

# On the Impact of Self Assessment on EFL Learners' Receptive Skills Performance

Hassan Shahrakipour

## **Abstract**

*It is believed that self assessment can influence learners' learning achievement (Dearing, 1997; Falchikov, 1997; Stefani, 1998; Taras, 2001, 2002). it can also influence language learners' writing and speaking performance ((Alibakhshi & Abbaszadeh, 2012). However, the impact of self assessment on EFL learners' listening and writing was not studied yet. This study was an attempt to see whether self assessment can influence EFL learners' receptive skills. It also tried to see whether it has the same effect on intermediate and beginner language learners. In doing so, 120 Iranian language learners were selected through convenience sampling and divided into beginner and intermediate language learners. The experimental groups received a treatment on self assessment and they self assessed their outcomes. All groups received a posttest. The data were analyzed through Acnova test. The results showed that self assessment significantly improved EFL learners' receptive skills. However, the effect value of self assessment for listening skill was less than reading skill. Also, self assessment effect value on beginner groups was less in comparison with intermediate groups.*

**Key words:** self assessment, language learners, EFL, receptive skills

## **Introduction**

Assessment is an integral part of every educational system through which we evaluate learners' achievement on basis of the material taught (Alibakhshi & Abbaszadeh, 2012). In learner-centered pedagogy which considers learners as active participants in education and learning, the task of evaluation or assessment is given to the students. They take part in the process of evaluation or what is called self-assessment. Bailey (1998) defines self-assessment as “procedures

by which the learners themselves evaluate their language skills and knowledge” (Bailey, 1998: 227).

What is obvious in the self-assessment approach to evaluation is the active participation of learners in the learning and evaluation process. This participation brings more motivation to the learners as they see themselves to be responsible for their own learning. It also brings autonomy which in long term makes learners life-long independent learners. They will become able to make judgment about their own learning, seek their own weakness and strength and be aware of their knowledge.

Oscarson (1997; 1989) believes in the importance of learners' responsibility and argue that assessment is not the sole responsibility of the teacher but rather it is a mutual responsibility of both learners and teachers. This mutual responsibility will result in a democratic educational system and learning context. Not only the learners but also the teachers and the institute make benefits of self-assessment.

Many scholars believe that self-assessment helps learners in learning language but what is disappointing is that self-assessment has not been attempted in many places yet. For example, the washback validity of self-assessment on EFL learners has not been adequately investigated. In other words, we are not sure whether it has positive or negative impacts on improving learners' reading and listening skills. This research attempts to study the impact of self-assessment on the reading and listening performances of beginner and intermediate EFL learners.

### **Research Questions and Hypotheses**

The main objective of the present study was to find out whether self-assessment has any significant impact on EFL learners' reading and listening beginner and intermediate.

Q1: Does self-assessment significantly improve EFL learners' reading performance?

Hypothesis: Self-assessment does not significantly improve EFL learners' reading performance.

Q2: Does self-assessment significantly improve EFL learners' EFL learners' listening performance?

Hypothesis: Self-assessment does not significantly improve EFL learners' listening performance.

### **Review of Literature**

Many researches in the 1980s were concerned with the development of self-assessment instruments and their validation (e.g., Oscarson, 1997; Lewkowicz & Moon, 1985). As a result, many approaches including pupil progress cards, learning diaries, log books, rating scales and questionnaires were developed (Boud, 1986; Dearing, 1997; Falchikov, 1997; Stefani, 1998; Taras, 2001, 2002). It helps learners to become autonomous and to be aware of their learning and reflect on their development (Freeman & Lewis, 1998; Boud, 2000). McDonald and Boud (2002, 2003) found that when learners assess their own learning, their learning will be promoted to a high extent. Higgins, Hartley, and Skelton (2001) and Ivanic, Clark, and Rimmershaw (2000) mention that for the development of self-regulation in learners both teacher' and learner' feedback on the learning process are needed and that this ability to give feedback must be promoted in both. Jewah et al. (2004) also confirm that learners must occupy an important and active role in the process of giving feedback.

Some scholars (Rowntree, 1987; Boud, 2000; Taras, 2001) mention that the use of self-assessment in some area like England and Wales is very uncommon and it is probably because they see it as inconsistent with the conventional forms of assessments.

In the same vein, Carton (1993) discusses how self-assessment can become part of the learning process. He describes his use of questionnaires to encourage learners to reflect on their learning objective and preferred modes of learning. He also presents an approach to monitoring learning that involves the learners in devising their own criteria; an approach that he believes helps learners to become more aware of their own cognitive processes.

Accordingly, Butler and Li (2001) investigated the effectiveness of self-assessment among young EFL learners. They found some positive effects of self-assessment on the students' English

performance as well as their confidence in learning English, though the effect sizes were small. The study also found that teachers and students perception about the effectiveness of self-assessment are different depending on their teaching/learning contexts. A number of interesting insights were also discovered through interviews with teachers. They have been asked about the best way to utilize self-assessment as part of foreign language instruction in contexts wherein teacher-centered teaching has been traditionally valued.

In line with the above scholars, Alderson (2005) investigated the importance of self-evaluation in the second language class today, and stressed the advantages of having students keep a regular journal. Taking the methodological framework offered by the Communicative Approach to Language Teaching as a starting point, the dynamic interdependence of purpose, methodology and evaluation within the curriculum were studied. In this sense, formative or ongoing evaluation becomes one of the most practical assessment techniques for controlling our students' progress as well as the effectiveness of our teaching program. Self-evaluation has a number of additional advantages regarding both the affective implication of students in assessing their own learning processes and to their participation in class management.

The fact that training and feedback have some influences on the accuracy of self-assessment has been investigated by many researchers (Stefani, 1998; Orsmond, Merry, & Reiling, 2000; Taras, 2001; Patri, 2002). Having enough training before doing self-assessment is believed to be effective (Hanrahan & Isaacs, 2001; Taras, 2002; Li, 2001). Some researchers gave some recommendations for doing a better self-assessment. Lejk and Wyvill (2001), for example, recommend a holistic approach rather than a category-based approach. Other studies, such as Blatchford (1997, p.6), concluded "a significant association between self-assessments and attainments in both English/reading and mathematics". Taras (2001) reported that the active participation of learners and the teacher' experience will enhance the process of self-assessment.

Different approaches to self-assessment have been investigated. Fallows and Balasubramanyan (2001), for example, reported that compulsory training combined with multiple ratings offer

many benefits. Motivation also plays a significant role in the accuracy of self-assessment. Al Fallay (2001) concluded that those learners who have integrative motivation do more accurately in assessing themselves than those with instrumental motivation. The former group was also seen to be less apt to reflect overestimation than the latter one. Moreover, AlFallay acclaimed that language proficiency also influence on the accuracy of self-assessment inducing that those with higher proficiency were more accurate than those with lower proficiency. He founded that high proficiency learners to some extent underestimate their performances while the lower proficiency level often overestimates their performances.

Similar results were reported by Davidson and Henning (1985), and Heilenmann (1990). Some researchers claimed that motivation intensity is an important factor in the accuracy of self-assessment (Livesey, 1992; Morton et al., 1999). A positive relationship between self-esteem and some other personality traits has also been reported (Calderon, 1991; Collins, 1993; Lindholm-Leary & Borsato, 2002). Researchers (Birckbichler et al., 1993; Heilenman, 1991; Wesche et al., 1990) have mentioned that learners are able to self-assess their achievements accurately. Stankowv (1998) reported that students are often overconfident on the tests of vocabulary and general knowledge.

Brantmeier (2006, 2005 a, 2005 b) reported that levels of self-assessed abilities positively correlated with levels of enjoyment. The study also produced significant effects for both self-assessed ability and enjoyment on written recall, but no such effects were reported on multiple-choice questions. These studies lend support to the hypothesis that self-assessment can be accurate for placement. Oscarson (1997) claimed that, "it seems to be fairly commonly agreed that the question of accuracy and appropriateness of self-estimates of proficiency depends, to a considerable degree, on feature of context and on the intended purpose of the assessment in each individual case" (p.8).

Harutyunyan and Gasparyan (2003) investigated the possibility of integrating students' self-assessment into the evaluation process of the Intensive English Program (IEP) for students at the American University of Armenia (AUA) to raise the students' awareness of their strengths and weaknesses in different language learning areas and to prepare the students for autonomous English language learning.

Most studies have involved older subjects such as college students (Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000; Topping, 1998) and in-service staff (e.g., Jones & Fletcher, 2002; Saavedra & Kwun, 1993). That is, little research has focused on the effects of self and peer-assessments in primary and middle schools. Significant differences between the characteristics of adolescents and adults suggest that studies should specifically investigate whether self- and peer-assessment are suitable for younger students.

According to Matsuno (2009), many researchers have reported high correlations between students and teacher-assessment, while other studies have reported low correlation between them. The work of Pierce et al. (1993) is based on school aged learners in a French immersion program in Canada. Learners assessed themselves against two criteria: by comparing themselves with a native speaker and by reflecting on the difficulty they experienced with everyday tasks in French. Results were compared against learners' results on proficiency tests of the four skills. The researchers concluded that self-assessment is not a reliable indicator of proficiency. However, as they mention, many of the subjects have little or no access to the target language or native speakers outside the classroom. In effect, it would be difficult for them to imagine how they would perform. In a comparison of a test of Dutch as a second language for adult learners and a self-assessed version of the same test, Janssen-van-Dieten (1989) found the self-assessed version less reliable although earlier studies and her pilot studies had been more encouraging. For her, the value of self-assessment is "its positive influence on the learning process" (Janssen-van-Dieten, 1989: 44). Thomson (1996), in studying learners of Japanese as a foreign language, also felt very positive about using self-assessment despite finding considerable diversity in the accuracy of self-making.

Other studies have reported that self-assessment is reliable. Bachman and Palmer (1989), for example, found that members of a multilingual, multicultural group of adult learners of English as a foreign language in the US were able to reliably self-rate themselves for their communicative language abilities. Another example of success with self-assessment has been reported in Blanche (1990)

study wherein the ability of a group of adult learners of French as a foreign language to estimate their own speaking ability. He concludes that “the overall accuracy of the self-evaluation... is impressive” (Blanche 1990: 226). Variability in sample size, age of subjects, cultural and educational backgrounds, target language, the test format, the education context and the criteria against which self-assessment is compared all affect reliability. What is comforting is that even when the results against reliability, researchers maintain the value of self-assessment. One way to validate individual self-assessments is to have teacher randomly check some of the results. This would encourage learners to be honest and realistic in their self-making and would contribute to accreditation. In addition, regular random checking would provide a clearer understanding of the reliability issue (Gardner and Miller, 1999). Xiao and Lucking (2008) examined the validity and reliability of student generated assessment scores. The findings indicated that the validity and reliability of student generated rating scores were extremely high. Al Fallay (2004) investigates the role of some selected psychological and personality traits of learners of English as a foreign language for accuracy in their self- and peer-assessments. The study also shows that long periods of practice and sufficient feedback have a positive effect on the accuracy of self-assessment. He also maintains that students with low self-esteem are the most accurate in assessing their performance, whereas learners with instrumental motivation are the least accurate (Al fallay, 2004). Sung et al. (2005) show that significant consistency is found between the results of student self- and peer-assessments and the results of teacher assessment.

Dlaska and Krekeler (2008) investigated the reliability of self-assessments of pronunciation skills and attempted to understand the causes of difficulties. In this study, 46 advanced learners of German assessed their own articulation of different speech sounds in comparison with the sounds produced by a native speaker. In 85% of all cases the assessments of the raters and the self-assessments were identical. However, the learners only identified half of the number of speech sounds which the raters believed to be inaccurate. The study therefore concluded that even experienced L2 learners seem to find it difficult to self-assess correctly their pronunciation skills.

Oscarson (1997) sums up progress in the area of self-assessment by reminding us that research in self assessments is fairly new. He concludes that there are still many problems remained. For instance, learner goals and interpretations need to be adjusted with external necessities. Also self- assessment is not self-explanatory; it must be introduced slowly and learners need to be guided and supported in their use of the instruments.

### **Participants**

The participants of this study were 120 Iranian EFL learners of English language majoring in English translation and teaching English as a foreign language (TEFL). The participants, with the age range of 20-25, and with no control of sex, were selected through convenience sampling from among the language learners in the Islamic Azad University at Rudehen, Tehran North, and Karaj Branches based on the results of a placement test administered. 60 of the participants were beginner in terms of language proficiency because their scores on the entrance examination to university were below 25 percent. Also, 60 junior students who were taking reading and speaking courses (III) were selected as intermediate language learners. The participants of each level of proficiency were divided into two groups: control and experimental. Learning achievement of the experimental groups was self-assessed by the learners themselves. However, the control groups' achievements were assessed by teachers. At the end of the semester, all the participants took part in a post test including a reading and listening parts.

### **Placement test**

The placement test is a kind of test used to measure the general language ability of the learners in different skills. The placement test used in this study was general language test consisting of 80 items adapted from Longman paper and pencil tests. It contained reading and listening parts. Reliability of the test was administered through KR-21. The reliability index of the test was .78 which seemed to be acceptable.



**Post-test**

This test consisted of two parts: listening (40 items) and reading (40 items). This test was also adapted from Longman paper and pencil TOEFL test. The participants of each group received the test under the same conditions (time, rubric). Reliability of the test estimated through KR-21 approach was shown to be .75 which was reported as an acceptable reliability index.

**Procedure**

Having selected the participants, the teachers labeled them as either control or experimental groups. For listening courses taken by both intermediate and beginner participants, the teachers were recommended to make use of tactics for listening textbooks. For reading courses, the teachers used two different textbooks for extensive reading and one common textbook for intensive reading (Peterson reading flash book). In reading course, the main focus was on teaching reading skills to language learners. The only difference between the control and experimental groups was the way they were assessed. Experimental groups received treatment on self assessment and they were given instructions to self assess their listening and reading using self assessment report sheets. During the treatment period, different techniques of self-assessment were introduced by the teacher. The teacher defined each technique in details and asked learners to practice the technique for the next coming section. Whenever necessary, the teacher provided the learners with enough information and counsels. At the beginning of the training the teacher gave support to each step taken by the learners and as the learners become more and more proficient in using the self-assessment techniques, the teachers' support became less to make the learners more autonomous and independent. After a 15 week treatment all groups received a same listening and reading test.

**Data analysis**

Due to the design of the study, (two pretest, two dependent variables, and two independent variables and two groups), analysis of covariance was used because it permits researchers to statistically control for differences on the pretest so that posttest differences would not be due

to initial differences before training. As listening and reading posttests were treated as two different dependent variables, we had to run two-way Ancovas, one for beginner groups and one for intermediate groups.

### Results of the Study

ANCOVA was used to analyze the data because of two-group pre-test/post-test design. The scores on the pre-test were treated as a covariate to 'control' for pre-existing differences between the control and experimental groups. The assumptions of ANCOVA including the issues of unequal sample sizes, outliers, normality, homogeneity of variance, linearity, homogeneity of regression and reliability of covariates were all checked, in this study only the key assumptions : homogeneity of variance, and homogeneity of regression , in addition to the descriptive statistics and inferential statistics are presented.

**Table 1: Homogeneity of regression**

Source	Dependent Variable	Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Beginner groups	Listening pretest	.371	1	.371	.21	.64	.004
	Reading pretest	5.387	1	5.387	3.08	.08	.054
Intermediate Groups *	Listening pretest	1.580	1	1.580	.93	.33	.017
	Reading pretest	.147	1	.147	.08	.77	.002

In the output obtained from this table, the only value that we are interested in is the significance level of the interaction term (shown above as Group\* listening pretest, etc). The Sig. level for the interaction between beginner groups and both listening and reading pretest is greater than .05 (.64 and .08), so interaction is not statistically significant, indicating that we have not violated the assumption. The results also showed that the interaction between

intermediate groups and both listening and reading pretests is not significant (sig value = .33 and .77). Therefore, it could be strongly argued that difference between control and experimental groups in the dependent variables is due to the impact of treatment, in this case self-assessment.

**Table 2 : Levene's Test of Equality of Error Variances**

Test	F	df1	df2	Sig.
Beginner listening	2.285	1	58	.136
Beginner reading	5.475	1	58	.09
Intermediate listening	1.285	1	58	.15
Intermediate reading	7.475	1	58	.23

Levene's Test of Equality of Error Variances was used to check for homogeneity of variance among the control and experimental groups. As the results in the above table show, the Sig. values (.136, .09, .15, and .23) were greater than the Alpha value (.05); therefore, the Null Hypotheses were not rejected. Thus the groups were not significantly different from each other so the variances were equal and the assumption of ANCOVA was not violated.

**Table 3: Descriptive statistics of intermediate and beginner groups**

Groups	Tests	Mean	SD	N
Beginner Control	Listening	17.1	2.32	30
	Reading	19.43	2.21	30
Beginner Experimental	Listening	20.1	2.39	30
	Reading	24.56	2.23	30
Intermediate Control	Listening	24.1	2.15	30
	Reading	25.8	2.17	30
Intermediate Experiential	Listening	27.3	2.14	30
	Reading	33	2.47	30

As the results in table 4 show, the mean scores of beginner control and experimental groups on listening posttest were shown to be 17.1 and 20.1 with stand deviations of 2.35 and 2.17, respectively. Therefore, the mean score of the beginner experimental group was higher than that of control group. The results also indicated that mean scores of control and experimental groups (beginners) on reading posttest were 19.43 and 25.46 , respectively.

Moreover, the results indicated that the mean scores of intermediate experimental groups on listening and reading were shown to be 29.1 and 32.33, respectively. However, the mean scores of control groups on listening and reading tests were shown to be 25.3 and 26, respectively. A two-way Acnova test was run to check whether the difference between control and experimental groups was significant or not. The results are shown in table 4.

**Table 4: The results of Acnova tests for beginner and intermediate learners**

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Beginner groups	Reading	246.214	1	246.214	147.9	.000	.70
	Listening	198.705	1	198.705	109	.000	.60
Intermediate groups	Listening	195.797	1	195.797	116.6	.000	.67
	Reading	773.096	1	773.096	160.8	.000	.74

The results in table 4 show that the difference between the mean scores of beginner language learners on both listening ( $F= 147.9$ ,  $\text{sig} = .000$ ) and reading test ( $F= 109$ ,  $\text{sig} = .000$ ) was significant. The results also indicate that the partial Eta squared of reading was shown to be .70 but that of listening test was .60 which is smaller than the effect of reading test. Therefore, it could be argued that the null hypotheses of the study were safely rejected at p value of less than .001 and the alternative hypotheses were confirmed. That is, self assessment can significantly improve beginner language learners.

Moreover, the results in table 4 (second column) show that the difference between intermediate control and experimental groups on listening test was significant (  $F= 116.6$ ,  $\text{sig.} = .000$ , and  $\text{Eta} = .67$ ). Also, the difference between control and experimental groups on reading was significant ( $F= 160.8$ ,  $\text{sig.} = .000$ ,  $\text{Eta} = .74$ ). Therefore, the null hypotheses which stated that self assessment does not significantly improve intermediate language learners' reading and listening were safely rejected.

### **Discussions**

The present study tested the hypotheses that self assessment does not significantly improve Iranian beginner and intermediate EFL learners' listening and speaking skills. In doing so, participants in experimental groups in both beginner and intermediate level received a 15 session treatment in which they got familiar with self-assessment and its techniques. In addition, they learned how to apply self-assessment in the process of their language learning particularly in reading and listening skills.

The data of the study for both beginner and control participants were analyzed through ACNOVA. From the two -way ACNOVA, several interesting findings were revealed. First, the results of the study showed that there was a significant difference between the mean scores of the beginner participants on both listening and reading posttests. The mean of the experimental groups were significantly higher than the mean scores of the control groups. Therefore, it could be strongly argued that self assessment had significant impact on beginner language learners' reading and listening performance. The findings are therefore consistent with the findings of Stefani, 1998; Orsmond et al., 2000; Taras, 2001; Patri 2002; Hanrahan and Isaacs, 2001; Li, 2001; Smith et al., 2002; and Taras, 2002; ALfallay, 2004; Sung et al., 2005; Dlaska & Krekeler ,2008).

The results are also consistent with the findings of Alibakhshi and Abbasszadeh ( 2012) who found that self assessment significantly improve speaking and writing performance. However, they believed that self assessment has the same impact on both intermediate and beginner language learners.

The results also revealed that the Partial Eta Squared of the reading posttest for beginner participants was .70; whereas, the Partial Eta squared of the listening test was .60. This value also indicates how much of the variance in the dependent variable is explained by the independent variable. If **partial** eta squared value is converted into a percentage by multiplying by 100, then the number 70 is obtained. Therefore, we are able 70 percent of the reading posttest is explained by self assessment, while self assessment can explain only 60 percent of the variance of listening posttest. Therefore, it could be strongly argued that the impact of self assessment on reading skill is higher than its impact on listening skills.

Such a difference is due to the fact listening comprehension is different from reading skill (Nunan, 2003). While listening and reading have some similarities, there two major differences between them. First, listening happens in real time and people listen and have to comprehend immediately without going back, reviewing, and looking up the meaning of unknown words. Second, despite being a receptive skill, it often happened in the midst of a conversation which requires productive and spoken responses (Nunan, 2003).

The results also showed that self assessment had significant impact on improving intermediate language learners listening and speaking skills. Partial Eta Squared of reading and listening tests were .74 and .67, respectively. Therefore, it could also be argued that the impact of self assessment on reading skill of intermediate students is also higher than its effect on listening skills because self assessment could explain 74 percent of the variance of reading posttest; whereas, it could explain only 67 % of the variance of listening posttest

Another interesting finding of the present study which was not reported by related studies such as Alibakhshi and Abasszadeh (2012) was the difference between Partial Eta Squared of posttests taken by beginner students and the tests taken by intermediate language learners. Such a controversy is either due to difference between productive and receptive skills (Nunan, 2003; Harmer, 2009) or due to the affective factors and psychological states of beginner and intermediate language learners as well as the rate of their dependence on the teachers. As stated by Richards and Rodgers (2001), beginner language learners are more dependent on the teachers than advanced

language learners. That is why, beginner language learners in comparison with intermediate or advanced language learners cannot benefit a lot from self-assessment technique.

Therefore, it could be strongly argued that despite the importance of self assessment on any educational system, teachers must somehow support beginner language learners and direct and monitor the self assessments done by the language learners. The difference between the effect of self assessment on listening and reading is deeply rooted in many other factors. The findings of this study are consistent with McDonald and Boud (2003) who found that when learners assess their own learning, their learning will be promoted to a high extent.

The results of the current study are also supported by Butler and Li (2005) who investigated the effectiveness of self-assessment among young EFL learners and found some positive effects of self-assessment on the students' English performance as well as their confidence in learning English. Moreover, in line with a few recent studies (e.g., Black & William, 1998; Pellegrino & Chudowsky, & Aglaser, 2001) it could be argued that formative self assessment has a significant positive effect on students' learning.

### **Pedagogical Implications**

First, self-assessment is believed to encourage increasing sophistication in learner awareness, helping learners to make accurate judgment on their own abilities; acquire how to do an evaluation that covers the whole learning process; and see errors as something helpful. It was also seen as a potentially useful tool for teachers since it provides the information on learning styles, on areas needing remediation and feedback (Barbot, 1991).

Second, textbook and curriculum designers who provide teaching materials for Iranian English learners should pay more attention to learners' needs and include more examples and varieties of the self-assessment in text books, teaching materials and stuff like that, so that the learners have more opportunities for practicing and using self-assessment.

Finally, this project can provide more insights and ideas for further research and can help the future researchers in their work.

### Suggestions for Further Research

While doing every study some other subjects may take the attention of researchers for future. Self-assessment could have a significant impact on other skills such as speaking and writing and may have a relation with some other elements such as gender, age, personality type, etc. So the researchers suggest the following subjects for further research:

1. The Impact of Self-assessment on productive skills
2. The impact of Age on the Use of Self-assessment Strategies.
3. The relationship between personality type and the use of self-assessment techniques.
4. The willingness of language teachers for self-assessing their own performances.
5. The willingness of English language teachers for replacing teacher-made test with students' self-assessment.
6. The relation between the gender and the use of self-assessment technique.

### References

- Alderson, C.A. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. Continuum, New York, NY.
- Alibakhshi, G., Abbasszadeh, S. (2012). The impact of self- assessment on Iranian EFL learners' writing and speaking. Unpublished master thesis. Yasouj university, Yasouj, Iran.
- AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, 32(3), 407–425.
- Bachman, L. F. & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bailey, K. (1998). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-79.
- Barbot, M. J. (1991). New approaches to evaluation in self-access learning (trans. form. French). *Études de Linguistique Appliquée*, 79, 77,-94.



- Black, P., Wiliam, D.(1998). Assessment and classroom learning. *Assessment in Higher Education* 5 (1), 7–74.
- Blatchford, P(997). Students' self assessment of academic attainment: accuracy and stability from 7 to 16 years and influence of domain and social comparison. *Educational Psychology* 17, 354–360.
- Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: the DLIFLC study. *Language Testing*, 7(2), 202-29.
- Birckbichler, D., Corl, K., Deville, C. (1993). The dynamics of language program testing: implications for articulation and program revision. In: *The Dynamics of Language Program Direction*. Heinle & Heinle, Boston, MA.
- Boud, D.( 2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education* 22 (2), 151–167.
- Brantmeier, C.(2006). Advanced L2 learners and reading placement: self-assessment, computer-based testing, and subsequent performance. *System* 34 (1), 15–35.
- Brantmeier, C., 2005a. Nonlinguistic variables in advanced L2 reading: learner's self-assessment and enjoyment. *Foreign Language Annals* 38 (4), 493–503.
- Brantmeier, C., 2005b. Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension. *The Modern Language Journal* 89 (1), 37–53.
- Brown, D. H. 2004. *Language Assessment: Principles and Classroom Practices*. London: Pearson Education, Inc.
- Brown, G., Bull, J., Pendlebury, M., 1997. *Assessing Student Learning in Higher Education*. Routledge, London.
- Butler. Y, Li. J, 2003. The effect of self-assessment among young learners of English. University of Pennsylvania.
- Calderon, M., 1991. Promoting Language Proficiency and Academic Achievement through Cooperation. ERIC Document, ERIC #: ED436983.
- Carton, F. (1993). Self-evaluation at the heart of learning. *Le Français dans le Monde* (special number), 28-35.
- Collins, A.(1993). A study of the Provision of Modern Language Pupils with Special Educational Needs. Unpublished M.ED. dissertation, The Queen's University of Belfast.

- Dearing, R.(1997). *Higher Education in the Learning Society*. London: HMSO.
- Dlaska, A, and Krekeler, C. (2008). *Self-assessment of Pronunciation*. Surrey university press.
- Davidson, F., Henning, G., 1985. A self-rating scale of English difficulty. *Language Testing* 2, 164–169.
- Dickinson, L. (1987). *Self-instruction in language learning*. London: Cambridge University Press.
- Falchikov, N. (1997). Why do lecturers involve students in assessment? In: Paper delivered at the 2nd North Umbria Assessment Conference, Encouraging Partnership in Assessing Learning, 3–5 September, University of Northumbria, Newcastle.
- Freeman, R., Lewis, R., 1998. *Planning and Implementing Assessment*. Kogan Page, London.
- Fallows, S., Balasubramanyan, C., 2001. Multiple approaches to assessment: reflections on the use of tutor, peer and self-assessment. *Teaching in Higher Education* 6, 229–246.
- Gardner, R., 1985. *Social psychology and second language learning*. Arnold, London.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287–322.
- Falchikov, N., Boud, D., 1989. Student self-assessment in higher education: a meta-analysis. *Review of Educational Research* 59 (4), 395–430.
- Graham, S., 2004. Giving up on modern foreign languages? Students perceptions of learning French. *The Modern Language Journal* 88 (2), 171–191.
- Hanrahan, S., Isaacs, G. (2001) . Assessing self- and peer-assessment: the students' views. *Higher Education Research and Development* 20, 53–70.
- Harmer, J. ( 2009). *The practice of English language teaching*. Pearson Longman.
- Heilenman, K. (1991). Self-assessment and placement: a review of the issues. In: Teschner, R.V. (Ed.), *Assessing Foreign Language Proficiency of Undergraduates*, AAUSC Issues in Language Program Direction. Heinle & Heinle, Boston, pp. 93–114.

- Higgins, R., Hartley, P., Skelton, A. (2001). Getting the message across: the problem of communicating assessment feedback. *Teaching in Higher Education* 6 (2), 269–274.
- Ivanic, R., Clark, R., Rimmershaw, R., 2000. What am I supposed to make of this? The message conveyed to students by tutors' written comments. In: Lea, M.R., Stierer, B. (Eds.), *Student Writing in Higher Education: New Contexts*. SHRE/Open University Press, Buckingham.
- Jones, L., Fletcher, C. (2002). Self-assessment in a selective situation: an evaluation of different measurement approaches. *Journal of Occupational and Organizational Psychology*, 75, 145–161.
- Jewah, C., Macfarlane-Dick, D., Matthew, R., Nicol, D., Ross, D., Smith, B.(2004) Enhancing Student Learning through Effective Formative Feedback. In: *The Higher Education Academy*. LTSN, London.
- Janssen-van Dielen, A. (1989). "The development of a test of Dutch as a foreign language: the validity of self-assessment by inexperienced subjects". *Language Testing*, 6 (1): 30-46.
- Lejk, M., Wyvill, M.( 2001). Peer assessment of contributions to a group project: a comparison of holistic and category-based approaches. *Assessment and Evaluation in Higher Education* 26, 62–72.
- Li, L. ( 2005). Some refinement on peer assessment of group projects. *Assessment and Evaluation in Higher Education* 26, 5–18.
- Lindholm-Leary, K., Borsato, G.( 2002). Impact of Two-Way Immersion on Students' Attitudes toward School and College. *Eric Digest*. ERIC Document # ED464541.
- Livesey, D. (1992). An Application of the Theory of Reasoned Action for Relating Attitude, Social Support and Behavioral Intention in an EFL Setting. Paper presented at the Annual meeting of the Teachers of English to Speakers of Other Languages 26th, Vancouver, BC, Canada, 3–7 March, 1992.
- Lewkowicz, J.A., MOON, J. (1985). Evaluation, a way of involving the learner. In J.C. Alderson (Ed.), *Lancaster Practical Paper in English Language Education* (Vol. 6: Evaluation), 45-80. Oxford: Pergamon Press.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75–100.
- Morton, L., Lemieux, C., Diffey, N., Awender, M. (1999). Determinants of withdrawal from the bilingual career track when entering high school. *Guidance and Counseling* 14, 1–14.

- McDonald, B., Boud, D. (2003). The Impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations. *Assessment in Education* 10 (2), 209–220.
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations. *Assessment in Education*, 10, 209–220.
- Nunan, D. (2003). *Practical English language teaching*. Singapore: McGraw Hill.
- Oscarson, M. (1989). Self-assessment of language proficiency: rationale and applications. *Language Testing*, 6(1), 1-13.
- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham & D. Corson (Eds.), *Language testing and assessment* (Vol. 7, 175-87). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6 (1), 1–13.
- Orsmond, P., Merry, S., Reiling, K. (2000). The use of student derived marking criteria in peer and self assessment. *Assessment and Evaluation in Higher Education* 25, 23–38.
- Patri, M., 2002. The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing* 19, 109–131.
- Pellegrino, J.W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academic Press.
- Peirce, B. M., Swain, M., & Hart, D. (1993). Self-assessment, French immersion, and locus of control. *Applied Linguistics*, 14, 25-42.
- Richards, J.C., & Rodgers, T.S., (2001). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.
- Shrauger, J.S., & Osberg, T.M. (1981). The relative accuracy of self-predictions and judgments by others of psychological assessment. *Psychological Bulletin*, 90, 322-351.
- Stefani, L.J., 1998. Assessment in partnership with learners. *Assessment and Evaluation in Higher Education* 23 (4), 339– 350.
- Salters, J., Neil, P., Wright, M., 1998. Modern languages and special educational needs: a

- Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, 78, 450–462.
- Stefani, L. A. J. (1994). Peer, self, and tutor assessment: relative reliabilities. *Studies in Higher Education*, 19, 69–75.
- Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H. T. (2005). The design and application of a Web-based self- and peer-assessment system. *Computers and Education: An International Journal*, 45, 187–202.
- Sung, Y.T., Chang, K.E., Yu, W.C., & Chang, T.H. Enhancing teachers' learning and reflection through structured digital portfolios. *Journal of Computer Assisted Learning*(2009), doi:10.1111/j.1365-2729.2009.00312.x.
- Stankowv, L., 1998. Calibration curves, scatterplots, and the distinction between general knowledge and perceptual tests. *Learning and Individual Differences* 8, 28–51.
- Taras, M., 2001. The use of tutor feedback and student self-assessment in summative assessment tasks: toward transparency for students and for tutors. *Assessment and Evaluation in Higher Education* 26, 605–614.
- Taras, M., 2002. Using assessment for learning and learning from assessment. *Assessment and Evaluation in Higher Education* 27, 501–510.
- Teweles, B., 1995. Motivation as a two-sided coin: motivational differences between college-level Chinese and Japanese learners of EFL. *Texas Papers in Foreign Language Education* 2, 1–22.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249–276.
- Thomson, C.K. (1996). "Self-assessment in self-directed learning: issues of learner diversity". In PEMBERTON, R.; LI, E.; OR, W.; PIERSON, H. (eds.). *Taking control: autonomy in language learning*. Hong Kong: Hong Kong University Press.
- Todd, R. W. 2002. "Using Self-Assessment for Evaluation", *English Teaching Forum* 40(1): 16–19.
- Upshur, J., 1975. Objective evaluation of oral proficiency in the ESOL classroom. In: Palmer, L., Spolsky, B.K. (Eds.), *Paper on Language Testing 1967–1974*. TESOL, Washington, DC.

- Von Elek, T. (1987). A test of Swedish as a second language: an experiment in self-assessment. In: Li, Y., Fok, A., Lord, R., Low, G. (Eds.), *New Directions in Language Testing*. Oxford University Press, Oxford, pp. 47–57.
- Wesche, M., Morrison, F., Ready, D., Pawley, C.(1990). French immersion: postsecondary consequence for individuals and universities. *Modern Canadian Language Review* 46, 430–451.
- Xiao, Y., Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education*, 11(4), 186–193