مطالعه‌ی مقایسه‌ای بین دو روش بازخورد در مهارت نوشتاری فراگیران زبان خارجی در ایران

پرویز مفتون

علیرضا زارع اکباتانی

چکیده

تحقیق حاضر به مقایسه‌ی عملکرد نوشتاری فراگیران ایرانی زبان خارجی در حالی که دو روش بازخورد یعنی: بازخورد با صدای ضبط شده (ATF) و کمترین نشانه‌گذاری (MM) را دریافت می‌کردند، می‌پردازد. ۸۴ شرکت‌کننده مرد از میان ۱۳۶ فراگیر زیر متوسط مهارت یادگیری زبان، که خود توسط مؤسسه‌ی کیش انتخاب شده بودند، در این تحقیق حضور داشتند. کلاس‌های مورد بررسی به‌صورت تصادفی به دو گروه تقسیم شدند: گروه دریافت کننده‌ی ATF و گروه دریافت‌کننده MM که هر گروه شامل ۴۲ نفر افراد واجد شرایط می‌باشند که مجددا توسط یک آزمون توسط یک آزمون همسان‌سازی، برای یکسان بودن افراد مورد بررسی قرار گرفته‌اند. در طول مدت این تحقیق ۳۶ جلسه شرکت کنندگان موظف بودند تا پاراگراف‌های توصیفی ۱۵۰ کلمه‌ای در مورد هشت موضوع را نگارش کنند. یکی از گروه‌ها در متن‌شان بازخورد ATF و دیگر گروه بازخورد MM دریافت کردند. دو آزمون برای مقایسه‌ی عملکرد نوشتاری گروه‌ها استفاده شد: یکی «پیش آزمون» که قبل از انجام آزمایش برگزار شد و دیگری «پس آزمون» که بعد از آزمایش انجام پذیرفت که هر کدام شامل نوشتن پاراگراف‌های ۱۵۰ کلمه‌ای سر کلاس در حد فاصل زمانی ۴۰ دقیقه‌ای بود. تحلیل داده‌ها و حساب آماری نشان داد که: الف، گروه دریافت‌کننده‌ی ATF در پس آزمون بهتر عمل کردند. ب. گروه دریافت کننده‌ی MM در پس

آزمون هیچ نوع پیشرفتی نسبت به پیش آزمون نشان ندادند. ج. گروه دریافت کننده‌ی ATF در پس آزمون بهتر از گروه دریافت‌کننده‌ی MM عمل کردند.

واژه‌های کلیدی: نشانه‌گذاری تحلیلی، بازخورد با ضبط صوت، نشانه‌گذاری کلیتی، کمترین نشانه‌گذاری، چند مصحح، نگارش.

A Comparative Study of Two Feedback Methods on Iranian EFL Learners' Writing Skill

Parviz Maftoon[*]

Ali Reza Zareh Ekbatani[*]

Abstract

The present study was conducted to compare the Iranian EFL learners' writing performances when receiving two different feedback styles: Audio taped feedback (ATF) and minimal marking (MM). Eighty-four male participants of this study were chosen from among 136 early intermediate EFL learners who were selected through the cluster sampling of students at Kish Language Institute. The classes under study were randomly divided into two groups: the ATF Group and the MM Group, each with 42 eligible participants, who were again checked to be homogeneous through a homogeneity test. During the course of this study, i.e.,

[*] Iran University of Science and Technology

[*] Islamic Azad University, Tehran Central Branch

36 sessions, the participants were assigned to write expository paragraphs of about 150 words on eight writing topics. One of the groups received ATF on their scripts, and the other group received MM. Two tests were used to compare the writing performances of the groups: a pre-test prior to the treatment and a post-test after the treatment, both in the form of in-class expository paragraphs of about 150 words under a time limit of 40 minutes. Data analyses and statistical calculations indicated that (a) the ATF Group performed better on the post-test, (b) the MM Group's performance on the post-test did not show any improvement over the pre-test, and, lastly, (c) the ATF Group performed better than the MM Group on the post-test.

Key Words: *Analytic Marking, Audiotaped Feedback, Holistic Marking, Minimal Marking, Multiple Raters, Writing*

Introduction

Giving feedback has always been a salient feature of the teaching profession. An appropriate feedback style can have a major influence on students' motivation for future learning. Whether what is given as feedback to students is quickly attended to, easily comprehended, and permanently remembered has been of great concern to most educators, as well as

teachers, of EFL writing. As Grabe and Kaplan (1996) contend, "students can be positively motivated to explore many areas of knowledge and personal creativity through supportive and constructive responses to their writing" (p. 377). It has been emphasized that appropriate feedback can have the positive effect of producing in the learner "a sense of reader awareness" (Muncie, 2000, p. 52) and "a sense of audience" (Boughey, 1997, p. 131) which students appear to lack.

Apparently, teachers play a leading role in facilitating learners' progress by providing learners with proper and reassuring feedback. Most writing teachers consider feedback, in general, and marking, in particular, a boring and unrewarding task. To make marking or any other type of feedback an effective tool in the writing classroom, teachers have to adopt inspiring feedback methods that encourage students to reconsider, revise, and rework their drafts. In this way, learners will be actively involved in the correction of their errors, rather than being passive viewers of the teacher's notes and corrective comments, usually marked in red.

To accomplish the foregoing proposition, Hyland (1990) advocates two effective styles of giving feedback: Minimal marking and taped commentary which he describes as two "interactive feedback styles" (p. 279). After all, "feedback must be interactive to be genuinely effective" (Hyland, 1990, p. 285).

With respect to minimal marking that employs a set marking codes, White and Arndt (1991), among others, suggest that throughout a writing course students can be provided with a checklist which consists of a series of correcting codes, like the list below, as a framework for them to check

their work and think about where they have, for any reason, failed to convey their message, and try to make necessary corrections in their next draft. If they are not successful in their attempts, they should ask their peers and, as a last resort, their teacher for advice.

S= subject missing

V= verb form error

A= article error

T= tense error

SV= subject-verb concord error

Adv= adverb order error; misplaced or missing adverb

Adj= adjective order error; misplaced or missing adjective

Prep= preposition error

Ss= sentence structure error

The above symbols (White & Arndt, 1991, p. 173) or any other correcting codes as such can be written in the margin of the students' scripts or above the error.

Hyland (1990) asserts that he has found this approach of guiding language learners to detect and repair their own errors very successful. He claims that "students are able to correct up to three quarters of their errors without further prodding, and the experience seems to help them avoid the same problems later" (p. 281). In fact, through minimal marking (MM), EFL learners are led towards "self-correction" which emphasizes self-discovery in the language learning process while activating the learners' linguistic competence and enhancing language awareness (Makino, 1993).

On the other hand, addressing the stimulating and constructive application of a recently–developed technique, i.e., audiotaped feedback (ATF) on writing, Boswood and Dwyer (1996) maintain, "the medium [and the method] that teachers choose for giving students feedback had far-reaching effects on the impact of their comments" (pp. 20-21). Kroll (as cited in Celce-Murcia, 1991) explains that:

> Some teachers provide all their feedback orally by asking students to submit a cassette tape with their draft. This method probably works best when the teacher silently reads a student's paper and makes comments directly into the tape recorder while marking some accompanying numbers or symbols on the student's text. For ESL students, this method has the advantage of providing more extensive feedback than that likely to be made in writing, as well as allowing the student to replay the tape as many times as necessary to understand and benefit from the teacher's comments. Once the teacher has learned to use this technique, it probably takes less time to complete taped remarks about a paper than it would to put them in writing. (p. 259)

According to Hyland (1990), the technique of recorded commentary is particularly helpful when learners' responses to feedback are desired. The importance of ATF has been emphasized by many educators (Hays, 1978;

Clark, 1981; McAlpine, 1989; Patrie, 1989; Hyland, 1990; Boswood & Dwyer, 1996) who stress the effectiveness and productiveness of teachers' taped commentaries.

Minimal marking and audiotaped feedback are only two of the various approaches to feedback. Lynch (1996) points out that language teachers have to employ a battery of styles to provide comments on the students' scripts if higher achievement is desirable. However, in an attempt to make the present investigation more manageable, this study focused on only two methods: MM and ATF.

There has been little research done, at least in Iran, to examine and promote this wide range of writing feedback styles and their impact on EFL learners' writing ability. Such a situation is perhaps due to educational constraints, the local teaching culture, or the nature of language learners. But one can justifiably raise this question: Considering all these, then which feedback styles can be most productive for Iranian EFL learners? This general question was later expressed more specifically in the form of the following research questions.

This study aimed at answering three questions as follows:

Q1: Does the writing performance of the participants receiving ATF on their writing assignments differ significantly on the pre-test and post-test?

Q2: Does the writing performance of the participants receiving MM on their writing assignments differ significantly on the pre-test and post-test?

Q3: Is there any significant difference between the writing performance of the participants receiving ATF and that of those receiving MM on their writing assignments?

Research Hypotheses

In order to investigate the above-mentioned research questions empirically, the following null hypotheses were stated:

H_0 (1): There is no statistically significant difference between the pre-test and post-test mean scores of the participants who receive ATF on their writing assignments.

H_0 (2): There is no statistically significant difference between the pre-test and post-test mean scores of the participants who receive MM on their writing assignments.

H_0 (3): There is no statistically significant difference between the post-test mean scores of the participants who receive ATF and the mean scores of those who receive MM on their writing assignments.

Research Method

Participants

The participants of this study were randomly selected from among early intermediate EFL learners studying at one of the branches of Kish Language Institute in Tehran. The type of sampling employed in this study was cluster sampling. That is, the unit of selection did not involve individuals, but a group of individuals being randomized from larger groups, including all early intermediate EFL learners, studying at all divisions of the institute, to smaller ones, i.e., to boys' and girls' divisions, then to branches, and finally to classes.

The resultant sample comprised seven classes with 136 early-intermediate EFL male learners in total. To ensure that the learners were at the expected language proficiency level and made up a more homogeneous sample, the researchers administered a validated teacher-made achievement test based on Headway Pre-Intermediate, the course book that the participants had all finished successfully about one month earlier. Then the learners whose scores were within one standard deviation above and one standard deviation below the mean on this test were chosen as the participants of this study.

Out of 136 language learners who took the test, 84 learners met the required criterion and were, therefore, considered to be the eligible members of the sample.

Instrumentation

The instruments utilized in the present study included (a) a validated teacher-made test as a means of homogenizing the participants regarding their language proficiency level, and (b) two one-paragraph expository compositions of about 150 words each, one serving as the pre-test and the other as the post-test.

Homogeneity Test: The item format of the homogeneity test was multiple-choice. After the planning, preparing, item writing, and reviewing stages, the prepared test items were pre-tested (a) to determine item characteristics, i. e., item facility, item discrimination, and choice distribution, and (b) to determine test characteristics, i.e., reliability, validity, and practicality.

Originally, 115 four-choice items were included in the test, and the test itself was divided into three sub-tests of Vocabulary, Grammar, and Reading Comprehension, with the initial number of items in each sub-test being 40, 40, and 35, respectively. However, through pilot-testing done with a group of 44 language learners whose characteristics were very similar to those of the target group, but from another branch of the institute, the newly-written test items were revised on the basis of item analysis. As a result, poor items were discarded, and the test was shortened from 115 items to 80 items with the Vocabulary Section containing 20 items, Grammar Section containing 35 items, and Reading Comprehension Section containing 25 items, with a time allocation of 15, 25, and 35 minutes for each sub-test, respectively.

Test of Written English: Throughout this study, all the participants were assigned to write expository paragraphs, each with the length of approximately 150 words, in response to the topics which were selected from the list of topics presented in the TOEFL 2000-2001 Information Bulletin for Computer-Based Testing, published by Educational Testing Service.

On the whole, there were ten topics assigned. Two topics were utilized as the pre-test and the post-test topics with a time limit of 40 minutes, and the remaining eight were the writing topics on which the participants composed their paragraphs and received the relevant feedback in the course of the treatment phase. To control possible factors that might otherwise affect the results of the study and for purposes of comparability, the writing

environment and the procedure for data collection were made uniform for both groups.

The topics on which the participants were required to write in the pre-test and the post-test sessions were, in order, as follows:

Some people think that teenage children should get part-time jobs while they are still at school. Other people believe that school is a full-time job itself and that teenagers shouldn't have part-time jobs while they are at school. Which idea do you agree with? Use specific reasons and examples to support your answer.

Some people think that childhood is the happiest time of life. Other people believe that childhood is not the happiest time of life and that the joy of being an adult is much more than that of childhood. Which idea do you agree with? Use specific reasons and examples to support your answer.

Scoring: Both impressionistic and analytic approaches to marking were applied in this study. This was for the most part due to lack of consensus on one single marking method that claims reducing marker errors contributes to the reliability and validity of the test. As a result, based on the claim that multiple markings improve the reliability of marking English essays (Weir, 1990), each script was subjected to more than one judgment by adding the multiple marks of three independent markers together, using both holistic and analytic methods of marking.

Accordingly, the assessment of the scripts was made by three raters, marking independently. The raters were experienced teachers who were well-accustomed to marking writing assignments. Nevertheless, in a briefing session, the raters were given detailed instructions, as well as sample writings to mark. This session was held to ensure consistent grading among the raters. In the scoring process, there were no indications as to which scripts had been written by the ATF Group and which had not. Not only had the names of the participants been removed, but also the scripts written by the ATF and MM groups were mixed; the only means of identification on each script was a number which also appeared on a separate sheet for recording the student writers' grades.

To assess the writers' scripts in the pre-test and the post-test analytically, the researchers employed the Jacobs' Composition Profile as the marking scheme. This profile is known to be "one of the most widely used analytical scales" (Reid, 1993, p. 235).

Final scores were reached by adding the individual raters' scores awarded first holistically, then, with an interval, analytically, with each of these marks being out of 100, and next by averaging the two sets of the three raters' scores for each script.

The inter-rater reliability for the marking of the scripts through the two multi-rater procedures (i.e., holistic and analytic marking methods) was computed using the average correlations among the three raters in the pre-test and the post-test (see Table 1). The reliabilities calculated for both holistic and analytic marking methods were 0.81.

Table 1

Correlation of the Three Raters' Scores on the Pre-test and Post-test

Raters	Pre-test		Post-test	
	Holistic	Analytic	Holistic	Analytic
R_1R_2	.78	.77	.84	.83
R_2R_3	.83	.82	.83	.85
R_1R_3	.79	.77	.84	.83

Procedure

The type of sampling employed in this study was cluster sampling. The resultant sample comprised seven classes, with a total of 136 early intermediate EFL learners. In the next step, a validated teacher-made test on the basis of Headway Pre-Intermediate was administered to choose those whose scores were within one standard deviation above and below the mean, in the interest of having more homogeneous a sample and ensuring that the participants were at the expected language proficiency level. The result indicated that from 136 students, 84 were eligible enough to be counted as the participants of this study. Once the random assignment of the classes to the ATF and MM groups was done, the number of eligible participants in both groups turned out to be the same (i.e., each comprised 42 subjects).

The design of this study was the pretest-posttest nonequivalent-groups design, one of the quasi-experimental designs. To become confident of the existence of no statistically significant difference between the language proficiency level of the participants in the ATF and MM groups, on the

basis of their homogeneity-test mean scores (see Table 2), a statistical test, i.e., an ANOVA F test was run.

Table 2

MM and ATF Groups' Homogeneity Test Means, Variance, etc.

Group	Mean	Standard Deviation	Variance	N
MM	48.14	5.31	28.20	42
ATF	48.95	5.02	25.21	42

Table 3 displays the result of the F-test comparison

Table 3

F-test Result: Comparing the ATF and MM Groups' Means

F observed	df	F critical
1.12	1 & 82	3.96

The F-observed value was 1.12, which at 1 and 82 degrees of freedom was lower than the critical value of F, i.e., 3.96, at the .05 level of significance. The result of this statistical test proved that there was no statistically significant difference between the ATF and MM groups from the beginning with respect to their language proficiency level.

The present study was conducted over two school terms, i.e., 40 sessions; the first two sessions were taken for paragraph writing instruction, one session for the pre-test, one for the post-test, and the remaining 36 sessions were devoted to the treatment.

Prior to the administration of the pre-test, for two sessions, the participants in both the ATF and MM groups were given detailed identical instruction and practice, all by one teacher, on how to write well-formed paragraphs on expository topics. It was then that the participants in both groups were assigned to write an in-class expository paragraph of about 150 words on a given topic as the pre-test within the examination time limit, i.e., 40 minutes. Soon after the completion of the pre-test, the participants' scripts, 84 scripts, were scored first holistically and then, with an interval, analytically by three raters, marking independently. The reliability of the three raters' scores on the pre-test with two rating procedures, i.e., holistic and analytic marking methods, was 0.81, obtained through averaging the correlations among the three raters in the pre-test (see Table 4).

Table 4

Correlation Matrix for the Scores Given by the Three Raters on the Pre-test

		Rater 1		Rater 2		Rater 3	
		Hol.	Ana.	Hol.	Ana.	Hol.	Ana.
Rater	Hol.	--	.92	.78	.75	.79	.74
1	Ana.		--	.79	.77	.81	.77
Rater	Hol.			--	.94	.83	.83
2	Ana.				--	.81	.82
Rater	Hol.					--	.88
3	Ana.						--

Note. All the coefficients are significant at .000 level of significance.

The treatment procedures were carried out within 36 sessions. In the course of the 36-session interval between the pre-test and the post-test, the participants in both groups wrote paragraphs on eight expository topics out-of-class under no time limit. The only difference between the treatment of the ATF and MM groups was, in fact, in the type of feedback they received on their scripts.

The participants in the ATF Group received audiotaped feedback on their scripts. This procedure required every member of the group to hand in a blank cassette tape, together with his writing. After collecting the scripts and the tapes, the teacher, one of the researchers, read the scripts and gave comments directly into the tape recorder. While giving comments, the teacher made sure to provide the participants with hints to give them the opportunity to repair the parts in which communication breakdown had occurred. For the ease of reference and quick detection of the part commented on, some signs and numbers were put on the participants' drafts. In the next session, the scripts, together with the tapes, were returned to the participants for their reconsideration.

It should be mentioned that the feedback to each erroneous point was given as it appeared in the script, rather than reading the paper all through before recording comments. Emphasizing the effectiveness of this technique, Hyland (1990) justifies it by reasoning that "this is because the writer can see how someone actually responds to their writing as it develops, where the ideas get across, where confusion arises, where logic or structure breaks down" (p. 283).

In tandem with the treatment given to the ATF Group, the participants in the MM Group were receiving minimal marking feedback that involved underlining in red the parts in which communication breakdown had happened, using a series of correcting codes and symbols in the margins or above the error. Needless to say, the members of the MM Group were well familiar with those codes and symbols because they had been provided with a complete list of marking codes after the pre-test.

To ensure that the participants in both groups gave due attention to the feedback given to them, the teacher had the participants go back and reconsider their own scripts, together with the relevant feedback, and return their second drafts only after having done the necessary changes. An added incentive for both groups was that they were told their grades would be averaged and counted as their writing grades in their final scores at the end of the term; no grade was recorded for any assignment until the modified, edited script was submitted. It is noteworthy that the members in both groups reappraised their assignments after receiving the relevant feedback and handed in their second revised copy of their writings before the assignment of a new topic.

At the end of the two-term period, i.e., in the 40th session, when the participants in both groups had already submitted the revised versions of their scripts on the eighth topic, not counting the pre-test topic, they sat for the post-test. The post-test required them to write another in-class expository paragraph of about 150 words on a given topic in 40 minutes. In fact, the participants wrote their paragraphs under a situation identical with the pre-test.

Soon after the completion of the post-test, the 84 scripts were scored first holistically and then, with an interval, analytically by the same three raters, marking independently. The reliability of the three raters' scores on the post-test with two rating procedures was 0.83, obtained through averaging the correlations among the three raters (see Table 5).

Table 5

Correlation Matrix for the Scores Given by the Three Raters on the Post-test

		Rater 1		Rater 2		Rater 3	
		Hol.	Ana.	Hol.	Ana.	Hol.	Ana.
Rater 1	Hol.	--	.90	.84	.81	.84	.82
	Ana.		--	.82	.83	.83	.83
Rater 2	Hol.			--	.88	.83	.83
	Ana.				--	.81	.85
Rater 3	Hol.					--	.87
	Ana.						--

Note. All the coefficients are significant at .000 level of significance.

Testing the Null Hypotheses

In order to investigate the possible differences between the mean scores of the ATF and MM groups on the pre-test and the post-test scored holistically and analytically, a multivariate analysis of variance was carried out. Table 6 shows the results.

Table 6

MANOVA: Pre-test, Post-test by Group

Source of Variation	Sum of Squares	DF	Mean Squares	F observed	P	F critical
Test	2236.72	1	2236.72	61.47	.000	3.96
Group x Tests	1035.05	1	1035.05	28.45	.000	3.96
Within Cells	2983.73	82	36.39	--	--	--

Note. = Data inapplicable, hence, unavailable.

As Table 6 demonstrates, the F-observed value for the effect of the tests, i.e., the pre-test and the post-test, is 61.47. This amount of F at 1 and 82 degrees of freedom is much greater than the critical value of F at .05 level of significance. It can, therefore, be concluded that the difference between the pre-test mean score and the post-test mean score is statistically significant. In other words, the participants performed better on the post-test.

The F-observed value for the effect of the group by tests factor, i.e., the interaction between the two variables, is 28.45 which at 1 and 82 degrees of freedom is much greater than the critical value of F, i.e., 3.96, at .05 level of significance. According to Kinnear and Gray (1994) "if the ANOVA F test gives significance, we know there is a difference somewhere among the means, but that does not justify us in saying that any particular comparison is significant" (p. 98). They add that "further analysis is necessary to localise whatever differences there may be among the individual treatment means" (p. 98). As a result, a Scheffé's test was utilized to pinpoint the precise location of the differences. The results of the post-hoc Scheffé's test

indicated that the participants in the ATF Group performed better than those in the MM Group on the post-test (see Table 7).

Table 7

Scheffé's Test for the Comparison of Pre-test, Post-test by Group

No.	Comparisons	Mean	Observed Difference	Critical Difference	Significance
1	Post-ATF Vs Pre-ATF	71.33 59.07	12.26	6.88	*
2	Post-ATF Vs Pre-MM	71.33 61.69	9.64	6.88	*
3	Post-ATF Vs Post-MM	71.33 64.02	7.31	6.88	*
4	Post-MM Vs Pre-ATF	64.02 59.07	4.95	6.88	No
5	Post-MM Vs Pre-MM	64.02 61.69	2.33	6.88	No
6	Pre-MM Vs Pre-ATF	61.69 59.07	2.64	6.88	No

Note. * denotes significant difference at .05 significance level.

Investigation of Null Hypothesis 1. In accordance with the result of the first comparison, it becomes evident that there is a statistically significant discrepancy between the pre-test and the post-test mean scores of the participants in the ATF Group who received audiotaped feedback on their writing assignments. This finding is not in conformity with the claim of Null Hypothesis 1. Accordingly, this hypothesis is statistically refuted.

Taking into account the mean score of the ATF Group on the pre-test and on the post-test (i.e., 59.07, and 71.33, respectively), one can easily see that the participants have had a much better performance on the post-test. This obvious significant change in the writing performance of the ATF Group from the pre-test to the post-test can, therefore, be attributed to the type of feedback they received on their scripts in the course of this study.

Investigation of Null Hypothesis 2. In light of the fifth comparison of the Scheffé's test, it can be maintained that Null Hypothesis 2 cannot be statistically rejected. In other words, there is no statistically significant difference between the pre-test and the post-test mean scores of the members in the MM Group whose writing assignments were minimally marked. By virtue of this finding, it can be claimed that MM technique does not lead to a higher writing performance, at least as far as the present study is concerned.

Investigation of Null Hypothesis 3. As it is determined by the third comparison of the Scheffé's test, the difference between the post-test mean score of the participants in the ATF Group (i.e., 71.33) and that of the participants in the MM Group (i.e., 64.02) proved to be statistically significant. Therefore, it follows that the ATF Group who received

audiotaped feedback outperformed the MM Group whose writing assignments were minimally marked. This finding is against the suggestion of Null Hypothesis 3 which denied the existence of any discrepancy in the mean scores of the two groups on the post-test; hence, this null hypothesis is also rejected.

Given the presence of no statistically significant difference between the two groups' initial writing performances on the pre-test, as indicated in the sixth comparison of the Scheffé's test (see Table 7), the noticeable superiority of the ATF Group on their post-test writing performance to that of the MM Group could well be as a consequence of the type of feedback the ATF Group received on their scripts. Along the same line, it can be claimed that the inferiority of the M M Group is, in all likelihood, due to the inefficacy of the type of feedback they received.

Conclusion and Implications

This study set out to determine the effect of audiotaped feedback and minimal marking feedback on the writings of Iranian EFL learners. The findings of the study support the contention that ATF can have real impact on the students' ability to write, but MM is not very effective in this regard. The results also echo the earlier findings of Hays (1978) and Kirschner (1991, as cited in Boswood & Dwyer, 1996) who claim that ATF is more productive than MM, while the time it takes to use the ATF technique or to use MM technique is not very different.

Based on the findings of the present study, the researchers arrived at the conclusion that revision should form an essential part of the pedagogy of

writing. As a result of the procedures built into the writing of each task, the participants of this study seemed to have grown convinced that revision is a necessary part of the process of writing. The fact that it helped them see the "learning process in action" (Flower & Hayes, 1981, p. 386) was clear from their expression of the need for revising the second draft in line with the teacher's feedback on the tape. Some of them became, in fact, disappointed because they did not have to write a third draft of their scripts.

Receiving audiotaped feedback seems to have encouraged the participants to write reader-based texts. Besides, building ATF into the instruction of writing may be useful because the participants of this study found ATF encouraging and personal, as well as a refreshing departure from traditional writing feedback methods, often expressed through a symbol system. Cohen and Cavalcanti (1990, as cited in Boswood & Dewyer, 1996) hold that such traditional comments have nearly always had negative impacts on the learners' writing achievement. However, those who received ATF were increasingly enthusiastic about the assignment in each stage and expressed that they learned more from the papers they wrote in the course of this study than any others they had already written.

The results of this study may also be helpful for syllabus designers and those involved in materials development for writing classes to adopt strategies which necessitate teachers treating EFL learners' scripts not as a final product but as a basis for more future learning. As this study showed, this can lead to even more appreciation of the students of the writing process and, as a consequence, to their higher writing achievements.

The observations made throughout the study alerted the researchers to the fact that not only can ATF lead to a higher writing performance, but it also can enhance learners' listening skills. As Boswood and Dwyer (1996) contend, "[students'] investment in their writing is a strong incentive for them to listen, a motivation strengthened by the novelty of the medium itself" (p.22). They further note that "unlike textbook-based ESL listening activities, [via ATF,] the listening task itself is transparent and subordinated to the authentic need to gather information, an excellent opportunity for skills development" (p. 22). Undoubtedly, additional research is needed to either support or challenge such an assumption.

In fact, the concept of feedback on writing deserves wider consideration as a way of increasing learner autonomy and teacher responsiveness to individual needs during the process of learning to write. Since the available literature concerning feedback on EFL writing performance is scanty, interested readers are invited to delve into other aspects of the issue, not dealt with in the present study.

References

Boswood, T., & Dwyer, R. H. (1996). From marking to feedback: Audiotaped responses to student writing. *TESOL Quarterly, 5(2)*, 20-23.

Boughey, C. (1997). Learning to write by writing to learn: A group-work approach. *ELT Journal, 51 (2)*, 126-134.

Celce-Murcia, M. (1991).*Teaching English as a second or foreign language*. New York: Heinle & Heinle Publishers.

Clark, T. D. (1981), Cassette tapes: An answer to the grading problem. *ABCA Bulletin, 44,* 40-41.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32 (4),* 365-387.

Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing.* New York: Longman.

Hays, J. (1978). *Play it again, Sandra: The use of tape cassettes to evaluate student compositions.* Paper presented at the annual meeting of the conference on college composition and communication. (Eric Document No. ED 162332).

Hyland, K. (1990). Providing productive feedback. *ELT Journal, 44 (4),* 279-285.

Kinnear, P. R., & Gray, C. D. (1994). *SPSS for Windows made simple.* London: Lawrence Erlbaum Associates, Publishers.

Lynch, T. (1996). *Communication in the language classroom.* Oxford: Oxford University Press.

Makino, T. (1993). Learner self-correction in EFL written compositions. *EFL Journal, 47 (4),* 337-341.

Mc Alpine, L. (1989). Teacher as reader: Oral feedback on ESL student writing. *TESL Canada Journal, 7,* 62-67.

Muncie, J. (2000). Using written teacher feedback in EFL composition classes. *ELT Journal, 54 (1),* 47-53.

Patrie, J. (1989). The use of the tape recorder in an ESL composition programme. *TESL Canada Journal, 6,* 87-89.

Reid, M. J. (1993). *Teaching ESL writing.* New Jersey: Prentice Hall.

TOEFL. (2000). *TOEFL: Information bulletin for computer-based testing.* New York: Educational Testing Service.

Weir, C. J. (1990). *Communicative language testing.* London: Prentice Hall.

White, R., & Arndt, V. (1991). *Process writing.* Harlow: Longman.