Full Length Research paper

Self-assessment and the impact on language skills

Jila Naeini

Department of English, Aliabad Katoul Branch, Islamic Azad, University, Aliabad Katoul, Iran.
Email: naeini.j@gmail.com; Tel.: 0098 173 622 4500

Alternative assessment is a new assessment culture that values processes of learning and the unique performance of individual. Self assessment, as an alternative assessment process, encourages learners to take greater responsibility for their own learning. To this aim, the researchers selected a total of 121 out of 150 EFL (English as a Foreign Language) university students who had been given a language proficiency test. The researcher divided the subjects randomly into two groups of experimental and control. The instrument of this experimental research included writing and speaking pretests and posttests as wells as the Jacob's writing scoring scale profile and Weir (1990)'s Criteria for Oral Test. The results indicated the significant improvement in the learners' writing ability applying the writing self-assessment check list in experimental group. In addition, the results of descriptive statistics conducting the series of independent T-Tests, Two-way ANOWA, and correlational analyses indicated the significant effects of treatment and outperformance of experimental group in all components of writing and speaking posttests. Therefore, it can be concluded that self assessment acted as a booster for language skills.

Keywords: Assessment, alternative assessment, self-assessment

INTRODUCTION

As one form of alternative assessments, self-assessment has recently gained significant attention in foreign language education as one measure of learners' language competencies. The growing interest in self-assessment in recent years is tied to the increasing interest in learner-centered language teaching and self-regulated language learning (Nunan, 1988; Bloom, 2007). In Vygotskyan terms, self-regulated learners are seen as exercising more control and autonomy over the process and outcome of their own learning when compared with other-regulated learners (Cameron, 2001). Some researchers have offered the benefits of applying self-assessment (Coombe and Canning, 2002; Liang, 2006). Such benefits include. Promoting learning, raising learners' awareness of their own learning, improving the goal orientation of individual learners, expanding the range of assessments into affective domains, reducing the burden of assessment placed on teachers, and improving learners' autonomy. It has also been demonstrated that self- assessment can be employed effectively not only for formative purposes but also for placement (Hargan, 1994; Leblanc and painchaud, 1985;

Brantmeier, 2006). Many benefits of applying self-assessment have been offered and studied. Though, there is no study regarding the effects of self-assessment of one skill, say, writing on the other skills. Therefore, this study attempted, firstly, to investigate the effects of self-assessment on the writing ability of ELT students and secondly to explore whether self-assessment of writing skill has any impacts on the improvement of the learners' speaking ability.

Alternative assessment and psychometric paradigm

According to Gipps (1994), the traditional psychometric testing model was essentially one limitation in which measurement attributes are the properties of the individual and are considered to be fixed. This notion is, then, one of the major limitations of the psychometric approach. In contrast, assessment supports learning and aims to help the individual to develop his/her own learning. Assessment is considered to be enabling rather than limiting. Another feature of psychometrics is the interpretation of scores in relation to norms. Norm-referencing grades an individual's performance in

relation to that of his/her peers, that is in terms of relative performance rather than their absolute performance.

Assessment as an alternative model of testing is a more dynamic one in which the students' learning potential is the focus in a dynamic and positive sense. In assessment, we need to be able to assess what the pupils already know, their learning strategies, their current knowledge and how they control their own learning. In addition, interactive or scaffolded assessment will indicate not only what the pupils know and can do but also what they nearly can do. As assessment has moved into the foreground of education, it has required all teachers to be able to make effective use of certain knowledge and skills. In addition to being ubiquitous and consequential, most language teachers are routinely expected to assess and respond purposefully to their students' written and oral communication. These ordinary pedagogical functions involve specific expertise and informed judgments. Cumming (2009) also argues that, experienced instructors of English as a second or foreign language typically use 27 different types of decision-making behaviors while evaluating a single composition. Furthermore, Harlen and Winter (2004) introduce ideas of 'assessment for learning' as a means whereby teachers can make their classroom assessment more directly focused on learners' development and can actively involve learners in this process. Accordingly, the following section discusses the value of assessment for learning.

Assessment for learning

 Educational assessment largely concerns attempts to develop new and better measurement techniques. So the great focus has been on 'how to assess' in the search for equity, transparency, legitimacy and, indeed, utility that issues of fitness for purpose in terms of who is being assessed (Broadfoot, 2005). A recent research review by the (UK-based Assessment Reform Group, 2002) found that:
• Feedback on assessments has an important role in determining further learning.
• Students evaluate their own work all the time, and how they do this depends on the classroom assessment climate.
• High-stakes assessment can create a classroom climate in which transmission teaching and highly structured activities predominate and favor only those students with certain learning dispositions.
• The validity of tests as useful indicators of students' attainment is challenged by the narrowness of the instruments and the way in which students are trained to answer the questions.
• Students dislike both selection and high-stakes tests, they, show high levels of test anxiety (particularly girls), and they prefer other forms of assessment.

Therefore, it can be concluded that learning is constru-

cted and controlled by pupils. According to Dann (2002), without teachers' willingness to engage with the curriculum and pupil's developing range of cognitive competencies and experiences, learning will not proceed. Dann (2002) emphasizes on the contribution of learners in the process of assessment in order for assessment to result in development and learning. Accordingly, self assessment as a form of alternative assessment supports authentic assessment and causes learners to be autonomous in terms of being a self-learner and a self-assessor (Tavakoli, 2010).

Self-assessment and Language Skills

Investigations concerning self-assessment in language learning have examined the value of self-assessment in proficiency testing with participants of all ages in different language skills and the results with these learners show that self-assessment positively correlates with language abilities. For example, Finch and Taeduck (2002) examined applying self-assessment as a valuable additional means of improving oral abilities. They developed a test focusing on the improvement in spoken English of 1700 Freshman University students over an academic year (64 hours). This was administered and evaluated using established oral-test criteria. They looked at improvement rather than level of achievement, and the Conversation-English course taken by the students was the basis of the test. Results showed that: 1) preparation for the test necessitated active spoken participation in lessons, 2) lessons tended to utilize task-based communicative teaching methods, 3) the means became the end - the test was not only a reason for developing oral skills, but also a means of achieving that goal. Furthermore, Tavakoli (2010) attempted to investigate the relationship between performance testing and alternative assessment. Tavakoli (2010) found high correlations among self-rating and self-classroom assessment, teacher-rating and teacher-classroom assessment, and self-assessment and teacher-classroom assessment. He concluded that alternative assessment such as self-assessment is likely to be as reliable and as valid as performance testing. In the field of self-assessment, Javaherbakhsh (2010) studied the impacts of self-assessment on Iranian EFL learner's writing skill. In order to fulfill the purpose of his study, he selected 73 participants from 105 students of advanced level of English in Zabansara and Marefat Institutions in Iran by means of a TOEFL test. He concluded that the administered self-assessment techniques to the experimental group improved writing skill significantly.
The effects of self-assessment have also been studied on the languages other than English. For example, Coronado-Aliegro (2006) investigated the influence of a continuous self-assessment component on the self-efficacy of undergraduate students studying Spanish as a foreign language.

One hundred and four undergraduate students (62 in treatment group and 42 in control group) participated in this experimental study. All participants completed Spanish as a Foreign Language Self-Efficacy Questionnaire during the second week of the semester and during the final week of the semester. Participants in the treatment group also completed weekly Self-Assessment Questionnaires throughout the semester. The results showed that Spanish undergraduate students' self-efficacy seemed to be heightened significantly more with continuous self-assessment than without it.

As it has been discussed, during the last 10 years there has been a surge of interest in self-assessment methodologies in foreign/second language education and various projects have been undertaken in different parts of the world and several reports on the theoretical and practical implications of using self-assessment techniques have emerged. It is also ideal if the findings of all the empirical studies on the effectiveness of self-assessment have been consistent. However, there was no study regarding the impacts of self-assessment in one skill on the other skills. Hence, the present study primarily attempts to examine the effects of self-assessment on the improvements of the components of writing skill. This study also intends to explore the relationship between self-assessment in writing, as a booster, on the speaking skill of the Iranian ELT university students.

MATERIALS AND METHODS

The Study

Given the importance of self-assessment and the impacts on language skills and the outcomes in language learning, the following research questions were raised:
1) Is there a significant relationship between the learners' self-assessment scores and their writing performance?
2) Does the subjects' writing self-assessment have any impacts on their speaking performannce?
This study was conducted in the autumn 2009. The subjects were ELT University students selected from three different universities in Iran.

Participants

The homogenous subjects of the present study selected from the initial 150 female and male ELT Iranian University students during the academic year of 2008-2009 at three universities. They were all first-year undergraduate ELT students taken the advanced writing course.

Instruments

The following sets of tasks and tests were employed in this study:

The intermediate Nelson Language Proficiency Test (1977)

It consisted of two parts: A cloze passage and some discrete-point items. These two parts were some modified versions of the original Nelson Test of Language Proficiency. The Cronbach's reliability for the pilot Nelson intermediate test was estimated at .85.

The writing check list

To score the subjects' compositions, a rather analytical (objective) procedure was taken. The criterion used was taken by (Jacobs et al., 1981, cited in Hughes 2003). According to this scale every composition must be read five times, and each time only one factor should be taken into consideration. The five factors were: 1. Content, 2. Organization, 3. Vocabulary, 4. Language Use, and 5. Mechanics.

Writing pretests and posttests

A pretest and a post-test of the writing tests were conducted to evaluate the subjects' writing performance. The participants were given general topics for the pre-tests and post-tests. They were given 30-45 minutes. The students' writings were evaluated analytically.

The speaking checklist

In order to evaluate the subjects' speaking ability prior and after the program the (Weir, 1990)'s assessment criteria for the oral test was applied. The check list consists of 6 sections: 1.appropriateness, 2. adequacy of vocabulary for purpose, 3. grammatical accuracy, 4. Intelligibility, 5. Fluency, 6. relevance and adequacy of content. In order to match the writing and speaking data, the data collected were reduced into 5 categories: 1. content relevance, 2. intelligibility, 3. vocabulary, 4. language use, 5. fluency. Therefore, speaking ability of the subjects was evaluated analytically.

The Speaking Pretest and Posttest

In order to assess the participants' general speaking ability, two versions of IELTS interview tests were utilized prior as well as the end of the term. The test took the form of a face to face interview between one participant and one examiner. The participants were assessed on their use of spoken English to answer short questions, speak at length on a familiar topic, and also to ask questions and interact with the examiner.

Procedure
First, the modified Intermediate Nelson Proficiency test was administered to the whole population. The Nelson scores of the subjects were then entered into

Table 1. Descriptive Statistics Pretest of writing and speaking by Groups

| | Levene's test | | | | |
	F sig.	t	(two-tailed)	df	Mean Dif
Equal Variances Assumed (Writing)	1.309 .255	.861	.391	119	.36185
Equal Variances not Assumed (Writing)		.838	.404	93.001	.36185
Equal Variances Assumed (Speaking)	1.544 .217	1.495	.138	119	.60420
Equal Variances not Assumed (Speaking)				90.825	.60420

According to the results shown in table 1, there is not any significant difference between the experimental and control groups' mean scores on the pretest of writing and speaking.

SPSS program for the statistical purposes. They were plotted on a bar chart to study the distribution pattern of the scores and at the same time a descriptive report was taken from SPSS to get the mean and Standard Deviation (SD) of the scores to decide on the final homogenous groups.

A pretest of writing and speaking were administered to evaluate the subjects' writing and speaking performance prior the program. The compositions were marked using an analytic method. For analytic scoring, the ESL composition profile by Jacobs et al (1981) was used. This profile consists of five traits, which tap different features of a written text by a set of descriptors corresponding to different quality levels. The five traits and their maximum number of points are Content 30, Organization 20, Language Use 20, Vocabulary 20, and Mechanics 5. The assessments of the learners' writing tests were done by three EFL instructors, including the researcher. The researcher briefed the other two raters. In order to calculate the inter-rater reliability of the sets of scores given by the three raters, the coefficient correlation (Pearson Product Moment) was used for the paired raters, and to examine the consistency of multiple ratings the researcher used the approach suggested by Bachman (1999). Therefore, the coefficient alpha was computed, treating the independent ratings as different parts. Then, the subjects were divided into two groups, the experimental and the control groups. The subjects in experimental group were given the same checklist and a score sheet to score their own writings. The researcher briefed the subjects. After each writing assignment, there would be a discussion: The teacher's scores and the students' scores were compared and the subjects were given feedback. In control groups, the subjects were given the same topics and assignments and the course materials were the same as those of experimental groups the only difference was that the control group subjects were not given the checklist. Therefore, their writing papers were rated just by their teachers based on the same checklist. During the academic semester, the writing assignments in experimental groups were rated by the teachers and the subjects themselves and the teacher would give them appropriate feedback regarding their errors. Finally, the very last session of the term all the subjects (in both experimental and control groups) participated in the posttest of writing and speaking and the collected data were analyzed applying different statistical analysis procedures.

In order to assess the learners' speaking performance, two versions of IELTS interview tests were utilized prior as well as the end of the term (as the pretest and post test of speaking ability). The tests took the form of face to face interviews between one participant and one examiner. The interviews were recorded and reassessed with the same interviewer and the intrarater reliability of the speaking scores was calculated. The speaking performance of the participants were assessed analytically applying Weir (1990)'s assessment criteria for the oral test.

DATA ANALYSIS AND THE RESULTS

In order to see whether the experimental and the control groups were homogenous, the final modified version of the Intermediate Nelson Proficiency Test was administered to the whole population (experimental and control groups). Out of 150 subjects who took the Nelson test, 72 of them were selected for the experimental group and 49 students for the control group. Then, two independent T-Tests were administered to examine the writing and speaking performance of the subjects prior the program. For the results and the explanations see table 1 above.

It should be noted that the subjects' writing papers were rated analytically by three raters and the inter-rater reliability of .78 indicated the high agreement among the three raters for the pre-test of writing. In addition, the speaking performance of the subjects was reassessed by the same interviewer and the intra-rater reliability of .86 also showed the high

Table 2. Two-Way ANOVA Posttest of Speaking by Group and Writing Ability and Descriptive Statistics   of Posttest of Writing and speaking

Source	df	Mean square	F	Sig.	Descriptive Statistics	Writing level
Group	1	135.868	36.687	.000	Group (Writing)	Group (Speaking)
Writing Level	1	74.928	20.232	.000	Experimental 18.021	High 17.657
Group*Writing level	1	1.144	.309	.579	Control 15.190	Low 15.555

This amount of F-value is higher than the critical value of F at 1 and 117 degrees of freedom, i.e. 3.92. Based on these results it can be concluded that there was a significant difference between the experimental and control groups mean scores on the posttest of speaking. In addition, as the descriptive statistics results displayed in table 2, the experimental group with a mean score of 18.02 outperformed the control group on the posttest of speaking.

Table 3. Descriptive Statistics Components of Posttest of Writing by Groups

Components	Group	N	Mean	SD	SEM
Content	Experimental	72	24.7917	2.77279	.32678
	Control	49	20.7551	3.23735	.46248
Organization	Experimental	72	20.3472	2.46759	.29081
	Control	49	15.7347	2.80458	.40065
Vocabulary	Experimental	72	18.0694	2.20962	.26041
	Control	49	15.3673	2.46385	.35198
Language use	Experimental	72	21.0694	3.22546	.38012
	Control	49	16.5306	3.14975	.44996
Mechanics	Experimental	72	6.9444	.57870	.06820
	Control	49	3.1837	.75480	.10783

As the results in table 3 reveal, the experimental group outperformed the control group on all the components of the posttest of writing.

agreement. Based on the mean score of the students on the posttest of writing, i.e. 16.62, the students were divided into two groups of high and low writing ability groups. The students whose scores were below 16.62 were considered as low ability group and the rest of the students formed the high ability group. A two-way ANOVA was, then, run to compare the mean scores of the experimental and control groups from two high and low speaking ability groups on the posttest of speaking. The F-observed value for comparing the experimental and control groups' mean scores on the posttest of speaking was 36.68 (See Table 2 above for the results).

In order to compare the performance of the participants in the components of the writing skill from pretest to post test, five separate independent T-tests were run (The results are shown in table 3 above). Finally, in order to compare the mean scores of the experimental and control groups on the components of the posttest of speaking five separate independent t-tests were run (See table 4 below).

DISCUSSION

This paper reported the outcomes of a study which examined the usefulness of a self-assessment procedure involving the use of self-assessment writing checklist as guides for learners to evaluate their own writing performances. The procedure was developed because a review of the literature revealed no precedent for the impacts of self-assessment of writing on the learners' oral skill. The review of the literature also revealed that in the investigation of the improvement of the writing skill, there were no attempts to examine the effects of self-assessment on the improvement of writing and speaking sub skills (components). Then, the results of the study indicated the outperformance of the experimental

Table 4. Descriptive Statistics Components of Posttest of Speaking by Groups

	Group	N	Mean	SD	SEM
Content	Experimental	72	26.6528	2.67044	.31471
	Control	49	21.4082	3.39066	.48438
Intelligibility	Experimental	72	21.0278	2.41458	.28456
	Control	49	15.7959	2.77624	.39661
Vocabulary	Experimental	72	19.1250	2.48906	.29334
	Control	49	15.9388	2.62526	.37504
Grammar	Experimental	72	18.3056	2.64649	.31189
	Control	49	15.4082	2.82045	.40292
Fluency	Experimental	72	7.0417	.61524	.07251
	Control	49	3.1837	2.37762	.33966

As displayed in table 4 all of the t-values are higher than the critical value of 1.96 at 119 degrees of freedom. Based on these results displayed in table 4 it can be concluded that there were significant differences between the experimental and control groups' mean scores on the components of the posttest of speaking. Descriptive statistics shown in table 4 indicate the outperformance of the subjects in the experimental group on the all 5 components of the posttest of speaking.

group over the control group in both writing and speaking. The data analyses also revealed that the experimental group scores significantly differed from the scores of the subjects in control group in all the components both in writing and speaking. This is an evidence for the treatment effect. Hence, self-assessment is suggested to be applied as a booster for the other skills. In other words, the study began with the intention of developing a means whereby language learners would be able to monitor their learning by evaluating their performance on writing tasks. Then, it became evident that this was ultimately resulted in the improvement of the other language skill, say, speaking.

Conclusion and Pedagogical Implication

According to the available literature, self-assessment has been widely used in many fields. In this study, self-assessment was applied in the treatment group to provide self-guidance and reflection. The results revealed that self-assessment motivated students to be more candid and forward with what they considered as their problems in the course. The pattern of self-assessment followed by increased motivation to achieve literacy resulted in higher performance of the subjects in the other productive language skill, namely speaking. Furthermore, self-assessment seems to work by boosting learners' motivation and self-esteem. As seen in this study, self-assessment in the FL classroom can also have other goals, such as enabling learners to assess their total achievement at the end of a course or course unit, or as a positive influence on the overall learning process and language skills. We could then use self-assessment as part of the overall learning process to

help learners understand their behaviors, helping them recycle what they have learned, and at the same time boosting their self-esteem and self-motivation. Self-assessment can also be used to provide the learners with an end-of-course view of their learning process as well as a step-by- step account of it.

With self-assessment no longer do students have to wait for the teacher to tell them how well they are doing and what they need to do next. The teacher remains generally the more knowledgeable and experienced person in the classroom, but the goal for students to move toward and perhaps even beyond the teacher's level of competence. Placing value on learners' knowledge helps them feel more capable of playing a larger role in their own learning (Jacobs and Farrell, 2003). This type of assessment is believed to be more effective than teacher assessment. According to Lee and Ridley (1998). The tenet of the argument in favor of this is that it allows learners to be the architects of their own learning resulting in increased interests, motivation and confidence.

Suggestions for Further Research

Although results of this study were statistically significant, and the research design and instruments were sound, the study was not without limitations. These limitations must be considered when attempting to generalize results to educational settings. For example, this study was limited to first-year undergraduate ELT students. Even though the results may be applicable to this population, caution is needed when generalizing the results to other populations. Since different English levels emphasize different skills, the link between self-assessment and

the skills other than writing and speaking not focused on in this study. Therefore, as the recommendations for future research, it would be worth and helpful to investigate if there is any relationship between self-assessment and the receptive skills of ELT students. The other limitation of this study pertains to the applied assessment tools. The instruments applied in this study were the Jacobs et al (1981)'s Writing Profile for rating writing skill and Weir (1990)'s Assessment Criteria for the Oral Skill. When conducting this type of study, it would be important to create new instruments, adjusting them accordingly.

REFERENCES

Assessment Reform Group (2002). *Testing, motivation and learning*. Cambridge: University of Cambridge.

Bachman LY (1999). *Fundamental considerations in language testing.* Oxford: OUP.

Bloom ME (2007). Tensions in a non-traditional Spanish classroom, Lang TR 11: 85-102.

Brantmeier CI (2006). Advanced L2 learners and reading placement: Self-assessment, CBT, and subsequent performance. Syst 34:15-35.

Broadfoot PM (2005). Dark alleys and blind bends: testing the language learning, Language Testing 2: 122-141.

Cameron LY (2001). *Teaching languages to young learners*. Cambridge: CUP.

Coombe CH (2002). *Using Self-Assessment in the Classroom: Rationale and Suggested Techniques.* Retrieved in September 7, 2010 from: http://www3.telus.net/linguisticsissues/selfassess2.html.

Coombe CH, Canning CH (2002). *Using self-assessment in the classroom: Rationale and suggested techniques.* Retrieved July 7, 2008 from: http://www3.telus.net/linguisticsissues/selfassess2.html

Cumming AL (2009). What needs to be developed to facilitate classroom-based assessment? TESOL QA 3: 515-520.

Dann RU (2002). Promoting assessment as learning: Improving the learning process. London: Routledge Falmer.

Finch AN, Taeduck HY (2002). *Oral Testing and Self-Assessment - The way forward?* Retrieved September 6, 2010from: http://www3.telus.net/linguisticsissues/oraltesting.html

Gipps CA (1994). Beyond testing: Towards a theory of educational assessment. London: The Falmer Press.

Harlen WY, Winter JA (2004). The development of assessment for learning: Learning from the case of science and mathematics, Language TEST 3: 390-408.

Hargan, NO (1994). Learner autonomy by remote control. Syst 22: 455-462.

Hughes AR (2003). *Testing for language teachers.* Cambridge: CUP.

Javaherbakhsh MO (2010). The impact of self-assessment on Iranian EFL learners' writing skill, English LT 2: 213-219.

LeBlanc RA, Painchaud GI (1985). Self-assessment as a second language placement instrument. TESOL QU 19: 673-687.

Liang JO (2006). *Overview of self-assessment in the second language writing classroom.* Retrieved July 8 2008 from: http//www.secondlanguagewriting.com/documents/Overview.

Nunan DA (1988). *The learner-centered curriculum*. Cambridge: CUP.

Tavakoli MA (2010). Investigating the relationship between Self-assessment and teacher assessment in academic contexts: A Case of Iranian university students. Asian EFL JO *1:* 234-260.

Weir CY (1990). *Communicative language testing.* New York: Prentice Hall.