Comparing Teacher-, Peer-, and Self-Assessment for Vocabulary Learning

Hamid Marashi

ahmuya@yahoo.com, hamid.marashi@iauctb.ac.ir

Associate Professor of Applied Linguistics

Islamic Azad University at Central Tehran

Elnaz Kafayi

MA in TEFL

Islamic Azad University at Central Tehran

Abstract

The present study was an attempt to investigate the comparative effect of teacher-, peer-, and self-assessment on EFL learners' vocabulary achievement. Accordingly, 90 female English language learners selected out of an initial 140 intermediate EFL learners based on their performance on a sample piloted PET took part in the study. All three groups were taught by the same teacher using the same course book. Moreover, they received the same hours of instruction and teaching aids in the same physical environment; therefore, the most significant point of departure in the three experimental groups in the present study was the form of feedback and assessment presented in the classroom. Each group enjoyed its own specific feedback throughout the treatment period. The findings of this study revealed that the participants' vocabulary improved significantly in the self-assessment and peer assessment groups compared to the teacher assessment group. Meanwhile, the learners in the peer- and self-assessment groups bore no significant difference with one another. The findings of this study could be used by EFL teachers and syllabus designers to develop efficient vocabulary teaching procedures.

Keywords: vocabulary learning; assessment; self-assessment; peer assessment; teacher assessment

Introduction

Vocabulary learning is and indeed has always been one of the major concerns in foreign language teaching and learning. The reason underlying this importance is of course very straightforward: "The building blocks of language learning and communication are not grammar, function, notions, or some other unit of planning and teaching but lexis, that is, word and word combination" (Richards & Rodgers, 2001, p. 132). It is thus no wonder then that the ELT literature is simply overwhelmed by an ever-growing array of studies on vocabulary acquisition (e.g., Avila & Sadoski, 1996; De la Fuente, 2002; Genç, 2004; Gu, 2010; Hassani, Sadeghi, & Mohammadi, 2017; Khabiri & Pakzad, 2012; Knight, 1994;

Marashi & Azarmi, 2012; Marashi & Hatam, 2009; McKeown & Curtis, 2014; Naeimi & Foo, 2013).

Research into vocabulary acquisition is of course unceasing and more and more techniques and methods are being investigated as to how the acquisition process could be expedited and facilitated. One approach in ELT which has taken a crucial role in switching from the teacher-centered paradigm to the learner-centered one is assessment (Bachman & Palmer, 2010). Assessment is a means of instruction and continuous evaluation (Bernhardt, 2013) through which teachers not only monitor learning but also endeavor to improve it (Birjandi, & Ahmadi, 2013). To this end, assessment is the process of gathering and discussing information from multiple and diverse sources whereby assessment results are used to enhance subsequent learning (Huba & Freed, 2000; Leung, 2014). A multitude of studies on the different modes of assessment, namely peer assessment, teacher assessment, and self-assessment has been carried out (e.g., Brown, Dewey, & Cox, 2014; Cheng & Warren, 2005; Gómez Sará, 2016; Lim, 2007; Puegphrom & Chiramanee, 2013; Wang, 2016) with many such studies also in the Iranian EFL context (e.g., Ariafar & Fatemipour, 2013; Ashraf & Mahdinezhad, 2015; Baleghizadeh & Masoun, 2014; Fahimia & Rahimi, 2015; Ghaslani, 2015; Sarani & Izadi, 2016; Weisi & Nabi Karimi, 2013).

In line with what has been discussed so far, this study sought to investigate the effect of peer-, teacher-, and self-assessment on EFL learners' vocabulary learning. Accordingly, the following research question was formulated:

- *Is there any significant difference among the effect of teacher-, peer-, and self-assessment on EFL learners' vocabulary achievement?*

Review of the Related Literature

Vocabulary

Vocabulary knowledge is crucial for fluent language use and is indeed viewed by many to be the single most important aspect of foreign language learning (Nation, 2001). This importance is not only a priority in the eyes of L2 learners but also often "a priority by teachers as well" (Knight, 1994, p. 1) since most language teachers – if not all – would assert that vocabulary is the main bounding block of all four language skills that should be mastered by language learners (Schmidt, 2014). Learning the L2 vocabulary and recalling it during the use of language as Nation (2010) stated is "a crucial and fundamental issue for language acquisition" (p. 12). Harmer (2001) has also asserted that vocabulary learning is the most significant task an L2 learner should engage in because vocabulary is the core of any manifestation of the language and is indispensable in language learning.

When encountering a word for the first time, the information about it is also connected to information from the context (Gu, 2010; Smith, 2004). A full and flexible knowledge of a word thus involves understanding the core meaning and how it changes in different contexts (Thornbury, 2002). To this end, one way to see the overall task of vocabulary learning is through the distinction between *knowing* a word and *using* it; in other words, the purpose of vocabulary learning should include both remembering words and the ability to use them automatically in a wide range of language contexts when the need arises (Erten & Tekin,

2008; Martinez & Schmidt, 2015). Vocabulary teaching techniques, therefore, should include strategies for using as well as knowing a word (Baracraft, 2004; Richards & Renandya, 2002).

Schmidt (2005) details the meaning of a word by elaborating that each word is in essence described as the collection of properties it bears. He specifically identifies the seven features of "register, association, collocation, grammatical behavior, written form (spelling), spoken form (pronunciation), and frequency" (p. 5) alongside meaning as the composite which represents word knowledge. Naturally, knowing and using a word efficiently, he argues, necessitates not only learning its meaning but also mastery of these seven aspects.

In addition to the process of learning words, vocabulary size is also an indicator of how well L2 learners can perform all the four language skills of speaking, reading, listening, and writing (Blachowiez & Fisher, 2004; Martinez & Schmitt, 2015; Nation, 2010). According to Nation (2006), knowledge of around 3000 word families is the threshold needed for tapping other language skills. Without this threshold, learners encounter problems understanding the language they are exposed to (Alderson & Banerjee, 2002; Read, 2004; Tschirner, 2004; Zimmerman, 2005).

Undoubtedly, all second language learners and their teachers are virtually well aware of the fact that learning L2 includes the learning of large numbers of words correctly (Carter, 1998; Laufer & Hulstijn, 2001; McCrostie, 2007) but how to accomplish this task is often their great concern. According to De La Fuente (2002), "It is crucial to determine the most effective way of teaching vocabulary and identify the most efficient means to promote effective acquisition of vocabulary" (p. 88).

Assessment

Assessment refers to any method, strategy, or tool a teacher may use to collect evidence about student progress toward achievement of established goals (Kunnan, 2013). In educational systems, assessment is considered highly significant as it may influence the learning and teaching process. In fact, when assessment is one of the main factors in the language classroom, it paves the way for the feedbacks and revisions which in their own turn can improve learning (Alias, Masek, & Salleh, 2015). Furthermore, "Through meaningful engagement of students in the learning process, assessment can affect motivation. Assessment would also enhance instruction by helping the teacher recognize students' weaknesses and strengths" (Orsmond, Merry, & Reiling, 1996, p. 240).

Teacher Assessment

Assessment appears in different typologies, one such being the trichotomy of teacher assessment (TA), peer assessment (PA), and self-assessment (SA). In TA, teachers rely on assessment as a process that can act as a monitor of and a guide to the progress of EFL teaching programs (Nicol & Macfarlane-Dick, 2006). TA can be used for administrative purposes such as classifying students and selecting them for further education or training (Ableeva & Lantolf, 2011). Advocates of TA such as Amo and Jareno (2011) argue that TA

can be used for motivating students to participate more in the teaching activities and that this mode of assessment "is a reliable means of getting the information required for evaluating educational interventions or variables in the educational system" (p. 41). These data can be considered as indicators for the appropriateness of any of the teaching process components – teaching method, syllabus, teacher education, etc. (Baleghizadeh & Masoun, 2014). Thus, it can be the basis of the evaluation process assumed after each teaching phase to make decisions as to exclude ineffective interventions in the teaching process or suggest new ones in the re-teaching process (Noonan & Duncan, 2005).

Peer Assessment

PA, in which learners assess the work of their classmates, is a kind of learning that lets learners provide feedback on each others' work (Falchikov, 2001) and marks may be awarded by students or negotiated with teachers (Liu & Carless, 2006). These assessments bring a closer relationship among teachers and students as well as helping students develop useful skills in academic and professional areas (Spiller, 2012).

PA is an alternative of the assessment process that involves learners' participation and can reflect the effectiveness of their feedback and cooperation as well as enhance their awareness of self-learning and self-esteem (Vu & Dall'Alba, 2007; White, 2009). Glasson (2009) suggested that in PA, learners have to determine the standard or quality of the performance of the learners of the same academic level, plan the learning process together, and indicate their peers' strengths and weaknesses. PA has also been considered an important part of the writing process that helps improve writing ability. The merits of PA have been well documented (e.g., Conrad & Goldstein, 1999; Gómez Sará, 2016; Min, 2006; Nowrozi Larsari & Sadegh Oghli, 2016).

Self-Assessment

In an attempt to evaluate learners' progress, teachers may be inclined toward the socioculturally oriented types of assessment in the forms of PA and SA (Boud, 2013; Lindblom-Ylänne, Pihlajamäki, & Kotkas, 2006). As Glasson (2009) argues, learners become aware of their ability while assessing their own language proficiency and their progress. By recognizing their weaknesses, learners would then be able to seek help with these areas in which they need to improve (Mican & Medina, 2015) and they would ultimately see how close or how distant their actual language proficiency is from the level they wish to achieve (Andrade & Du, 2007).

SA has been discussed as an important tool for autonomous language learners in that it can be regarded both "as a testing device leading to accreditation and as a device for personal self-monitoring" (Gardner, 2000, p. 49). There is ample evidence in the ELT literature that SA can be effective as it can play a significant role in helping learners to monitor their learning progress and that their evaluative abilities could be enhanced (e.g., Brew, 1995; Butler & Lee, 2006; Iraji, Enayat, & Momeni, 2016; Kayler & Weller, 2007; Shams & Tavakoli, 2014).

Method

Participants

The main participants of the present study were 90 female English language learners aged between 15 and 18, selected out of an initial 140 intermediate EFL learners based on their performance on a language proficiency test (all those whose scores fell one standard deviation below and above the mean). The participants of the study were thus divided into three homogenous groups with 30 learners in each. These three groups were randomly assigned as the peer assessment group (PAG), teacher assessment group (TAG), and self-assessment group (SAG). Another 30 learners sharing almost the same English language background with the main participants of the study participated in the piloting of the tests utilized in this study.

Furthermore, the two researchers who enjoyed inter-rater reliability (r = 0.82, p = 0.0001 < 0.05) scored the participants' writing papers.

Instrumentations and Materials

Preliminary English Test (PET)

A sample PET was administered for the participant selection process as described above. For the assessment of parts two and three of the writing section, the researchers used the PET general mark scheme used as a rubric for a summative score. According to this scale, the criteria include language range, variety, and complexity of message communication, grammatical structure, vocabulary, spelling, punctuation, content points, length, and target reader. Furthermore, the reliability of the test during the piloting and main administration stood at 0.84 and 0.93, respectively.

Test of Vocabulary

To make sure that the participants were not familiar with the vocabulary to be taught, a teacher-made test of vocabulary was administered at the outset. This multiple-choice test comprising 45 items was developed by the researchers based on the words in the learners' textbook (described below) which were taught during the treatment procedure. The test was administered at the outset to make sure that the items were not known by the learners.

After the instruction, the learners in the three instructional groups received the same test as the posttest of vocabulary. Prior to this final administration, however, the posttest was piloted and five faulty items were removed after item analysis leaving 40 items in the posttest. The learners were given 30 minutes in each administration of the test to respond to the items.

Course Book

Units 6-9 of Book 2 of the intermediate level book of *Top Notch* (Saslow & Ascher, 2011) series were used in the present study. Top Notch is an innovative four-level series for adults and young adults, taking students from beginning to intermediate levels and based on the Common European Framework of Reference for languages, the series covers A1 to B1. The series presents natural language in authentic contexts and explicitly develops conversation strategies so that learners could speak with fluency and confidence.

Procedure

Once the three groups – TAG, PAG, and SAG – were established as described above and the vocabulary test was administered, the treatment commenced. All three groups were taught by the same teacher (one of two researchers) using the same course book. Moreover, the 90 learners received the same hours of instruction (12 sessions of 90 minutes each during six weeks) and teaching aids in the same physical environment; therefore, the most significant point of departure in the three experimental groups in the present study was the form of feedback and assessment presented in the classroom. Furthermore, in all three groups, the dynamic assessment model of Poehner and Lantolf (2005) was integrated.

To shed light on the actual procedure, one example is described below: The first topic in the first lesson of all three groups was making friends and the covered vocabularies were TV shows, clothes, food, and weekend activities. Also, the students were to cover an article about giving advice. The students were very interested in this issue so the teacher/researcher decided to use it for an interview. Accordingly, each student wrote a series of questions about the topic and had to answer them.

The feedback process in the three groups was of course different. In the TAG, the teacher herself provided the feedback and assessed the students' process of learning. Throughout the instruction process, the teacher monitored the students' progress and provided feedback on their strengths and weaknesses in the TAG. The teacher/researcher in this study provided the same hints for all learners and gave feedback from the most implicit to the most direct and explicit based on the needs of the learners in each of the three groups.

In the PAG, the teacher informed the learners of the role they had to play in the classroom to either score each others' performances or rate one another's language development. The students were trained to develop checklists and use them in the classroom to keep records of their classmates' performances. The peers provided feedback if there were any mistakes and they acted as mediators. The learners were divided into small groups of four to six and each group worked together both in the instruction phase and the assessment phase. The students were provided feedback by their peers in the group regarding their questions and answers. The teacher ultimately checked their work and only corrected the mistakes by writing the correct forms.

In the SAG, the learners were firstly familiarized with the concept of SA and they were trained to develop SA checklists and scored their own performance based on the checklists they had developed. Subsequently, they were asked to say how much they were ready for the coming steps and procedures. Also, they were asked to assess themselves at the end of each

session of the classroom and see how well they had learned the vocabularies taught. Of course, they received relative feedback by the teacher, something which was decreased as the learners increased in the quality of their SA. After answering the questions, the learners themselves checked the answers to see if they had been correct or not. Then they evaluated themselves as being excellent, good, bad, or in need of more work.

At the end of the treatment in the three experimental groups, all the learners took the same vocabulary test as the posttest.

Results

Participant Selection

To select the participants required in this study, the researchers used a sample PET – as described earlier. Prior to the actual administration, the test was piloted to make sure that it could be used confidently for this screening. Table 1 shows the descriptive statistics of this administration. As can be seen, the mean and standard deviation of the scores stood at 48.34 and 11.76, respectively.

Table 1

Descriptive Statistics of the PET Administration

	N	Minimum	Maximum	Mean	Std. Deviation
PET Administration	140	22	73	48.34	11.755
Valid N (listwise)	140				

Posttest

The researchers administered the vocabulary posttest among the three experimental groups once the treatment was completed. Table 2 below shows the descriptive statistics of this administration disaggregated by the three groups. As is seen in the table, the mean and standard deviation of the TAG were 29.93 and 4.78, respectively, while those of the PAG

stood at 33.53 and 2.65, respectively. As for the SAG, the mean was 34.23 and the standard deviation 4.21.

Table 2

Descriptive Statistics for the Posttest in the Three Groups

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
TAG	30	17	39	29.93	4.785	-.796	.427
PAG	30	27	38	33.53	2.649	-.660	.427
SAG	30	22	40	34.23	4.207	-.788	.427
Valid N (listwise)	30						

Testing the Hypothesis

To verify the null hypothesis of the study, i.e. there is no *significant difference among the effect of TA, PA, and SA on EFL learners' vocabulary achievement*, the researchers set out to conduct a one-way ANOVA on the scores. Prior to this, the normality of the distribution of these scores within each group had to be checked. Going back to Table 2, the skewness ratios of all three groups fell within the acceptable range of ±1.96 (-1.84, -1.54, and -1.86) thus signifying that the score distributions in all groups represented normality. Table 3 below shows that the variances among the three groups were not significantly different ($F_{(2,87)}$ = 5.408, p = 0.24 < 0.05). Hence, running a one-way ANOVA was legitimized.

Table 3

Levene's Test of Homogeneity of Variances

F	df1	df2	Sig.
3.886	2	87	.24

Subsequently, Table 4 below shows that the mean scores of the three groups on the vocabulary posttest bore a significant difference ($F_{(2,87)}$ = 10.063, p = 0.0001 > 0.05). Hence, the null hypothesis of the study was rejected meaning that the practice of TA, PA, and SA has a significantly different impact on EFL learners' vocabulary.

Table 4

One-Way ANOVA of the Mean Scores of the Three Groups on the Posttest

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	319.400	2	159.700	10.063	.000
Within Groups	1380.700	87	15.870		
Total	1700.100	89			

To identify which groups had significant differences with one another, both Tukey HSD and Scheffe post hoc tests were run.

Table 5

Multiple Comparisons

	(I) Group 1	(J) Group 1	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1	2	.700	1.029	.775	-1.75	3.15
		3	4.300*	1.029	.000	1.85	6.75
	2	1	-.700	1.029	.775	-3.15	1.75
		3	3.600*	1.029	.002	1.15	6.05
	3	1	-4.300*	1.029	.000	-6.75	-1.85
		2	-3.600*	1.029	.002	-6.05	-1.15
Scheffe	1	2	.700	1.029	.794	-1.86	3.26
		3	4.300*	1.029	.000	1.74	6.86
	2	1	-.700	1.029	.794	-3.26	1.86
		3	3.600*	1.029	.003	1.04	6.16
	3	1	-4.300*	1.029	.000	-6.86	-1.74
		2	-3.600*	1.029	.003	-6.16	-1.04

*. The mean difference is significant at the 0.05 level.

Table 5 above shows that while the mean of the scores of the participants in the SA and PA groups bore no significant difference with one another, those who underwent the TA instruction had significantly lower results than the participants in the SA and PA groups.

Discussion

The findings of this study revealed that the participants' vocabulary improved significantly in the SA and PA groups compared to the TA group. Meanwhile, the learners in the SA and PA groups bore no significant difference with one another. The present findings are in line with those of a good number of previous studies focusing on the effects of SA and PA on the EFL/ ESL development (e.g., Brown et al., 2014; Lindblom-Ylänne et al., 2006; Mican & Medina, 2015; Noonan & Duncan, 2005; Spiller, 2012; Wang, 2016). A number of studies also in Iran verify this finding (e.g., Ariafar & Fatemipour, 2013; Ashraf & Mahdinezhad, 2015; Baleghizadeh & Masoun, 2014; Salimi, Kargar, & Behjat, 2014; Shams & Tavakoli, 2014; Weisi & Nabi Karimi, 2013).

A number of studies such as those of Andrade and Du (2007) and Lim (2007) which compared the effects of SA and PA on learners' oral proficiency development found that PA could be more beneficial to intermediate level learners as through this form of assessment, learners cooperatively help each other and provide one another with corrective feedback. The findings of the present study, however, revealed no significant difference between SA and PA in this regard. This difference might lie in the specific nature of oral proficiency and speaking in which peer correction could be more useful, while the development of vocabulary might not necessarily follow that modality.

The present study also takes support from Amo and Jareno's (2011) study on SA and TA as active learning methods that revealed that "third-year students are more critical when they evaluate themselves because their self-evaluation scores were always lower than the teacher assessment score" (p. 44). This signifies that fact that SA can help learners improve their L2 learning abilities even more than TA.

In the process of instruction in the three groups, the researchers observed that the learners' assessment became a more inclusive procedure in which they were able to participate and not merely rely on the teacher's judgment. Furthermore, both PA and SA have more novelty for EFL learners (Weisi & Nabi Karimi, 2013) while TA seems to have lost its novelty in the EFL classrooms as the learners prefer to be engaged in something new. This engaging in a new activity may have well been a source of additional motivation for the learners which per se facilitates and enhances learners' language learning.

Another reason for the lack of success of TA in this regard can be pursued in the active role that PA and SA provide the learners with (Wang, 2016). Through PA and SA, learners develop more consciousness toward classroom discussions as they are expected to evaluate either their own performance or that of their peers. Accordingly, they get more involved in the assessment process and focus on it as a method of learning and performance development.

Conclusion

The positive impact of PA in EFL classrooms paves the way for providing an atmosphere in which learners improve their second language skills eagerly in a cooperative mode (Amo & Jareno, 2011). At the same time, employing user-friendly tasks aiming at facilitating the retention of vocabulary items through word identification and code-breaking as SA techniques while reading the texts accompanied with enriched texts have also been recorded as being effective (Laufer & Hulstijn, 2001).

SA and PA could both be employed by second language teachers to make the learners more aware of what they are dealing with. The assumption is that these assessment types and their related tasks can facilitate learning (Ashraf & Mahdinezhad, 2015; Lim, 2007) and learners enjoy a cooperative mode in the language classroom and pay attention to their peers' development. In this way, cooperation and collaboration will be energized and competition will be minimized (Azarnoosh, 2013). Through classroom discussions ensued from SA and PA issues, learners combine classroom discussions and develop an acceptable level of L2 vocabulary (Gómez Sará, 2016).

English teachers and learners could employ both PA and SA in their classes to facilitate learning. This could contribute to the enrichment of classroom interactions and would help learners' subsequent L2 development.

Syllabus designers and materials developers in the ELT domain also could employ the findings of the present study and those of similar ones to present tasks in which learners' awareness toward L2 vocabulary, through assessment types, is enhanced. Such tasks may help the learners move towards PA and SA.

Last but not least, the researchers would recommend the following two cases for further research:

1. All the learners participating in the present study were females and teenagers; the gender and age of the learners could be changed in another study of the same type to see whether the two serve as decisive variables or not.

2. The present study compared the impacts of PA, SA, and TA in the L2 vocabulary classroom through an immediate posttest. The residual effects of SA and PA on the development of L2 vocabulary could be examined in another study. Future studies might consider examining to explore whether and how long-term these effects could actually be. Various intervals might affect the retention issue differently. A semi-longitudinal study of the concept of L2 vocabulary in the SA and PA supported contexts on a specific group of learners can perhaps reveal if this theory energizes retention of vocabulary items in the learners' mentality or not.

References

Ableeva, R., & Lantolf, J. P. (2011). Mediated dialogue and the micro-genesis of second language listening comprehension. *Assessment in Education*, *18(3),* 133-149.

Alderson, J. C., & Banerjee, J. (2002). State of the art review: Language testing and assessment (part two). *Language Teaching, 35*(2), 79-113.

Alias, M., Masek, A., & Salleh, H. D. (2015). Self, peer and teacher assessments in problem based learning: Are they in agreements? *Social and Behavioral Sciences, 204,* 309-317.

Amo, E., & Jareno, F. (2011). Self, peer and teacher assessment as active learning methods. *Research Journal of International Studies, 18,* 41-48.

Andrade, H. & Du, Y. (2007). Student responses to criteria-referenced self-Assessment. *Assessment and Evaluation in Higher Education, 32*(2), 159-181.

Ariafar, M., & Fatemipour, H. R. (2013). The effect of self-assessment on Iranian EFL learners' speaking skill. *International Journal of Applied Linguistics & English Literature, 2* (4), 25-32.

Ashraf, H., & Mahdinezhad, M. (2015). The role of peer-assessment versus self-assessment in promoting autonomy in language use: A case of EFL learners. *Iranian Journal of Language Testing, 5* (2), 110-120.

Avila, E., & Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning*, *46*, 379-395.

Azarnoosh, M. (2013). Peer assessment in an EFL context: Attitudes and friendship bias. *Language Testing in Asia, 3*(11), 12-24.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice.* Oxford: OUP.

Baleghizadeh, S., & Masoun, A. (2014). The effect of self-assessment on EFL learners' self-efficacy. *TESL Canada Journal, 31*(1), 42-58.

Baracraft, J. (2004). Effects of sentence writing in L2 lexical acquisition. *Second Language Research, 20,* 303-334.

Bernhardt, V. (2013). *Data, data, everywhere: Bringing all the data together for continuous school improvement*. London: Routledge.

Birjandi, P., & Ahmadi, H. (2013). Authenticity in second language assessment: A social-constructivist perspective. *Advances in Asian Social Science*, *4*(3), 899-903.

Blachowiez, C. L., & Fisher, P. (2004). Vocabulary lessons. *Educational Leadership, 61*(6), 66-69.

Boud, D. (2013). *Enhancing learning through self-assessment (2nd ed.).* London: Routledge.

Brew, A. (1995). What is the scope of self-assessment? In D. Boud (Ed.), *Enhancing learning through self-assessment* (pp.48-63). London: Kogan Page.

Brown, N. A., Dewey, D. P., & Cox, T. L. (2014). Assessing the validity of can-do statements in retrospective (then-now) self-assessment. *Foreign Language Annals, 47*(2), 261-285.

Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal*, *90*(4), 506-518.

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives*. New York: Routledge.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, *22*(1), 93-121.

Conrad, S. M., & Goldstein, L. M. (1999). ESL student revision after teacher-written comments: Text, contexts, and individuals. *Journal of Second Language Writing, 8*(2), 147-179.

De la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary. *Studies in Second Language Acquisition, 24*, 81-112.

Erten, I. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System, 36*, 407-422.

Fahimia, Z., & Rahimi, A. (2015). On the impact of self-assessment practice on writing skill. *Social and Behavioral Sciences, 192,* 730-736.

Falchikov, N. (2001) *Learning together: Peer tutoring in higher education*. London: Routledge.

Gardner, D. (2000). Self-assessment for autonomous language learners. *Links and Letters, 7*(1), 49-60.

Genç, B. (2004). New trends in teaching and learning vocabulary. *Journal of Çukurova University Institute of Social Sciences*, *13*(2), 117-126.

Ghaslani, R. (2015). The effect of self-assessment on Iranian EFL learners' reading comprehension skill. *Journal of Academic and Applied Studies (Special Issue on Linguistics & English Literature), 5*(3), 1-9.

Glasson, T. (2009). *Improving student achievement: A practical guide to assessment for learning*. Carlton, AU: Curriculum Corporation.

Gómez Sará, M. M. (2016). The influence of peer assessment and the use of corpus for the development of speaking skills in in-service teachers. *HOW, 23*(1), 103-128.

Grigorenko, E. L., & Sternberg, R. G. (2002). *Dynamic testing: The nature and measurement of learning potential*. New York: Cambridge University.

Gu, Y. (2010). Learning strategies for vocabulary development. *Reflections on English Language Teaching, 9*(2), 105-118.

Harmer, J. (2001). *The practice of English language teaching (3rd ed.)*. Edinburgh: Pearson Education.

Hassani, M. T., Sadeghi, B., & Mohammadi, M. (2017). The impact of games on EFL Iranian learner's vocabulary proficiency in second grade of high school. *Journal of Teaching English Language Studies, 6*(1), 18-30.

Huba, M. E., & Freed, J. E. (2000). *Learner-centered assessment on college campuses: Shifting the focus from teaching to learning*. Needham Heights, MA: Allyn & Bacon.

Iraji, H. R., Enayat, M. J., & Momeni, M. (2016). The effects of self-and peer-assessment on Iranian EFL learners' argumentative writing performance. *Theory and Practice in Language Studies, 6*(4), 716-722.

Kayler, M., & Weller, K. (2007). Pedagogy, self-assessment, and online discussion groups. *Educational Technology and Society*, *10*(1), 136-147.

Khabiri, M., & Pakzad, M. (2012). The effect of teaching critical reading strategies on EFL learners' vocabulary retention. *Journal of Teaching Language Skills, 31*(1), 73-106.

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal, 78*(3), 285-299.

Kunnan, A. J. (2013). *Validation in language assessment*. London: Taylor & Francis.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*, 1-26.

Leontiev, A.N. (1981). *Problems of the development of mind. Moscow*: Progress Press.

Leung, C. (2014). Exploring language assessment and testing. *ELT Journal*, *2*(1), 123-140.

Lim, H. (2007). A Study of self-and peer-assessment of learners' oral proficiency. *Language and Linguistics, 2*(2), 169-176.

Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self, peer, and teacher assessment of student essay. *Active Learning in Higher Education, 7*(1), 51-62.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11(*3), 279-290.

Marashi, H., & Azarmi, A. (2012). The comparative effect of presenting words in semantically related and unrelated sets in intentional and incidental learning contexts on Iranian EFL learners' vocabulary learning. *Journal of Second Language Teaching and Research, 1*(2), 71-89.

Marashi, H., & Hatam, G. (2009). Using tasks in a content-based program: Does it enhance vocabulary learning? *Iranian Journal of Teaching English as a Foreign Language and Literature, 1*(4), 73-91.

Martinez, R., & Schmitt, N. (2015). Vocabulary. In Biber, D. &Reppen, R. (Eds.), *Cambridge handbook of English corpus linguistics* (pp. 120-132). Cambridge: Cambridge University Press.

McCarthy, M., McCarten, J., & Sandiford, H. (2013). *Touchstone* (2nd ed.). Cambridge: Cambridge University Press.

McCrostie, J. (2007). Examining learner vocabulary notebooks. *ELT Journal, 61*(3), 246-255.

McKeown, M. G., & Curtis, M. E. (2014). *The nature of vocabulary acquisition*. New York: Psychology Press.

Mican, A. D., & Medina, L. C. (2015). Boosting vocabulary learning through self-assessment in an English language teaching context. *Assessment and Evaluation in Higher Education, 2*, 13-35.

Min, H. T. (2006). The effect of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15,* 118-141

Naeimi, M., & Foo, T. C. V. (2013). The study of direct vocabulary learning strategies in reading comprehension: The case of Iranian context. *Journal of Higher Education Theory and Practice, 2(1*), 95-98.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press**.**

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82.

Nation, I. S. P. (2010). *Second language vocabulary development.* New York: Newbury.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31(2*), 199-218.

Noonan, B., & Duncan, R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research and Evaluation, 10*(17), 1-17.

Nowrozi Larsari, V., & Sadegh Oghli, H. (2016). On the effect of self-assessment and peer-assessment on Iranian EFL learners' learner autonomy. *Journal of Language and Linguistics, 2* (1), 26-31.

Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education, 21*(3), 239-250.

Poehner, M., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research, 9*(3), 233-265.

Puegphrom, P., & Chiramanee, T. (2013). The effectiveness of implementing peer assessment on students' writing proficiency. The 3rd International Conference on Humanities and Social Sciences. April 2, 2011, Faculty of Liberal Arts, Prince of Songkla University. *Proceedings: Factors Affecting English Language Teaching and Learning.* Songkla, Thailand: Prince of Songkla University Press.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146-161.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.

Salimi, M., Kargar, A., & Behjat, F. (2014). Writing self-assessment and its impact on EFL learners' writing ability. *International Journal of Language Learning and Applied Linguistics World (IJLLALW), 6* (4), 587-604.

Saslow, J., & Ascher, A. (2011). *Top notch*. London: Pearson.

Sarani, A., & Izadi, M. (2016). Diagnosing L2 receptive vocabulary development using dynamic assessment: A microgenetic study. *Journal of Teaching Language Skills, 35*(2), 161-189.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64(*4), 913-951.

Shams, N., & Tavakoli, M. (2014). The effect of peer, self, and traditional assessment on Iranian EFL learners' L2 reading comprehension. *Journal of Applied Linguistics and Language Research*, *1*(1), 29-44.

Smith, F. (2004). *Understanding reading*. New York: LEA.

Spiller, D. (2012). Assessment matters: Self-assessment and peer assessment. *Teaching Development Journal, 2*, 1-18.

Thornbury, S. (2002). *How to teach vocabulary*. Harlow: Longman.

Tschirner, E. (2004). Breadth of vocabulary and advanced English study: An empirical investigation. *Electronic Journal of Foreign Language Teaching*, *1*(1), 27-39.

Vu, T. T., & Dall'Alba, G. (2007). Students' experience of peer assessment in a professional course. *Assessment and Evaluation in Higher Education, 32*(5), 541-556.

Wang, W. (2016). Using rubrics in student self-assessment: student perceptions in the English as a foreign language writing context. *Assessment and Evaluation in Higher Education*, *2*(2), 1-13.

Weisi, H., & Nabi Karimi, M. (2013). The effect of self-assessment among Iranian EFL learners. *Procedia – Social and Behavioral Sciences, 70*, 731-737.

White, E. (2009). Student perspectives of peer assessment for learning in a public speaking course. *Asian EFL Journal – Professional Teaching Articles, 33*, 1-55.

Zimmerman, K. (2005). Newly placed versus continuing students: Comparing vocabulary size. *TESL Reporter, 38*(1), 52-60.