The Effect of Portfolio Assessment on the Development of Metadiscourse Awareness in EFL Learners' Writing in the Academic Context

Mansoor Tavakoli
University of Isfahan
Mr.tavakoli14@gmail.com

Zahra Amirian
University of Isfahan
Amirian_z@yahoo.com

Abstract

This paper aimed at probing the implementation of portfolio assessment in the writing classroom in an attempt to examine its effect(s) on EFL learners' metadiscourse awareness. It addressed the following questions: Does portfolio assessment as a teaching technique have positive wash back effect on the participants' achievement in their writing ability in an academic context? And to what extent do the students develop metadiscourse awareness in their writing by the treatment of portfolio assessment?

The participants were the students of English literature enrolled for their composition course. After being homogenized for their proficiency level, they were randomly divided into an experimental group (EG) and a control group (CG). As the treatment, portfolio assessment was employed as the teaching technique for the experimental group. Data was then subjected to different statistical procedures. The results of data analysis revealed that the participants in the EG outperformed those in the CG with regard to the achievement in their overall writing ability. Second, based on chi-square results, participants in the EG used metadiscourse markers more correctly and efficiently compared with those in the CG. The results of the present study may have some implications for teaching of writing to EFL learners.

Keywords: Achievement, Metadiscourse knowledge, Portfolio assessment, Performance testing, Writing ability

Introduction

In order to assess the effect of teaching on students' learning a second language, different test methods such as paper and pencil tests and performance tests have so far been used. Teachers usually apply such tests to examine the outcome or (product) of learning. But how do they assess the process of learning simultaneously when the students are learning something? This fundamental question led most researchers to call for a new paradigm in language teaching and assessment whereby the learner would be able to learn and assess the process of learning (Teasdale & Leung, 2000; Lynch, 2001; Leung & Lewkowicz, 2006). So, the need for a paradigm shift from "positivism" to "constructivism" attracted the attention of researchers in the field (e.g., Lynch, 2005; Leung & Lewkowicz, 2006).

This movement was accelerated as we entered the "constructivist post-method era" in which assessment is regarded as part of learning since in the process of assessing, the learner tries to learn something (Clapham, 2000; Hancock, 1994; Omalley & Valdez Pierce, 1996). Influenced by this shift of paradigm from testing the outcome to assessing the learning process, language assessors proposed a number of alternatives to testing such as self-assessment and portfolio assessment (Genesee and Upshur, 1996; Hamp-Lyons 1996). According to Harris (1997), self-assessment is a key learning strategy for autonomous language learning, enabling students to monitor their progress and relate learning to individual needs. One of the fundamental elements of self-directed language learning is the opportunity for learners to assess their own progress and thus help them to focus their own learning; therefore, self-assessment is rightly seen as one of the pillars of learner autonomy. Henner-Stanchina and Holec (1985, p. 98) state that through self-assessment, "learners simultaneously create and undergo the evaluation procedure, judging their achievement in relation to themselves against their own personal criteria, in accordance with their own objectives and learning expectations".

Another advantage of self-assessment is providing the learners with opportunities for reflective learning and ongoing formative assessment. Self-assessment can help learners to realize that they have the ultimate responsibility for their own learning. It helps learners to think about what they need to do in order to get better marks. In other words, "By encouraging such individual reflection, self-assessment can begin to make students see their learning in *personal* terms" (Harris, 1997, p. 3). As Chamot and O'Malley point out (1994, p. 119), "self-rating requires the student to exercise a variety of learning strategies and higher order thinking skills that not only provide feedback to the student but also provide direction for future learning".

In fact, because of the potential impact that alternative assessment has on classroom instruction in assessing the process of learning, there is much for it to flourish in the field of language assessment.

Among such alternatives to testing, portfolio assessment was singled out in this study because it might have potential effect on instruction. So, this paper aimed to explore the effect of portfolio assessment on the students' writing ability, especially on their knowledge and use of metadiscourse markers.

Review of the literature

Assessing the academic achievement of every student is a necessary part of class activities, especially writing that presents a challenge for teachers and learners throughout the classroom period. This study focuses on one type of alternative assessment system, portfolio assessment.

The theoretical foundation that underpins this study is the notion of "constructivism" which refers to more recent views on teaching and learning proposing that all individuals learn by constructing information about the world and by using active and dynamic mental processes (Omalley & Valdez Pierce, 1996). This idea led to several alternatives to testing (Omalley & Valdez Pierce, 1996; Brown, 2004; Leung & Lewkowicz, 2006). Following this line of investigation, many researchers have come to recognize that alternative assessment is an important means of gaining a dynamic picture of students' academic and linguistic development. "Alternative assessment refers to procedures and techniques used within the context of instruction which can be easily incorporated into the daily activities of the school or classroom" (Hamayan, 1995, p. 213). In contrast to traditional testing, through alternative assessment, students are evaluated on what they integrate and produce rather than on what they are able to reproduce and recall (Hamp-Lyons, 1996).

Meanwhile, Portfolios have been associated with alternative assessment not only in general education but more particularly in second language education as well (Darling-Hammond, 1994; Hamayan, 1995; Shohamy, 2001). The literature reveals a degree of controversy and confusion concerning the use of portfolio assessment as an alternative to traditional testing. It is argued that portfolio assessment is more than merely one of many homogeneous alternatives in assessment (Brown & Hudson, 1998). As further argued by Lynch and Shaw (2005, p. 264), "…. the portfolio, as an exemplar of alternative assessment, represents a different paradigm or culture that requires an approach to validity evidence (i.e., to establishing the trustworthiness of the

inferences made from the assessment process) differing in certain critical aspects from the approach used in traditional testing."

Portfolio assessment as one of the alternatives to testing is defined as the systematic collection of student work measured against predetermined scoring criteria. These benchmarks may include scoring guides, rubrics, check lists, or rating scales (O'Malley & Valdez Pierce, 1996). Portfolio assessment is a systematic collection of a variety of teacher observations and student products, collected over time, that reflect a student's developmental status and progress (cited in Shabban, 2001, p. 30). In fact, the popular format of authentic assessment, according to Chapman (1990), is portfolio assessment, in which students complete a body of writing over a prolonged period of time. The goal of portfolio assessment is to integrate writing into the teaching of all subject areas.

Genesee and Upshur (1996) provide a plausible reason for the application of portfolio assessment: "A portfolio is purposeful collection of students' work that demonstrates to the students and others their efforts, progress, and achievements in given areas" (p. 99). They maintained that the primary value of portfolios is in the assessment of student achievement because they provide a continuous record of students' language development that can be shared with others.

Moreover, many researchers (e.g., Genesee and Upshur, 1996; Upshur and Turner 1998; Kormos 1999; Papajohn 1999; Lynch, 2001; Khoshsima, 2006 to just mention a few) investigated the effect of portfolio assessment on learning process in classroom context. Their findings depicted plausible impact on the assessment of students' classroom performance either in writing or speaking. For instance, Elahinia (2004) investigated the effect of portfolio assessment on Iranian EFL learners' writing achievement. She found that portfolio assessment had a significant effect on writing performance of the participants. Moreover, she found that there was a correlation between portfolio assessment scores and scores on the final exam writing test. Also, the participants of the study had a positive attitude toward their writing experience (i.e. portfolio assessment).

Marefat (2004) investigated the nature of students' comments, their reactions to teacher's comments and their views on portfolio use in an email based EFL writing class. She found that most of the students limited themselves to comment on their spelling and grammar errors and few of them commented on organization and content. Meanwhile, the majority of the participants found portfolio approach a positive and refreshing opportunity. Also some students developed a personal understanding of their learning process.

In the same way, in order to investigate the wash back effect of alternative techniques on Iranian students' writing ability and reading comprehension, Khoshsima (2006) conducted an experimental study. His findings indicated that using oral and written portfolios affected the participants' writing ability and reading comprehension.

Similarly, Paesani (2006) found that students perceived the value of the writing portfolio project. In their critical commentaries, the students noted that the portfolio project was valuable to the development of their skills in writing as well as their grammatical competence. As Paesani (2006) noted, the themes that emerged from the commentaries reflect some of the main objectives of the course: "a process-oriented approach to writing, the simultaneous learning of skills and content, and the contextualized study of grammar" (p. 626).

In an interesting study, Lam and Lee (2009), while taking a formative function of portfolio to writing assessment, specifically investigated how the formative potential of portfolio assessment can be better applied in the EFL writing classroom. Their findings indicate that although students responded positively to the formative aspects of portfolio assessment, they still preferred summative grading and tended to believe that grades were the best way to inform their current standards of writing.

Following this line of investigations, the current study has taken both the formative and summative functions of portfolio assessment in the classroom so as to clearly portray the writing ability of the students through their real performances during the course and at the end of the instruction. This research specifically focuses on participants' use of "meta-discourse markers". Metadiscourse is "discourse about discourse" (Vande Kopple, 1985) and refers to the author's linguistic manifestation in a text to "bracket the discourse organization and the expressive implications of what is being said" (Schiffrin, 1980, p. 231). Textual metadiscourse refers to terms such as text connectives, code glosses, illocution markers, etc. Interpersonal meta-discourse is pertaining to ideas, including validity and modality markers, attitude markers, and commentaries (Vande Kopple, 1985). Pronouns such as *I, me, us, we, you*, conjunctions such as *indeed, of course, perhaps, unfortunately, seems*, and *phrases* such as *I felt that, it seems to me, I believe, according to my point of view* are some examples of metadiscourse markers. Some "discourse markers" are used to indicate relations between segments of discourse ("*and,*" "*because,*" "*on the other hand*"), interpersonal relations ("*sorry, but,*" "*you know,*" "*as a friend*"), and cognitive attitudes toward what is being said ("*I mean,*" "*in a sense,*" "*certainly*"). "Linguistic action verbs" are used to describe the social

actions performed in discourse ("she *asked*," "don't *threaten* me"), and, in some cases, simultaneously to carry out those actions in "performative utterances" ("I *promise*," "I *tell* you"). "Reported speech" (direct or indirect quotation) purports to represent for some present purpose something that was said previously (Lucy 1993, pp. 18-21).

Various studies have examined the role that discourse markers play in helping readers and writers achieve the communicative function of the texts. For instance, Dafouz-Milne (2008) sought to explore the role that metadiscourse markers play in the construction and attainment of persuasion. Her findings suggested that both textual and interpersonal metadiscourse markers are present in English and Spanish newspaper columns, but that there are variations as to the distribution and composition of such markers, specifically in the case of certain textual categories (i.e. logical markers and code glosses).

Simin and Tavangar (2009) attempted to look at foreign language learners' written products from a pragmatic perspective, focusing on the use of metadiscourse markers. Based on the results, they inferred that the more proficient learners are in a second language, the more they use metadiscourse markers. Also, it appeared that explicit instruction has a positive effect on the correct use of metadiscourse markers.

Following the above-mentioned studies, the present research tries to investigate the effect of portfolio assessment on using textual and interpersonal markers in terms of their frequencies and appropriacy of occurrence in the participants' compositions. By frequency, it means the number of metadiscourse markers correctly used by the participants in either experimental group or comparison group. Appropriacy of metadiscourse means that whether the participants who received portfolio assessment could develop a plausible understanding of that knowledge or awareness in writing their composition.

Purpose and Research Questions

The purpose of this study was to investigate the implementation of portfolio assessment in an academic context in an attempt to examine its effect on the students' writing ability, especially their metadiscourse awareness.

Taking the above purposes into consideration, the present study addressed the following research questions:

1) Does portfolio assessment as a teaching technique have positive wash

back effect on the participants' achievement in their writing ability in an academic context?

2) To what extent do EFL students develop metadiscourse knowledge (awareness) in their writing by the treatment of portfolio assessment?

The tentative statements that can be formulated in the form of null hypotheses about the outcome of the study are that portfolio assessment does not have any significant effect on the participants' overall writing ability nor does it have any impact on improving their metadiscourse knowledge.

Methodology

In order to provide plausible answers to the aforementioned questions, a quasi-experimental study was conducted which will be described and delineated below.

Participants

The population from which the participants were selected included sophomore students of English literature from the University of Isfahan, Faculty of foreign languages who had enrolled for the writing courses. The participants were 86 and were distributed into two classes based on their registration for the term. By applying different elicitation procedures, their writing performances were closely observed during the term. These procedures are described and discussed below.

Instrumentation

At the outset of the term, an Oxford Placement Test (OPT) was administered in order to neutralize the potential effect that the participants' proficiency might have on the outcome of the study. Then, the descriptive statistics of the scores obtained on OPT were computed (See Table 1). The mean and standard deviation were 69.25 and 2.80, respectively. Those participants who scored above 60% were considered as qualified for the experiment, and the data obtained by those below the standard were excluded from the final analysis. After applying this criterion, 60 participants remained for the study who were randomly distributed between the two classes. So, one class was randomly assigned to the experimental group (EG), receiving the treatment (using portfolio assessment as a teaching technique) and the other group (CG) was exposed to the traditional method of teaching and testing writing.

Table 1

OPT Scores

Test	Mean	SD	Variance	N
OPT	69.25	2.80	19.49	60

The experiment

As mentioned in the previous section, 60 students of the English writing course were randomly distributed in two classes of 30. For the control group, the instructor dealt with teaching writing using the traditional techniques such as explicitly talking about paragraph development, the role of connectives in developing ideas and attitudes, use of topic sentences, use of major and minor supporting sentences, and so on. However, for the experimental group, the teacher asked the students to build up portfolios for themselves as a technique for both teaching and assessing their writing progress. These portfolios contained all samples of their works including both classroom assignments and homework as well as self-assessment records. Self-assessment records were the students' self-ratings of their own works based on the criteria defined by the teacher at the beginning of the treatment (See Table 2. below). In each session, students were asked to write about an argumentative topic which is one of the portfolio assessment techniques. Then, the teacher observed their classroom assignments and gave them some guidelines to review their compositions. The students received some assignments and tasks including various topics concerning the population growth, life in dormitory, choice of marriage partners, use of the Internet, importance of academic life, and many other argumentative subjects and were asked to write about them in about 150-200 words for the next session they came to the class. After that, some criteria were given to students to help them judge their own tasks (self-rating criteria, See Table 2). In each session, the instructor spent almost half of the class time to check the students' written works and provided them with useful feedback about their development in using metadiscourse markers. He was concerned with how well they would be able to convey their ideas, thoughts and attitudes by applying their interpersonal knowledge, and what types of organization they developed in their compositions by using textual markers such as connectives. In the next session, the instructor collected the students' assignments in order to

observe how much they progressed in terms of their overall writing ability as well as their metadiscourse knowledge in comparison with previous versions. During the course in the experimental group, each student had ample opportunities receiving feedback from the instructor and being checked four or five times.

Based on the information provided by the participants about their own strengths and weaknesses, a profile analysis as to the classification of their problems was done. That is, the instructor categorized the participants according to the problems they had in writing: those who had problems with main idea and theme; those who had problems with organization ; those who had trouble with conventions; and finally those who had difficulty with metadiscourse awareness.

Then, the instructor scored the students' writing portfolios using a rating system with the analytic criteria borrowed from Chapman (1990). These criteria were used with some modifications as shown in Table 2. below.

Table 2

Rating System

Scores/Criteria	Descriptions
5/Focus	Is the main idea, theme, or point of view clear and consistently maintained?
5/Support/Elaboration	Are arguments and conclusions adequately supported and explained?
5/Organization	Is the logical flow of ideas clear and connected?
5/Conventions	Are standard English conventions (spelling, grammar, punctuation) properly followed?
5/Complexity	How many words, phrases and sentences are embedded in the composition?

In addition to discrete scoring, the assessment of students' portfolios also included a focused, holistic score integration, which reflected how well the students as a whole accomplished the assignment. The instructor used holistic marks such as highly-being developed, well-being developed, fairly-being developed, and not-being developed in order to account for the writing ability of the participants in qualitative terms. Of course, the instructor considered the student's entire portfolios not just single assignments. This rating system emphasizes stages of development, and writing ability is described qualitatively from highly being developed to not being developed. The participants' written portfolios were closely observed by the instructor later in the course to examine whether participants were able to develop metadiscourse knowledge in their writing ability.

For data collection procedure, two samples of writing were taken from the participants in both EG and CG. The first one was taken exactly at the second week of the term, and the second one obtained at the end of the experiment. The first sample is normally called pre-test, and the second one, post-test. For the first administration, the first two argumentative topics (mentioned below) were given to the students in both classrooms, and they were asked to write their ideas about one of them in about 150 to 200 words in the classroom.

The second two topics (mentioned below) were given to the participants in the second administration. The topics given to the participants were attempted to be more communicative so as to elicit authentic information concerning the use of metadiscourse elements. Again, the participants were asked to write their ideas about one of the two topics in about 150 to 200 words in the classroom.

The selected topics are as follows:

- It is high time men ceased to regard women as second-class citizens. What do you think?

- World governments should conduct serious campaigns against smoking. How?

- Parents are too permissive with their children nowadays. What do children think?

- Examinations exert a destructive influence on education. Can you suggest a better way?

Scoring

Based on the analytic benchmarks suggested by Chapman (1990) with some modifications, the two samples of writing composition were scored. The criteria considered in scoring were the presence or absence of clear thesis statements and topic sentences, paragraphing, overall organization, complexity of sentences, and the style of argumentative reasoning, as well as such mechanics of writing as grammar, spelling, and punctuation. The two raters (researchers) used interval scaling for scoring the compositions; they assigned 5 to each criterion such as focus, support/elaboration, organization, conventions, and complexity if fully observed in the participants' compositions. They also considered the quantity and quality of the participants' use of textual and interpersonal metadiscourse in terms of organization of the text and transfer of their ideas and attitudes. T-units were also used to determine the complexity of the compositions in terms of embedded phrases and clauses. In order to avoid the subjectivity of scoring, the two raters followed the same rating system (mentioned in Table 2. above). As shown in Table 3 below, the inter-rater reliability obtained from the first and second samples are 0.82 and 0.89, respectively. The obtained data were submitted to a series of statistical analyses, which will be described and discussed in the following part.

Table 3

Inter-rater Reliability

Raters	Mean	SD	R
1st Sample	37.5	6.48	00.82
2nd Sample	43	5.83	00.89

Results

As mentioned above, this study aimed at examining the effect of portfolio assessment on the participants' achievement in writing in academic context, especially focusing on metadiscourse knowledge (awareness). It should be noted that, portfolio assessment here was used as a technique for teaching writing to EFL students and it was not considered in its broader sense as a

separate system of assessment.

In order to investigate the null hypotheses (mentioned above), a series of statistical procedures such as t-test and chi-square were run. The descriptive statistics were computed in order to provide the average mean scores for both experimental and control groups. The obtained data are described and summarized in table 4 below.

Table 4

Descriptive Statistics: Pre-test and Post-test Concerning CG and EG

Groups	Mean	SD
CG Pre-test	19.5	4.32
CG Post-test	20.10	**3.40**
EG Pre-test	19.7	4.54
EG Post-test	22.53	3.56

The descriptive statistics in Table 4. indicates that there is a difference between the mean scores of the pre-test and the post-test in the control and experimental groups. The participants' performance on the second sample was better than their performance on the first one. Specifically, Table 4. reveals that the participants' writing abilities in the experimental group seem to have improved after being exposed to the treatment in the classroom.

In order to investigate the first null hypothesis, an independent sample t-test was run. The t-observed value for the comparison of the control and experimental groups' mean scores on the test is 2.70. As shown in Table 5, this amount of t-value exceeds the t-critical. It can be claimed that there is a significant difference between the two groups mean scores on the test, so the first null hypothesis is rejected.

Table 5

A Comparison of the Post-test Mean Scores: Control and Experimental *Groups*

Observed t	D.F.	t-Critical
2.70	58	2.01

N = 60; *p = < 0.05

The findings imply that, the experimental group with a mean of 22.53 outperformed the control group whose mean is 20.10. As a result, based on the analyses, it can be said that the application of portfolio assessment as the teaching procedure in writing proved to be significantly effective and helped the participants in their overall writing ability in the experimental group.

As for the second question, which aimed at investigating the effect of portfolio assessment on metadiscourse awareness in the participants' writing, the chi-square test was applied. This statistical procedure was used in order to examine the frequency of metadiscourse markers such as textual and interpersonal markers used correctly and appropriately by the participants. It should be noted again that only those textual and interpersonal markers which were correctly and efficiently used as to contribute to the theme of the writing were counted, and through percentage analysis they would be analyzed.

Through the analysis of a corpus of 120 essays written by the participants, the number of metadiscourse markers, textual and interpersonal, appropriately used by the participants was counted. The obtained data were described according to frequencies and percentages (See Table 6. below). Then, a comparison was made between frequencies and percentages of meta-discourse markers in order to observe if there would be any meaningful difference between the comparison and experimental groups in terms of proper use of them. To do so, two sets of chi-squares were run to find the difference between CG and EG concerning the appropriate use of metadiscourse markers.

Table6

Frequency and percentage of meta-discourse markers: Textual and interpersonal

Groups- Meta-discourse	Total frequency	Percentage
CG Textual	331	46.2
CG Interpersonal	235	45.6
EG Textual	385	53.8
EG Interpersonal	280	54.4

The chi-square observed value for comparing the experimental and control groups' appropriate use of the textual markers is 4.07. As Table 7 below depicts, this amount of chi-square value exceeds the critical value of chi-square, i.e., 3.84. It can be claimed that there is a significant difference between the numbers of textual markers produced by the two groups. As shown in Table 6 above, the experimental group produced 53.8% of correct textual markers while the control group produced 46.2%.

Table 7

Chi-square Textual Markers

Observed chi-square	D.F.	Critical chi-square
4.07	1	3.84

$N = 60$; *$p = 0.05$

The chi-square observed value for the comparison of the experimental and control groups use of the inter-personal markers is 3.93. As observed in Table 8 below, this amount of chi-square value was greater than the critical value of chi-square, i.e. 3.84. It can be claimed that there is a significant difference

between the number of inter-personal markers produced by the two groups in terms of both efficiency and appropriacy. The experimental group produced 280 (54.4%) of inter-personal markers while the control group produced 235 (45.6%).

Table 8

Chi-square Interpersonal Markers

Observed chi-square	D.F.	Critical chi-square
3.93	1	3.84

N = 60; *p = 0.05

Discussion

Concerning the first null hypothesis, the finding confirmed significant difference between experimental and control groups. That is, portfolio assessment affected the participants' overall writing ability to a large extent in the experimental group. By further observing the portfolio assessment, it can be said that it not only provides improved information about students' achievement in writing but also makes a positive influence on teaching and student learning. This explanation confirms the argument by Hancock (1994) and Omalley & Valdez Pierce (1996) that alternative assessment has a useful backwash effect on teaching and learning. The results can also be explicated in the sense raised by Genesee and Upshur (1996). That is, using portfolio assessment in second language classroom can have a very specific focus, such as writing, or broad focus that includes examples of all aspects of language development.

Furthermore, step by step observation of the written portfolios of the students in the experimental group indicates that the holistic ratings they received after the instructor's assessment were highly correlated to their scores obtained on the last writing sample given to them as post-test. This finding implies that participants in the experimental group had a significant achievement in their writing ability through using portfolio assessment.

As shown in Tables 7. and 8. above, the results obtained through running the chi-square tests for comparing the frequency of correctly using metadiscourse

markers in the participants' compositions revealed that the students in EG employed both textual and interpersonal markers more properly than those in CG. Similarly, the participants in experimental group produced 53.8% textual markers and 54.4% interpersonal markers correctly in their essays, which are greater than those used in control group, i.e., 46.2% and 45.6%, respectively. Based on this analysis, it can be concluded that the implementation of portfolio assessment in class affected the participants' writing abilities in experimental group to the extent that they could use metadiscourse markers (textual and interpersonal) more frequently and efficiently than those in the control group. This will be more clarified below when the qualitative analysis of metadiscourse knowledge observed in some randomly selected compositions are discussed.

Moreover, through qualitative analysis of the participants' compositions obtained at the end of the experiment, a number of interesting points concerning the use of metadiscourse markers both textual and interpersonal can be raised. It was found that good compositions, which favored higher scores, included more metadiscourse markers. The higher frequency of using appropriate metadiscourse markers led to higher ratings of these compositions.

Conversely, lower use of these markers in lower-scored compositions indicates that paying less attention to how the audience will perceive the text would lead to poor rating of these texts by the raters.

Of course, there were some other low-scored compositions in the control group in which a lot of metadiscourse markers were used. This overuse made the compositions unnecessarily wordy and difficult to follow. On the other hand, there were some higher-scored compositions in the experimental group with fewer metadiscourse markers. This point would imply that making a balance between the content of the composition and the use of metadiscourse markers is essential. These findings have been supported by other researchers (e.g., Crismore, 1984; Vande Kopple, 1985; Hyland, 1998).

Conclusion

Based on findings of this study and the related discussion, it can be concluded that:

1) The use of portfolio assessment in the classroom in an EFL context affected participants' achievements in their overall writing ability and hence led them to create types of discourse appropriate to academic settings.

2) The metadiscourse awareness of the participants in the experimental group proved to be significantly enhanced compared with the comparison group.

3) The use of alternative assessments (portfolio assessment in case of this study) demonstrated a more useful backwash effect on instruction than the traditional methods such as paper-and-pencil tests.

Two immediate implications are implied by the findings obtained in this study. First, the application of alternative assessment procedures such as portfolio assessment in classroom can be highly beneficial; this is when assessment is integrated with instruction. Second, portfolio assessment is really an authentic assessment and is a productive and useful tool for assessing the students' progress in class performance since the characteristics of the classroom tasks in portfolio assessment correspond to the characteristics of tasks in the target language use (TLU) situation.

In fact, findings of this study would propose an integrative model of assessment for classroom application with performance testing such as writing and speaking or other types of tests. The findings obtained would also support the application of such instruments in our writing and speaking courses in academic context. This can be a significant starting point toward the integration of instruction and assessment.

In the long run, the challenges (whether to assess the students' performance in writing based on alternative assessments on the one hand or to measure their ability of writing using such traditional measures on the other) that language teachers and testers faced would seem to be endless until a compromise is reached. That is, in order to come to sound and unbiased decisions regarding the learners' classroom language behaviors, language testers and assessors should move toward a multi-level system of evaluation that can provide multiple sources of information. This has been, indeed, the concern of most researchers at the turn of the century (e.g., Teasdale & Leung, 2000; Shohamy, 2001; Lynch, 2005; Leung & Lewkowicz, 2006; Lam & Lee, 2009). In other words, educators would require both quantitative information and qualitative description about language ability in order to better understand the meaning of scores obtained by students.

Like many other studies, this one also had some limitations. The first limitation is concerned with the design of the study, i.e., quasi-experimental design, the results of which cannot be generalized. Perhaps a study with a more scientific design may come to more significant results than what were obtained

here. The second limitation is to do with the nature of portfolio assessment used here. As mentioned before, a portfolio is a collection of a student's work, experiences, exhibitions, self-ratings, commentaries, etc. accumulated over time. However, since in this study, portfolio assessment was used as a technique for teaching writing to EFL students, such integral features as self-rating or commentary were neglected. More comprehensive studies can be conducted which take consideration of these significant features of portfolio assessment. The third limitation concerns with the low sample size which is also a threat to the generalizability of the findings. Finally, although alternative assessments such as portfolios can provide teachers and assessors with useful and authentic results, they are time-consuming to be employed for large scale administrations.

References

Brown, D. H. (2004). *Language assessment: Principles and classroom practices*. Pearson education, Inc.

Brown, J. D., & Hudson, T. D. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*, 653–675.

Chamot, A.U. and O'Malley, J.M.. (1994). *The CALL A Handbook: Implementing the Cognitive Language Learning Approach. Reading*,

MA: Addison Wesley.

Chapman, C. (1990). Authentic Writing Assessment. *Practical Assessment, Research and Evaluation*, *2*,7.Retrieved from

http:ericae.net/pare/getvn.asp?v=2&n=7

Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics, 20,* 147-161.

Crismore, A. (1984). The rhetoric of textbooks: Metadiscourse. Curriculum Studies, *16*, 279-296.

Elahinia, H. (2004). *Assessment of writing through portfolios and achievement tests.* Unpublished M.A thesis, Teacher Training University, Iran.

Dafouz-Milne, E. (2008). The pragmatic role of textual and interpersonal Metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse *Journal of Pragmatics*. *40*, 1, 95-113

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review, 64*, 5–30.

Faghih, E. and Rahimpour, S. (2009). Contrastive rhetoric of English and Persian written texts: metadiscourse in applied linguistics research articles. *Rice Working Papers in Linguistics*, *1*, 92-107.

Genesee, F. and Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.

Hamayan, E.V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics*, *15*, 212-226.

Hamp-Lyons, L. (1996). Applying ethical standards to portfolio assessment of writing in English as a second language. In M.Milanovic and N. Saville (eds.) *Performance testing, cognition and assessment*. (151-162). Cambridge: Cambridge University Press.

Hancock, C. R. (1994). Alternative assessment and second language study: What and why?, *The Ohio State University* Retrieved from http://*www.ERIC_Digests.com*

Harris, M. (1997). Self-assessment of language learning in formal

settings. *ELT, 51*, 1. 1-8.

Henner-Stanchina, C. & Holec, H. (1985). Evaluation in an autonomous learning scheme. In P. Riley (ed.). *Discourse and Learning*. London: Longman.

Hyland, K. (1998). Boosting, hedging and negotiation of academic knowledge. *Text*, *18*, 349-82.

Intaraprawat, P., & Steffensen M. S.(1995). The Use of Metadiscourse in Good and Poor ESL Essays. *Journal of Second Language Writing*,*4*, 3,253-272.

Khoshsima, H. (2006). *The Washback Effect of Alternative Assessment Techniques on Students' Reading Comprehension and Writing Ability.*

Unpublished PhD Dissertation. University of Isfahan, Faculty of Foreign Languages.

Kormos, J. (1999). Simulating conversations in oral proficiency assessment: A conversation analysis of role play and non-scripted interviews in language exams. *Language Testing*, *16,* 163-188.

Lam, R. and Lee, I. (2009). Balancing the dual functions of portfolio assessment**.** *ELTJournal, 63,* 4, 1-12.

Leung C. and Lewkowicz. J.(2006). Expanding Horizons and Unresolved Conundrums: Language Testing and Assessment. *TESOL Quarterly*,*40,*1, 211-234.

Lucy, J. A. (Ed.). (1993). *Reflexive language: Reported speech and meatpragmatics.* Cambridge and New York: Cambridge University Press.

Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18,*351-72.

Lynch, B. & Shaw, P. (2005). Portfolios, Power, and Ethics. *TESOL Quarterly,* 39, 263 – 297.

Marefat, F. (2004). Portfolio revisited. *Iranian Journal of Applied Linguistics, 7,* 2, 79-96.

McNamara, T. (1996). *Measuring second language performance.* London: Longman.

O'Malley, J. M., and Valdez Pierce, L. (1996). *Authentic assessment for English language learners*: *Practical approaches for teachers*. New York: Addison-Wesley.

Paesani, K. (2006). Exercises de style: Developing multiple competencies through a writing portfolio. *Foreign Language Annals,39 ,* 4, 618-39.

Papajohn, D. (1999). The effect of topic variation in performance testing: case of the chemistry TEACH test for international teaching assistance. *Language Testing, 16*, 1, 52-81.

Schiffrin, D. (1980). Meta-Talk: Organizational and Evaluative Brackets in Discourse. *Sociological Inquiry: Language and Social Interaction*,*50*, 199-236.

Shaaban, K. (2001). Assessment of young learners. *English Teaching Forum, 39*, 4, 16-22.

Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, *18*, 4, 373–391.

Simin, Sh. and Tavangar, M. (2009). Metadiscourse knowledge and use in Iranian EFL writing. *Asian EFL Journal, 11*, 1, 230-255.

Teasdale A. and Leung, C. (2000). Teacher assessment and psychometric theory: a case of paradigm crossing? *Language Testing, 17*, 2, 163–184.

Vande Kopple, W. J. (1985). Some exploratory discourse on Meta-discourse. *College Communication and Composition*, *36,* 82-93

Valette, R. (1994). Teaching, testing and assessment: Conceptualizing the relationship. In C. Hancock (Ed.),*Teaching*, *testing and assessment*: *Making the connection*(pp.1-42).Lincolnwood,IL: National Textbook Company.

..