory 50* (12): 3250-3264.