The Effect of Diagnostic Assessment on EFL Learners' Performance on Selective and Productive Reading Tasks

Fateme Nikmard[1], Kobra Tavassoli[2*]

[1]ELT Department, Karaj Branch, Islamic Azad University, Karaj, Iran,
fatemehnikmard@gmail.com

[2*]ELT Department, Karaj Branch, Islamic Azad University, Karaj, Iran,
kobra.tavassoli@kiau.ac.ir

Abstract

The main purpose of diagnostic assessment is to predict test-takers' strengths and weaknesses and use the information to provide suitable feedback to them (Jang & Wagner, 2014). Accordingly, the present study focused on investigating the effect of diagnostic assessment on selective and productive reading tasks to help English as a foreign language (EFL) learners improve their performance on these tasks. The quasi-experimental pretest-posttest design was used and 60 participants were chosen based on the results of a Nelson proficiency test. The learners were chosen from four intact classes who were then divided into two groups of control and diagnostic assessment, each group comprising of two classes. Then, a modified version of the reading section of Key English Test (KET) (2015) with 28 selective and 27 productive items was used in the two groups as the pretest. The diagnostic assessment group received a feedback-based treatment on four reading tests during sixteen sessions of instruction. However, in the control group, the learners focused on reading texts and the corresponding tasks during the course. At the end of the semester, another modified version of KET reading test (2014) with 28 selective and 27 productive items was administered as the posttest. The results showed the diagnostic assessment group showed a significant improvement on both selective and productive tasks. Teachers may be the most beneficiaries of the present study as they can find useful information about their students' strengths and weaknesses through using diagnostic assessment to help them improve their reading comprehension ability.

Keywords: Assessment, Diagnostic Assessment, Productive Tasks, Reading Comprehension, Selective Tasks

1. Introduction

Almost all teachers have heard their students nagging about the areas they had problems with while taking an exam. The students usually want to know their errors to avoid them in subsequent exams, but are they well-equipped to overcome their errors? Do they really know the best way of amending them? In fact, it has been argued by Jang and Wagner (2014) that the traditional assessment is not appropriate to provide such information needed to help students because there are not any special kinds of feedback to be presented to the students to help them solve their problems. That is why the focus has shifted into different new forms of assessment, especially diagnostic assessment, to overcome such problems. The main goal of diagnostic assessment is to predict a test-taker's strengths and weaknesses in a test and use the information to provide the best kind of feedback needed to improve learning (Jang & Wagner, 2014). In spite of the fact that diagnostic assessment is very important in educational contexts, there is not enough literature about the process of putting this kind of assessment into practice. Of course, there have been some studies about diagnostic assessment in first language (L1) contexts but not much in second or foreign language contexts (Harding, Alderson, & Brunfaut, 2015).

On the other hand, in second or foreign language contexts, the reading skill becomes extremely important because it is the first priority of language learners who read in a language they cannot understand fully and their problems are mainly language related rather than reading related (Alderson, Haapakangas, Huhta, Nieminen, & Ullakonoja, 2015). Hence, to diagnose reading difficulties in a second or foreign language context, the teacher or the learner has to diagnose language problems, too. If a learner has difficulties with reading, the best way of resolving the difficulties is to find out the sources of the difficulties (Nation, 2009). One such useful way is to use a well-defined technique, such as diagnostic assessment as a special way of providing feedback, to deal with the problem, especially while performing on different types of reading comprehension tasks including selective and productive tasks. The purpose of this study was therefore to identify the effect of providing diagnostic assessment on EFL learners' performance on selective and productive reading comprehension tasks. Accordingly, the following research questions were proposed:

1. Does diagnostic assessment have any significant effect on EFL learners' performance on selective reading comprehension tasks?

2. Does diagnostic assessment have any significant effect on EFL learners' performance on productive reading comprehension tasks?

3. Is there any significant difference between EFL learners' performance on selective and productive reading comprehension tasks using diagnostic assessment?

In what follows, a brief review of the related literature on the main variables of this study, including reading comprehension, different kinds of reading tasks, and diagnostic assessment is presented.

2. Literature Review

2.1. Reading Comprehension

Reading is the most important academic language skill (Grabe & Stoller, 2001). It is the process of identifying and comprehending the meaning of a number of words, finding their grammatical relationship, and building a coherent meaning of the whole text (Uso´-Juan & Martinez-Flor, 2006). "It is a complicated skill because it requires the combination of "attention, memory, perceptual processes, and comprehension processes" (Kern, 1989, p. 135). In fact, reading is a meaning-making process including an interaction between the reader and the text in which readers use mental activities to construct meaning from the text (Goodman, 1988). The reading process is an active cognitive system operating on printed material to comprehend the text (Chastain, 1988).

According to Oakhill, Cain, McCarthy, and Nightingale (2013), a text is a combination of words, sentences, and paragraphs. To understand a text, a reader has to first recognize the words within it, that is, vocabulary knowledge is an essential part of understanding a text. In other words, knowing the meaning of words is a great help in comprehending a text. The point, however, is that there are also other skills which should work in coordination to make the reader able to understand a text, e.g., some higher-level skills such as inference making, comprehension monitoring, and understanding structures are involved in comprehending at text.

Because of the importance of the reading skill, a lot of research has been done on various aspects of reading comprehension. For instance, Cai and Kunnan (2018) examined the inseparability of content knowledge from reading ability in language for specific purposes. To do so, they used 1,491 Chinese nursing students as their participants to predict the variance of the domain-general reading factor and at the end they concluded that content knowledge cannot be separated from reading ability in language for specific purposes as it is a good tool to predict the variance of the domain-general reading factor.

Reading comprehension has also been identified as a good means of helping learners' vocabulary learning. Serrano and Huang (2018) working on

learning vocabulary through assisted repeated reading found out that intensive work on reading causes the learners to pick up the vocabulary faster whereas practice on reading with time intervals causes greater long-term retention.

The effect of first language (L1) on the learners' reading ability in a second language (L2) was also investigated by Kim (2018) who conducted a study on 136 university mixed proficiency Korean English language learners to complete pre- and post-reading sorting maps in L1 and L2. According to the outcomes of the study, low proficiency learners were better able to understand L1 post-reading sorting maps while high proficiency readers were equally good in both L1 and L2 post-reading sorting maps.

2.2. Reading Comprehension Assessment Tasks

The most common question about assessment is about its necessity, a question which Geoffrey and Roberts (1999) provided an explicit answer for. They believed that assessment is a vital part which cannot be detached from learning. Without assessment, a teacher cannot understand the students' level, their strengths, and their weaknesses. The only way of making effective decisions is to know the current condition of learners. There are other benefits in assessment, one of which is that teachers can reflect upon their teaching techniques to be sure whether they are useful or not. However, the kind of assessment is different based on the kind of task which is being used in assessment.

There are several types of reading comprehension tasks which can be used for assessment purpose, such as selective, productive, and interactive tasks (Suvorov & Hegelheimer, 2014). Brown and Abeywickrama (2018) defined selective tasks as the ones intended to assess the readers' identification of lexical, grammatical, or discourse features of language within a very short stretch of language. Both bottom-up and top-down processing may be used in these tasks. Some examples of selective tasks are picture-cued tasks, matching, true/false, multiple choice, etc. Productive tasks on the other hand include longer stretches of texts from one word to several paragraphs on one or more pages. In this kind of reading task, the reader is supposed to interact with the text to negotiate meaning. Here, top-down processing is usual but sometimes bottom-up processing may be necessary. Some examples of this kind are fill in the gaps with a word of their own, answering questions using a limited number of words, finding a word for the definition given, etc. (Suvorov & Hegelheimer, 2014). Finally, interactive tasks are those of a longer length with which the reader has to interact and understand the meaning. Schema or the background knowledge someone brings to the reading task is an important factor in answering this kind of reading task. Some examples are impromptu reading plus comprehension

questions, or editing of longer texts (Brown & Abeywickrama, 2018). The focus of the current research was on selective and productive reading tasks.

2.3. Diagnostic Assessment

Diagnostic assessment aims at identifying a learner's strengths and weaknesses in the parts the assessment and instruction are based on and, later on using the information obtained to help the student's learning and guides the instruction (Jang & Wagner, 2014). It is a kind of assessment dependent on feedback which gives learners the data they need to reflect on their learning in order to take remedial action.

Alderson, et al. (2015) asserted that there are some significant differences between diagnostic tests and other types of language tests: (1) In diagnostic assessment, the teacher is both a diagnostic tester and a diagnostic test user. (2) The test-taker is a foreign language learner. (3) Modifying the teaching process is the main purpose of the test. Finally, (4) the linguistic content is defined by the curriculum. They further stated that diagnostic testing aims at chasing the implementation of the curriculum to provide feedback to both teachers and learners.

Although most of the definitions describe both students' strengths and weaknesses as equally important in diagnostic assessment, in the real context of the classroom, as Alderson, et al. (2015) mentioned, more attention is paid to weaknesses and the kind of feedback needed to be provided based on them. In fact, the main responsibility of diagnostic assessment is to provide the necessary information about the progress of the learners. It has been pointed out that feedback should be of various kinds, that is, it is not good to emphasize on correctness more than needed or to give just positive or negative kind of feedback. Rather, it is better for the teacher to use a variety of different kinds of feedback (Harding, et al., 2015; Jang & Wagner, 2014).

In fact, diagnostic process starts with finding problems that a learner has, and this problem becomes clear using some exercises or a quick test. After diagnosing the problem and finding its reason, it is the time to do something to solve it. To make sure the prediction has an effect on the learning process, the best way is to exchange the information about the problem. This information is called feedback which is an important factor in learning. There are different sources of feedback such as the teacher, a grade on a certificate, a peer's gesture, and so forth (Alderson, et al., 2015). According to Jang and Wagner (2014), diagnostic feedback refers to the information provided to the learners to make them able to reflect on their performance with the purpose of receiving the appropriate remedy.

Using standard diagnostic tests is a good way to find the source of the learners' errors but the number of the tests developed so far is very limited.

In addition, as most of them are computer-based and low-stakes, they are not appropriate to be used in the classroom context as normal tests for the purpose of diagnosing students' weaknesses. Some of the most famous ones are DIALANG, Diagnostic English Language Needs Assessment (DELNA), and Diagnostic English Language Tracking Assessment (DELTA). However, there are some steps which can be taken to do diagnostic assessment in the classroom context. Harding, et al. (2015) proposed a framework that can be used as the reference for the steps to be taken to do diagnostic assessment (Figure 1).
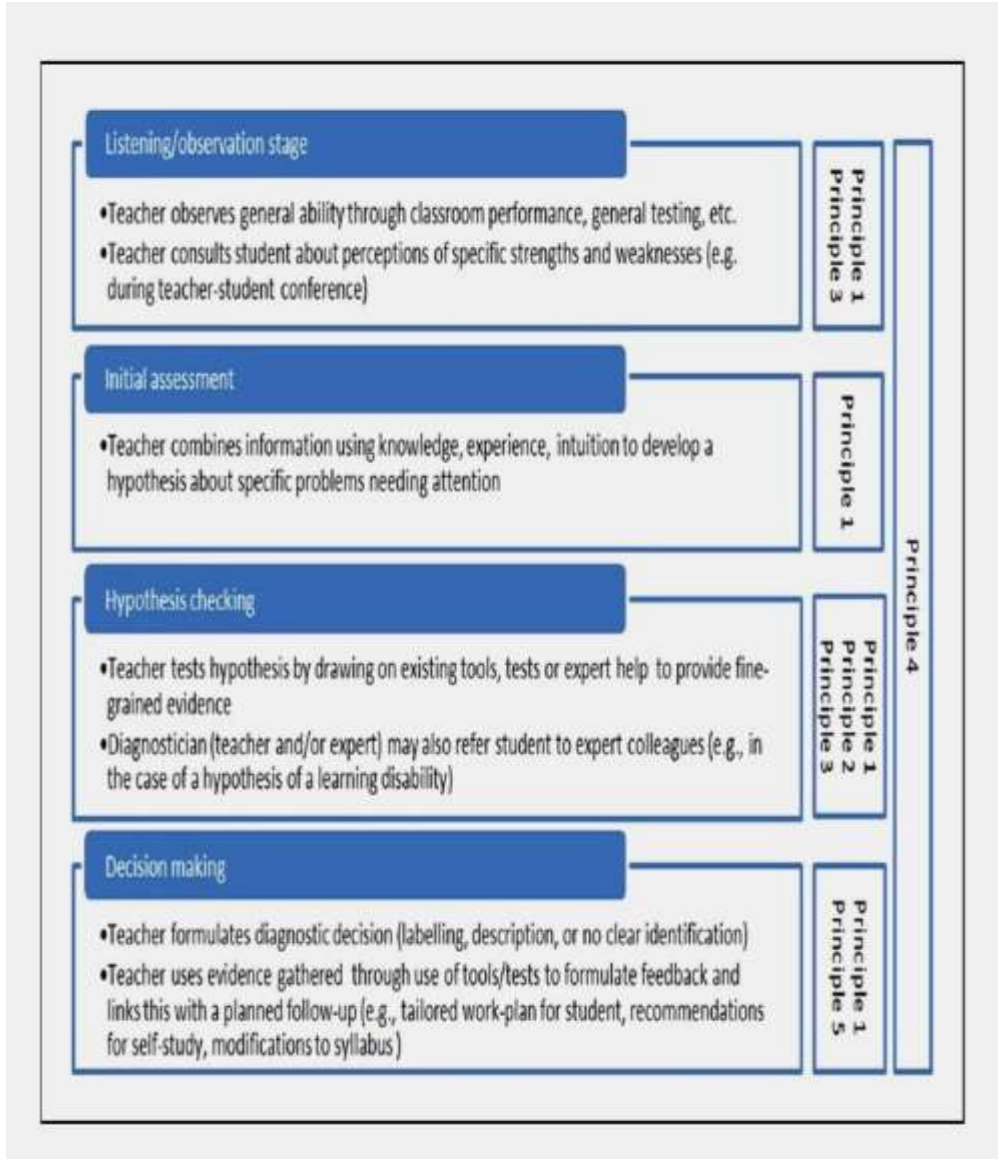

Figure 1. The Diagnostic Process (Harding, et al., 2015, p. 319)

1. *Listening/Observation stage:* It is the phase in which the teacher tries to find out some information about the learners' general ability through observing their performance in the classroom, on a general test, in a teacher-student(s) conference, etc.

2. *Initial assessment:* Here, using the information obtained at the previous stage and his/her experience, the teacher tries to come to a conclusion about the problems which need further attention.

3. *Hypothesis checking:* Now, the teacher tests the hypotheses using an expert help or a test to check the learners' progress.

4. *Decision making:* This is the last step in which the teacher come up with a final decision about the real causes of the learners' strengths and weaknesses. Now, s/he can choose the best feedback needed.

Diagnostic assessment has also been investigated by different researchers around the world. Doe (2015) conducted a study using diagnostic assessment and found out that the best parts of it are that it makes the learners able to be aware of the sections they are weak or strong at, a knowledge which would help them improve their learning, and it provides a supportive atmosphere that can make the students' progress faster. Mazloomi and Khabiri (2016) also tried to check the effect of diagnostic assessment on the participants' writing ability. Their conclusion was that diagnostic assessment of writing through dynamic self-assessment is a useful way of giving feedback to learners which can be introduced as a helpful technique to improve the learners' writing ability.

3. Method

3.1. Participants

The participants of this research were 60 learners from among an initial 78 elementary level female EFL students with the age range of 18-40 years old who were chosen based on their performance on a proficiency test. That is, after all the participants took the Nelson Proficiency Test, the researchers chose those students whose scores fell between one standard deviation above and below the mean. The researchers used non-random sampling or what Best and Kahn (2006) called convenience sampling. It is also worth mentioning that the researchers used four intact classes, each group consisting of two classes. However, to have homogeneous participants, the learners whose scores were not in the appropriate range were excluded from the study.

3.2. Materials and Instruments

A number of different instruments were used in this research.

3.2.1. A Nelson test

This was used to check the proficiency level of the participants to ensure homogeneity. As Haywood and Lidz (2007) stated having homogeneous groups is an important part of doing research on groups of participants since it reduces the amount of variability among the participants' performance and makes the results more reliable. The test is a 150 version of the test which is designed to check the proficiency level of elementary EFL learners. Here, 150 means the learners had 150 hours of English training and the test consists of 50 multiple choice questions.

3.2.2. Key English Test (KET)

Two KET tests were used as the pretest (KET, 2015) and posttest (KET, 2014) in the present research. Alderson (2000) presented a brief introduction to this kind of test which says it is an appropriate test for the learners of the elementary level as they are supposed to be able to read simple texts available in everyday life or when travelling abroad. The topics used in these tests are usually about ordering food in restaurants, shopping, starting a conversation and maintaining it, etc. Hence, the kinds of texts used are often notices, forms, newspapers, and so forth. As the number of the selective and productive tasks were not parallel, the researchers adapted some of them, that is, they changed seven multiple choice questions to short answer questions, which as Alderson (2000) described are questions with short responses of a few words, and made the tasks almost half selective half productive. That is, as each KET reading test consists of four sections along with 55 questions, after this adaptation, there were 28 selective and 27 productive items in both the pretest and the posttest. It should be mentioned that the reliability of these adapted tests were also checked and are reported in the next section.

The following is an example of a selective kind of item. For this item, the examinees had to choose the right answer based on their understanding of the text.

Thomas asked Inger to marry him when they were on a boat.

A. Right   B. Wrong   C. Doesn't say

On the other hand, the following item is an example of a productive item for which the examinees had to respond each item with no more than three words. The answers had to be derived from the text.

Where do Thomas and Inger live?

3.2.3. Practice Tests

Four tests with the same format as the pretest and posttest, including both selective and productive items, were also used during the term in the diagnostic assessment group.

3.3. Procedure

Seventy-eight elementary level female EFL students with the age range of 18-40 years old, from four intact classes, participated in this research and took the Nelson Proficiency Test appropriate for the number of hours they studied English. Then, 60 students whose scores were one standard deviation above and below the mean on the test were selected as the actual participants. The four classes were assigned to two groups, an experimental group called the diagnostic assessment and a control group. Each group consisted of two classes and 30 students.

In the diagnostic assessment group, using the processes mentioned in Figure 1, the researchers gave the learners the pretest to check their initial ability to do selective and productive reading comprehension tasks at the beginning of the study. To find out the main areas of the learners' weaknesses, the researchers then gave them the first practice test. After correcting the tests and identifying the students' weaknesses, one of the researchers, the teacher of the classroom, held a conference with the learners to talk about their weaknesses and to get informed about the learners' ideas about the reasons of their difficulties. Next, the teacher put all the hypotheses together to reach an overall hypothesis about the problem. After that, the hypotheses were checked with the potential tools for diagnostic assessment of reading to be sure they were right. Finally, finding the right assumption, the teacher provided the detailed feedback and follow-up support. The feedback was both individual and in groups. There were two points here. First, the feedback could not be a general suggestion like extensive reading. It should have been a specific point focusing on the exact problem. Second, if the teacher found out that her hypothesis was not right before the stage of giving feedback, she had to start over and try other potential problems based on another conference with the learners.

The strategies used in this group were derived from Nicol and Macfarlane-Dick (2006) and Alderson, et al. (2015) which were:

•The teacher asked the learners to talk about the kinds of feedback they would rather receive after handing their assignments in.

•The teacher asked learners to specify exactly the parts of the assignment they had difficulty doing.

•The teacher characterized the incorrect answers along with a clear explanation of the reason for why they were wrong.

•The teacher tried to teach learners to make sure about the correctness of their answers, e.g., to ask this question at the end of doing a task: Why am I certain about the correctness of the answer to this question.

•In addition to presenting feedback continuously to the learners, after each two sessions of feedback, another practice test was used to make the researchers sure of the learners' progress and other weaknesses/strengths. Eventually, at the end of the process, there was another KET test (the posttest) which was implemented to test the learners' achievement on selective and productive tasks based on the feedback they received throughout the study.

•The second group acted as the control group, as a criterion to measure the degree of the progress of the participants of the experimental group. Comparing the results of the two groups, it was possible to understand whether diagnostic assessment had any significant effect on the performance of the learners on selective and productive reading tasks or not. The learners of this group followed the typical routine of reading classes in language institutes and focused on different reading strategies such as scanning and skimming the texts in order to answer the related items. Moreover, the students in this group were asked to do some selective and productive tasks such as true/false/doesn't say, filling in the gaps with an appropriate answer or word from the text, answering questions using one to three words, and so forth after each reading text. This way, they were familiar with various selective and productive tasks. It should be noted that the difficult words of each text were taught in advance, that is, before starting reading. This group also participated in the same pretest at the beginning of the term and the same posttest at the end of it, similar to the experimental group.

4. Results and Discussion

4.1. Results

An important factor to be checked at the beginning of any research is the normality of the data. Checking the normality is necessary due to the fact that without having a set of normal scores, the results obtained from using the statistical analysis to check the hypotheses are not guaranteed to be generalizable. Therefore, Tables 1 and 2 are provided to show the normality of the results of the Nelson Proficiency Test, and the pretest and posttest of the two groups which are calculated through one-sample Kolmogorov-Smirnov test.

Table 1

One-sample Kolmogorov-Smirnov Test of the Nelson Proficiency Test of the Participants

		Proficiency test of the control group	Proficiency test of the diagnostic group
N		30	30
Normal Parameters	Mean	35.53	35.40
	SD	4.38	4.46
Asymp. Sig. (2-tailed)		.25	.39

As it can be seen in Table 1, the scores on the Nelson test in both groups are normal because both the asymptotic two-tailed levels of significance are higher than .05.

Table 2

One-sample Kolmogorov-Smirnov Test of the Pretest and Posttest of the Participants

		Pretest of the control group	Posttest of the control group	Pretest of the diagnostic group	Posttest of the diagnostic group
	N	30	30	30	30
Normal Parameters	Mean	41.80	42.60	38.03	48.00
	SD	4.53	4.74	4.69	3.91
Asymp. Sig. (2-tailed)		.06	.26	.36	.37

Upon checking the results provided in Table 2, it becomes obvious that the pretest and posttest scores are all normal as all of their asymptotic two-tailed levels of significance are higher than .05. Hence, the most suitable formulae to be used here are the parametric formulae. However, it should be mentioned that since the significance value reported for the pretest of the control group, .06, is very close to the .05 critical value, the researchers used the parametric formulae and interpreted the results cautiously.

Before doing any further analyses, it is necessary to check whether the participants of the two groups had the same level of English knowledge at the beginning of the study. To do so, the researchers used an independent-samples t-test, the outcomes of which are presented in Tables 3 and 4.

Table 3

Descriptive Statistics of the Nelson Proficiency Test of the Control and Diagnostic Groups

Group	N	Mean	SD
Control Group	30	35.53	4.38
Diagnostic Group	30	35.40	4.46

As it is shown in Table 3, the mean scores of the participants of the two groups on the Nelson test, i.e. 35.53 and 35.40, are very close to each other, which means they did not have a difference in their knowledge of English at the beginning of the study. However, to check whether the difference between the two groups' English knowledge is statistically significant, an independent samples t-test was conducted. The results are presented in Table 4.

Table 4

Independent-Samples T-Test of the Nelson Proficiency Test of the Control and Diagnostic Groups

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	T	df	Sig. (2-tailed)
Nelson Proficiency Test	Equal variances assumed	.00	.95	.11	58	.90

Using an independent-samples t-test, the outcome of which is presented in Table 4, the researchers checked the difference between the two groups' performance on the Nelson test. The sig. value reported for the t-test for equality of means is .90 which is bigger than the standard .05 level and confirms that there was not a significant difference in the participants' English knowledge at the beginning of the study.

The next important point to be checked is the reliability of the pretest and posttest. As it has been stated by Pallant (2011), reliable scales play a very important role in each study. One of the most common ways to check reliability is to use Cronbach's alpha coefficient which should be above .7 to call a test reliable. Table 5 shows the reliability of the pretest and posttest of the two groups in this study.

Table 5

Reliability of the Pretest and Posttest

	Pretest	Posttest
N of Items	58	58
Control Group	.76	.76
Diagnostic Group	.78	.78

According to Muijs (2004), reliability values above .7 are considered reasonable. Hence, since the value of Cronbach's alpha for the pretest and posttest of both groups are above .7, the conclusion was that the tests used in this study had acceptable reliability.

Now, it is the time to answer the first two research questions of the study using two repeated-measures two-way ANOVAs. To start with, the

descriptive statistics of the learners' performance in selective and productive tasks in the pretest and posttest are presented in Table 6.

Table 6

Descriptive Statistics of the Selective and Productive Tasks of Pretest and Posttest of the Control and Diagnostic Groups

		Selective Tasks in Pretest	Selective Tasks in Posttest	Productive Tasks in Pretest	Productive Tasks in Posttest
N		30	30	30	30
Control Group	Mean	22.97	23.00	18.70	19.60
	SD	2.84	2.44	3.23	3.57
Diagnostic Group	Mean	22.90	24.80	15.10	23.23
	SD	2.35	2.31	3.34	2.44

The information in Table 6 shows that the performance of all the participants in selective tasks has been better than the productive tasks on the pretest. That is, their weakness was basically in doing productive tasks rather than selective tasks.

The mean score of selective reading tasks of participants in the control group has changed from 22.97 to 23.00 which is not a lot and their performance on productive reading tasks has changed from the mean score of 18.70 to 19.60 which shows that they had a slightly more progress in productive tasks.

Checking the mean scores of the two types of reading tasks for the diagnostic assessment group shows that they have improved a lot, especially in the productive tasks. The mean scores of their performance on selective reading tasks in the pretest and posttest, which are 22.90 and 24.80 respectively, shows their progress in doing this sort of task. In addition, the mean scores of their performance on productive tasks in the pretest and posttest, which are 15.10 and 23.23, also suggests a great improvement in this kind of task.

Checking the mean scores of the selective tasks in the pretest of the two groups, i.e., control and diagnostic groups, which are 22.97 and 22.90 respectively, and the mean scores of the selective tasks in the posttest of the two groups that are 23.00 and 24.80 makes it clear that there is a slight change in the progress level of the control group whereas the diagnostic assessment group's performance shows a high improvement.

Moreover, upon checking the mean scores of the productive tasks in the pretest of the control and diagnostic assessment groups that are 18.70 and 15.10 respectively, it was seen that both groups' performance improved on

the posttest according to their posttest mean scores which are 19.60 and 23.23 for the control and diagnostic assessment groups. However, the improvement of the diagnostic assessment group has been much more.

To see whether the improvements have been significant or not, two repeated-measures two-way ANOVAs as well as a MANOVA were run. Because there were two groups, i.e., control and diagnostic groups, whose participants' performance was repeatedly measured on two types of reading comprehension tasks, selective and productive tasks, the researchers decided to use two repeated-measures two-way ANOVAs since it is the appropriate kind of analysis to be run in such cases (Hinton, Brownlow, McMurray, & Cozens, 2004; Page, Braver, & Mackinnon, 2003). Tables 7 and 8 and Figures 2 and 3 are dedicated to the results of these analyses.

Table 7

Repeated-Measures Two-way ANOVA of Selective Tasks in the Pretest and Posttest of the Control and Diagnostic Groups

Effect		Value	F	Sig.	Partial Eta Squared
Selective tasks	Pillai's Trace	.11	7.26	.00*	.11
Group			2.61	.11	.04
Selective * Group	Pillai's Trace	.10	6.77	.01*	.10

In Table 7, the level of significance of within-subject factor, that is selective kind of task, is reported as .00 which is smaller than the critical .05 level of significance. Therefore, the conclusion is that there is a significant difference in the participants' performance on selective reading comprehension tasks from pretest to posttest. It means the treatment given to the learners had a great effect on their performance. In addition, based on Cohen's (1988) categorization of partial eta squared values cited in Pallant (2011), in which .01=small effect, .06=medium effect, and .138=large effect, the respective partial eta squared value, .11, shows a medium effect size of the learners' performance on the selective tasks.

The Sig. value reported for the between-subject factor, that is grouping, is .11 which is larger than .05 and means that the reading ability of the participants of the two groups on selective tasks in either the pretest or the posttest was not significantly different from each other.

However, more important of all, the level of significance calculated for the interaction of within-subject factor and between-subject factor is reported as .01 which is also smaller than the critical .05 level of significance. Hence here, the conclusion is that there is a significant difference in the progress of the two groups on selective reading comprehension tasks from the pretest to the posttest, but the difference in their progress is not the same

in the two groups. As it is shown by the Partial Eta Squared value, .10, the size of this effect is medium as well.

To better understand these results, Figure 2 is presented to reconfirm what has been mentioned so far. As it is clear in this figure, although the participants of the two groups had almost the same performance on the selective tasks in the pretest, the diagnostic assessment group's performance was improved a lot in the posttest while the control group's performance did not improve a lot.
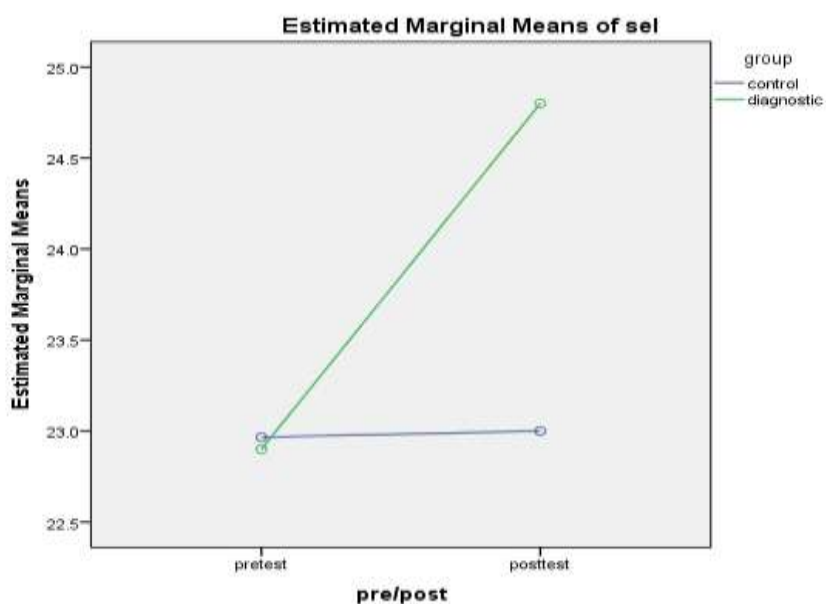

Figure 2. Differences between the Pretest and Posttest of Selective Reading Comprehension Tasks of the Two Groups

Moreover, to check the amount of the effect of diagnostic assessment on productive tasks of the two groups of the study, Table 8 and Figure 3 are presented.

Table 8

Repeated-Measures Two-way ANOVA of Productive Tasks in the Pretest and Posttest of the Control and Diagnostic Groups

Effect		Value	F	Sig.	Partial Eta Squared
Productive tasks	Pillai's Trace	.67	119.48	.00*	.67
Group			.00	.98	.00
Productive * Group	Pillai's Trace	.56	76.61	.00*	.56

In Table 8, the level of significance of within-subject factor, that is productive tasks, is reported as .00 which is smaller than the standard .05

level of significance. Therefore, the conclusion is that there is a significant difference in the participants' performance on productive reading comprehension tasks from pretest to posttest. It means the treatment given to the learners had a great effect on their performance. In addition, the partial eta squared reported is .67 that shows the large effect size of the learners' performance on the productive tasks.

The Sig. value reported for the between-subject factor, grouping, is .98 which is larger than .05 and means that the reading ability of the participants of the two groups on productive tasks in either the pretest or the posttest was not significantly different from each other.

However, the level of significance calculated for the interaction of within-subject factor and between-subject factor is also reported as .00 which is again smaller than the standard .05 level of significance. Therefore, the conclusion is that there is a significant difference in the progress of the two groups on productive reading comprehension tasks from the pretest to the posttest, but the difference in their progress is not the same in the two groups. As it is shown by the Partial Eta Squared value, .56, the size of this effect is large as well.

In addition, Figure 3 is also presented to reconfirm what has been mentioned about productive tasks. As it is obvious from the figure, the members of the two groups performed not well on the pretest, but differently on the posttest. Specifically, the diagnostic assessment group's performance improved much more in the posttest while the control group's performance did not improve that much.
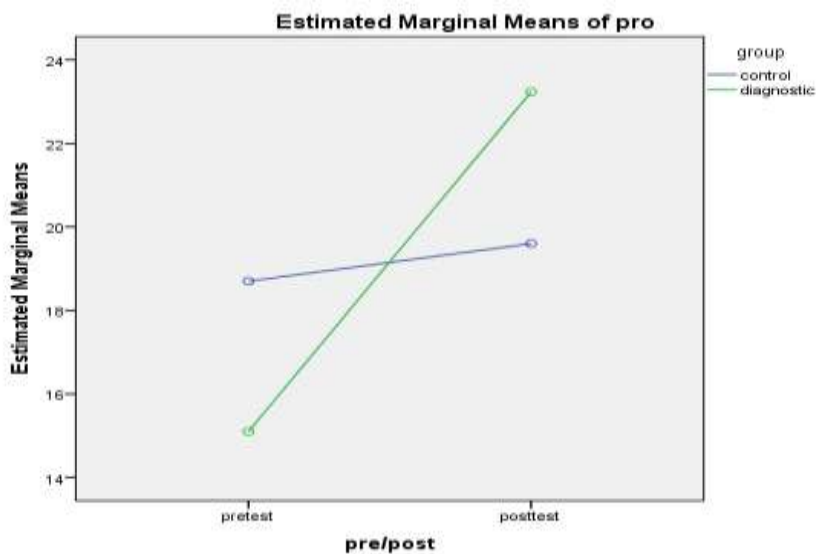

Figure 3. Differences between the Pretest and Posttest of Productive Reading Comprehension Tasks of the Two Groups

An important point which was seen in the learners' performance on the pretest was that almost all the learners had a great degree of difficulty doing the productive tasks rather than the selective ones. Findings of the present research, however, made it clear that diagnostic assessment was effective to improve the learners' performance on both kinds of tasks and especially on productive tasks.

To sum up the results of the present study, it should be noted that although there has been a difference between the participants' performance on selective and productive tasks in the pretest and the posttest, the progress from the pretest to the posttest has not been so significant in the control group whereas the other group, i.e. diagnostic assessment, benefitted a lot from the treatment as there has been a considerable amount of improvement in their performance on selective and productive reading comprehension tasks from pretest to posttest. That is to say the feedback in the diagnostic assessment group caused a significant improvement in the participants' performance on the two types of reading tasks, especially the productive tasks.

The first research question of the study received a positive answer as the results of Tables 6 and 7 and Figure 2 showed that diagnostic assessment had a significant effect on EFL learners' performance on selective reading comprehension tasks.

Furthermore, the second research question also received a positive answer as the results of Tables 6 and 8 and Figure 3 showed that diagnostic assessment had a significant effect on EFL learners' performance on productive reading comprehension tasks.

Finally, to answer the third research question, a MANOVA was used to check the two groups' performance on selective and productive kinds of tasks in pretest and posttest. MANOVA was chosen as the researchers wanted to find out the underlying relationship between the dependent variables in combination (i.e., selective and productive tasks), with respect to the independent variable(s) (i.e., the two groups) (Hinton, et al., 2004). Using this analysis, it is possible to check the participants' improvement in the two groups as well as their performance in relation to each other from pretest to posttest. The outcomes of the analysis are presented in Tables 9 and 10.

Table 9

Multivariate Test of the Pretest and Posttest of the Control and Diagnostic Groups

Effect		F	Sig.	Partial Eta Squared
Wilks' Lambda test	Group	1.93	.15	.03
	Time	29.99	.00*	.34
	Time * Group	19.26	.00*	.25

Table 9 shows the multivariate test of the pretest and posttest in general. Upon checking the row specified to grouping in Table 9, it becomes obvious that there is no significant difference between the performances of the two groups as the sig. value reported is .15 which is larger than the standard .05 level. Partial eta squared reported for this item is .03 which is a sign of the small effect size of the treatment used in the groups.

On the other hand, the sig. value reported for time, which refers to the interval between pretest and posttest, is .00 and smaller than the standard .05 level, which means performance of the participants of the two groups was significantly different from pretest to posttest which is considered large as the partial eta squared is .34 in this case.

Here too, the data presented in the third row is of more importance as it shows the interaction of time and grouping. The sig. in this row is reported as .00 which is smaller than the standard level and again it means there is a significant difference in the progress of the two groups from pretest to posttest. The point to be made is that the amount of this difference in progress was not the same in the two groups. According to the partial eta squared reported as .25, the amount of this effect has been large.

Finally, Table 10 below is presented to check the two groups' progress from pretest to posttest in more details.

Table 10

MANOVA on the Selective and Productive Tasks of the Pretest and Posttest of the Control and Diagnostic Groups

Source	Measure	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	Selective	22.53	1	22.53	3.60	.06	.03
	Productive	.00	1	.00	.00	.97	.00
Time	Selective	28.03	1	28.03	4.48	.03*	.03
	Productive	612.00	1	612.00	60.49	.00*	.34
Time *	Selective	26.13	1	26.13	4.18	.04*	.03
Group	Productive	392.40	1	392.40	38.78	.00*	.25

Upon checking the sig. value of selective tasks specified for grouping in Table 10, i.e. .06, it becomes clear that there was no significant difference between the two groups' performance in selective tasks. The partial eta squared value reported also confirms the point as it is .03 and shows a small effect size. Moreover, the sig. value of the productive tasks of the two groups is .97 which is also bigger than the standard .05 level, which means there was not a significant difference between the two groups' performance in productive tasks either. Here too, the size of this effect is small as the partial eta squared value reported is .00.

On the other hand, in case of time (i.e., improvement from pretest to posttest), it should be said that the two groups had a significantly better performance in both selective and productive tasks from pretest to posttest as the sig. value reported for these tasks are .03 and .00 respectively. The size of the effect of time for selective tasks was small due to the partial eta squared value which is .03, but the effect size of time for productive tasks is large as the partial eta squared reported is .34.

Once more, the last row's information is far more important as it reports the interaction of time and grouping. Here too, the conclusion is that there was a significant progress in the two groups' performance from pretest to posttest in both selective and productive tasks as the sig. values reported for these tasks are.04 and .00 respectively and both are smaller than the standard .05 level. According to the partial eta squared value for selective tasks which is .03, this effect size is small whereas it is large in the case of productive tasks due to the partial eta squared value which is .25. This shows the much better performance of the participants of the diagnostic assessment group in productive tasks in the posttest in comparison to the control group and in comparison to the participants' progress in selective tasks.

Eventually, to answer the third research question of the study, it was found that there was a significant difference in EFL learners' performance on selective and productive reading comprehension tasks using diagnostic assessment according to the outcomes of Tables 9 and 10. Even though the participants of the diagnostic assessment group had a better performance in both selective and productive tasks on the posttest, the progress was significantly better in productive tasks.

4.2. Discussion

To sum up the results of the present research, it should be mentioned that the first and second research questions received a positive answer. In other words, it was found that diagnostic assessment had a significant positive effect on EFL learners' performance on both selective and productive types of reading comprehension tasks.

Moreover, the third research question also received a positive answer. Although diagnostic assessment was seen to be effective on improving the learners' performance in both types of selective and productive reading comprehension tasks, it was more effective on improving the learners' performance on productive reading tasks.

The results of this study are in line with similar recent studies. For example, Zandi (2018) carried out a study to investigate the effect of diagnostic assessment on selective and productive listening tasks and she also found that the participants had a better performance in both selective and

especially productive listening tasks using diagnostic assessment. In another similar study, Ardin (2018) investigated the effect of diagnostic assessment on descriptive and narrative kinds of writing. She also concluded that diagnostic assessment is an effective kind of assessment to help improve the learners' writing ability in both kinds of writing. Furthermore, Kazemi (2018) studied the effect of diagnostic assessment on improving the speaking skill of EFL learners and found out that diagnostic assessment has also a significant effect on the improvement of EFL learners' speaking ability.

Diagnostic assessment has also been the subject of Yi's (2017) research who examined four cognitive diagnostic assessment models to examine the test takers' response behavior. The study used the function of the four models to come up with some conclusions about the response behavior of examinees who took four forms of TOEFL reading and listening comprehension sections. That is, the cognitive diagnostic assessment models were used to determine what the scores can reveal about the processing of L2 comprehension skills.

Kim (2015) also carried out a study on cognitive diagnostic assessment with the aim of diagnosing the strengths and weaknesses of second language learners' reading ability. The results showed that cognitive diagnostic assessment was a useful tool in identifying the learners' strengths and weaknesses which should be taken into account by teachers whose objective is to enhance the reading comprehension ability of students and developing appropriate instructional materials.

Interestingly, diagnostic assessment was also found to be effective in identifying and improving the learners' mathematical ability (Shim, Hudzaifah, Shakawi, & Azizan, 2017). Whatever strengths and weaknesses Shim, et al. (2017) diagnosed through diagnostic assessment in their study manifested itself in the results of the final examination of the students. That is, diagnostic assessment acted as a good technique helping both learners and teachers to be well aware of the main points that needed remedial action.

Moreover, selective and productive kinds of tasks have been the subject of Esfandyari's research (2019) who investigated the effect of self-assessment and peer-assessment on young EFL learners' performance on selective and productive reading comprehension tasks. Through her study, she found out that both types of assessment, i.e., self-assessment and peer-assessment, had considerable positive effects on young EFL learners' performance on both types of selective and productive reading comprehension tasks. She also found that there was not a significant difference between the impact of self-assessment and peer-assessment on young EFL learners' performance on the mentioned tasks. In other words, the

two types of assessment had almost similar degrees of influence on the students' performance.

Furthermore, Tavassoli and Nikmard (2019) did a study on selective and productive reading comprehension tasks to find out whether dynamic assessment has any significant influence on EFL learners' performance on these reading comprehension tasks. They also concluded that performance on these two kinds of reading tasks can be improved using dynamic assessment. In other words, dynamic assessment is another useful technique to help learners improve their performance on selective and productive reading comprehension tasks.

As it was mentioned earlier, the basis of diagnostic assessment is finding out the learners' strengths and weaknesses, especially weaknesses, to be remedied afterwards (Alderson, et al., 2015; Jang & Wagner, 2014). The remedies, then, can be in any form dependent on the kind of problem diagnosed by the teacher. An important point is that diagnostic assessment has to provide the essential information on the progress the learners are making (Alderson, et al., 2015).

Based on the results of this study and similar studies on the effect of diagnostic assessment on EFL learners' performance, it can be concluded that diagnostic assessment is a very good way of finding and solving the learners' problems and helping them improve their language knowledge. It makes the teachers aware of the most useful kinds of feedback effective for each level to help learners solve their problems. Moreover, as the present study showed, diagnostic assessment can help teachers to talk to the learners about their weaknesses or even their strengths and to provide them the necessary feedback. The results of this study showed that diagnostic assessment is a good way of helping learners to perform better on their final exams after receiving the necessary feedback on different kinds of tasks throughout the semester.

Since selective and productive reading comprehension tasks are the most common types of tasks used in different tests, if a learner is helped to overcome his/her potential problems on such tasks, his/her English knowledge can be improved. Therefore, the results of this study show the importance of working on these two types of tasks, especially productive ones, throughout the semester.

To learn a language, a learner needs to be able to both understand and produce the language. Understanding language is, in fact, manifested through the learners' good performance on selective tasks whereas producing language is mostly manifested through productive tasks. That is, if a learner performs well on selective tasks, it is a sign of his/her good understanding of

the language and his/her good performance on productive tasks indicates his/her ability to produce the language well.

Eventually, according to the outcomes of the present research which showed that diagnostic assessment has a considerable effect on both the learners' understanding and production of language, it is highly suggested to use this technique in various language classes.

However, it should be mentioned that since the participants of the present research consisted of only female EFL learners, strong claims cannot be made since the results may not be generalizable to male EFL learners.

5. Conclusion and Implications

This study was planned to investigate the effect of diagnostic assessment on EFL learners' performance on selective and productive reading comprehension tasks using the pretest/posttest quasi-experimental design and also availability sampling for choosing the participants.

The next point to be made is about the limitations of the study one of which was related to the selection of the participants. The researchers had to non-randomly select the participants of the study from among female learners at a language institute, so there were only female EFL learners in this study. Another limitation of the present study was that improving the participants' reading comprehension ability on the two specific types of tasks cannot be restricted to 16 sessions of practice which was the case in the present study. A considerable amount of time is necessary to help learners to get familiar with the kinds of feedback which are useful for them through diagnostic assessment in both kinds of selective and productive tasks. On the other hand, this research was delimited by availability sampling of EFL learners at the elementary level as the researchers had more access to them.

The main conclusion of the present study is that diagnostic assessment has a significant effect on the amount of the learners' improvement as they can learn different strategies to be used in diverse situations afterwards. Moreover, diagnostic assessment is also a good guide for EFL teachers, materials developers, and specifically testers as it gives them ideas to apply to make their job more efficient. Diagnostic assessment helps teachers identify the most useful kind of feedback effective for each student to solve their problems. In addition, through this kind of assessment a strong link can be made between those learners who easily get the points and those who need more practice, which is a great help to teachers.

The usefulness of diagnostic assessment can provide materials developers with good ideas about how to arrange the flow of their course-books to find out the learners' problems easier and quicker. That is, after

using diagnostic assessment several times, it becomes obvious which points are appropriate to be presented to the learners at the beginning, in the middle, or at the end of the course-books. Diagnostic assessment also helps materials developers to choose the best themes for each level to help learners' fast progress as the number of new vocabulary and new grammatical points necessary to talk about are closely related to the topic of discussion.

Testers might be the most beneficiaries of the present study as they would find the best ways of developing not only achievement tests but also other tests as a means of getting to know more and more about the major weaknesses of learners at each level to help them overcome their problems and perform better on subsequent tests.

Finally, future studies are recommended to the interested researchers. This study only concentrated on elementary level female EFL learners whereas the intermediate and advanced levels or male learners, or even a mixture of male and female learners could be the focus of another study to see whether diagnostic assessment has the same effect on different levels and genders or not. Another suggestion is that diagnostic assessment can be used along with other kinds of assessments, for example, formative assessment, self-assessment, peer-assessment, etc. to compare and contrast the influence of each type of assessment and to find the most and least effective assessment type.

References

Alderson, J. C. (2000). *Assessing reading*. New York: Cambridge University Press.

Alderson, J. C., Haapakangas, E. L., Huhta, A., Nieminen, L., & Ullakonoja, R. (2015). *The diagnosis of reading in a second or foreign language.* New York, NY 10017.

Ardin, M. (2018). *The effect of diagnostic assessment vs. dynamic assessment on EFL learners' descriptive and narrative writing*. Unpublished MA Thesis, Islamic Azad University, Karaj, Iran.

Best, J. W., & Kahn, J. V. (2006). *Research in education (10th ed.)*. Boston: Pearson Education.

Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices (3rd ed.)*. USA: Pearson Education.

Cai, Y., & Kunnan, A. J. (2018). Examining the inseparability of content knowledge from LSP reading ability: An approach combining bifactor-multidimensional item response theory and structural equation modeling. *Language Assessment Quarterly, 15*(2), 109-129.

Chastain, K. (1988). *Developing second language skills theory and practice*. USA: Harcourt Brace Jovanovich Publishers.

Doe, C. (2015). Student interpretations of diagnostic feedback. *Language Assessment Quarterly, 12*(1), 110-135.

Esfandyari, S. (2019). *The effect of self-assessment and peer-assessment on young EFL learners' performance on selective and productive reading comprehension tasks.* Unpublished M.A. Thesis, Islamic Azad University, Karaj, Iran.

Geoffrey R., & Roberts, G. R. (1999). *Learning to teach reading*. Paul Chapman Publishing Ltd.

Goodman, K. (1988). The reading process. In P. L. Carrell, J. Devine, & D. E. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 11-21). New York: Cambridge University Press.

Grabe, W., & Stoller, S. L. (2001). *Teaching and researching reading*. Harlow: Longman.

Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing, 32*(3), 317-336.

Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice clinical and educational applications.* Cambridge: Cambridge University Press.

Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2004). *SPSS explained.* London and New York: Routledge.

Jang, E. E., & Wagner, M. (2014). Diagnostic feedback in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment, Volume II: Approaches and development* (pp. 157-175). John Wiley & Sons, Inc.

Kazemi, N. (2018). *The effect of dynamic vs. diagnostic assessment on EFL learners' speaking ability*. Unpublished MA Thesis, Islamic Azad University, Karaj, Iran.

Kern, R. G. (1989). Second language reading strategy instruction: Its effects on comprehension and word inference ability. *The Modern Language Journal, 72*(2), 135-149.

Kim, A. Y. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing, 32*(2), 227-258.

Kim, K. (2018). Visualizing first and second language interactions in science reading: A knowledge structure network approach. *Language Assessment Quarterly, 14*(4), 328-345.

Mazloomi, S., & Khabiri, M. (2016). Diagnostic assessment of writing through dynamic self- assessment. *International Journal of English Linguistics*, *6*(6), 19-31.

Muijs, D. (2004). *Doing quantitative research in education with SPSS*. London: Sage Publications Ltd.

Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. Taylor & Francis e-Library.

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.

Oakhill, J., Cain, K., McCarthy, D., & Nightingale, Z. (2013). Making the link between the vocabulary knowledge and comprehension skill. In M. A. Britt, S. R. Goldman, & J. F. Rouet (Eds.), *Reading from words to multiple texts* (pp.101-114). New York, NY 10017.

Page, M. C., Braver, S. L., & Mackinnon, D. P. (2003). *Levine's guide to SPSS for analysis of variance (2ⁿᵈ ed.).* London: Lawrence Erlbaum Associates.

Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS for windows (4ᵗʰ ed.)*. Crows Nest NSW 2065 Australia.

Serrano, A., & Huang, H. Y. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text. *TESOL Quarterly, 52*(4), 971-994.

Shim, G. T. G., Hudzaifah, A. M., Shakawi, A., & Azizan, F. L. (2017). Relationship between students' diagnostic assessment and achievement in a pre-university mathematics course. *Journal of Education and Learning, 6*(4), 367-371.

Suvorov, R., & Hegelheimer, V. (2014). Computer-assisted language testing. In A. J. Kunnan (Ed.). *The companion to language assessment, Volume II: Approaches and development* (pp. 61-80). John Wiley & Sons, Inc.

Tavassoli, K., & Nikmard, F. (2019). The effect of dynamic assessment on EFL learners' performance on selective and productive reading comprehension tasks. *Journal of Foreign Language Research (JFLR), 9*(2), 445-478.

Uso'-Juan, E., & Martinez-Flor, A. (2006). *Current trends in the development and teaching of the four language skills.* Wiley & Sons, Inc.

Yi, W. S. (2017). Probing the relative importance of different attributes in L2 reading and listening comprehension items: An application of cognitive diagnostic models. *Language Testing, 34*(3), 337-355.

Zandi, G. (2018). *The effect of diagnostic assessment vs. dynamic assessment on EFL learners' performance on selective and productive listening comprehension tasks*. Unpublished MA Thesis, Islamic Azad University, Karaj, Iran.

Bibliographic information of this paper for citing:

Nikmard, F., & Tavassoli, K. (2020). The effect of diagnostic assessment on EFL learners' performance on selective and productive reading tasks. *Journal of Modern Research in English Language Studies*, *7*(1), 79-104.