Efficacy of Self-Assessment in Written Production among High School Students

Yoko Suganuma Oi

Abstract—The purpose of the present study is to find the efficacy of high school student self-assessment of written production. It aimed to explore the following two research questions: 1) How is topic development of their written production improved after student self-assessment and teacher feedback? 2) Does the consistency between student self-assessment and teacher assessment develop after student self-assessment and teacher feedback? The data came from the written production of 82 Japanese high school students aged from 16 to 18 years old, an American English teacher and one Japanese English teacher. Students were asked to write English compositions, about 150 words, for thirty minutes without using dictionaries. It was conducted twice at intervals of two months. Students were supposed to assess their own compositions by themselves. Teachers also assessed students' compositions using the same assessment sheet. The results showed that both teachers and students assessed the second compositions higher than the first compositions. However, there was not the development of the consistency in coherence.

Keywords—Feedback, self-assessment, topic development.

I. INTRODUCTION

STUDENT assessment is effective to motivate students and raise their self-awareness [1]. It is also useful for teachers to save time and encourage students to participate in class more positively. Two types of assessment, self-assessment and peer assessment, have been investigated by the researcher, focusing on which assessment type is more reliable and effectual for high school students in terms of class management and the improvement of English proficiency. The results presented that both assessments could develop students' motivation and self-awareness, but only self-assessment partly showed agreement with teacher assessment. Peer assessment did not present the agreement with teacher assessment. Students commented that peer assessment stimulated their competitiveness and encouraged them to develop themselves to be evaluated higher by classmates. However, peer assessment made students tense and students did not rely on peer assessment. Besides teachers needed to censor students' comments and evaluation before feedback, so it became another burden for teachers. For it is assumed that students were still short of maturity as evaluators compared to adult learners. On the other hand, it is hypothesized that self-assessment could relax students' tension much more than that of peer assessment [1]. So it is meaningful to examine the efficacy of high school student self-assessment to adopt it more usefully in class.

Yoko Suganuma Oi is with Graduate School of Education, Waseda University, 1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo, CO169-8050, Japan (e-mail: yokosuganuma@suou.waseda.jp).

II. PREVIOUS STUDIES

A. Self-Assessment

Reference [2] investigated the usefulness of self-assessment as a second language placement instrument. Reference [3] also proved that self-ratings could be reliable and valid measures of communicative language abilities. Other researchers stated benefits and suggestions on self-assessment. For instance, [4] presented six advantages of using self-assessment: 1) promotion of learning 2) raising level of awareness 3) improving goal-orientation 4) expansion of range assessment 5) sharing assessment burden 6) beneficial post-course effects. Reference [5] also showed three implications of self-assessment. The first one was the development of self-reflection on their performance and learning process. The second was the aid to train students to become better raters and learners. The final implication was the students' psychological factors. The students in [5] tended to evaluate themselves lower and it was caused by their affective factors

Though some researchers [6], [7], doubted that learners could evaluate their own proficiency correctly and might overstate or downgrade their achievement, [8] insisted that self-assessment is fundamental to effective European Language Portfolio (ELP) use, showing the usage of "can-do" statements.

The reliability of student self-assessment has been researched, focusing on the agreement between teacher assessment and student self-assessment. For instance, [9] presented that even 141 young children who are in three infant schools could rank themselves with a high degree of agreement with teacher assessment. On the other hand, [10] investigated the self-assessment of French proficiency made by 500 Grade 8 students and indicated that (1) self-assessment of language proficiency correlated only weakly with objective measures of language proficiency; (2) self-assessment measured on specific tasks were more highly correlated with tested proficiency than were global self-assessment measures.

B. Feedback

There are a variety of options for error feedback, from direct correction of error to some fairly indirect and less informative approaches. According to [11], teachers must choose the need of students and goals of writing task: 1) Error treatment, including error feedback by teachers, is necessary component of L2 writing instruction. 2) Teachers should provide indirect feedback that engages students in cognitive problem-solving as they attempt to self-edit based upon the feedback that they have received. 3) Different types of errors will likely require varying treatments. 4) Students should be required to revise (or at least self-edit) their texts after receiving feedback. 5) Supplemental

grammar instruction can facilitate progress in accuracy if it is driven by student needs and integrated with other aspects of error treatment. 6) The maintenance of error feedback can heighten student awareness of their weakness and of their improvement.

Error correction feedback has been considered that it improves aspects of writing accuracy, although theoretical competence, writing fluency, and writing complexity are largely unaffected by written corrective feedback [12]-[16]. Whereas, [14] suggested that correction is ineffective or harmful and harms students' fluency and their overall writing quality.

Reference [16] insisted that written feedback should be encouraged rather than error correction feedback, because discourse-level feedback is more valuable to promote students' thinking and learning. Reference [17] presented that 1) written feedback is more effective than oral feedback for writing development. 2) Commenting is more effective than error location. 3) Focus on form and content seems to be more effective than an exclusive focus on form.

The present study aims to examine the topic development, not the accuracy of language use in written production; therefore, written feedback was mainly given to students. And it also focused on discourse level feedback, for it is more useful to improve the topic development of students' written production.

C. Topic Development

The objective to teach written production has been recognized to further enhance students' abilities to evaluate facts and opinions from multiple perspectives and communicate through reasoning and a range of expressions. So the necessity to teach how to develop a topic has been more strongly focused after the enforcement of new national government guidelines in Japan [18].

Besides a style of writing by oriental students in English has been compared to "approach by indirection, turning and turning in a widening gyre" [19]. Reference [20] also stated that the Japanese essays written in English seemed to be disorganized and illogical, filled with nonrelevant material, developed incoherently with statements that remain unsupported. Though the shortage of concrete data in both studies has been indicated [21], it is a fact that Japanese students have had difficulties mastering rhetorical patterns at the level of discourse, not the sentences [22]. It has been also discussed that Japanese writers tend to hesitate to express their initial positions, so the readers must wait for the final paragraph to understand the writers' conclusion. Reference [23] characterized Japanese organizational pattern as "quasi-inductive". Reference [24] presented that a characteristic of Japanese writing is "induction". Therefore, it is important for Japanese students to learn how to develop a topic in English composition and be aware of what part is insufficient in their written production in terms of cohesive device and overall organization. For the researcher believes that self-awareness and self-reflection work to develop learner autonomy and encourage students to revise their written production by themselves [25].

III. RESEARCH QUESTIONS

The two research questions considered in this study are:
1. How is topic development of their written production improved after self-assessment and teacher feedback?
2. Does the consistency between student self-assessment and teacher assessment develop after self-assessment and teacher feedback?

IV. DATA COLLECTION

A. Participants and Procedure

The data for the analysis came from the written production of 82 Japanese high school students. Their ages ranged from 16 to 18 years old. One American English teacher and one Japanese English teacher also participated in the study.

Before the research, every student received the instructions of writing patterns and comments about their previous English compositions. They also practiced to assess their English compositions by themselves. Teachers gave a lecture on Topic Development. Students were asked to write an essay on the guided writing topic, "What will you do after you graduate from high school? Why? Use specific reasons and details to support your decision" about 150 words for thirty minutes without using dictionaries. Students were not informed of the topic before writing.

Students assessed their written production after writing, using the same assessment sheet as teachers. The assessment sheet consisted of big three components: "Structure", "Coherence", and "General Evaluation". "Structure" is composed of "Introduction", "Body", "Conclusion". There are also "Discourse Markers" and "Coherence" components. The scores of each small component are 3 points, and "General Evaluation" is 5 points. So the full mark of this assessment sheet is 20 points. Table I presents the scores of teacher assessment. After the first writing session, students were given written feedback and the score report in terms of topic development by teachers.

After two months interval, students were asked to write English compositions on the following topic: A foreign visitor has only one day to spend in your country. Where should this visitor go on that day? Why? Use specific reasons and details to support your choice". Students and teachers assessed students' second written production using the same assessment sheet as the first writing. They were asked to answer questionnaires related to the self-assessment after the experiment. During two months intervals, students took a school regular English writing class using government authorized textbooks.

TABLE I
TEACHER ASSESSMENT OF TOPIC DEVELOPMENT

	N	MAX	MIN	MEAN	SD
The first written production	82	20	0	12.01	3.95
The second written production	82	20	2	13.20	4.02

B. Inter-Rater Reliability between Two Teachers' Assessment

The inter-rater reliability between the assessments by two teachers was calculated before the two teachers' total sum. It

was found that there was a high correlation between the assessments by two teachers. To confirm the reliability of two teachers' assessment, the inter-rater reliability between the assessments by six teachers was also calculated. It was found that there was a high correlation between the assessments of teachers except one teacher. So the researcher judged that two teachers' assessment was reliable and decided to analyze the data based on the results of two teachers, a native English teacher (NET) and a Japanese English teacher (JET) who was class teachers.

TABLE II
INTER-RATER RELIABILITY BETWEEN TWO TEACHERS' ASSESSMENT

Components	Introduction	Body	Conclusion	Discourse Markers	Coherence	Total Score
Kendall's coefficient	.970**	.939**	.867**	.965**	.888**	.922**

*p<.05 **p<.01

C. Teacher Feedback

Two teachers categorized 82 written productions into 8 patterns, referring to four components. The components were composed of Introduction, Body, Conclusion, and Discourse Markers in terms of Topic Development. Chart I presents the categorization of topic development patterns.

The first and second patterns did not have any points which should be evaluated. The written production of Pattern 1 had too few sentences. Pattern 2 did not follow the task or instruction of written production. On the other hand, the third pattern showed the presence of a topic sentence in the introduction. Both Patterns 4 and 5 had a topic sentence in the introduction, and topic development was also found in the body. However, discourse markers were not adequately used and the main idea was not restated in the conclusion. The difference between Pattern 4 and 5 was the degree of topic development in the body. Pattern 6 showed that the characteristics of the introduction included the introduction, body and conclusion, but it did not show an adequate usage of discourse markers. On the other hand, Pattern 7 met the requirements of introduction, body, and discourse markers, but it did not have a conclusion. Final pattern was Pattern 8. It showed every component.

TABLE III
THE COMPONENTS OF TOPIC DEVELOPMENT PATTERNS

	Introduction	Body	Conclusion	Discourse Markers
Pattern 1				
Pattern 2				
Pattern 3	✓			
Pattern 4	✓			
Pattern 5	✓	✓		
Pattern 6	✓	✓	✓	
Pattern 7	✓	✓		✓
Pattern 8	✓	✓	✓	✓

Students were given written comments in English and also given the above topic development patterns to help students to understand the shortage of compositions. After the experiment, students' individual Topic Development pattern and written comments by teachers were given to each student.

V. DATA COLLECTION

A. The Difference of Scores between the First and Second Writing

Both scores of structure and coherence which were assessed by two teachers increased. Tables III and IV present the difference of scores between the first and second compositions.

TABLE IV
DIFFERENCE OF STRUCTURE SCORES ASSESSED BY NET AND JET BETWEEN THE FIRST AND SECOND COMPOSITIONS

	-4	-3	-2	-1	0	1	2	3	4	SUM
NET	0	0	10	24	24	16	6	1	1	82
JET	0	3	11	22	23	17	3	3	0	82

TABLE V
DIFFERENCE OF COHERENCE SCORES ASSESSED BY NET AND JET BETWEEN THE FIRST AND SECOND COMPOSITIONS

	-4	-3	-2	-1	0	1	2	3	4	SUM
NET	0	2	12	26	25	12	5	0	0	82
JET	0	3	13	27	23	11	4	1	0	82

About 41 to 43 percent students increased in the scores of structure. About 28 to 29 percent students did not show any change in scores. About from 28 to 29 percent students decreased in scores.

About 48 to 52 percent students increased in the scores of coherence. The scores of about 28 to 30 percent were not changed at all. About 19 to 20 percent students reduced scores.

B. Students' Response about Questionnaires

Students were asked to answer two questions related to the self-assessment and topic development patterns after the experiment. The questions were as follows:
1. Do you think that self-assessment of organization and topic development of English composition is effective to improve English proficiency?
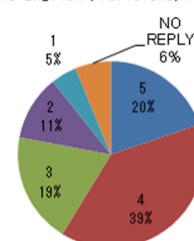2. Do you think that the patterns of Topic Development are useful to develop English proficiency?

Concerning the first question, 59% students replied "very effective" or "effective". About the second question, 73% students replied that it was very useful or useful.
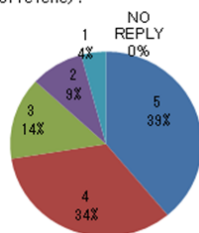

Fig. 1 Distribution of Students' Response 1

Fig. 2 Distribution of Students' Response 2, Note. 5 Very effective (useful), 4 Effective (useful), 3 So so, 2 Not so effective (useful), 1 Not effective

C. The Consistency between Student Self-Assessment and Teacher Assessment

Based on the difference between teacher assessment and students' self-assessment, the Wilcoxon signed-rank test was conducted to see the differences between two data. If the absolute value of the z scores was higher than 1.96, we could say that there was a significant difference between two data. In this case, we can ignore the minus and plus signs.

The results of the first writing and second writing were similar, because the absolute values of the z scores were less than 1.96. It means that the self-assessment of structure was not different from teachers' assessment. That is to say, students could assess the structure of their written production as well as teachers did. On the other hand, neither of the z scores of coherence indicated less than 1.96. Therefore, students could not assess the coherence as well as teachers.

TABLE VI
THE WILCOXON SIGNED-RANK TEST OF THE FIRST WRITING BETWEEN SELF-ASSESSMENT AND TEACHER ASSESSMENT

	Structure		Coherence	
	T	Z	T	Z
Students' self-assessment-NET assessment	54	-.740	36.00	-5.116
Students' self-assessment -JET assessment	61	-1.218	86.50	-5.037

p<.05 T (the sum of ranks of the smaller of the two ranks); *z* (Wilcoxon signed-rank test *z*-score); < 1.96

TABLE VII
THE WILCOXON SIGNED-RANK TEST OF THE SECOND WRITING BETWEEN SELF-ASSESSMENT AND TEACHER ASSESSMENT

	Structure		Coherence	
	T	Z	T	Z
Students' self-assessment-NET assessment	64	-1.250	104.00	-4.318
Students' self-assessment -JET assessment	73	-.245	127.00	-4.121

p<.05 T (the sum of ranks of the smaller of the two ranks); *z* (Wilcoxon signed-rank test *z*-score); < 1.96

VI. DISCUSSION

A. RQ 1: How Is Topic Development of Their Written Production Improved after Self-Assessment and Teacher Feedback?

Both teachers' assessment showed that about 40 percent students developed the scores of structure after self-assessment and teacher feedback. About 50 percent students also showed increase in the scores of coherence. Therefore, students improved both the scores of structure and coherence. During the research, some students mentioned that students needed to use more specific advice on topic development. Others also replied that the concrete examples of topic development were also helpful for them to understand their ideal topic development. However, one student mentioned that the instruction of topic development was different from error correction of language use, so it is difficult for students to understand the shortage of their components, even if teachers gave them feedback.

According to the results of descriptive replies for the questionnaires, some students commented the importance of self-reflection and self-analysis especially about language use such as spelling and grammar. For self-assessment could help students to correct errors in spelling; though, the present study did not focus on error correction. They also insisted on the importance of self-assessment, since they could be aware of incoherence of written production.

Whereas some students showed lack of confidence in self-assessment about topic development, because they had not received training in organizing written production yet. Students usually took a lesson which focused on a sentence rather than paragraphs and organization. Students indicated the necessity to study how to organize paragraphs and topics.

The students' written production was categorized into eight patterns. The most frequent pattern was Pattern 8. It dominated about 40% of all the compositions. After the experiment, students' individual topic development pattern was given to each student. Some students commented that the presentation of the patterns of topic development could help them to think or decide the organization of written production and how to develop a topic, because they could prepare or make a plan for a next chance. They also commented that they could overview the usage of discourse markers. The presentation of Topic Development patterns was useful to be aware of the lack of elements in their written production. Therefore, students supported the effectiveness of topic development patterns, because it could work as a guide on how to organize English written production.

On the other hand, some of the students seemed to be unsatisfied with eight patterns, because they insisted that it should be necessary to subdivide Topic Development patterns more specifically to indicate students' written production more accurately. That is to say, a more specific device is needed to develop assessment criteria about the topic development.

B. RQ2: Does the Consistency between Student Self-Assessment and Teacher Assessment Develop after Self-Assessment and Teacher Feedback?

It was found that students could assess structure as well as teachers. On the other hand, any consistency between teacher assessment and student assessment was not found in coherence assessment in both the first and second writing. However, as the scores of coherence increased higher than that of structure, it is assumed that students could develop the ability of coherence. Moreover, students could assess structure from the first writing. According to the replies of descriptive questionnaires, students did not have confidence in assessing coherence. Therefore, it is assumed that there was not consistency between teacher assessment and student self-assessment, though the assessment scores of coherence increased compared to those of structure. In conclusion, student self-assessment of structure could be reliable. So students could assess structure as well as teachers. However, it needs consideration how to use self-assessment on coherence. It is useful to let students be more objective and revise their written production by themselves, but it was not proved that teachers can adopt it as class assessment.

VII. CONCLUSION

It is possible for high school students to assess structure in their written production as well as teachers, but it is difficult for them to assess coherence as well as teachers. It is also possible to adopt the assessment of structure in class as assessment, but it was difficult to use self-assessment of coherence. Self-assessment is effective to improve students' topic development in their written production, because students could reflect by themselves and be objective to revise it. Students could increase the scores of topic development, even though students could not assess coherence. Teacher feedback also helps students to be aware of the shortage of their written production. Teacher feedback of the present study mainly discussed written comments instead of the error correction of language use. Students supported this type of feedback to improve topic development, but some of them mentioned the necessity of more specific indication of their shortage. Self-assessment of written production was conducted twice during two months, so longitudinal research for a longer span is expected to prove the efficacy of self-assessment.

ACKNOWLEDGMENT

REFERENCES

[1] Oi, S. Y. (2012). *A Study of Student Evaluation and Teacher Evaluation* Unpublished master's thesis, Waseda University, Tokyo.
[2] LeBlanc, R. & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19*, 673-87.
[3] Bachman, L.F. & Palmer, A.S. (1981). The construct validity of the FSI oral interview. *Language Learning, 31*, 67-86.
[4] Oskarsson, M. (1989). Self-assessment of language proficiency: Rational and applications. *Language Testing, 6* (1), 1-13.
[5] Cheng, Y. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research, 12*(2), 254-255.
[6] Kent, D. (1980). Self-Grading Versus Instructor Grading. *Journal of Educational Research, 73*(4), 207-211.
[7] Boud, D. & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education, 18*, 529-549.
[8] Little, D. (2002). The European Language Portfolio: Structure, origins, implementation and challenges. *Language Teaching, 35*(3), 182-189.
[9] Crocker, A. C. & Cheeseman, R. G. (1988). The Ability of Young Children to Rank Themselves for Academic Ability. *Educational Studies, 14*(1), 105-110.
[10] Peirce, B., Swain, M., and Hart, D. (1993). Self-Assessment, French Immersion, and Locus of Control. *Applied Linguistics, 14*(1), 25-42.
[11] Ferris, D. R. (2004). *The "Grammar Correction" Debate in L2 Writing: Where are we, and where do we go from here? (and what do we do in the meantime...? Journal of Second Language Writing. 13*(1), 49-62.
[12] Hartshorn, K. J., Evans, N. W., & Merrill, P. F. (2010). Effects of Dynamic Corrective Feedback on ESL Writing Accuracy. *TESOL QUARTERLY. 44*(1), 84-109.
[13] Truscott, J. & Hsu, A. P. (2008). Error correction, revision, and learning. *Journal of Second Language Writing 17*(4), 292-305.
[14] Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing. 13*(4), 337-343.
[15] Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing. 14*(3), 191-205.
[16] Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing. 17*(2), 102-118.
[17] Kepner, C. G. (1991). An Experiment in the Relationship of Types of Written Feedback to the Development of Second-Language Writing Skills. *The Modern Language Journal. 75*(3), 305-313.
[18] Biber, D., Nekrasova, T., & Horn, B. (2011). The Effectiveness of Feedback for L1-English and L2-Writing Development: A Meta-Analysis. *TOEFL iBT ^{TM} Research Report. 14*.
[19] The Ministry of Education, Culture, Sports, Science and Technology. (2011). Designated for Course of Study for Senior High Schools. Retrieved June 27, 2013, from http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/eiyaku/1298353.htm
[20] Kaplan, R. (1966). Cultural Thought Patterns in Inter-Cultural Education. *Language Learning, 16*, 1-20.
[21] Harder, B. D., & Katz-Harder, H. (1982). Cultural Interface and Teaching English Composition in Japan. *The English Teacher's Magazine, 7. XXXI*, 4 (pp. 19-23). Tokyo: Taishu-kan.
[22] Oi, K. (1986). Cross-cultural Differences in Rhetorical Patterning: A Study of Japanese and English. *The Japan Association of College English Teachers 17*, 23-48.
[23] Hinds, J. (1990). Inductive, deductive, quasi-inductive: Expository writing in Japanese, Korean, Chinese and Thai. In U. Connor & A. M. Johns (Eds.), *Coherence in writing: Research and pedagogical perspectives* (pp. 87-109). Alexandria, VA: Teachers of English to Speakers of other Languages (TESOL).
[24] Kubota, R. (1998). An investigation of Japanese and English L1 essay organization: Differences and similarities. *The Canadian Modern Language Review, 54*, 475-507.
[25] Oi, S. Y. (2013). A Pilot Study of Self-Evaluation and Peer Evaluation. *Selected Papers of the 17th Conference of Pan-Pacific Association of Applied Linguistics, 1*-11.