

EFL Learners' Proficiency and Roles of Feedback:
Towards the Most Appropriate Feedback for EFL Writing

Yukio TONO and Ken KANATANI

Tokyo Gakugei University

ARELE

Volume 6

— Annual Review of English Language Education in Japan —
July 1995 所載論文

Published by

THE FEDERATION OF ENGLISH LANGUAGE EDUCATION SOCIETIES IN JAPAN

EFL Learners' Proficiency and Roles of Feedback: Towards the Most Appropriate Feedback for EFL Writing

Yukio TONO and Ken KANATANI
Tokyo Gakugei University

Abstract

The purpose of this paper is to investigate the effect of different types of feedback (direct correction, stamping, underlining) on the development of EFL learners' writing ability. The subjects (2nd- and 3rd-year junior high and 2nd-year senior high students) participated in the four-month program in which they wrote six in-class essays with different feedback. The results show that there was a significant relationship between feedback types and learners' grades. Especially direct teacher correction, which was found to be no more effective than the other feedback methods in the case of junior-high subjects, showed a significant difference in effect for the group of senior high students.

1. Introduction

In EFL teaching in Japan, the image of L2 writing has been shifting from writing as a means of grammar learning to a means of communication. The Ministry of Education issues the so-called Course of Study, which regulates the syllabus at junior and senior high schools, and in 1994, a new subject called "WRITING" has been introduced into high school. They emphasize the importance of writing for communication, like free composition, and essay writing. The students are expected to be fluent in writing, too. This "transplant" has yet to bear full fruit in our country, although researchers such as Chaudron (1984) and Kroll (1990) have convincingly urged new directions.

One of the obstacles for this new approach to writing is the teachers' tremendous work load at school. Many teachers think that it would be nice if they could have students express their ideas freely in writing and give each student words of encouragement or some comments on grammatical or stylistic errors. However, the teachers are simply too busy to do all of these

things. This is one of the reasons why EFL teachers are reluctant to give students free composition tasks.

Recently, however, the effect of teacher feedback has been questioned and the use of peer feedback and other means of evaluation are found to be preferable or at least equally effective. The following picture emerges from the recent research findings:

- (1) Peer feedback and conferencing are generally regarded as more ideal forms of feedback in comparison to teachers' direct correction. (Rose 1982; Zamel 1985; Urzua 1987; Keh 1990)
- (2) Conferencing and detailed direct error correction by the teacher would be extremely difficult to implement in a large class like that of Japan. (Charles 1990)
- (3) The opinions differ as to the effect of teacher feedback. The previous research shows that while some studies indicate that direct correction is useful for grammatical error correction (Cardelle and Corno 1981), other studies show that direct correction is no more effective than other modes of editing (Lalande 1982; Robb, Ross and Shortreed 1986; Hatori et al. 1990). Some scholars even claim that it may have some negative effect on the learners' subsequent writing quality and their attitudes toward L2 writing (Hendrickson 1981; Semke 1984).

If we know that there is not so much difference in the effect from the ways the feedback is provided and the teacher becomes free from all those editing tasks, then she can encourage students to write as much as they can. This intensive writing practice will be quite beneficial for their learning. Especially in such an acquisition-poor environment as in Japan, teachers are to provide learners with optimal opportunities for input. At the same time, learners should always be encouraged to give output in the target language in a natural communicative setting. Free or task-based writing in L2 is an excellent way to meet this need especially in such a classroom situation as the one in Japan.

This paper is based upon a series of studies, supported by the Grant-in-Aid for Scientific Research (C) offered by the Ministry of Education, Science and Culture, investigating the effects of different feedback on Japanese EFL students' development of writing ability. We have already found in the last research project (Hatori et al. 1990) that there was no significant difference in the quality and quantity of writing among the following three types of feedback; direct correction, underlining, and stamping. That was the case for high school students whose academic level was about in the middle. This time, we are particularly interested in the relationship between the effectiveness of different feedback and the learners' learning experience.

The purpose of this paper is to provide additional supporting evidence for the claim that direct correction of surface errors by the teacher would not bring about any benefit that counterbalances the time spent on it. In the last project (Hatori et al. 1990), we had senior high school students as subjects, which intended to supplement the data of junior college students obtained by Robb, Ross, and Shortreed (1986). This project, however, pursued the effect of feedback in the developmental

context of writing proficiency. Therefore, both junior and senior high school students took part in this study. Most of the researches, except for Shizuka (1993), focused on the subjects of the same proficiency level. The present study tried to investigate how the manipulation of types of feedback influences the quality of EFL writing by the learners of different language learning experience. In the next section, the hypotheses for the present study will be proposed and the research design will be discussed.

2. Method

Hypothesis

The following hypotheses were formed on the basis of our research questions:

- 1) There is no difference in the efficacy of the following three types of feedback: direct teacher correction, stamping and underlining.
- 2) Hypothesis 1 holds true irrespective of the students' learning experience: in this case, 2nd- and 3rd-grade junior high and 2nd-grade senior high students.

If we could verify these two hypotheses, it will lend more support to our assumption that the amount of feedback does not need to be in proportion to that of work given to the students and we could encourage teachers to invite their students to more exposure to and the use of English inside and outside the school curriculum (Hatori et.al. 1990: 5).

Subjects

A total of 400 Japanese EFL students (120 2nd grade junior high; 120 3rd grade junior high; and 160 2nd grade senior high) participated in this study. All of them were students of junior and senior high schools attached to the same national university. Their academic level was basically high because of the nature of the schools, as compared with ordinary junior and senior high school students. Almost all the students went onto higher educational institutions after graduation, which shows their academic level was far above average by the public school standard.

Design and Procedure

Each intact class was assigned to one of the three treatments: direct correction, underlining, or stamping. During the second and third terms in their school year, a total of six composition tasks were given to the subjects. Instruction and task contents were basically the same as those in Hatori et al. (1990) although some minor modification was made for composition topics. The students were given a writing task (see Appendix B. for the sample task) and had to write their essay in about twenty minutes during a class period. The rest of the class hour was spent on regular classroom activities. Students were first asked to read the task sheet and then wrote

their opinion in English.

Every time the compositions were complete, the papers were sent by the instructors to Tokyo Gakugei University, where the compositions were recorded on computer database. Feedback was then given according to the feedback types by our research assistants (graduate students at our department) under close guidance of the project team. Once the feedback was given, the papers were returned to each student. Every class was required to write six in-class essays during four months. Although every student was encouraged to take part in the activities, about 25% of all the students failed to hand in a part of their essays.

Data Analysis

The project team employed the following criteria for fluency and accuracy measures respectively:

- Fluency measure : a) Text length
- b) Mean T-unit length
- Accuracy measure : c) Mean error-free T-unit length
- d) Error-free T-unit/T-unit ratio
- e) Grammatical errors/100 words

The measurement devices related to T-units have been one of the most popular and powerful developmental indices both in L1 and L2. Although it was difficult to count T-units in the essays of our project, we believe that this system will give more credibility to our research results. As we counted T-units, we followed the guideline below:

- (1) One-word utterances were counted as T-units: e.g. "Great!", "Really?"
- (2) Unintelligible phrases or passages were not counted as T-units.
- (3) Any T-unit without any morphological or syntactic errors was counted as error-free T-units.
- (4) Spelling mistakes were ignored when counting error-free T-units.
- (5) Punctuation was also disregarded. Otherwise, it might seriously affect the result. We basically followed what the students wrote down in the original paper.

After all the numerical data was obtained, three-way ANOVA (feedback types × grades × compositions) was performed to see the differences among the means and the interaction of the independent variables.

3. Results

Since we not only have three main independent variables, namely, the types of feedback, learning experience (operationally defined as subjects' academic grades) and time (defined as six

compositions), but we also have several different levels of dependent variables, i.e. the six different composition measures, the presentation of our results will naturally become quite complicated. The descriptive statistics of each composition measure is listed in Appendix A. Table 1 summarizes the results of three-way ANOVA on each composition measure. In this section, we will briefly describe the results for each composition measure respectively and we will go on to discuss the interaction of the three independent variables and the effect on each composition measure, followed by some implications for future study.

Table 1. The results of three-way ANOVA on five composition measures

Effects	TTL	MT	MEFT	EFT/T	GERR
F.	3.93*	14.76**	8.49**	1.35	5.43**
G.	22.01**	89.49**	43.19**	.43	11.40**
C.	30.05**	28.90**	34.90**	1.34	7.08**
F.xG.	3.88**	3.89**	10.61**	1.37	16.35**
F.xC.	.84	.24	.49	1.04	1.95*
G.xC.	3.17**	.72	1.71	1.05	1.29
F.xG.xC.	2.05	.46	1.09	1.17	.49

TTL = total text length; MT = mean T-unit length; MEFT = mean error-free T-unit length; EFT / T = error-free T-units / T-units ratio; GERR = grammatical errors / 100 words; F. = feedback; G. = Grade; C. = composition

* p.<.05; ** p.<.01

Total Text Length

The results in Table 1 show that the main effects of all the three independent variables are significant. As for the main effects of the feedback types, the results of ANOVA indicate that the STAMP group wrote longer essays than the CORRECTION group and they wrote longer essays as their academic grades increased. The results, however, did not confirm the positive effect of extensive writing on writing quantity. As was mentioned earlier, we used different topics for every task, and did not ask the subjects to rewrite the drafts. Therefore, the number of words in the text may change depending on what kind of topic was given, or whether the subjects liked the topic or not. There was a serious decline of total text length in composition No. 5 & 6. One of the reasons may be that these two composition tasks were given in January, which was right after the winter holidays and the students might have been reluctant to work on the task.

Two-way interaction effect was significant especially between feedback types and grades, and between grades and compositions. The subjects produced longer essays as they were provided with correction and underline as feedback while the STAMP group showed some decline as they moved from JH3 to SH2. This may indicate that the effect of stamps may not be so positive for the higher graders. The significant effect of relationships between grades and compositions shows that the second-year senior high school students wrote more than the other two groups

when we look at each composition by grade.

Three-way interactions were significant, too. This is mainly due to the interactions between feedback and grades, as is indicated in the results of other measures. When the subjects were in junior high school, the text length increased as they received stamps for feedback. At senior high school level, however, the UNDERLINE group or the CORRECTION group outperformed the STAMP group. This also indicates that the subjects' interest in the task or the feedback tend to change as the grades increase.

Mean T-unit Length

The main effects of the three independent variables turned out to be significant again, but it is interesting to note that the group orders between total text length and mean T-unit length are different as follows:

Total text length: STAMP > UNDERLINE > CORRECTION

Mean T-unit length: UNDERLINE > CORRECTION > STAMP

The STAMP group produced longer essays in total length, but actually their sentence structures were not very complicated as compared with the other two groups. At least we could be fair to say that the CORRECTION group did not perform better than other two groups.

The two-way and three-way interaction effect was not significant except for the relationship between feedback and grades. The mean T-unit length became significantly longer as the grades increased, which was quite natural in terms of their developmental stages in SLA.

Mean Error-free T-unit Length

The main effects (feedback, grade, composition) were all significant. The UNDERLINE group performed significantly better than the STAMP and the CORRECTION groups. This is also confirmed by the grammatical error percentage. It is likely that underlining as feedback may be better than the other two modes of feedback; for stamping is simply childish for senior high school students and correction is sometimes just too much.

The interaction effect was only significant between feedback and grade. Other interaction effects were made rather weak primarily due to the effect of compositions. The students' performance varied in different composition tasks and this made it difficult to interpret each interaction effect.

Error-free T-unit/T-unit Ratio

The ratio of error-free T-units to total T-units has been a quite powerful index for measuring accuracy in writing in L1 and L2. The results, however, indicate that the effect was not significant in any of the variables and their interactions. Although this is not significant, it is interesting to note that there is a steady increase in this ratio especially from Composition No. 3 through No. 6 in the case of Junior High 3 and Senior High 2. As we mentioned earlier, the last

two compositions were written one month after the other four. There was a gradual decline in T-unit length especially in the last two compositions, which implies that the students might have been discouraged to write more, but even though their drafts became shorter and less complex, they became more accurate. This may lend support to our claim that intensive writing practice may suffice to cure surface errors in the long run. Further research is definitely needed to confirm this claim.

Grammatical Error Percentage

The main effects of the three independent variables were all significant. The results of the simple feedback effect show that the CORRECTION group has higher grammatical error percentages than the other two (CR= 6.80, ST= 6.20, UL= 5.74), but the interaction effects between feedback types and grades show a different picture. As for the CORRECTION group and the UNDERLINE group, the grammatical error percentages steadily decreased as they went on to the higher grades (CR: 8.05% → 6.97% → 4.28%; UL: 7.06% → 6.20% → 4.65%). The STAMP group, on the other hand, showed excellent percentages in JH2 (4.60%), but error percentages went up to 7.35% in JH3 and stayed until SH2 (6.90%). We could argue stamping is again effective for junior high students, but that it is no more effective for learners in higher grades. It is also noteworthy that teacher's direct correction seemed to work highly effectively in the case of this group of subjects. Hatori et al (1990) claimed that there was no difference in those three types of feedback by investigating writing behaviors of local public high school students. Their academic level was much lower than the present subject group, so this difference in the subjects' proficiency levels might cause this kind of conflicting findings.

4. Discussion

The purpose of the present study has been to show that direct teacher correction of surface error will not be so effective as compared with other feedback types. The results indicate that the main effect in feedback types was significant and that it was greatly influenced by the subjects' grades. The effect of writing essays repeatedly did not show any meaningful result.

The underline groups outperformed the other groups in terms of mean T-unit length and mean error-free T-unit length. This result at least shows that direct correction was not as effective as mere underlining in order to make the learners use more complex structures. In the last research by Hatori et al.(1990), the underline group performed better than the correction group although there was no statistically significant difference. This time we obtained a statistically significant result to confirm the previous findings. However, the stamp groups, which also outperformed the correction group in Hatori et al.(1990), did not perform so well this time. The interaction effect between feedback types and grades was very significant. The stamp groups outperformed the other two groups only in JH-2. As the students' grades went higher, the effect of stamping

decreased. This result may have something to do with the academic level of the students in this particular study. In Hatori et al.(1990), two local public high school students were used as subjects and their academic levels were not as high as those in the present study. Colorful stamps attracted the students in the last research group, but mere stamping may not have been so attractive for academically successful learners like those in the present study. Stamping may be useful for the younger students (such as JH-2 in our study) or those who are academically average or below average.

At least underlining will be beneficial for the learners to identify the error locations. If the academic level of the students is very high, simple underlining might better serve the purpose of their self-editing. We need to find out more about this relationship between the students' proficiency level and the optimal feedback in the future studies.

Further, a detailed investigation of the data according to the grammatical error percentages revealed that the correction groups outperformed the other two groups in Senior High 2. Although the mean error-free T-unit length became shorter in the last two compositions, a steady rise was observed in terms of grammatical error percentage. There may be two possibilities. First, this phenomenon just describes a matter-of-course fact: the simpler the structure becomes, the more accurately the learner can produce it. Secondly, there may be another possibility that direct teacher correction has some role in L2 writing. Shizuka(1993), for example, found that teacher correction was very effective in producing high quality drafts although the modification did not seem to be carried over to subsequent writing. In other words, direct correction has some immediate effect though it has no positive carry-over effect. In the same way, we can claim that direct teacher correction may be useful in order to improve the grammatical accuracy of L2 writing, but we should note that this result only holds true for the students with more learning experience (or higher grades). This suggests that there will be some relationship between the types of feedback and the students' proficiency levels. Although lacking in statistical significance in the interaction effect due to the defective composition variable effect, these data ought not be dismissed as superficial.

There are certain limitations to generalizing from the results achieved in this study because we could not control fully all of the confounding variables. One limitation involved the selection of the composition topics. We tried best to choose the topics which hopefully motivated the subjects to write more, but the result showed some apparent change of total text length depending upon the composition topics. As we mentioned earlier, our study did not focus on the learners' editing processes. Our main concern is that the learners' writing ability develops along with the increased amount of practice. In order to see any visible contribution to the learners' achievement in writing, the number of composition tasks should have been much larger and the period spent for the research, much longer. This kind of one-shot time-series design and a future longitudinal case study should go hand in hand in order to fully understand the role of feedback in L2 writing.

A second limitation involved the implementation of the actual writing tasks. Even though we had four months in total for the program, we had the subjects write only six times. Besides, we had almost one whole month break during the winter holidays, which seemed to discourage the subjects from participating in the program. The sharp declines in the last two tasks suggest that the program should have been carried out without any break.

5. Conclusion

In the last research project, we concluded, "the types of feedback does not seem to have any clear influence on the quality and quantity of at least some senior high school students". (Hatori et al. 1990:46) The present research findings show a little different picture. It suggests that teacher correction may have its own uses and advantages, which was in agreement with the account by Keh (1990), Jacobs and Zhang (1989) and Shizuka (1993). The question is whether this advantage is only limited to a small number of gifted learners or not. As far as the present study on the gifted learners and the last research on the average senior high students are concerned, it is likely that direct teacher correction can only be effective for such gifted learners. There is room for further investigation in this issue.

Further research on the qualitative aspect of writing, especially each writer's use of lexical and grammatical items is now underway. We hope that a systematic analysis of the L2 writers' data both from quantitative and qualitative viewpoints will shed more light on the development of L2 writing ability and the way teachers are involved in that process.

References

- Cardelle, M. and Corno, L. 1981. Effects on second language learning of variations in written feedback on homework assignments. *TESOL Quarterly*, 15(3): 251-261.
- Charles, M. 1990. Responding to problems in written English using a student self-monitoring technique. *ELT Journal*, 44 (4): 286-293.
- Chaudron, C. 1984. Evaluating writing: effects of feedback on revision. *RELC Journal*, 15(2): 1-14.
- Hatori, H., Itoh, K., Kanatani, K., and Noda, T. 1990. *Effectiveness and Limitations of Instructional Intervention by the Teacher. -Writing Tasks in EFL-* Tokyo: The Ministry of Education, Science and Culture.
- Hendrickson, J. M. 1981. *Error analysis and error correction* (Occasional Papers No. 10). Singapore: SEAMEO Regional Language Centre.
- Jacobs, G. and Zhang, S. 1989. Peer Feedback in Second Language Writing Instruction: Boon or Bane? Paper presented at the 1989 Annual Convention of the American Educational Research Association, March 27-31, 1989. (ERIC Document Reproduction Service No. ED 306 766)

- Kanatani, K., Itoh, K., Noda, T., Tono, Y., and Katayama, N. 1993. *The Role of Teacher Feedback in EFL Writing Instruction*. A report of a grant-in-aid of scientific research (C). Tokyo: The Ministry of Education, Science and Culture.
- Keh, C.L. 1990. Feedback in the writing process: a model and methods for implementation. *ELT Journal*, 44(4): 294-304.
- Kroll, B. (ed.) *Second Language Writing: Research insights for the classroom*. Cambridge University Press.
- Lalande, II, J. F. 1982. Reducing composition errors: an experiment. *Modern Language Journal*, 66: 140-149.
- Robb, T., Ross, S., and Shortreed, I. 1986. Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20(1): 83-95.
- Rose, A. 1982. Spoken versus written criticisms of student writing: some advantages of the conference method. *College Composition and Communication* 33(3): 326-330.
- Semke, H. 1984. Effects of the red pen. *Foreign Language Annals*, 17: 195-202.
- Shizuka, T. 1993. Effects of different editing methods on EFL writing quality at high school level. *JACET Bulletin*, 24:139-158.
- Urzua, C. 1987. You stopped too soon: second language children composing and revising. *TESOL Quarterly*, 21(2): 279-304.
- Zamel, V. 1985. Responding to student writing. *TESOL Quarterly*, 19(1): 79-101.

Appendix A. Descriptive Statistics of the Composition Measures

Mean:

	FEEDBACK:			GRADE:		
	CR	ST	UL	CR	ST	UL
TTL	49.68	53.73	51.26	46.40	53.80	54.90
MTUL	7.43	7.39	8.05	6.89	7.33	8.80
MEFT	6.84	6.93	7.41	6.71	6.59	7.94
EFTT	.66	.69	.79	.70	.75	.70
GERR	6.80	6.20	5.74	6.51	6.86	5.24

	COMP:					
	1	2	3	4	5	6
TTL	61.08	49.81	60.81	50.06	43.62	44.21
MTUL	8.13	6.75	7.81	8.42	8.11	6.66
MEFT	7.88	6.03	7.16	7.89	7.52	5.94
EFTT	.76	.67	.63	.66	.68	.89
GERR	4.71	6.97	6.86	6.02	5.95	6.76

Note: The descriptive statistics of the subgroups cannot be listed because of space limitation. Please see Kanatani et al (1993) for more details.

Appendix B. Sample composition task

新年明けましておめでとう。今年もよい年だといいですね。

お年玉はたくさんもらえましたか？かなりいい線行っただと言う人、ちょっと不景気(?)という人などいろいろでしょう。ところで、もらったお年玉で何をしますか？

前から何かを買いたいと思って貯金している人、旅行に行く人などさまざまでしょう。君はどうします？

前川千鶴さんと篠田豊君はこんな使い方をするそうです。

前川千鶴

I love skiing. But I don't have スキーの板, gloves or ski wear. I always went skiing in winter, but I never had my own ski boots or ストック. I always had to rent them and pay a lot of money. So I save OTOSHIDAMA to buy 板, ストック, and boots! It's exciting to think that I can use my own ski セット.

篠田豊

I want to use OTOSHIDAMA and travel all over Japan by train. I'm a train マニア. I collect all kinds of train tickets, pictures, and 電車の部品など. I travel a lot and take pictures of many trains and stations. This year I want to go to Kyushu. There are many good trains in Kyushu.

