Journal Writing and the Question of Transfer of Skills to Other Types of Writing

Peter M. Duppenthaler
Tezukayama Gakuin University

This report looks at the effect of journal feedback and the possible transfer of skills from journal writing to in-class compositions. Although no statistically significant justification for belief in a positive transfer-of-skills effect was found, only the meaning-focused feedback group was found to have made a steady increase in the number of error-free clauses over time, in both their journal entries and in their in-class writing samples.

本稿は英語日記に対するのフィードバックの効果、ならびに英語で日記をつけることで得たスキルが、授業の中でのライテイングへ転化されるかどうかを検証する。スキルの有効な転化がある、という見解を裏付ける統計的に有意な確証は得られなかったが、意味に中心をおいたフィードバックを受けたグループについては、誤りのない文節の数の着実な増加が、日記への書き込みにも、授業の中でのライテイングのサンプルにもみとめられた。

Introduction

One of the most frequently cited reasons for using journals in educational settings is that they provide opportunities for authentic, meaningful communication that is focused on the message rather than the form. In addition, there seems to be a fairly strong belief in the idea that "when the learner participates in meaningful communication with a native

speaker of the target language and with a focus on the message rather than the forms of the language, competence with the forms will follow" (Kreeft, 1984, p. 1).

Although a number of articles and books have discussed the advantages of using journals (Davies, 2001; Fulwiler, 1987; Gutstein, Batterman, Harmatz-Levin, Kreeft, & Meloni, 1983; Holmes & Moulton, 1997; Kelen, 2001; Kreeft, 1984; Kresovich, 1988; Leki, 1985; Peyton, l990a; Peyton, l990b; Peyton & Reed, l990; Staton, 1987; Werderich, 2002), only a few researchers have investigated the question of improvement in journal writing over time (Barba, 1992; Casanave, 1993, 1994; Duppenthaler, 2002a; Kreeft, 1984), and even fewer have investigated the possibility of a transfer-of-skills effect from journal writing to other types of writing (see for example Abdel Fattah, 1993; Hirose & Sasaki, 2000; and Ross, Shortreed, & Robb, 1988). Do the gains, for example in fluency or accuracy, reported in journals carry over into other types of writing? As noted by Takahashi (1993), "the question remains whether this teaching technique [i.e., journals] leads to transfer of skills to other areas of learning" (p. 85), a question that, despite some encouraging evidence, remains largely unanswered.

The one-year study reported here involved the use of journals with a group of 99 second-year students at a Japanese girls' high school and investigated the effect of three different types of feedback: (a) meaning-focused feedback, (b) positive comments, and (c) error-focused feedback, on (1) students' journal entries, (2) possible positive transfer effects on in-class compositions, and (3) motivation. This report focuses on the second item, possible positive transfer effects on in-class compositions. (For information on other aspects of the study see Duppenthaler 2002a and 2002b.)

Methods

Participants

All second-year students in the school are divided into one higher-level class, one middle-level class, and three lower-level classes based on their performance during their first year of high school. The students in the three lower-level classes are assigned to individual classes on the basis of alphabetical order. The 99 participants in this study consisted of the students in the three lower-level classes. The main reason for using the three lower-level classes was that they constituted the largest group of students at one ability level. Several procedures, which will be

explained later, were carried out to ensure the equivalency of the three treatment groups.

Materials

Fifteen in-class writing sheets were developed. Each sheet included a simple set of instructions in English, a four-frame picture sequence that the students were to use as the basis of a story, the first line of the story, and space to write the story and record the number of words written.

Picture sequences were selected for their clear story line and because writing stories about them did not necessarily require prior knowledge of the subject, which might have affected the students' writing. In addition, a number of researchers (Ishikawa, 1995; Ross, Shortreed, & Robb, 1988; Rousseau, Bottge, & Dy, 1993) have used similar picture sequences.

The particular four-picture sequence sets used in this study were taken from a set of commercially available pre-first grade *Eiken* (STEP, Society for Testing English Proficiency) study guides (Akao, 1997). The four-picture sequence sets are used as prompts for the speaking part of the pre-first grade STEP test.

The Japanese English teachers at the school and an American consultant with an MA in TESOL and several years of teaching experience at a Japanese high school were asked to select three sheets which they felt were equal in difficulty and which the participants would find interesting and not too difficult. These three sheets were used for the three in-class writing samples.

Procedures

During the first week of school, the three intact classes of second-year students were given a cloze test that I developed in conjunction with the same group of Japanese English teachers and the American consultant mentioned above. The reliability of the test was found to be acceptable for blocking purposes (split-half adjusted reliability was .82) given that the students are a rather homogeneous group of individuals and that "the more homogeneous the group is with respect to the trait being measured, the lower will be the reliability coefficient" (Ary, Jacobs, & Razavieh, 1990, p. 280). The cloze scores were used to block the 99 students into three treatment groups of 33 students each, with each of the three original intact classes almost equally represented in each treatment group. The three groups were Group 1, the meaning-

focused feedback group; Group 2, the positive comments group; and Group 3, the error-focused feedback group.

During the second week of school, the participants filled in a bilingual pretreatment questionnaire, which I also developed with the group mentioned above (see Duppenthaler 2002b for samples of both the pretreatment and posttreatment questionnaires). It consisted of ten questions designed to establish the students' familiarity with and exposure to English outside of their regular classes. It was used to check for any pretreatment differences among the treatment groups. An analysis of the questionnaire responses indicated that there were no significant differences among the three treatment groups.

The first of the three in-class writing assignments was given during the second week of school prior to any journal writing, which did not begin until the fifth week of school. This assignment was used (a) to establish a base-line estimate of the students' in-class writing ability, and (b) to further determine if there were any group differences prior to the outset of the study.

The students were given 35 minutes to complete the writing assignment. They were told, in Japanese, to read the instructions and to write as much as they could. No other instructions were given to the students. Students were allowed to use dictionaries. No student was absent on the day of the first in-class writing assignment. The sheets were collected and given to me. I photocopied them, and returned them to the school so that they could be returned to the students. All of the students' writing samples were typed up, checked for accuracy, and run through the spell checker program in Microsoft Word. Any spelling errors were corrected. This was done so that accurate estimates of vocabulary and readability could be obtained from the two computer programs (i.e., VocabProfile and RightWriter), which were used to analyze the writing samples. Any words in Japanese were left unchanged. There were very few of these and they were usually either names of persons or places.

The number of clauses and error-free clauses in each sample were counted by hand. The American consultant mentioned earlier and I rated the papers independently in order to establish interrater reliability. Prior to counting clauses and error-free clauses, the two raters met and agreed on a set of criteria to be used (see Appendix A). The interrater reliability for the number of clauses was .96 and for the number of error-free clauses was .98. The same procedures and raters were used for the second and third in-class writing samples.

To simplify the computer data processing, all of the writing samples in one treatment group were combined into one file, resulting in three files, one for each treatment group. Each of the three files was then processed separately. The total number of words was counted using Microsoft Word's word count function. The VocabProfile computer program was used to check vocabulary and the RightWriter program for readability.

At the beginning of the fifth week of school, the participants began keeping their weekly journals. On average, students wrote between 40 and 50 words each week. All journal entries were written at home on the students' own time.

The students completed the second in-class writing assignment in week 24. One student in each of the three treatment groups was absent; however, these students wrote their samples after school during the same week under the same conditions as the other students. The same procedures as in the case of the first in-class writing sample were then followed. The interrater reliability was .99 for the number of clauses and 1.0 for the number of error-free clauses.

The third in-class writing assignment was given during the 37th week of school, the last week of the treatment period. The third in-class writing assignment was carried out in exactly the same manner as the previous two. This time two students were absent from each treatment group; however, they wrote their samples after school, under the same conditions, during the same week. The interrater reliability was .99 for the number of clauses and .99 for the number of error-free clauses.

A bilingual posttreatment questionnaire was given one day after the third in-class writing assignment. The questionnaires were collected and passed on to me for analysis. The same procedures as those used in the case of the pretreatment questionnaire were used. An analysis of the questionnaire responses indicated that, with regard to extracurricular English activities, no significant differences among the three treatment groups had occurred during the course of the year.

As briefly mentioned earlier, the three types of feedback given in the students' journals were (a) meaning-focused feedback (often referred to as "dialog journals") in which I engaged in an ongoing, interactive dialog with the participants; (b) positive comments, in which I responded with phrases such as "well done," "that was an interesting story," and so on, and occasional short positive comments on a few scattered entries (these were provided in order to ensure that, as in the case of the other two treatment groups, the students were aware of the fact that their entries

were being read); and (c) error-focused feedback in which I corrected all errors, in red, with no revision required on the part of the participants.

"Improvement" was operationalized as a significant increase over time in (a) *quantity*, as measured by the number of words per composition; (b) *accuracy*, as measured by the number of error-free clauses per composition; and (c) "*quality*" as measured by a battery of six variables: (1) number of clauses, (2) Token%1, (3) Token%2, (4) Token%3, (5) TokenNot, (6) Read (i.e., Flesch-Kincaid Readability index).

As stated above, clauses were counted by hand. Variables 2 through 5 (i.e., Token%1, Token%2, Token%3, and TokenNot), were obtained by running the VocabProfile computer program, which allows the user to determine "the percentage of words a learner uses at different vocabulary frequency levels in her writing, or, put differently, the relative proportion of words from different frequency levels" (Laufer & Nation, 1995, p. 311). The program compares "a text [i.e., the students' writing samples] against vocabulary lists to see what words in the text are and are not in the lists, and to see what percentage of the items in the text is covered by the lists" (VocabProfile manual, n. d., p. 1).

The "lists" in this case are (a) Token%1, the most frequent 1,000 words of English, (b) Token%2, the second most frequent 1,000 words, (c) Token%3, "words not in the first 2,000 words of English but which are frequent in upper secondary school and university texts from a wide range of subjects" (VocabProfile manual, n. d., p. 3), and (d) TokenNot, words not found in any of the other three lists. The sources of these lists are *A General Service List of English Words* (West, 1953) for the first 2,000 words (i.e., Token%1 and Token%2), and "The University Word List" (Nation, 1990) for words in Token%3.

The Flesch-Kincaid Readability index was obtained by using Right-Writer (version 3.1), a commercially available computer grammar/style program. A Readability Index is designed to indicate the level of education a reader will need in order to understand a given text. (For more on readability see Duppenthaler, 2000.) The formula for this index is as follows: Grade Level = (.39 x ASL) + (11.8 x ASW) -15.59; where ASL (Average Sentence Length) is the number of words divided by the number of sentences in the sample, and ASW (Average number of Syllables per Word) is the number of syllables divided by the number of words in the sample. "Extensive testing of RightWriter's readability calculation shows an average error of less than 2%. This is usually lower than the error rate for calculations made by human operators" (RightWriter User's Manual, 1990, p. 7-5[sic]).

Clauses are introduced in the second year of junior high school and were selected rather than t-units for a number of reasons. According to Richards, Platt, and Weber (1985) a t-unit is "the shortest unit which a sentence can be reduced to, and consisting of one independent clause together with whatever dependent clauses are attached to it" (pp. 299-300). Although a number of studies have been carried out using t-units, several researchers (Gaies, 1980; Ishikawa, 1995; Laufer & Nation, 1995; Pery-Woodley, 1991), have pointed out problems with the use of t-units. In addition, in a study with low-level Japanese learners, Ishikawa (1995) recommended against their use when attempting to examine "the efficacy of different experimental treatments for low-proficiency EFL writing" (p. 68).

Researchers often disagree about what constitutes correctness. Some (especially Larsen-Freeman & Strom, 1977) would consider a writing sample to be error-free only if correct in every respect. However, low-proficiency writers, such as those who took part in this study, often make mistakes, and requiring perfectly correct samples would amount to holding learners to impossibly high native-speaker standards. As a result, a number of concessions were made (see Appendix A).

Analysis

The major research question was to determine if there were significant group differences in the students' writing at the end of the study. The goal of the analysis was to create a linear combination of eight dependent variables to maximize mean group differences. The eight dependent variables consisted of: the total number of words, number of clauses, number of error-free clauses, the Flesch-Kincaid readability index, and the four vocabulary indices. The independent variable was group assignment to one of three feedback treatment groups.

Assumptions

Procedures related to the identification of possible outliers, the evaluation of the assumptions of normality of sampling distributions, homogeneity or variance-covariance matrices, linearity, and multicollinearity were carried out following recommendations found in Tabachnick and Fidell (1996). The results were found to be satisfactory (for details see Duppenthaler, 2002a). The overall alpha level of the study was set at .05; however, a Bonferroni type adjustment was made in order to guard

against inflated Type I errors. The adjusted alpha for all Multivariate Analysis of Variance tests (MANOVA) in the full study was set at .005 (the original alpha level of .05 divided by 10, the total number of Mulitvariate Analysis of Variance tests in the main study). In addition an adjustment was made in the alpha level for all Univariate F tests. In this case, the adjusted alpha for the Mulitvariate Analysis of Variance tests (.005) was divided by the number of dependent variables (DVs). In the case of the in-class writing samples the adjusted alpha was set at .000625 (.005/8 DVs).

In-class Writing Samples

An examination of the first in-class writing samples showed that no student had included any Token%3 vocabulary items. As there was no variance for this item, it had to be dropped from the analysis because the statistical program used will not run when a variable with no variance is present. Dropping this item had no effect on the overall analysis because it contributed nothing to indicating any group difference; all of the groups were exactly the same in that not one of the students had written any Token%3 vocabulary items. A one-way MANOVA was performed using the remaining seven dependent variables. Group assignment was used as the independent variable. No significant differences were found among the three groups at p = .6825.

In the case of the second in-class writing samples a one-way MANOVA was performed using the eight variables of interest as the eight dependent variables and group assignment as the independent variable. No overall significant difference was found at p = .0176. In the case of the third in-class writing samples, no overall significant differences were found at p = .0146.

Time Series

As mentioned above, there were no significant differences in the three in-class writing samples; however, plotting the variables over time (See Appendix B) did result in some interesting findings. Appendix B consists of graphic representations of the information shown in Table 1 where TotW = Total number of words, C = number of clauses, EFC = number of error-free clauses, Read = Flesch-Kincaid Readability Index, Tok%1 = Token%1, Tok%2 = Token%2, and TokNot = TokenNot. Let us now look at each of these in turn.

Table 1. Means for the Three In-class Writing Samples (*N* = 99)

Sample	Group	TotW	C	EFC	Read	Tok%1	Tok%2	Tok%3	TokNot
first	1	58.67	4.90	2.33	2.96	85.70	7.71	0.00	4.91
first	2	67.36	4.82	2.24	2.93	85.64	8.40	0.00	5.17
first	3	60.97	4.39	2.15	2.66	86.52	8.58	0.00	4.90
second	1	94.42	9.24	3.66	3.87	79.84	10.46	1.29	2.35
second	2	98.76	7.88	3.13	3.67	82.04	11.00	1.41	2.52
second	3	85.55	6.85	1.67	4.06	82.34	9.64	1.60	2.90
third	1	82.12	10.52	4.09	1.78	81.95	4.01	0.15	1.77
third	2	79.24	9.03	2.91	2.25	80.76	3.71	0.28	3.13
third	3	82.27	10.21	3.55	2.58	86.67	2.96	0.18	4.11

In the line graph showing the Total Number of Words (see Appendix B), we can see that all three groups wrote more in the second writing sample than in the first writing sample and then fewer in the third than in the second. This was most likely the effect of the topic of the picture sequence. For some reason students were able to write more about the second writing assignment's set of pictures than the other two sets. All three groups follow the same up-down pattern and all three group means are almost identical. Treatment did not seem to have any noticeable effect on the total number of words (i.e., fluency).

In the Clause graph, we can see that all three groups made fairly steady progress in the number of clauses they wrote, and again, all three group means are very similar. Treatment did not seem to have any noticeable effect on the total number of clauses. However, with regard to the Error-free Clauses graph (i.e., accuracy), Group 1 (meaning-focused feedback) is the only group to have made a steady increase in the number of error-free clauses over time. Treatment did seem to have some positive effect on accuracy, especially as the students only wrote one entry each week. Admittedly the difference is not statistically significant but it is nonetheless very encouraging. Whereas all three groups were writing more clauses over time, the meaning-focused group was the only one to show consistent improvement. This was also found to be the case in the journal entries themselves in which Group 1 outperformed the other two groups in both (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 (see Duppenthaler, 2004).

In the Readability graph we can see that once again, all three groups show a similar pattern and that the means are very close. The increase in readability in the second sample was probably related to the high percentage of Token%3 vocabulary in this sample as shown in the Token%3 graph. Remember that readability will increase with the number of words in a sentence and the number of syllables in those words. Level 3 vocabulary items would tend to have more syllables than level 1 vocabulary items.

Token%1 and TokenNot show the highest degree of variability among the groups. In the third in-class writing sample, Group 3 used far more TokenNot words than Group 1, even when taking into account Japanese vocabulary items and TokenNot words which were the same for each group. This may indicate a tendency on the part of Group 3 to use the basic words they knew and to supplement them with words that they looked up in a dictionary, while Groups 1 and 2 tended to use slightly more Token%2 vocabulary items than Group 3. It is tempting to interpret this difference as an indication that Groups 1 and 2 had acquired a higher level (i.e., Token%2 vocabulary) of working vocabulary than Group 3. However, it might also indicate either a lack of desire for risk-taking, which might be result of Group 3's treatment (i.e., error-focused feedback), or simply a smaller working vocabulary. A longer study might be able to shed some light on this point.

Limitations

Because each treatment group was made up of a similar portion of students who were enrolled in classes that 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. However, there are some methodological limitations that were pointed out to me after the completion of the study and that should be mentioned here. One was the overall problem of topic. This could have been avoided by altering the in-class writing assignment topics instead of giving them to all the participants at the same stage. If one-third of each group had written on one topic at each stage, perhaps more clear-cut results of feedback effects would have been found. Although this would have been very desirable, given that all the students had been blocked into treatment groups

rather than using intact classes, and that I was neither their teacher nor on site, I felt that the logistics of this would been impossible to control. I therefore opted not to do this. I could have also used a holistic measure of quality to check the validity of the six quantitative variables. A holistic measure would have also provided another way of looking at the data. Lastly, it has been pointed out that the Bonferroni adjustment might be too strict. I also had questions about this and consulted an internationally recognized statistician on this point. His advice was to leave the adjustment rather than to change it. This simply goes to show that the more advice one seeks before one undertakes a study the better off one will be and that once a decision has been made one has to live with it.

Conclusion

The vast majority of the literature on the use of journals in educational settings seems to conclude that the use of journals, especially those that focus on meaning, contributes to improvement in the 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 journals actually leads to any quantifiable changes over time.

As we have seen, there were no statistically significant differences among the three groups in the first, second or third in-class writing samples. Therefore, we cannot say that the use of journals, regardless of feedback type, resulted in any significant transfer-of-skills effect. However, the in-class time series graphs indicated some suggestive tendencies in terms of differences among the three groups.

An examination of the Error-free Clauses graph showed that over the three samples, Group 1 was the only group that made a steady increase in the number of error-free clauses over time. A look at the Clauses 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 either that positive comments are not a particularly effective type of feedback or that the students had not "bought into the treatment," meaning that my occasional short positive comments on a few scattered entries (which were provided in order to ensure that, as in the case of the other two treatment groups, these students were aware of the fact that their entries were being read) were not enough to convince them of the reader's interest in what they were writing.

Although the overall findings of the full study reconfirm, to some extent, the positive effects of meaning-focused feedback in journal writing within that genre, no statistically significant justification for belief in a positive transfer-of-skills effect to other types of writing, as represented in this case by in-class compositions, was found. This may be the result of time constraints. This lack of clear evidence of transfer might also be 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-of-skills 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. In addition, it was impossible in the study to have a group of students not writing journals and this also complicates the issue. However, the fact that only the meaning-focused group made a steady increase in the number of error-free clauses over time, in both their journal entries and in their in-class writing samples, might at least be taken as an indication of the value of this type of treatment over the other two, and indicate a possible difference in transfer of skills. This is, of course, only a possibility but nevertheless one that would certainly seem to merit further study.

Acknowledgements

Peter Duppenthaler is a Professor in the Communication Department, Faculty of Literature at Tezukayama Gakuin University. His research interests include the use of student journals in language teaching, and motivation and strategy use.

References

Abdel Fattah, A. E. G. (1993). *The influence of dialogue journals and other practicum*. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

Akao, F. (Ed.). (1997). *Eiken jun 1 kyu zen mondaishu '97* [STEP pre-first grade comprehensive study guide, 1997]. Tokyo: Obunsha.

Ary, D., Jacobs, L. C., & Razavieh, A. (1990). *Introduction to research in education* (4th ed.). Fort Worth, TX: Harcourt Brace College Publishers.

Barba, C. (1992). *The effect of dialogue journal writing on the speaking ability of students of Spanish as a foreign language.* Unpublished doctoral dissertation, The Pennsylvania State University.

Casanave, C. P. (Ed.). (1993). *Journal writing: Pedagogical perspectives.* 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.

Davies, C. (2001, Winter). Journal writing: A response to Christopher Kelen. *On Cue, 9*(3), 19-20.

Duppenthaler, C. (2000). Readability measurements of some English readers used in Japanese high schools. *Baika Review, 33,* 35-45.

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 Journal, 24*(2), 29-53.

Duppenthaler, P. (2004). The effect of three types of feedback on the journal writing of Japanese EFL students. *JACET Bulletin, 38,* 1-17.

Fulwiler, T. (Ed.). (1987). *The journal book.* Portsmouth, NH: Heinemann.

Gaies, S. J. (1980). T-unit analysis in second language research: Applications, problems, and limitations. *TESOL Quarterly, 14,* 53-60.

Gutstein, S. P., Batterman, H., Harmatz-Levin, C., Kreeft, J., & Meloni, C. (1983). *Using real English: Writing a dialogue journal.* U. S. Department of Education, National Institute of Education, (ERIC Document Reproduction Service No. ED 256 155)

Hirose, K., & Sasaki, M. (2000). Effects of teaching metaknowledge and journal writing on Japanese university students' EFL writing. *JALT Journal, 22*(1), 94-113.

Holmes, V. L., & Moulton, M. R. (1997). Dialogue journals as an ESL learning strategy. *Journal of Adolescent and Adult Literacy, 40*(8), 616-621.

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4*(1), 51-69.

Kelen, C. (2001, Summer). Journal writing: Some observations from a Hong Kong high school classroom. *On Cue, 9*(2), 11-14.

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.

Kresovich, B. M. (1988). The journal assignment in composition class at a Japanese university. *Ryukyu Daigaku Kyoiku Gakubu Kiyo, 33*(1), 71-83.

Larsen-Freeman, D., & Strom, V. (1977). The construction of a second language acquisition index of development. *Language Learning, 27*(1), 123-134.

Laufer, B, & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307-322.

Leki, I. (1985, September). Linking writing journals to writing in the ESL classroom: Part II. *TECFORS, VIII*(4), 43-48.

Nation, I. S. P. (1990). A university word list. In I. S. P. Nation (Ed.), *Teaching and Learning Vocabulary* (pp. 235-239). Boston: Heinle and Heinle.

Pery-Woodley, M. (1991). Writing in L1 and L2: Analyzing and evaluating learners' texts. *Language Teaching, 24(*2), 69-83.

Peyton, J. K. (l990a). 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.

Peyton, J. K. (Ed.). (l990b). *Students and teachers writing together: Perspectives on journal writing*. Washington, DC: TESOL.

Peyton, J. K., & Reed, L. (l990). *Dialogue journal writing with nonnative English speakers: A handbook for teachers*. Alexandria, VA: TESOL.

Richards, J., Platt, J., & Weber, H. (1985). *Longman dictionary of applied linguistics*. London: Longman.

RightWriter. (1990). *User's manual.* Carmel, IN: Que Software, a Division of Macmillian Computer Company.

Ross, S., Shortreed, I., & Robb, T. (1988). First language composition pedagogy in the second language classroom: A reassessment. *RELC Journal, 19*(1), 29-48.

Rousseau, M. K., Bottge, B. A., & Dy, E. B. (1993). Syntactic Complexity in the writing of students with and without mental retardation. *American Journal of Mental Retardation*, *98*(1), 113-120.

Staton, J. (1987). The power of responding in dialogue journals. In T. Fulwiler, (Ed.), *The journal book* (pp. 47-63). Portsmouth, NH: Heineman.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins College Publishers.

Takahashi, Y. (1993). Journal writing for Japanese students: What do they write and what do they learn? In C. P. Casavave (Ed.), *Journal Writing: Pedagogical Perspectives* (pp. 78-86). Institute of Languages and Communication, Keio University, SFC (Shonan Fujisawa Campus).

VocabProfile. (n. d.). *User's manual*. Wellington, New Zealand: Victoria University, English Language Institute. Retrieved from http://www.vuw.ac.nz/lals/staff/nation.aspx

Werderich, D. E. (2002). Individual responses: Using journal letters as a vehicle for differentiated reading instruction. *Journal of Adolescent & Adult Literary, 45*(8), 746-754.

West, M. (1953). *A general service list of English words.* London: 1953.

Appendix A

Criteria for Judgments of Correctness

(a) Punctuation, misplaced or omitted commas, misplaced or omitted punctuation used in or with direct quotations, and misplaced or missing apostrophes in plural possessives (e.g., women's' college) were disregarded; misplaced or missing apostrophes in contractions were counted as mistakes.

(b) Capitalization, uncapitalized proper nouns, and sentences not beginning with a capital letter were not counted as mistakes.

(c) Spelling errors were disregarded.

(d) Use of the native language which either rendered the clause or sentence incorrect when there was a common, one-word equivalent in English, which did not render the clause or sentence incorrect when expressing the concept in English would have required complex sentence-structure(s) or sophisticated cultural knowledge, which did not render the clause or sentence incorrect when the word (e.g., typhoon, tatami) was already fairly widely used in English-speaking countries, and which did not render the clause or sentence incorrect when it was a proper name (e.g., Mr. Suzuki, Umeda [a geographical area within Osaka City]).

(e) When two clauses were incorrectly joined, one clause was counted as incorrect.

(f) A sentence beginning with incorrect use of a conjunction was counted as two clauses, but the first one was counted as incorrect. Several examples of this type of concession are: (1) correlative conjunctions (either...or, neither...nor) when only one was used in an otherwise grammatically correct sentence; (2) subordinate conjunctions followed by an otherwise grammatically correct sentence; (3) conjunctive adverbs (after, still) followed by an otherwise grammatically correct sentence (I went to school. After I had breakfast). In these cases the remainder of the sentence was grammatically correct. In order not to invalidate that, the unit was counted as two clauses, with the first one being counted as incorrect.

(g) Blanks or missing words other than articles rendered the clause or sentence incorrect.

Appendix B
Time Series Graphs

Total Number of Words

Clauses

Error-free Clauses

Readability

Token% 1

Token% 2

Token% 3

Token Not