

A Study of the Effect of Three Different Types of Feedback on Writing: Part 3 – Discussion and Conclusions

By
Peter Duppenhaler

Introduction

This is the third and final part of a three-part series describing a one-year study involving the use of journals with a group of 99 second-year students at a Japanese girls' high school. As we have seen, the students in the study were very similar in both ability level and background to those at our school, and therefore the findings should be applicable to our students. Part 1 (Duppenhaler, 2003) was devoted to a brief introduction, the research questions, and information on the participants, site, materials, and procedures used in the study. Part 2 (Duppenhaler, 2004a) was devoted to the statistical analyses of the data. In Part 3, I will discuss the findings and limitations of the study, make suggestions for further study, and state my conclusions.

Discussion

As we saw in Part 1, the vast majority of the literature on the use of journals in educational settings, and the kinds of feedback given in them, seems to conclude that the use of journals, especially those that focus on meaning, contributes to improvement in students' writing. However, most of these studies report only the general impressions of teachers or students or both. Only a very few attempts have been made to determine if the use of feedback in journals actually leads to any objectively quantifiable changes over time (Kreeft, 1984; Peyton, 1990; Minjong, 1997; Duppenhaler, 2002a). Although there have been a few studies involving the use of journals in Japan (Robb, Ross, and Shortreed, 1986; Casanave 1994; Hirose & Sasaki, 2000), there have been no experimental studies comparing the effect of different types of feedback in those journals.

The study, reported on in this series of articles, investigated the effect of three different types of teacher feedback on students' journal entries, and possible positive transfer effects on their in-class compositions. The three types of feedback were (a) meaning-focused

feedback (Group 1), (b) positive comments (Group 2), and (c) error-focused feedback (Group 3). The major purpose of the analysis was to determine if there were significant group differences among the three treatment groups. This was done by using a linear combination of eight dependent variables to maximize mean group differences if they in fact existed. The eight variables were: number of words, number of clauses, number of error-free clauses, Flesch-Kincaid Readability Index, and four vocabulary indices. The independent variable was treatment group assignment with three levels (i.e., the three types of feedback). For more on feedback see Duppenhaler (2001a & 2001b). Clauses were used as a measure of quality because in her study of the writing of two groups of low-proficiency English language students in Japan, Ishikawa (1995) found that clauses rather than T-units, "best quantify the overall quality of each individual [writing] sample" (p. 68).

The three research questions of the study were: (a) Do students who receive meaning-focused feedback show a greater degree of improvement over time in their journal entries than students who receive either positive comments or error-focused feedback? (b) Do students who receive meaning-focused feedback show a greater degree of improvement over time in their in-class writing samples than students who receive either positive comments or error-focused feedback? (c) Do students who receive meaning-focused feedback show a greater degree of positive motivation than students who receive either positive comments or error-focused feedback?

In order to avoid the problem of group differences, always a possibility with intact classes, the students were blocked into three treatment groups, according to their scores on a multiple-choice cloze test. In addition, several other procedures were carried out in order to ensure that there were no significant differences among the groups prior to treatment.

First, after the groups were formed, all students were asked to complete a bilingual Pretreatment Questionnaire designed to ascertain their degree of extracurricular exposure to English. It contained both dichotomous and scaled responses. A logistic regression analysis of the dichotomous responses in this questionnaire showed no significant differences among the three groups. A Multiple Analysis of Variance (MANOVA) of the Pretreatment Questionnaire questions that were on a scale also showed no significant differences among the three treatment groups. Specifically, the analysis of the questionnaire served to indicate that there were no significant differences among the three treatment groups with regard to their degree of extracurricular exposure to English prior to the outset of the study. In addition, an analysis of exactly the same questions on a bilingual Posttreatment Questionnaire also indicat-

ed that no significant differences among the three groups had developed during the treatment period. An examination of the Posttreatment Questionnaire data indicated that there had been only very minor changes with regard to extracurricular English activities during the course of the year and that none of these had proved to be significant.

Second, the students were asked to complete the first of three in-class writing samples. A one-way MANOVA was performed using the eight variables of interest (minus Token%3 which lacked any variance in this particular sample): total number of words, number of clauses, number of error-free clauses, three of the four remaining vocabulary indices (i.e., Token%1, Token%2, and TokenNot), and the Flesch-Kincaid Readability Index as the dependent variables; and group assignment as the independent variable. No significant differences were found among the three groups.

The results of the above analyses all indicated that there were no significant differences among the three treatment groups at the outset of the study. In addition, each treatment group was formed by blocking the students, and each group contained an almost equal number of students from each of the original three intact classes. Because each treatment group was made up of a similar portion of students who were enrolled in classes which were taught by each of the teachers who taught the second-year students, I was able to control for course content, possible initial ability level differences among the students, and teacher and instructional differences that might have occurred during the students' regular course of study.

The possible novelty of the treatment (i.e., Hawthorne effect) would seem to have been eliminated by the fact that the treatment lasted for one year. In fact, an examination of the journal data indicated that there might have been some such effect during the first month of treatment as shown by an initial increase in the number of words each group wrote. However, this was followed by a decline in all groups which I assume indicated that any such effect had worn off. Although it cannot be statistically proven, it would seem to be difficult to argue that any novelty effect would last for an entire academic year.

However, there is some possibility of a Hawthorne effect between treatment groups. Obviously, some students must have had friends in different treatment groups and these students might have talked about their journals. In fact, I became aware of one such case, when a student in Group 2 (positive comments) wrote in her journal that I had written more in her friend's journal (Group 1, meaning-focused feedback) than in her journal and asked that I write more to her. Her exact words were:

I am sad because you didn't answer my writings. My friend says, "Why do your notebook no answer?"

I answered, I don't know, but I thought that I might write to little writing. So this time is writing more than before. Would you answer to my writing?

I wrote, *Your writing is good and interesting.*

This incident occurred early on in the treatment and only happened once; however, it does indicate that at least some students shared information about their journals with students in other groups. This is a potential problem in any study that seeks to compare different treatments within the same educational setting. Still, the teachers at the school seemed to feel that this was a rare case and that most of the students neither showed nor talked about their journals with others. With this in mind, it was felt that although it might be a problem, it was an unavoidable one in this case, and given the other precautions that had been taken, as well as the longitudinal nature of the study, it would not seriously damage the study's validity.

A second unforeseen problem had to do with how to deal with direct questions, from Group 2 (positive comments) and Group 3 (error-focused feedback) students, such as the following three examples:

I went to travel in Kusatsu with my friend. It's only one day. We went to there by train. Then we went from Kusatsu to Tsuge by train. We felt natural more and more. I want to travel again. How many time do you go to travel a year?

I wrote, *About two or three times a year. It's fun.*

My favorite writer is Jiro Akagawa. He writs mistery. I like "Mikeneko Homuzu" that written by him. Who is your favorite writer?

I wrote, *I like Agatha Christie.*

Is there a plan where you will go in the summer vacation?

I wrote back, *I don't know. I have not decided yet. I want to go somewhere with my family.*

There were 11 such requests, nine from Group 2 and two from Group 3 (out of a total of 1,452 possible entries from these two groups), spread out over the course of the year. I felt, for ethical reasons, that I should answer these direct questions. As you can see from my responses, I tried to make them short. Again, because of the rarity of this type of occurrence, I felt that it did not constitute a threat to the validity of the study. However, it is something that future researchers should keep in mind.

As can be seen from the above, a number of precautions were taken in order to ensure the similarity of the groups prior to the outset of the study, and to control for the possible effects of ability, course content, and teacher and instructional differences.

Reliability of the multiple choice cloze test was also investigated. The reliability of the test was found to be Cronbach alpha (.77), Split-half adjusted reliability (.82). Given the extreme likelihood of this being a very homogeneous group of students, and the fact that homogeneity tends to lower the reliability coefficient, it was felt that the level of reliability for the multiple-choice cloze test was acceptable for blocking purposes. Interrater reliability (Pearson r) for the variables counted by hand (i.e., clauses and error-free clauses) was .96 (clauses) and .98 (error-free clauses) for the first in-class writing sample and subsequently never lower than .99 for either in-class writing samples or journal entries. As all other variables were computer-generated only a small random sample in each treatment group was rerun through the computer programs to check for accuracy. No inaccuracies were found.

Finally, it should be noted that the results of a MANOVA performed on the first four journal entries (i.e., the first month of treatment) using the eight variables of interest as the dependent variables and group assignment as the independent variable indicated that there were no significant differences among the three treatment groups. As in the case of the first in-class writing sample, there were no significant differences among the three groups with regard to their first month of treatment.

A factor analysis of the journal data resulted in the identification of three factors with Eigenvalues greater than one. These factors were consistent with those that had been predicted to occur and accounted for 78% of the variance.

The above would seem to indicate that any significant differences among the three groups that might have developed during the course of the academic year could be attributed to the effect of the treatment the students received during that time rather than to any group differences that might have existed prior to the outset of the treatment period, or to group differences which might have been the result of differences in ability, course content, and teacher and instructional methods.

Having said this, let us now proceed to an examination of the results of the analyses focusing on each of the three main research questions in turn.

Research Question 1

Research Question 1: Do students who receive meaning-focused feedback show a greater degree of improvement over time in their journal entries than students who receive either positive comments or error-focused feedback?

"Improvement" was operationalized as a significant increase over time in the (a) *quantity*, as measured by the number of words per entry; (b) *accuracy*, as measured by the number of error-free clauses per entry; and three measures of (c), *quality*: (1) the *number of clauses* per entry, (2) *readability* as measured by the Flesch-Kincaid Readability Index, and (3) *vocabulary*, as measured by the four vocabulary indices.

The results of a MANOVA using the eight variables of interest contained in the 22 journal entries written during the course of the academic year as the dependent variables and group assignment as the independent variable indicated an overall significant difference at $p = .0000$. Univariate F tests indicated that the three significant differences were in the following areas: (a) the total number of words at $p = .0000$, (b) the number of clauses at $p = .0000$, and (c) the number of error-free clauses at $p = .0000$.

Post hoc analysis (Scheffé test) of the total number of words revealed a significant difference between Groups 1 (meaning-focused feedback) and 2 (positive comments) at $p = .0000$, and between Groups 1 and 3 (error-focused feedback) at $p = .0000$, but no other significant differences. An examination of the means showed that Group 1 had a significantly high-

er mean than either Group 2 or Group 3 (means: G1 = 2070, G2 = 1407, G3 = 1565). Group 1 had written significantly more words than either Groups 2 or 3.

Post hoc analysis (Scheffé test) of the total number of clauses again revealed a significant difference between Groups 1 and 2 at $p = .0000$, but no other significant differences. An examination of the means showed that Group 1 had a significantly higher mean than Group 2, and Group 3 had a higher mean than Group 2 (means: G1 = 168, G2 = 110, G3 = 137). Group 1 had written significantly more clauses than Group 2, and Group 3 had written more clauses than Group 2.

Post hoc analysis (Scheffé test) of the number of error-free clauses again revealed a significant difference between Groups 1 and 2 at $p = .0000$, but no other significant differences. An examination of the means showed that Group 1 had a significantly higher mean than Group 2, and that Group 3 had a higher mean than Group 2 (means: G1 = 82, G2 = 56, G3 = 67). Group 1 had written significantly more error-free clauses than Group 2, and Group 3 had written more clauses than Group 2.

Group 1 had shown a significant "improvement" in two of the three measures of improvement (i.e., *quantity*, *accuracy*, and *quality*). With regard to Research Question 1, meaning-focused feedback does seem to be more effective than either positive comments or error-focused feedback in facilitating improvement.

The question of "improvement over time," was addressed in two ways. First, pre/post analysis of the first four weeks indicated that there were no significant differences among the three treatment groups during the first four weeks of the treatment at $p = .7536$; and an analysis of the last four journal entries (i.e., the last month of journal entries) also indicated no overall significant difference among the three groups at $p = .0124$.

As can be seen from the above, no significant differences existed among the three groups during either the first four weeks or the last four weeks of journal writing. If we view the first/last four weeks as pre/post scores, we cannot say that there was a significant change over time. However, it should be kept in mind that there were, as we have seen, significant differences in the 22-entries analysis indicating that improvement was not in the form of steady growth but nonetheless did occur.

A second indication of "improvement over time" was found in the time series analysis, which, in general, indicated that all three groups started out almost the same, and then made some early progress which was followed by a decline, at which point Group 1 showed a tendency to outperform the two other groups, ending, in some interesting differences. For exam-

ple, the 22-entries journal graph for the total number of words showed that after having started out at almost exactly the same point in the first entry, Group 1 consistently wrote more than either of the other two groups (for samples of all of the graphs, see Duppenthaler, 2002a). In addition, Group 3 had a tendency to write more than Group 2. This same general pattern seems to hold for both the number of clauses and the number of error-free clauses – Group 1 is the highest, followed by Group 3, and then Group 2.

The 22-entries journal graph for the number of clauses, showed that although Group 1 started out between Groups 2 and 3, by the time of the third entry, Group 1 had overtaken both of the other groups and maintained this position for the remainder of the treatment with the exception of weeks 14 and 17. The pre/post analysis graph for this same variable showed that after an initial decline in all groups, Group 1 showed a steady increase in the number of clauses while Groups 2 and 3 made moderate progress only at the very end of the study and finished far below Group 1.

The same general pattern can be seen in the pre/post graph and also in the 22-entries journal graph for the total number of error-free clauses; all groups made some early progress followed by a decline, but then Group 1 tended to outperform the two other groups.

In summary, meaning-focused feedback (Group 1) does seem to be more effective than either positive comments (Group 2) or error-focused feedback (Group 3) in facilitating improvement in journal entries. Group 1 outperformed the other two groups in two of the three measures of improvement: (a) *quantity*, as measured by the number of words per entry; and (b) *accuracy*, as measured by the number of error-free clauses per entry. As to the three measures of (c), *quality*, Group 1 outperformed the other two groups in the *number of clauses* per entry, however, with regard to *readability*, and *vocabulary*, no group was significantly different from any other (also see Duppenthaler 2004c).

Research Question 2

Research Question 2: Do students who receive meaning-focused feedback show a greater degree of improvement over time in their in-class writing samples than students who receive either positive comments or error-focused feedback?

As stated above, the results of a MANOVA on the first in-class writing sample indicated that there were no significant differences among the three treatment groups at $p =$

.6825. No significant group differences were found in the first in-class writing sample which was collected just prior to the treatment period. The results of a MANOVA on the second in-class writing sample indicated no overall significant difference at $p = .0176$. The results of a MANOVA performed on the third in-class writing sample also indicated no overall significant difference at $p = .0146$. As we have seen, there were no significant differences in the first, second or third in-class writing samples. Therefore, we cannot say that there is steady improvement over time. The in-class time series graphs showed that all three groups generally tended to move in similar patterns; however, within this general pattern, there were some interesting differences in the second and third in-class writing samples. An examination of the error-free clause graph showed that over the three samples, Group 1 was the only group that had made a steady increase in the number of error-free clauses over time. A look at the clause graph showed that all three groups made progress in the number of clauses, but that Groups 1 and 3 outperformed Group 2 in the third in-class sample. Group 2's progress was also not as dramatic as that of Groups 1 and 3. This may indicate that positive comments are not a particularly effective type of feedback (for a more detailed discussion of this see Duppenthaler 2004b).

In summary, we cannot say that meaning-focused feedback is more effective than either positive comments or error-focused feedback in facilitating improvement over time in students' in-class writing samples. The difference between the positive effects seen in the journal entries (i.e., 22-entries) and the lack of a similar effect in the in-class writing samples may be the result of time constraints and the obviously different nature of the two types of writing.

Research Question 3

Research Question 3: Do students who receive meaning-focused feedback show a greater degree of positive motivation than students who receive either positive comments or error-focused feedback?

Questions 11 through 20 in the Posttreatment Questionnaire, which did not appear in the Pretreatment Questionnaire, were designed to determine (a) the degree of either positive or negative feelings the students had toward writing in their journals and (b) whether they felt the experience had been a positive one. A 5-point Likert scale was used for each question

(1 = strongly agree, 2 = agree, 3 = neither agree nor disagree, 4 = disagree, 5 = strongly disagree). Question 20 included space for a written response. Students were free to write in either Japanese or English. All of the students wrote comments. I coded these using the same 5-point Likert scale used for the other questions. The questions were as follows:

Question 11: I enjoyed writing in my journal.

Question 12: I think writing in my journal had a positive effect on my English.

Question 13: I would like to continue writing in a journal next year.

Question 14: I enjoy writing in English more now than I did a year ago.

Question 15: I think my writing is better now than a year ago.

Question 16: I can express myself in writing more easily now than a year ago.

Question 17: I think writing in my journal was a good experience for me.

Question 18: Writing in my journal made me want to study English more.

Question 19: I looked forward to getting my journal back each week.

Question 20: Has writing a journal changed your attitude toward English?

A one-way MANOVA of questions 11 through 20 indicated no significant differences among the three groups at $p = .0007$; however, in order to interpret the results of the post-treatment questionnaire let us look at Table 1 which shows the group averages for questions 11 through 20. Once again, the three groups are: Group 1 (meaning-focused feedback), Group 2 (positive comments), and Group 3 (error-focused feedback).

Table 1. Posttreatment Questionnaire Questions 11 through 20 Averages

Question	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
Group 1										
average ($n = 30$)	2.3	1.9	2.2	2.2	2.8	2.6	1.8	2.9	2.0	2.5
Group 2										
average ($n = 30$)	2.6	2.0	2.9	2.3	2.3	2.6	2.0	2.8	3.0	2.6
Group 3										
average ($n = 29$)	2.8	2.0	3.2	2.4	2.8	3.0	1.9	3.3	2.2	2.7

For purposes of interpretation, the following standard was used: 1.8 - 2.3 = agree, 2.4 - 2.8 = less strongly, 2.9 - 3.3 = neutral. Using this standard we can interpret the above averages for each question as follows:

Question 11: Group 1 most enjoyed writing in their journals, with the other two groups enjoying the journals, but to a lesser degree.

Question 12: All groups agreed that writing in their journals had a positive effect on their English.

Question 13: Group 1 agreed they would like to continue writing in a journal next year, while Groups 2 and 3 neither agreed nor disagreed.

Question 14: Groups 1 and 2 enjoyed writing in English more at the end of the year than a year earlier. This sentiment was shared to a slightly lesser degree by Group 3.

Question 15: Group 2 felt that their writing was better than a year earlier, and the other two groups agreed to a lesser extent that their writing had improved.

Question 16: Group 3 neither agreed nor disagreed that they could express themselves in writing more easily than a year earlier, but Groups 1 and 2 agreed to a slightly higher degree that they could do so.

Question 17: All groups felt that writing in a journal had been a good experience.

Question 18: Group 2 felt, to some extent, that writing in a journal made them want to study English more, but Groups 1 and 3 were neutral.

Question 19: Groups 1 and 3 looked forward to getting their journals back each week, but Group 2 was neutral on this.

Question 20: All groups agreed to some extent that writing journals had changed their attitude toward English. An examination of the written comments accompanying this question indicated that almost all of the comments were positive.

All three groups seemed to be positively disposed to journal writing. In addition, all three groups reported that they felt it had had a positive effect on their English, and that it had been a "good experience" for them (for more on motivation see Duppenthaler, 2002b).

Conclusion

Research Question 1 – Hypotheses

Hypothesis 1 (The number of words per journal entry over time will be significantly higher in the group which receives meaning-focused feedback.) was supported. The number of words per journal entry over time was significantly higher in the group which received meaning-focused feedback: in descending order the groups were meaning-focused, error-focused, positive comments (Group means: G1 = 2070, G3 = 1565, G2 = 1407).

Hypothesis 2 (The number of error-free clauses per journal entry over time will be significantly higher in the group which receives meaning-focused feedback.) was partially supported. The number of error-free clauses per journal entry over time was higher in the group which received meaning-focused feedback: in descending order the groups were meaning-focused, error-focused, positive comments (Group means: G1 = 82, G3 = 67, G2 = 56). However, the difference was only significant with regard to Group 1 and Group 2, and there was no significant difference between Group 1 and Group 3.

Hypothesis 3 (The number of clauses per journal entry over time will be significantly higher in the group which receives meaning-focused feedback.) was partially supported. The number of clauses per journal entry over time was higher in the group which received meaning-focused feedback. In addition, there was a statistically significant difference between the error-focused feedback group and the positive comments feedback group: in descending order the groups were meaning-focused, error-focused, positive comments (Group means: G1 = 168, G3 = 137, G2 = 110). However, the difference was only significant with regard to Group 1 and Group 2, and there was no significant difference between Group 1 and Group 3.

Hypothesis 4 (The Flesch-Kincaid Readability Index per journal entry over time will be significantly higher in the group which receives meaning-focused feedback.) was not supported. The mean Flesch-Kincaid readability index per journal entry over time of the meaning-focused feedback group was higher than the other two groups; however, not significantly so. In descending order the groups were meaning-focused, error-focused, positive comments (Group means: G1 = 3.17, G3 = 3.05, G2 = 2.96).

Hypothesis 5 (The use of Token%2 and Token%3 vocabulary per journal entry over time will be significantly higher in the group which receives meaning-focused feedback.) was not supported. The use of Token%2 and Token%3 vocabulary per journal entry over time was not significantly higher in the group which received meaning-focused feedback. There were no significant differences among any of the three treatment groups with regard to any of the vocabulary indices. There also does not seem to be any obvious pattern: in descending order the groups were Token%2 – positive comments, meaning-focused, error-focused, (Group means: G2 = 5.77, G1 = 5.74, G3 = 5.42); and Token%3 – error-focused, positive comments, meaning-focused (Group means: G3 = .873, G2 = .868, G1 = .818).

Research Question 2 – Hypotheses

Hypothesis 1 (The number of words per in-class composition over time will be significantly higher in the group which receives meaning-focused feedback.) was not supported. The number of words per in-class composition over time was not significantly higher in the group which received meaning-focused feedback. There were no significant differences among the three treatment groups in the first, second, or third in-class writing assignments with regard to the number of words.

Hypothesis 2 (The number of error-free clauses per in-class composition over time will be significantly higher in the group which receives meaning-focused feedback.) was not supported. The number of error-free clauses per in-class composition over time was not significantly higher in the group which received meaning-focused feedback. There were no significant differences among the three treatment groups in the first, second, or third in-class writing assignments with regard to the number of error-free clauses.

Hypothesis 3 (The number of clauses per in-class composition over time will be significantly higher in the group which receives meaning-focused feedback.) was not supported. The number of clauses per in-class composition over time was not significantly higher in the group which received meaning-focused feedback. There were no significant differences among the three treatment groups in the first, second, or third in-class writing assignments with regard to the number of clauses.

Hypothesis 4 (The Flesch-Kincaid readability index per in-class composition over time will be significantly higher in the group which receives meaning-focused feedback.) was not supported. The Flesch-Kincaid readability index per in-class composition over time was not significantly higher in the group which received meaning-focused feedback. There were no significant differences among the three treatment groups in either the first, second, or third in-class writing assignments.

Hypothesis 5 (The use of Token%2 and Token%3 vocabulary per in-class composition over time will be significantly higher in the group which receives meaning-focused feedback.) was not supported. The use of Token%2 and Token%3 vocabulary per in-class composition over time was not significantly higher in the group which received meaning-focused feedback. There was no significant difference in either Token%2 or Token%3 among the three treatment groups in the three in-class writing assignments.

Research Question 3 – Hypothesis

Research question 3's hypothesis (The degree of motivation, as measured by a post-treatment questionnaire, will be highest in the group which receives meaning-focused feedback.) was partially supported. The data from the Posttreatment Questionnaire would seem to indicate that journal writing, regardless of feedback type, is a motivating experience; but that one type of feedback is not "statistically" more motivating than any other. An examination of the data indicated that in general Group 1 was slightly more positively disposed towards getting their journals back (Question 19). However, it should also be noted that Group 3 also looked forward to getting their journals back, although to a slightly lesser degree. This may be related, as noted by Timson, Grow, and Matsuoka (1999), to Japanese students' apparent preference for error correction. Another possible interpretation might be that positive comments lack face validity with Japanese students even when the students are certain that the teacher is reading their journal entries. In summary, although all groups tended to find the experience motivating, there does seem to be some justification for the suggestion that meaning-focused feedback results in slightly greater overall positive motivation than the other types of feedback.

Limitations

The main limitation of this study is its generalizability. "Generalizability of a study is the degree to which the results can be said to be meaningful beyond the study" (Brown, 1988, p. 113). In other words, it seeks to answer the question – To what other persons, in what other settings or situations, would the results apply? The answer to this question depends on how conservative we wish to be. If we adopt a very conservative approach the results are only applicable to the particular group of individuals involved in the study, in the particular setting in which it was carried out, and at the particular time it was carried out. This is, of course, true for any study.

Brown (1988) states that the central question of any study is whether it is logical and meaningful. He notes that "[some extraneous] variables affect the logic of a study (internal validity) while others affect its meaningfulness (external validity)" (p. 36). In this study a number of steps were taken to minimize the effect of extraneous variables in order to "assure that the groups were representative and equivalent on all potentially related variables except

those under investigation" (Brown, 1988, p. 37). These precautions included blocking based on cloze test scores, a pretreatment questionnaire, and a pretreatment writing sample, none of which indicated any significant group differences.

As noted earlier there may have been a possible Hawthorne effect. In addition, although I do not think this was the case, there may have been a subconscious researcher expectancy. I feel that I made every possible effort to provide the appropriate type of feedback to each treatment group, and to treat all the students with equal consideration; however, as noted in the literature review, we are sometimes unaware of our own actions. Again, this is true for almost any study one can think of.

I did not meet the students and so I do not think that there was a halo effect or subject expectancy – the students may have wanted to please me but I think it unlikely in that the study lasted one academic year and this is too long for the students to maintain this state of expectancy (i.e., to make efforts to please someone they never met).

In short, I feel that the internal validity of the study is acceptable. On the other hand, the external validity, the generalizability of the findings to other situations, is limited and this should be kept in mind when interpreting the results.

Further Study

There is an obvious need for a replication of this study at a similar institution, preferably for an even longer period of time. Following this, a replication at a Japanese boy's high school and at a Japanese coeducational high school, as well as similar institutions in different cultural settings is called for. This would provide valuable data for comparing the effects of various types of feedback in different settings. Such studies would also lend themselves to an examination of possible gender and cultural differences. A study focused on comparing the effect of students with different starting abilities would also be of interest: we may find that one type of feedback is more effective with one level of student than another. Radecki and Swales (1988) reported that as students progress in their English language development they become less tolerant of their teachers' feedback roles.

Investigation would also be useful into attempting to find out if there is a relationship between the student's level of English or, possibly, time spent abroad in English-speaking environments, and feelings towards journal writing. Casanave (1993) mentions that,

A handful of other students, particularly returnee students [i.e., returning to

live in Japan after having lived abroad for a period of time] at the end of three semesters of English, remarked that they 'hated journal writing,' yet recognized that it benefited their English in a number of ways (p. 100).

A second area of interest would be to build on the work, in the field of ESL, of Dolly (1990), in which she investigated "partners' sharing of responsibility for conversation management [in dialogue journals]" (p. 318). She found that although native speakers' moves, such as introducing and extending topics and handling communication problems, remained consistent across journals, "the degree and style of participation varied immensely among the students" (p. 320). It would be interesting to attempt to determine the types of moves made by native-speaker English teachers and EFL Japanese students in dialogue journals, and to compare these to those of American teachers in Japan and Japanese, ESL students in America. Are certain moves more effective in one cultural setting than another?

Finally, some researchers claim (see, for example, Cobine, 1995) that journal writing can connect reading, writing, and discussion through activities that accommodate diverse learning styles and that further students' linguistic development. There is obviously a need to investigate the effects of journal feedback and writing on reading and discussion.

Implications for Language Pedagogy

It would seem that at least with regard to journal writing, meaning-focused feedback (Group 1) does seem to be more effective than either positive comments (Group 2) or error-focused feedback (Group 3) in facilitating overall improvement in journal entries over one academic year. This statement is of course limited to the context of the present study. Group 1 significantly outperformed both groups in (a) *quantity*, as measured by the number of words per entry. It significantly outperformed Group 2 and outperformed, but not significantly, Group 3 in (b) *accuracy*, as measured by the number of error-free clauses per entry. As to the three measures of (c), *quality*, Group 1 significantly outperformed Group 2 and outperformed, but not significantly, Group 3 in the *number of clauses* per entry, however, with regard to *readability* as measured by the Flesch-Kincaid Readability Index, and *vocabulary*, no group was significantly different from any other. This would seem to reconfirm the findings of several researchers (Sandler, 1987; Kepner, 1991, Casavave, 1994) that "an interactive, self-generated, cumulative and functional writing/reading exchange between teacher and student" (Gutstein, 1983, p.1) is related to both "improvement," as defined in this study, and high-

er student motivation. In addition, error-correction was found to be more effective than positive comments even though, after an extensive review of the literature, Truscott (1996) reported that grammar correction is both ineffective and harmful. Finally, positive comments, even when these included short personalized responses so that the students were aware that their entries were being read, were found to be the least effective type of feedback.

However, we cannot say with any certainty that meaning-focused feedback is more effective than either positive comments or error-focused feedback in facilitating improvement over time in students' in-class writing samples. This lack of transfer is probably related to the very different nature of the two types of writing, and it may be inappropriate to expect to find much, if any, transfer effect. In the journals the students have unlimited time to explore a topic of their choice, to think about what they want to say, to self-edit and to work on mechanics. The fact that students felt free to ask questions in the journals but not in the in-class writing assignments is another indication of how different the tasks are.

With regard to motivation, the data from the Posttreatment Questionnaire would seem to indicate that meaning-focused feedback does result in a slightly greater overall positive motivation than either positive comments or error-focused feedback. The meaning-focused feedback group was most positively disposed towards getting their journals back each week. However, it should also be noted that the error-focused feedback group also looked forward to getting their journals back, although to a slightly lesser degree, and that not one of the groups felt that the experience had been a negative one, in spite of the fact that all journal writing had to be done, in addition to regular homework assignments, at home. Whatever the case may be, there does seem to be some justification for the suggestion that meaning-focused feedback results in greater overall positive motivation than either positive comments or error-focused feedback; however, this interpretation must be made with some caution.

The overall findings of this study reconfirm, to some extent, the positive effects of meaning-focused feedback in journal writing within that genre; however, no substantial justification for belief in a positive transfer effect to other types of writing, as represented in this case by in-class compositions, was found. The findings also reconfirmed the positive effect journal writing, regardless of feedback type, has on motivation and the somewhat greater positive effect of meaning-focused feedback over the other two types of feedback.

Journals, as many have noted (Reed, 1988; Jones, 1988; Baskin, 1994) provide opportunities to connect with students in a personal, non-threatening way, opening up their world to

a teacher in ways that would not otherwise be possible. I "lived" with these students for an entire academic year, and although I never met them I felt closer to them than to most of the students in my own classes during that time. I will admit that I was still somewhat skeptical of the value of journal writing when I began this study; however, the meaning-focused feedback group wrote more, and what they wrote was both more complex and more accurate. Journal writing, and especially journal writing that incorporates meaning-focused feedback, does seem to be a worthwhile addition to any language program.

References

- Baskin, R. S. (1994). *Student feedback on dialogue journals*. National Institute of Education, Educational Resources Information Center (ERIC Document Reproduction Service No. ED 375 627).
- Brown, J. D. (1988). *Understanding research in second language learning*. Cambridge: Cambridge University Press.
- Casanave, C. P. (1993). Student voices: The insiders speak out on journal writing. In C. P. Casanave (Ed.), *Journal Writing: Pedagogical Perspectives* (pp. 95-115). Institute of Languages and Communication, Keio University, SFC (Shonan Fujisawa Campus).
- Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing*, 3(3), 179-201.
- Cobine, Gary R. (1995). Writing as a Response to Reading. ERIC Digest. Bloomington, IN: ERIC Clearinghouse and Reading, English, and Communication. (ERIC Document Reproduction Service ED 386 734).
- Dolly, M. R. (1990). Adult ESL students' management of dialogue journal conversation. *TESOL Quarterly*, 24(2), 317-321.
- Duppenthaler, P. (2001a). A review of studies on the effect of different types of feedback on student writing. *Studies and Essays*, 36, 1-16.
- Duppenthaler, P. (2001b). Helping students to reduce written errors. *RECL Guidelines*, 23(2), 24-26.
- Duppenthaler, P. (2002a). *Feedback and Japanese high school English language journal writing*. Unpublished doctoral dissertation, Temple University, PA.
- Duppenthaler, P. (2002b). The effect of three types of written feedback on student motivation. *JALT Journals*, 24(2), 130-154.
- Duppenthaler, P. (2003). A Study of the Effect of Three Different Types of Feedback on Writing: Part 1 – Research Questions, Participants, Site, Materials, and Procedures. *Studies and Essays*, 37, 1-21.
- Duppenthaler, P. (2004a). A Study of the Effect of Three Different Types of Feedback on Writing: Part 2 – Data Analysis. *Studies and Essays*, 39, 1-27.
- Duppenthaler, P. (2004b). *Journal Writing and the Question of Transfer of Skills to Other Types of Writing*. *JALT Journals*, 26(2), 171-188.
- Duppenthaler, P. (2004c). The effect of three types of feedback on the journal writing of Japanese EFL students. *JACET Bulletin*, 38, 1-17.
- Gutstein, S. P. (1983, March). *Using language functions to measure fluency*. Paper presented at the annual meeting of the Teachers of English to Speakers of Other Languages, Toronto, Ontario.
- Hirose, K., & Sasaki, M. (2000). Effects of teaching metaknowledge and journal writing on Japanese university students' EFL writing. *JALT Journal*, 22(1), 94-113.
- Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, 4(1), 51-69.
- Jones, P. M. (1988). *Knowing opportunities: Some possible benefits and limitations of dialogue journals in adult second language instruction*. Unpublished doctoral dissertation, School for International

Training, Vermont.

- 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(iii), 305-313.
- Kinoshita, A. Kyoto academia turns to private sector for help. (2000, June 5). *The Daily Yomiuri*, p. A5.
- Kreeft, J. E. (1984). *Dialogue journal writing and the acquisition of grammatical morphology in English as a second language*. Unpublished doctoral dissertation, Georgetown University, Washington, DC.
- Minjong, S. (1997). The effect of dialogue journal writing on writing quality, reading comprehension, and writing apprehension of EFL college students. Report. National Institute of Education, Educational Resources Information Center (ERIC Document Reproduction Service No. ED 410 766).
- Peyton, J. K. (1990). Dialogue journal writing and the acquisition of English grammatical morphology. In J. K. Peyton (Ed.), *Students and teachers writing together: Perspectives on journal writing* (pp. 65-97). Washington, DC: TESOL.
- Radecki, P., & Swales, J. (1988). ESL student reaction to written comments on their written work. *System*, 16(3), 355-365.
- Reed, L. (1988). Dialogue journals make my whole year flow: The teacher's perspective. In J. Staton, R. W. Shuy, J. K. Peyton, & L. Reed (Eds.), *Dialogue journal communication: Classroom, linguistic, social, and cognitive views* (pp. 56-72). Norwood, NJ: Ablex.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20(1), 83-93.
- Sandler, K. W. (1987). Letting them write when they can't even talk? Writing as discovery in the foreign language classroom. In Fulwiler, T. (Ed.), *The journal book* (pp. 312-322). Portsmouth, NH: Heineman.
- Timson, S., Grow, A., & Matsuoka, M. (1999). Error correction preferences of second language learners: A Japanese perspective. *JACET Bulletin*, 30, 135-148.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327-369.