

Self- and Teacher-Assessment in an EFL Writing Class¹

Autoevaluación y Evaluación Docente en una Clase
de Escritura de Inglés como Lengua Extranjera

Sasan Baleghizadeh and Tahereh Hajizadeh^{2*}

*Shahid Beheshti University, G.C.,
Allameh Tabataba'i University, Iran*

Abstract

The present study investigated how fifteen Iranian EFL learners developed the ability to self-assess their writings through having access to the rater's scores. The participants were supervised for four weeks as they went through their first experience in self-assessment. They were provided with a detailed evaluation sheet for assessing their work, and after each self-evaluation they were able to have access to the teacher-assigned scores. The results indicated a high correlation between self-assessment and teacher-assessment. It was revealed that students' self-assessment throughout the study turned out to be highly correlated with the teacher-assessment. It was also shown that the learners assessed different components of their writing in a manner comparable to that of the teacher. The findings confirmed that self-assessment could not only be viewed as a useful tool for evaluating learners' performance but also be regarded as an efficient instrument for developing their writing skill.

Keywords: self-assessment, teacher assessment, writing

Resumen

Este estudio investigó cómo 15 estudiantes iraníes de inglés como lengua extranjera desarrollaron la capacidad de evaluar sus escritos al tener acceso a las puntuaciones de los evaluadores. Los participantes fueron supervisados durante cuatro semanas al ser su primera experiencia en el proceso de autoevaluación. Se proporcionó una hoja de evaluación detallada a cada estudiante para que evaluara su trabajo y después de cada autoevaluación, los estudiantes pudieron tener acceso a las puntuaciones globales asignadas por el profesor. Los resultados indicaron una alta correlación entre la autoevaluación y la evaluación del profesor. Esto reveló que la autoevaluación de los estudiantes durante todo

¹ Received: October 4, 2013 / Accepted: April 18, 2014

² sasanbaleghizadeh@yahoo.com, t.hajizade@yahoo.com

el estudio resultó estar altamente correlacionado con la evaluación docente. También se demostró que los estudiantes evaluaron diferentes componentes de su escritura de una manera comparable a la realizada por el profesor. Los resultados confirmaron que la autoevaluación podría no sólo ser vista como una herramienta útil para evaluar el rendimiento de los estudiantes sino también podría ser considerada como un instrumento eficaz para desarrollar sus destrezas de escritura.

Palabras clave: autoevaluación, evaluación de la actividad docente, escritura

Resumo

Este estudo pesquisou como 15 estudantes iranianos de inglês como língua estrangeira e desenvolveram a capacidade de avaliar seus escritos ao ter acesso às pontuações dos avaliadores. Os participantes foram supervisionados durante quatro semanas ao ser sua primeira experiência no processo de autoavaliação. Foi proporcionada uma folha de avaliação detalhada a cada estudante para que avaliasse seu trabalho e depois de cada autoavaliação, os estudantes puderam ter acesso às pontuações globais designadas pelo professor. Os resultados indicaram uma alta correlação entre a autoavaliação e a avaliação do professor. Este revelou que a autoavaliação dos estudantes durante todo o estudo resultou estar altamente correlacionado com a avaliação docente. Também se demonstrou que os estudantes avaliaram diferentes componentes da sua escritura de uma maneira comparável à realizada pelo professor. Os resultados confirmaram que a autoavaliação poderia não só ser vista como uma ferramenta útil para avaliar o rendimento dos estudantes como também poderia ser considerada como um instrumento eficaz para desenvolver as suas destrezas de escritura.

Palavras chave: autoavaliação, avaliação da atividade docente, escritura

Introduction

Despite the numerous advantages of analytic scoring of productive language skills such as speaking and writing, there are still teachers who, due to time constraints and work pressure, prefer holistic scoring of their learners' performance through summative assessment. When applied to evaluating students' written compositions, this approach can result in potentially biased evaluation because more often than not teachers do not have clear criteria for marking the papers. Even worse, most of them tend to correct all grammatical and spelling errors with red pens which, as Peñaflorida (2002) aptly put it, "bleeds students' papers to death" (p.345). In order to overcome the limitations of summative assessment, particularly when it is done holistically, alternative assessments or "alternatives in

assessment” (Brown & Hudson, 1998, p. 57) has become the common practice of many teachers around the world. One such alternative assessment is self-assessment, which is in line with current learner-centered education (Brown, 2001) and can help students become independent learners (Blanche & Merino, 1989) and, in turn, lessens the burden on teachers (Oscarson, 1989).

Literature Review

While self-assessment came hand in hand with learner-centered approaches such as the communicative approach, it was not until the late 90s when the focus of language teaching shifted to promoting learner autonomy. According to Kumaravadivalu (1994), learner autonomy involves helping students to learn on their own, raising their awareness of their learning strategies, and encouraging them to self-direct their own assessment. This approach to assessment no longer places the teacher at the center of the evaluation process; rather, it prompts learners to take responsibility for assessing their own performance (Oscarson, 1989), paving the way for their improvement through reflection and action. In this way, assessment would surely play a positive role in the learners’ learning process (Roberts, 2006) and would help to increase their autonomy (Cresswell, 2000).

Brown and Hudson (1998) have argued that self-assessment is a type of “personal-response assessment” (p. 63) and define it as a kind of assessment that “require(s) students to rate their own language” (p. 65). Upshur (as cited in Heilenman, 1990) was one of the first to support the use of this kind of assessment in the measurement of second language abilities since he believed it is only the learner who knows how successfully he could use the language. Many advantages of self-assessment such as speed, direct involvement of learners, encouragement of autonomous learning (Brown & Hudson, 1998), and the possibility of enlarging the domain of language behavior sampled without substantially increasing the time and cost involved (LeBlanc & Painchaud, 1985) have been extensively dealt with in the literature. For example, Bachman and Palmer (1989) argued that “self-ratings can be reliable and valid measures of communicative language abilities” (p. 22). Likewise, Huerta-Macias (as cited in Brown & Hudson, 1998) has claimed that “alternative assessment (not just self-assessment) consists of valid and reliable procedures that avoid problems inherent in traditional testing including norming, linguistic, and cultural biases” (p. 55). Thus, students can even gain more insight into their strengths and weaknesses as writers if they monitor their own writing tasks (Myers, 2001).

Despite the foregoing advantages, not all research studies have confirmed the usefulness of this sort of evaluation. For example, Blanche and Merino (1989), who summarized major findings in self-assessment, argued that the accuracy of most students' self-estimates often varied depending on the linguistic skills and materials involved in the evaluations. Similarly, Davidson and Henning (1985) indicated that although classical reliability estimates of students' self-ratings might be reasonably high, "little confidence should be placed in these particular student self-ratings" (p. 176). This is mostly because students might not be well-trained in doing the self-ratings. Moreover, Heilenman (1990), who investigated the role of response effects (tendencies to respond to factors other than item content) in the self-assessment of second language ability, found out that both a measure of "acquiescence effects" and "overestimation effects" were present (p. 188). Thus, Heilenman realized that those students who had been learning English for two years or more were more likely to overestimate their performance. On the other hand, Matsuno (2009) found that Japanese EFL learners underestimated their performance when they were asked to self-assess their own writings, which was particularly true for high-achieving students. Hence, Matsuno (2009) concluded that "self-assessment was somehow idiosyncratic and therefore of limited utility as a part of formal assessment" (p. 75). Even when the criteria for assessment were set, the participants could not judge their performance in a manner comparable to that of the teachers (Patri, 2002).

In addition to the studies conducted to show whether self-assessment is a useful evaluation tool or not, many studies have been carried out to find out how teachers can help students become better evaluators of their own performance. For example, Roberts (2006) maintained that "in order to have higher correlation between self-assessment and teacher-assessment, we need to provide learners with guidance" (p. 3). Moreover, the result of Jafarpur and Yamini's (1995) study showed that training with self-assessment questionnaires could improve learners' skill to estimate their own language ability. Although without direct instruction, students' self-assessments have shown great improvement over time (Chen, 2008), a number of researchers like Oscarson (1989) emphasize that students do need training for improving their self-assessment.

The writing skill seems to be a good area for investigation when it comes to assessment either by teachers or students themselves. This is mainly because even if the raters assign the same score to a piece of writing, they might have arrived at it based on different criteria (Connor-Linton, 1995; Hamp-Lyons, 1995). Furthermore, even native

English speakers may not have the same reaction as non-natives toward the same written paper. For example, the results of Khalil's (1985) study revealed that native speakers found semantically deviant utterances more problematic than grammatically deviant ones. Native raters proved to be stricter than non-native raters (Kobayashi, 1992), and they gave lower ratings to content than language (Santos, 1988). Even researchers like Brown (1991), who found no statistically significant difference between the ratings given by English and EFL raters, admitted that these two groups arrived at these scores from different perspectives, suggesting that native English speakers focused more on cohesion when assessing a piece of writing, whereas their non-native counterparts attended more to organization.

In order to overcome the problem of subjectivity in holistic assessment of writing, analytic scoring has been proposed as an alternative (Heaton, 1988). According to Stiggins, Richard, Nancy, and Bridgeford (as cited in Perkins, 1983) "Holistic scoring calls for the reader to rate overall writing proficiency on a single rating scale, (but) analytic scoring breaks performance down into component parts (e.g. organization, wording, idea)" (p. 652). Bacha (2001) confirmed that adopting either of these techniques depended on the purpose of writing in EFL programs. She maintained that to provide learners with more specific feedback, analytic scoring would be more appropriate since holistic scoring tends to be highly subjective and lacks internal consistency due to shifting standards (Perkins, 1983). Other scholars such as Hamp-Lyons (1995) have pointed out that "holistic scoring system is a closed system" (p. 760) and no one can have access to points for different parts since the raters do not have a certain criterion for scoring a piece of writing. Cumming (1990), who discussed biases in holistic evaluations of ESL writings, claimed that "analytic scales may have the advantage of drawing raters' attention to specific aspects of students' composition" (p. 42). Studies have indicated that high inter-rater reliability has been obtained from analytic scoring (Bachman, 1990; Jacobs, Zingraf, Wormuth, Hartfiel, & Hughey, 1981; Perkins, 1983).

Apart from having clear criteria for assessment, all raters need training in assessment. Lumley and McNamara (1995) argued that even teachers who wanted to act as raters needed some training courses to make them internally consistent or "self-consistent" (p. 57). Jacobs et al. (1981) argued that the guidance helped to neutralize the differences in their judgment related to raters' backgrounds. Taking the above mentioned points into account, it is evident that training is essential, particularly for learners who want to practice self-assessment for the first time.

Methodology

Research Design

As mentioned previously, the results obtained about the efficiency of self-assessment are inconclusive. Given the fact that very little research has been done in this respect in the Iranian context, this point needs further exploration since cultural backgrounds of the participants could affect self-assessment results (Blanche & Merino, 1989). In the present study, Pearson-product moment correlation coefficient was used to investigate the relationship between self-assessment and teacher-assessment of students' writing. It was hoped that students' self-assessment would improve over time. In other words, it was predicted that during the first cycle of assessment, the correlation between the teacher-assessment and student self-assessment would be relatively low, but gradually students would learn how to assess themselves and this assessment of their own writings would progress to be a better approximation of the teacher's assessment. Hence, the correlation would be greater in the last cycle of assessment. To this end, the following research questions guided the study:

1. Is there a statistically significant relationship between teacher-assessment and learners' self-assessment in each cycle of assessment?
2. Does learners' self-assessment improve over time (as they have access to teacher-assigned scores)?
3. In the last cycle of assessment, do learners assess different components of their writing in the same way as the teacher did?
4. Does self-assessment lead to an improvement in learners' writing?

Participants

Twenty Iranian female EFL learners at the upper-intermediate level of English language proficiency with an average TOEFL score of 550 participated in this study. The participants were all females with an average age of 20, and were all taking a TOEFL preparation course at a private English language school in Tehran, Iran. Since the study lasted for several weeks, not all the participants were able to attend all the assessment sessions. Therefore, the researchers decided to include the individuals who took part in all sessions, as a result of which the number of participants was reduced to fifteen.

In addition to the learners, a native Iranian teacher (the second researcher of the present study) with five years of experience in teaching English as a foreign language was in charge of teaching paragraph writing to the students and rating their written assignments. Inasmuch as analytic scoring offers high inter-rater and intra-rater reliability, the rater assessed each paper only once.

Data Collection Instruments

In this study both the teacher and the students assessed the writings through a detailed evaluation sheet (see Appendix A). This evaluation sheet with five subscales (see Table 1) was taken from Jacobs et al. (cited in Bacha, 2001). However, the researchers thought that students would need to know how these five components (content, organization, vocabulary, language, and mechanics) would be broken down into smaller sub-scales while scoring their papers. To this end, following Matsuno (2009), these five criteria were clearly defined. Thus, for example, it became clear that content refers to sub-scales, such as the amount of writing, the development of the topic, and the relevance of the students' writing to the assigned topic. Similarly, it was known that organization refers to the opening, supporting sentences, closing and logical sequences of ideas in writing (see Table 1 for further details).

Table 1. Jacobs et al.'s Composition Profile

Content	30%
Organization	20%
Vocabulary	20%
Language	25%
Mechanics	5%

By using this detailed evaluation sheet, interval scales could be obtained which paved the way for using Pearson correlation coefficient formula. Working with this checklist was quite easy for the participants because instead of a main category like content, they had access to the subcategories which helped them in their assessment.

The second instrument was a questionnaire developed by the researchers (Appendix B). The questions were written in simple

language to ensure they would be easily comprehended. The main goals behind utilizing this questionnaire were to elicit the learners' attitude toward self-assessment, their beliefs about the usefulness of this kind of assessment, and their ideas on how they had self-assessed their performance. This instrument was primarily used to save time since obtaining the same amount of information through interviewing all the participants would be rather time-consuming.

The last instrument used was a semi-structured interview that the researchers conducted with a few of the participants. Although the research was mainly quantitative, it was felt that the use of an interview might shed more light on some dark points, such as the participants who would either overestimate or underestimate their performances.

The study was conducted at a private language institute and lasted for one month. Although the participants were upper-intermediate learners, some of them had problems in developing a paragraph. Therefore, their teacher (the second researcher) devoted the first session to providing them with some instruction on appropriate length, format, content, and organization of a paragraph. Then during the second session, the participants were introduced to the evaluation sheet (Appendix A). The second researcher explained what each category as well as the related subcategories meant and made sure students understood what they were expected to do. After realizing how to assign scores to different components, they were given a topic to write about on the same day and were asked to evaluate their writings two or three days later, trying to be as objective as possible. This interval would help the participants to detach themselves from their writings and enable them to be more critical of them. It is worth mentioning that the topics were related to what they had studied during the week, and an attempt was made to take the participants' interest into consideration in the process of topic selection.

After each writing and self-assessment, the second researcher, who was both the teacher and the rater, collected the papers and returned them the next session along with her evaluation (she used the same evaluation sheet for assigning scores). She did not write any comments either in the margin of the papers or the evaluation sheets. The participants were supposed to figure out for themselves why they had received a particular score. Furthermore, the second researcher asked them to read their writings one more time and reflect on the scores assigned by the rater. Then, the participants were given back their writings, which were evaluated two times (once by the learners and the other time by the rater) to have a chance to compare their self-

assessments with the teacher-assessment and to find out whether they had overestimated or underestimated their performances. This cycle of writing, self-assessment, and rater's assessment lasted for four weeks during which the participants wrote about four different topics.

Data Analysis and Interpretation

The strength of relationship between each self-assessment and the teacher-assessment was calculated using Pearson-product moment correlation coefficient. Since the study consisted of four cycles of self-assessment and teacher-assessments, the results of four correlations are included in the paper. The correlations were used in order to find out whether there was a relationship between the two types of assessments, and if yes whether an acquaintance with the teachers' assessment had positively affected their self-assessments or not. A matched t-test was used to compare the learners' first writings with the last ones. The result of this t-test could be used as an indicator of whether self-assessment had helped improve learners' writing ability or not. The means and standard deviations of different components of learners' last writing were compared. The purpose of this comparison was to find out if there was a difference between the way the rater and the students assessed the papers.

Results

The Correlation

Table 2 shows the results of four correlations. The magnitude of these correlations is larger than the critical value ($p < .01$). This indicates that the obtained results are statistically significant.

Table 2. Correlation between self-assessment and teacher-assessment in the four cycles of writing

First cycle	.63
Second cycle	.63
Third cycle	.71
Fourth cycle	.91

$p < .01$, $n=15$.

Before discussing the result of these correlations in length, it is worth showing the scatter plots of the four sets of scores (Figure 1), since they make the interpretation of the results much easier.

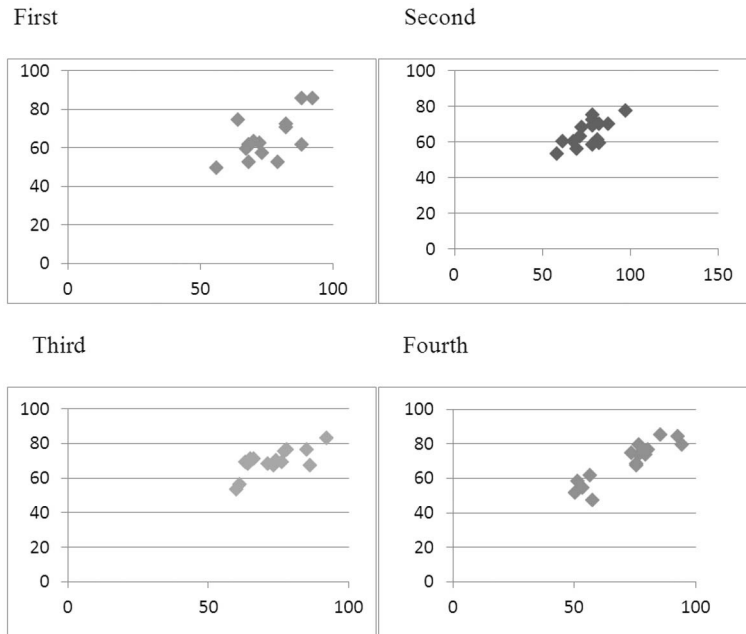


Figure 1. The scatter plots of four sets of scores obtained from self-assessment (horizontal axes) and teacher-assessment (vertical axes)

The scatter plots clearly show that there is a positive correlation between self-assessment and teacher-assessment. The amount of this correlation remained constant ($r=.63$) during the first and second cycles of assessment. This might have been due to the fact that the experience was new to the students. As the first and second scatter plots reveal, there were some outliers that affected the results of the correlations to a great extent. All of these outliers overestimated their writing performance.

108

The third scatter plot indicates a significant change in the way the participants assessed their writings; here we had only one outlier who overestimated her performance though the majority of the participants reported scores which were quite close to those of the rater. There were also four participants who underestimated their performance. Although the correlation was not a perfect positive one ($r=.71$), the improvement of the learners' self-assessment cannot be overlooked.

The result of the fourth cycle of assessment was quite unexpected ($r = .91$); the participants seemed to have developed an excellent skill in assessing their writings. The researchers assume that this great improvement might have been due to the fact that the second researcher asked the students to have a second meticulous look at their third papers and compare the scores they had given to themselves and those that the rater had assigned them. The motivation for such a decision came from the second researcher's observation. She realized that during the first two sessions, the participants took a quick look at the rater's scores and returned the papers in a minute or two. The extra time allocation has obviously had a great effect on raising the learners' awareness and consequently their self-assessment. This is rather surprising as the learners were not provided with any direct explanation or instruction.

The mean and standard deviations of the scores assigned by learners and the rater to different components of writing are displayed in Table 3. A comparison of these scores can provide us with the answer to the third research question. The scores that the participants gave to different parts of their last writing were quite similar to those assigned by the teacher. Thus, we can conclude that in the last cycle of assessment, there was not only a high correlation between teacher-assessment and students' self-assessment, but also a great similarity in the way learners and the rater assessed different components of writing.

Table 3. The means and standards deviations of the scores assigned by learners and the rater to different components of writing

Components of writing	Self-assessment		Teacher-assessment	
	<i>m</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Content	23.28	4.73	24	3.78
Organization	13.71	2.73	13.27	2.5
Vocabulary	14.43	2.31	13.35	3.43
Language	13	3.98	12.78	2.86

The Questionnaire

After evaluating the last writing, the participants were asked to complete the questionnaire (Appendix B). The answers shed more light on the processes which students went through in order to evaluate their performance. It was revealed that the experience of self-assessment was quite new to all the participants. Some of them claimed that they self-evaluated their performance in most of the tests, and their obtained scores had been close to what they had expected. This was particularly true about the participants who claimed to have been objective in their self-assessment.

The learners' attitudes toward the experiment varied. While all of them agreed that self-assessment was quite motivating, some of them found it a bit hard. Most of them believed that they were quite objective in their self-assessment. Many of them found the rater's score fair and indicated that having access to these scores had helped them a lot in their evaluations. Most of the participants claimed that they had learned a great deal about self-assessment during the experiment. Nevertheless, some of them maintained that teacher-assessment is more beneficial for the following reasons:

- When we learn a new grammar point and want to use it in our writing(s) we need someone to correct our mistake(s).
- My teacher is more knowledgeable and can correct my paper better.
- I'm still a student. I cannot judge my writing.

Others believed that when they evaluate their works objectively they could have a better image of their weaknesses and strengths:

- I think it is better to check my writing myself, because I can realize my problems better.
- Self-assessment helps me to find out my problems myself.

110

Almost all the participants argued that self-assessment could be more useful when it comes hand in hand with teacher-assessment. The researchers suppose that this attitude stems from the way Iranian students have been treated in schools. They have learned to regard the teachers as authority figures and they respect them as people who are capable of judging their performance.

The interesting point was that most of the participants found having access to the rater's scores more useful than having direct training in

self-assessment. They believed that participating in this experiment had not only improved their self-assessment abilities but also had helped them develop their writing skill. This claim is supported by the result of a matched t-test (see Table 4).

Table 4. Matched t-test on improvement of students' writing after self-assessment

	<i>n</i>	<i>mean</i>	<i>SD</i>	<i>t value</i>	<i>Df</i>	<i>p</i>
First writing	15	65.6	10.28	3.86	14	.001
Fourth writing	15	71.7	9.19			

p < .05

The Interview

As mentioned earlier, a semi-structured interview was carried out with the participants who had influenced the result of the correlations in one way or the other, i.e. those who had either over-estimated or underestimated their performance. One of the participants, who had assigned high scores to her writings, claimed that she did not pay much attention to the accuracy of her self-assessment and all she wanted was improvement of her writing skill. Her goal was to receive more or less the same scores she had given to her writing from the rater. She thought assigning higher scores to her writing meant that her writing was actually improving.

Another learner who had underestimated her performance believed that she had never been good at writing. She was quite modest and maintained that the rater had been quite generous; otherwise, her scores could not have been so high. The most interesting part of the interview was talking to the learner whose self-assessment scores were always quite close to those of the rater. She said that she had tried to be quite objective in her self-assessment. She also told the second researcher that whenever she wanted to assess her writing she imagined that it was someone else's paper, not hers.

Conclusions

This paper sought to take a closer look at self-assessment in the Iranian context. The study confirmed that there is a statistically significant positive correlation between teacher-assessment and self-assessment. This finding can be viewed as confirmation of the research carried out in different EFL contexts. However, the main contribution of this study to the literature lies in the fact that this research indicated that training learners simply through introducing them to checklists is not the only possible way to improve their ability to self-assess their performance.

Unlike previous studies, this research took students through four cycles of writing, self-assessment, and teacher-assessment. After each self-assessment, the learners were provided with the scores that the teacher assigned to different components of their writings. The comparison that students made between their self-assessment and teacher-assessment had a positive effect on the way they self-evaluated their subsequent writings.

The results of the correlational studies justified the improvement in learners' self-assessment. The magnitude of correlation between self-assessment and teacher-assessment remained constant during the first two cycles of assessment when the experience of self-assessment was still quite novel to the participants. However, the obtained correlation rose significantly in the third cycle of assessment since the learners had a better image of what self-assessment was about, and how they could assess their writings in the same way that the rater did. Assessing their last writing was much easier for the learners since having access to the teacher-assigned scores for their previous writings had taught them how to be objective and critical toward their own works, so the last self-assessment was the closest to teacher-assessment ($r = .91$). Consequently, it can be claimed that there is a direct relationship between training of students in self-assessment, which is done by providing the learners with their teacher-assessed papers, and the accuracy of their self-assessment scores. That is to say, the more the students ponder the scores assigned by the rater, the better they tend to assess their subsequent writings.

112

Another finding inferred from students' writings was the fact that their writing skill improved significantly toward the end of the experiment. This fact was not only noticed by the students themselves, but was also supported through statistical procedures. The result of the matched t-test indicated a statistically significant change in the learners' writing ability.

The last, but obviously not the least finding was that the participants evaluated their paragraphs almost in the same way that the teacher (rater) did. The scores that students assigned to different components of their last writing were quite close to those assigned by the teacher.

This study confirmed the effectiveness of self-assessment, as a kind of alternative assessment, in the context of Iran. The learners' assessment of their own writings not only proved to correlate with those of the rater, but also improved significantly over the course of the study. Nevertheless, the findings should be treated with caution since the research was carried out with only fifteen participants, all from the same institute. This indicates that further research is still needed before we can be fully sure about the beneficial effects of training students in self-assessment.

References

- Bacha, N. (2001). Writing evaluation: What can analytic versus holistic essay scoring tell us? *System*, 29(3), 371-383.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, 16(1), 14-29.
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39(3), 313-340.
- Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy* (2nd ed.). New York: Pearson Education.
- Brown, J. D. (1991). Do English and ESL faculties rate writing samples differently? *TESOL Quarterly*, 25(4), 587-603.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.
- Chen, Y. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12(2), 235-262.
- Connor-Linton, J. (1995). Looking behind the curtains: what do L2 composition ratings really mean? *TESOL Quarterly*, 29(4), 762-765.

- Cresswell, A. (2000). Self-monitoring in student writing: Developing learner responsibility. *ELT Journal*, 54(3), 235-244.
- Cumming, A. (1990). Expertise in evaluating second language composition. *Language Testing*, 7(1), 31-51.
- Davidson, F., & Henning, G. (1985). A self-rating scale of English: Rasch scalar analysis of item and rating categories. *Language Testing*, 2(2), 164-179.
- Hamp-Lyons, L. (1995). Rating nonnative writing: the trouble with holistic scoring. *TESOL Quarterly*, 29(4), 759-762.
- Heaton, J. B. (1988). *Writing English language tests*. New York: Longman.
- Heilenman, L. K. (1990). Self-assessment of second language ability: The role of response effects. *Language Testing*, 7(2), 174-201.
- Jacobs, H. J., Zingraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley: Newbury.
- Jafarpur, A., & Yamini, M. (1995). Do self-assessment and peer-rating improve with training? *RELC Journal*, 26(1), 63-85.
- Khalil, A. (1985). Communication error evaluation: Native speakers' evaluation and interpretation of written errors of Arab EFL learners. *TESOL Quarterly*, 19(2), 335-351.
- Kobayashi, T. (1992). Native and nonnative reactions to ESL compositions. *TESOL Quarterly*, 26(1), 81-112.
- Kumaravadivalu, B. (1994). The postmethod condition: (E)merging strategies for second/foreign language teaching. *TESOL Quarterly*, 28(1), 27-48.
- LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19(4), 673-687.
- Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing*, 12(1), 54-71.
- 114 Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75-100.
- Myers, J. L. (2001). Self-evaluations of the "stream of thought" in journal writing, *System*, 29(4), 481-488.
- Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6(1), 1-13.

- Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, 19(2), 109-131.
- Peñaflorida, A. H. (2002). Nontraditional forms of assessment and response to student writing: A step toward learner autonomy. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 344-353). Cambridge: Cambridge University Press.
- Perkins, K. (1983). On the use of composition scoring techniques, objective measures, and objective tests to evaluate ESL writing ability. *TESOL Quarterly*, 17(4), 651-671.
- Roberts, T. S. (2006). Self, peer, and group assessment in E-learning: An introduction. In T. S. Roberts (Ed.), *Self, peer, and group assessment in E-learning* (pp. 1-16). London: Information Science Publishing.
- Santos, T. (1988). Professors' reactions to the academic writing of nonnative speaking students. *TESOL Quarterly*, 22(1), 69-90.

Authors

***Sasan Baleghizadeh** is Associate Professor of TEFL at Shahid Beheshti University (G.C.) in Tehran, Iran, where he teaches courses in applied linguistics, syllabus design, and materials development. He is interested in investigating the role of interaction in English language teaching and issues related to materials development. His published articles appear in many international journals including *TESL Reporter*, *TESL Canada Journal*, *ELT Journal*, and *Language Learning Journal*.

***Tahereh Hajizadeh** holds an M.A. degree in TEFL from Allameh Tabataba'i University in Tehran, Iran. She is an experienced EFL teacher and is interested in issues related to classroom assessment.

Appendix A. Evaluation Sheet

1. Content	1-10
Amount	
Development of the topic	
Relevance to the topic	
2. Organization	1-5
Opening	
Supporting sentences	
Closing	
Logical sequencing	
3. Vocabulary	1-10
Range	
Word form/ word choice	
4. Language	1-10
Grammar	
Use of variety of structures	
5. Mechanics	1-5
Spelling	
Punctuation	

Appendix B. The Questionnaire

1. How long have you been learning English?
2. Have you ever tried to evaluate your performance in the writing part of an exam? If yes, was your estimation close to the mark you received?
3. Was this experiment new to you? If not, when and how did you experience something like this?
4. How did you feel when you had to evaluate your own works?
5. Did you try to be objective in your self-assessments? If yes, what did you do?
6. Did you think that the scores that the rater assigned to your works were fair? Why?
7. Do you think that comparing your self-assessments with teacher-assessments helped you evaluate your writings better? If yes, how?
8. After this experiment, do you like to be given the chance to assess your writings yourself? Why?
9. Do you think that self-assessment can help you learn better? Why?
10. Do you think that students need training for self-assessment? Why?