



Effect of Holistic vs. Analytic Assessment on Improving Iranian Intermediate EFL Learners' Writing Skill

Massumeh Hosseini¹, Bahram Mowlaie^{2*}

¹ Faculty of Persian Literature and Foreign Languages, Islamic Azad University, South Tehran Branch, Tehran, Iran

² Faculty of Persian Literature and Foreign Languages, Islamic Azad University, South Tehran Branch, Tehran, Iran

Received: 3 July, 2016

Accepted: 9 August, 2016

Abstract

In assessing foreign language writing, holistic and analytic scoring can be used to measure a variety of discourse and linguistic features. This study aimed to investigate the possible significant effect of analytic and holistic assessments on improving writing skill among Iranian EFL learners. For this purpose, two groups of intermediate EFL learners, after being homogenized, were divided into two experimental groups. In treatment phase, groups A's compositions were scored holistically while group B's compositions were scored analytically using Paulus's scoring rubric as the benchmark. The result of Paired t-tests revealed that both scoring methods caused statistically significant differences between pre- and post-test in both groups. However, the result of the independent samples t-test in post-test between the two experimental groups showed that whether the writings of both groups were scored analytically or holistically, the group, which received analytic scoring during the treatment outperformed the group, which received holistic scoring. This study could have pedagogical implications in that it could encourage writing teachers to take the findings into account to improve the quality of writing as the Cinderella skill in a foreign language instruction and avoid sweeping the writing assignments under the practicality carpet.

Keywords: Assessment, Analytic scoring, Composition, Holistic scoring, Writing skill

INTRODUCTION

With so much emphasis on writing in foreign language classes, it seems justifiable to make an effort to investigate the effectiveness of different methods for teaching this skill and different approaches to measure students' progress towards mastering it. The increasing significance of writing assessment lies in the point that it does not only evaluate students' writing samples, but it

rather gives us a picture of how the whole system has functioned and hence can help us to improve the deficiencies (Huot, 2003). Even in the first language learning, for many students writing proves to be daunting, but as Weigle (2002) points out, the whole process becomes even more challenging in EFL context where the basic linguistic elements, which are taken for granted by native speakers prove to be nothing less than a herculean task for most if not all learners who

*Corresponding Author's Email:
bahrammoelai@yahoo.com



must grapple with bits and pieces of what they have acquired to convey even their basic idea.

Due to the challenging nature of writing in foreign language, the effective teachers in EFL settings must try to find appropriate methods not only in teaching, but also in assessing it to improve EFL learners' proficiency in writing skill. This point can be dealt with more effectively if teachers are familiar with merits and demerits of different methods to correct their students' writing. To deal with the constructive notion of different scoring methods the next part elaborate on two main scoring methods of writing; analytic and holistic scoring.

Analytic scoring

This is a method of scoring a student's composition, which depends on a marking scheme that has been carefully drawn up by an examiner or a body of examiners. The procedure consists of an attempt of an attempt to separate the various features of a composition for scoring purposes, for example, content, organization, vocabulary, grammar, and mechanics. The use of separately delineated scales may guard against the possibility that raters will collapse categories during the rating process, since they must produce a separate rating for each scale (Astika, 1993). In this method of marking, each separate criterion in the mark scheme is awarded a separate mark and the final mark is the composition of these individual estimates.

In addition to showing students how their particular grades were determined, there are other advantages to analytical scoring as follows (as in Bacha, 2001; Heaton, 1975; Xi, 2007).

Because an analytical scale focuses on graders' scoring, the procedure is supposed to enhance the interrater reliability to permit a reliable score to be derived from summed multiple ratings. The next point is that it reveals the problem of uneven development of sub-skills, which might be conspicuous in EFL context in which writing assignment is swept under the carpet due to lack of time. The third point is that raters are compelled to consider aspects of performance, which they might other-

wise ignore. The fourth point is that training of raters is easier when there is an explicit set of analytic scales. This method of scoring also allows for more exact diagnostics reporting especially where skills may be developing at different rates. This can play an important role in remedial courses in focusing instructors or programme developers to concentrate on the weak points. Besides the points mentioned, analytical scoring can be used for correlation research, prediction, and programme evaluation. It can also serve as helpful guides to formative evaluation which is used to determine the degree of mastery of a given learning task and to pinpoint the degree of task not mastered.

Despite the positive points mentioned for analytic scoring, it has some disadvantages too. Bacha (2001), Heaton (1975), and Xi (2007) consider the following points as the short coming of analytical scoring method:

The first point is that it is time-consuming, so it is vulnerable to the same threats to reliability as holistic scoring and questionable validity. The next point is that this system of scoring isolates features from context, and therefore, lacks sensitivity to variations in purpose, speaker role, and conception of audience. Furthermore, the focus on specified aspects of the performance may divert raters' attention from its overall effect. The other problem is that writing is more than the sum of its parts. The last point refers to individual scales that may call for qualitative judgments that are difficult to make.

Holistic scoring

A method of scoring students' composition in which one or more evaluators read the paper without marking anything, and then rate the paper as a whole, assigning single scores based on total impression of composition as a whole text or discourse. The holistic scoring method is based on the theory that a whole piece of writing is greater than sum of its parts (Bacha, 2001).

The holistic method takes a positive approach to essay writing by asking the reader to concentrate on what the writer has done well, given

the testing condition under which the essay was written. According to Bacha (2001), Halleck (1995), and Jafarpur (1991), the following advantages can be considered for holistic scoring:

Since holistic scoring requires a response to the writing as a whole, test takers do not run the risk of being assessed solely on the basis of one lesser aspect (e.g., grammatical ability). Holistic scoring makes the direct testing of writing practical. A single global rating also tends to be more reliable than one from one from a rating scales consisting of several subclasses. The next positive feature of holistic scoring is that the approach generally puts the emphasis on what is done well and not on deficiencies. The approach allows teachers to explicitly assign extra or exclusive weight to certain assessment criteria. In addition to what mentioned, the greatest advantage of this method is its efficiency; in large scale writing, for example, raters can score substantial numbers of papers in a relatively short period of time.

Bacha (2001), Halleck (1995), and Jafarpur (1991) also consider the following points as disadvantages of holistic scoring method: In scoring holistically, the grader reads the composition, forms a general impression, and assigns a mark to that composition based on some standard. That standard may either be a model composition to which the reader has reference, or the general impression the reader has, based on experience in reading students' compositions. Such evaluation can therefore be highly subjective due to bias, fatigue, internal lack of consistency, previous knowledge if the student, and /or shifting standard from one paper to the next. Other disadvantages of the holistic scoring are the following: One score does not provide diagnostic information. It is also difficult to interpret the meaning of a composite score to the raters and to the score users. The other problem with holistic scoring is that the approach lumps together in one score what for a given test taker may constitute uneven abilities across sub-skills. Still the other problem is raters may overlook one or more aspects of writing performance. The next weak point is that

the approach penalizes efforts at development, since writers may display only novice ability with more complex forms, while those using simpler forms get higher ratings. Similar to the previous point, longer essays may get higher ratings.

Reducing a score to one figure also tends to reduce reliability. The last point is that effort at ensuring reliability may be at the expense of validity.

Based on the points mentioned above, the following questions were raised in the present study:

1. Does holistic scoring have any significant effect on improving the writing skill?
2. Does analytic scoring have any significant effect on improving the writing skill?
3. Which scoring method (i.e. Analytic vs. Holistic) produces more gains on writing skill?

METHODS

Participants

In the first stage of this study, 40 female learners aged 17- 24 were selected from among 60 intermediate EFL learners of Guyesh Language Institute, which is located in the city of Qom. Nelson Test (1976) was used to ensure the homogeneity of the participants in terms of English language proficiency. In the next stage, they were divided in two experimental groups, that is, each experimental group consisted of one class of 20 participants run by the researcher as the instructor. They attended English classes one and a half hours a day, three days a week for 12 sessions. They studied Summit A by Saslow and Ascher (2007). participants also attended all sessions of the class.

Instruments

The following instruments were used in this study. The Nelson test (1976) was used to homogenize the participants. It includes fifty multiple-choice questions. The PET test (writing section) was used as pre-test and post-test. As the pre-test, the students were asked to write a composition on the topic of the life in day of someone they knew in 100 words. The

post-test was an essay about 100 words on the topic of what they would do on their holiday in another country.

A pilot study was performed with 30 intermediate EFL learners with similar language proficiency status to the main participants of the study. Using KR-21 reliability statistics, the piloting results showed that the reliability of Nelson test was assessed 0.9; the reliability of PET writing pre-test and post-test was estimated 0.82 and 0.84 using inter-rater reliability method.

The next instrument used for analytic scoring of the essays, was a scoring rubric devised by (Paulus, 1999). It consisted of six categories of organization, development, cohesion, structure, vocabulary, and mechanics. In this scale, organization referred to the effectiveness of the thesis statement and unity of ideas, development referred to the appropriate use of examples and support, cohesion referred to the relationship of ideas to each other and the use of transitions, structure referred to syntax complexity and grammatical accuracy, vocabulary referred to clarity of meaning and the precision of the words used, and mechanics referred to spelling, punctuation, capitalization, and general formatting. The language proficiency related to each category was described in detail in ten points moving from minimum proficiency to the native like. Since the participants in the present study were in intermediate level, the related proficiency level was focused by the raters.

Design

Due to the restrictions associated with research in social studies in general and EFL in particular, the design of the preset study was quasi-experimental. It consisted of pre-test followed by the experiment followed by post-test. It involved pretest / posttest comparison of two experimental groups going through two distinct scoring procedures to see if the related scoring procedures had any statistically significant effect on the writing quality. It also involves post-test comparison of the two experimental groups to find out the possible relative superiority of one of the scoring

procedure as the independent variable on writing quality as dependent variable.

Procedure

The purpose of this study was to determine whether the assessment method (Analytic vs. holistic) had any significant effect on improving writing skill among Iranian Intermediate EFL learners. To this end, first, the Nelson test was administered to decide on general language proficiency of the participants of the two groups and to homogenize them. In the next stage, the participants were randomly assigned to two experimental groups to assure that there was no prior difference between the participants in the two experimental groups in their language proficiency. In the pre-test phase, the participants were asked to write a 100-word essay on a topic of interest which had been chosen from among ten topics. Next, 12 sessions of instruction for the two experimental groups began as follow. In each session, the participants wrote an article on a pre-selected topic of consensus, the topic and the moment of time as possible intervening variables were kept the same for the two experimental groups. After submission of their article, experimental group A's articles were scored holistically by two raters, but experimental group B's were scored analytically by two raters, using the Paulus (1997) scoring procedure. Prior to using this scoring rubrics for analytic scoring, the two raters who were experienced language teachers, scored up to five essays and points of disagreement were negotiated with them until a partial agreement was achieved. After scoring, the participants' articles were returned to them and they were asked to revise their essays based on the scores they had received either holistically or analytically. In holistic scoring this comprised of a single score but in analytic scoring each of the six categories mentioned in Paulus scoring rubric was assigned a separate scores so the participants had a better idea which part to focus more based on the feedback they had received. The important point to mention at this point is that the scores in each of the six categories might not have been a genuine

score of the related part. However, since different sections of each essays was scored separately, it was assumed that it could sensitize the participants more than the single holistic scores they received in group A.

RESULTS

Sixty intermediate EFL students took Nelson

Test to be homogenized. The descriptive statistics of the participants' scores on Nelson Test are presented in Table 1. As it can be seen in Table 2, the mean, median and mode of the Nelson scores were 24.73, 24.50, and 22 respectively. These central parameters are not very far away from each other implying that the scores are equally distributed around the mean, 24.73 obtained on Nelson test.

Table 1.

Descriptive Statistics for Nelson Homogenizing Test

N	Range	Mean	Median	Mode	SD	Skewness	Kurtosis
60	23	24.73	24.50	22	5.41	0.147	-0.670

Those 41 students whose score were one standard deviation, 5.41 above and below the mean, 24.73 (scores between 19 and 30) were chosen as homogeneous intermediate participants for this study. The normality of Nelson scores is approved since the amount of skewness, 0.147 falls within the range of +/- 1.96.

Before presenting the results of paired samples t-test that was used to see whether holistic scoring has any significant effect on improving the writing skill, the descriptive statistics are represented in Table 2. The mean of post-test ($\bar{x} = 23.92$, $SD = 1.68$) is larger than the mean on the pre-test of writing ($\bar{x} = 22.62$, $SD = 1.46$) because of the holistic scoring method. It must be noticed here that the average of the two raters in

each group was used to answer the research question of the present study.

Table 2.

Descriptive Statistics for Experimental Group A

Time	Mean	N	Std. Deviation
Pre-test	22.625	20	1.467
Post-test	23.925	20	1.688

The pre-test and post-test of writing scores in experimental group A was compared using paired samples t-test. The researchers applied paired t-test as the two sets of scores met assumptions of Parametric Analysis (interval data, independence of subjects, normality and homogeneity of variances). Table 3 shows the results of this t-test.

Table 3.

Paired Samples T-test to Compare Means on the Pre-test and Post-test of Writing in the Experimental Group A

Mean	SD	T	Df	Sig. (2-tailed)	95% Confidence Interval of the Difference	
					Lower	Upper
1.300	1.380	4.212	19	.000	.654	1.946

According to Table 3, paired t-test found a statistically significant increase in writing scores of treatment group from pre-test to post-test, $t(19) = 4.212$, $p = .000$, $p < .05$ (two-tailed), in which the t-observed, 4.21 was more than the t-critical, 2.09, and also the p value, .000 was less than .05. The mean increase in writing scores is 1.30 with a .95% confidence interval ranging from .654 to 1.946. Hence we can claim that the

holistic scoring influences the writing skill.

Therefore, the related null hypothesis stating lack of any significant effect of holistic scoring on improving writing is rejected.

In order to investigate if analytic scoring has any significant effect on improving the writing skill, paired t-test was used. Before presenting the results of this analysis, the descriptive statistics are represented in Table 4. Table 5 also shows

that the mean on the post-test of writing ($M = 25.42$, $SD = 1.59$) is greater than their mean on the pre-test of writing ($M = 22.78$, $SD = 1.99$) because of the analytic scoring method.

The results of paired t -test that was used to compare the pre-test and post-test of writing scores in experimental group B appear in Table 5.

Table 5.

Paired Samples T-test to Compare Means on the Pre-test and Post-test of Writing in the Experimental Group B

Mean	SD	T	df	Sig. (2-tailed)	95% Confidence Interval of the Difference	
					Lower	Upper
2.642	2.080	5.821	21	.000	1.695	3.589

According to Table 5, paired samples t -test found a statistically significant increase in writing scores of treatment group from pre-test to post-test, $t(21) = 5.82$, $p = .000$, $p < .05$ (two-tailed), in which the t -observed, 5.82 exceeds the t -critical, 2.08. The mean increase in writing scores is 2.64 with a .95% confidence interval ranging from 1.69 to 3.58. Therefore we can conclude that the analytical scoring improves the writing skill. By the same token, the related null hypothesis stating lack of any significant effect of analytic scoring on improving writing is rejected.

Additionally, in order to find out which scoring method (i.e. Analytic vs. Holistic) produces better gains on writing skill, independent samples t -tests were used. The main reason for running two independent samples t -tests was that once both group A and group B's posttests were scored holistically and the second time both were scored analytically to have a comprehensive and unbiased

Table 7.

Independent Samples T-test to Compare the Two Group' Writing Scores (Scored Holistically)

	Levene's Test for Variances		T-test for Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Diff.
Equal variance assumed	.006	.940	-3.729	39	.001	-2.051

Independent t -test (Table 8) detected a statistically significant difference in writing scores between the two experimental groups A and B scored holistically, $t(39) = 3.729$, $p < .05$, in which the t value, 3.729 exceeded the t critical, 2.02, and also the p value, .001 was below the selected significant level

Table 4.

Descriptive Statistics for Experimental Group B

Time	Mean	N	Std. Deviation
Pre-test	22.786	21	1.991
Post-test	25.429	21	1.599

result, not favoring either groups. Before discussing the results of independent samples t -test, the descriptive statistics are shown in Table 6. Based on Table 6, the mean writing for experimental group B ($\bar{x} = 25.97$, $SD = 1.82$) is greater than experimental group A ($\bar{x} = 23.93$, $SD = 1.68$).

Table 6.

Descriptive Statistics for the Two Group' Writing Scores (Scored Holistically)

Group	N	Mean	Std. Deviation
Experimental A	20	23.925	1.688
Experimental B	21	25.976	1.826

Independent samples t -test was run to compare two groups' writing mean scores, and the related results are laid out in Table 7. As Table 7 shows the two sets of scores had equal variances because the Sig. was .94 in Levene's Test, which is more than .05.

for this study, .05; as a result, it was concluded that the analytic scoring method produces better gains than holistic one on writing skill. In fact the two groups scored quite differently on the final test of writing.

The second independent samples t -test was run

to compare the mean scores of group A and B scored analytically this time. The descriptive statistics are given in Table 8. Based on Table 8, the mean scores of experimental group B (analytical) ($M = 25.42$, $SD = 1.59$) is more than experimental group A (Holistic) ($M = 23.35$, $SD = 1.55$).

The results of independent t-test conducted to compare two groups' mean scores, scored analytically, are represented in Table 9. Table 9 indicates that the hypothesis of equal variances was met

because the Sig., .74 in Levene's Test was more than .05.

Table 8.
Descriptive Statistics for the Two Groups' Writing Scores (Scored Analytically)

Group	N	Mean	Std. Deviation
Experimental A	20	23.350	1.556
Experimental B	21	25.429	1.599

Table 9.

Independent Samples T-test to Compare the Two Groups' Writing Scores (Scored Analytically)

	Levene's Test for Variances		T-test for Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Diff.
Equal variance assumed	.104	.749	-4.214	39	.000	-2.078

The results of the independent t-test showed that there was a statistically significant difference in writing scores between the two experimental groups A and B scored analytically, $t(39) = 4.21$, $p < .01$, in which the t value, 4.21 was above the t critical, 2.02. Therefore, it was found out that the analytic scoring method produces better gains than holistic one on writing skill when the result of both groups are scored analytically. In fact the two groups scored quite differently on the final test of writing.

DISCUSSION AND CONCLUSIONS

The findings of the present study lend support to (Rezaei & Lovorn, 2010) finding in that rubrics are highly helpful and effective for essay scoring. The results also agree with (Bitchener, Young, & Cameron, 2005); (Chandler, 2003) and (Liu, 2008) who have shown that teacher feedback on students' grammatical and lexical errors results in a significant improvement in both accuracy and fluency in subsequent writing of the same type over the same semester, disproving (Truscott, 2004) claim on the negative effect of error correction on fluency. It is also contrary to Soleimani and Rahmanian's (2014) finding that teacher feedback was less effective than peer feedback and self-assessment.

This might be because of one of the inherent

features of analytic scoring, which is clarity in providing feedback to the students and focusing their attention on the points of weakness. This is the very feature that according to Weigel (2002) can enable writing teachers to diagnose their students' weak points. In this manner the learners might have a better and clear idea of what to do to improve their writing. This might be one of the contributing factors to the point that most of the students affectively prefer their teachers' feedback to their peers' not only because of the position the teacher holds (Fathman & Whalley, 1990) but because they view them of immense value in improving their writing proficiency (Fujieda, 2007). Paulus (1999) also found that her EFL students prioritize and incorporate more teacher feedback in their revisions. She suggested that students have more confidence in teacher comments because the teacher is considered more experienced and more authoritative. Other reasons for the popularity of teachers' comments can be that their comments might be considered to be of better quality and more specific. They might be considered to be able to explain what the problems are, and make concrete suggestions for revision (Tsui & Ng, 2000). Each of these features is bound with the main features of analytic scoring mentioned above. Specifically, the clarity of what teachers want their students to do becomes possi-

ble by means of analytic scoring.

What mentioned above, might be a reason why Brown and Hudson (2002) believe that analytic scoring can be used in achievement tests to provide both students and teachers with the information on the extent students have achieved or failed to achieve a certain points in a course of instruction, thereby giving teacher a better view about the remedial courses or instructions about the points of interest. Therefore, as also pointed out by Bacha (2001), Heaton (1975), and Xi (2007) perspective analytic scoring can act as a diagnostics means for both students and teachers or even more broadly to material developers to consider the feedback from students and teachers as a very good source of points in their needs analyses. This is in opposition to holistic scoring in that because of its impressionistic nature; it cannot provide learners with a clear picture of their weak points (Weigle, 2002). Not only does holistic scoring seems to lack in providing viable corrective feedback to learners to better the quality of their writing, according to Cumming and Riazi (2000) because it unites multiple features of students' essays into one single score; it does not contribute much to the learning research.

An important point to consider here is that, emphasizing that analytic scoring must be preferred to holistic scoring in any situations is far from reality because of the related problems or drawbacks associated with analytic scoring. The most obvious challenge with analytic scoring is its being time consuming, which can be against the practicality issue, which can potentially be one of the important obstacles on implementing analytic scoring in crowded classes. Another related point is that having crystal clear bands with clear description for each one is easier said than done because its different levels might be confused quite easily even in the case of experienced or qualified raters. Perhaps due to its this inherent feature of analytic scoring (as stated by Finson & Ormsbee, 1998; Moskal, 2000) when there is an overlap between the criteria set for the evaluation of the different factors, a holistic scoring rubric may be preferable to an analytic scoring

rubric. This is what we can see in reality in most if not all of the writing classes where on the ground of practicality, teachers resort to holistic scoring rather than analytic one despite being aware of its advantages such as increasing teachers' uniformity in giving feedback and objective fairness in the evaluation process (Johnson & Hamp-Lyons, 1995). In fact, in spite of all these advantages associated with analytic scoring, practicality as stressed by Weigle (2001), analytic Scoring can tip the balance in the favor of holistic scoring in most of the writing classes. This is the reason why Weigle (2002) stated that holistic scoring may often be easier to be constructed and more efficient. Therefore the teachers can save time in their assessment procedures.

More Effectiveness of analytic scoring versus holistic scoring can be viewed from theoretical perspective too. Feedback from teacher to learner can fulfill the necessary conditions we can see in ZPD (Vygotsky, 1934) and scaffolding (Spada & Lightbown, 2006) better and more efficiently because in a gradual manner, it can lead learners to modify their writing by means of a clear and relevant feedback they receive on different parts of their writing from a more competent person, i.e. their teacher. Based on their progressive level, the number and amount of feedback can be decreased so that the learners can gradually achieve the required and desired autonomy. Holistic scoring seems to lack that gradual assistance emphasized in ZPD and scaffolding because it has to do with a single score.

Another theoretical concept, which can explain superiority of analytic over holistic scoring is Schmidt's (1990) noticing. Because of its inherent nature of analyzing the text, analytic scoring can help learners notice the gap between what they have produced and the desired level indicated by their teachers' feedback a point which is confirmed by (Gholami & Narimani, 2012) in their study on the effect of feedback on consciousness raising. In fact, noticing is a cognitive process that involves attending to linguistic form in the input learners receive and the output they produce. Therefore, in line with Schmidt and

based on the results of the present study, it can be argued that noticing their mistakes in different levels analyzed and indicated through analytic scoring is necessarily a conscious process and is prerequisite for learning to take place which manifested itself in the form of learners' modified and improved essays in the present study. Receiving and processing inputs, here analytic feedback, with a deliberate effort of paying attention to the linguistic forms of the inputs results in better learning than receiving inputs without making any efforts in noticing linguistic forms of the inputs, which inevitably happens in holistic scoring (Schmidt, 1990).

In line with drawing learners' attention to their error in analytic scoring, Ellis (1998) pointed out that negative feedback (error correction) helps students to notice the gap between their incorrect production and the correct version in the target language. Qi and Lapkin (2001), in a study investigating error feedback, as seen in analytic scoring, and the noticing of forms, have found that when noticing is combined with understanding, the impact on learning in L2 writing is greater. This is in line with the findings of the present study. It can be postulated that one of the reasons why analytic scoring group outperformed holistic one might have been the point that the feedback they received was more understandable for them so they knew what they were expected to do about them to improve the quality of their writing. As Qi and Lapkin (2001) asserted that learners' greater success, especially in analytic scoring group, might be attributed to greater level of metacognitive processing which in turn has resulted in the greater level of understanding. Therefore, noticing can act as the first stage in understanding the meaning of teachers' feedback in analytic scoring as language input. Therefore, it can be hypothesized that input processing was what learners in analytic group needed and they were provided with by the teacher in order to connect their writing level to the desired level and move towards gradual modification of their texts. This was achieved through processing instruction in the form of corrective-analytic feed-

back which can be achieved by explicit instruction, as was the case with the feedback, in this study (Ellis, 1998).

The learners' lack of achievement in holistic scoring in the present study could be attributed to the point stated by Chastain (1987) and Hadley and Reiken (1993) that without feedback, learners will not be able to monitor their progress adequately. Kroll (2001) further claimed that without feedback opportunities in a writing course, there are little reasons for students to endeavor for betterment. In the same line Kaplan (2010) stated that the way readers respond to a writer's text probably has more influence on a writer's motivation and progress than any other single feature of writing instruction. Therefore, teachers' feedback, by means of analytic scoring, can encourage students to write for specific audience and monitor different aspects of their writing with that audience in mind.

References

- Astika, G. G. (1993). Analytical assessments of foreign students' writing. *RELC Journal*, 24(1), 61-70.
- Bacha, N. (2001). Writing evaluation: what can analytic versus holistic essay scoring tell us? *System*, 29(3), 371-383.
- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of second language writing*, 14(3), 191-205.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*: Cambridge University Press.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of second language writing*, 12(3), 267-296.
- Ellis, R. (1998). Teaching and research: Options in grammar teaching. *TESOL Quarterly*, 39-60.
- Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form

- versus content. *Second language writing: Research insights for the classroom*, 9, 178-190.
- Finson, K. D., & Ormsbee, C. K. (1998). Rubrics and their use in inclusive science. *Intervention in school and clinic*, 34(2), 79-88.
- Fujieda, Y. (2007). Perceptual change toward peer response: How writers incorporated feedback into revisions. *Kyoai Gakuen Journal*, 7, 139-153.
- Gholami, J., & Narimani, E. (2012). Consciousness-Raising Through Written Corrective Feedback: The Case of Marked Third Person-s. *Research in Applied Linguistics*, 3(2), 49-66.
- Hadley, A. O., & Reiken, E. (1993). *Teaching Language in Context, and Teaching Language in Context--Workbook*: ERIC.
- Halleck, G. B. (1995). Assessing oral proficiency: A comparison of holistic and objective measures. *The Modern Language Journal*, 79(2), 223-234.
- Heaton, J. B. (1975). *Writing English language tests: a practical guide for teachers of English as a second or foreign language*: Longman Publishing Group.
- Huot, B. (2003). *Rearticulating writing assessment for teaching and learning*: University Press of Colorado.
- Jafarpur, A. (1991). Can naive EFL learners estimate their own proficiency? *Evaluation & Research in Education*, 5(3), 145-157.
- Johnson, D. M., & Hamp-Lyons, L. (1995). Research on the Rating Process: Rating Nonnative Writing: The Trouble with Holistic Scoring. *Tesol Quarterly*, 29(4), 759-762.
- Kaplan, R. B. (2010). *The Oxford handbook of applied linguistics*: Oxford University Press.
- Kroll, B. (2001). Considerations for teaching an ESL/EFL writing course. *Teaching English as a second or foreign language*, 3, 219-232.
- Liu, Y. (2008). The effects of error feedback in second language writing. *Arizona working papers in SLA & Teaching*, 15(1), 65-79.
- Moskal, B. (2000). Assessment Resource Page.[Online] 13 de abril de 2001< <http://www.mines.edu/Academic/assess.Resource.htm>.
- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of second language writing*, 8(3), 265-289.
- Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of second language writing*, 10(4), 277-303.
- Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing writing*, 15(1), 18-39.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 206-226.
- Soleimani, H. & Rahmanian, M. (2012). Self-, peer-, and teacher-assessments in writing improvement: A study of complexity, accuracy, and fluency. *RALs*, 5(2), 128-148.
- Spada, N., & Lightbown, P. M. (2006). *How languages are learned*: Oxford University Press.
- Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of second language writing*, 13(4), 337-343.
- Tsui, A. B., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of second language writing*, 9(2), 147-170.
- Vygotsky, L. (1934). 1986. Thought and language. *Trans. A. Kozulin*. Cambridge, MA: Harvard University Press.
- Weigle, S. C. (2002). *Assessing writing*. New York: Cambridge University Press.
- Xi, X. (2007). Evaluating analytic scoring for the TOEFL® Academic Speaking Test (TAST) for operational use. *Language Testing*, 24(2), 251-286.

Biodata

Massumeh Hosseini received her bachelor's and master's degree in the field of Teaching English as a Foreign Language (TEFL) from Islamic Azad University, South Tehran branch. She has been teaching English in different institutes at different levels of proficiency. Her main areas of research interest are in relation to testing and assessment in language teaching.

Email: masihosseini11@yahoo.com

Bahram Mowlaie is an assistant professor at Islamic Azad University, South Tehran Branch. He has published articles on testing, writing, teacher attitude, and morphology at different journals. His topics of research interest are mainly testing and vocabulary.

Email: bahrammowlai@yahoo.com