International Journal of Research in English Education (IJREE)

Original Article **Published online: 20 September 2019.**

Iranian EFL Learners' Perceptions toward Paper Assessment in Mid-term and Final Exams in an English Language Institute

Zohreh Zare Toofal[1], Reza Vaseghi[1], Mohsen Zare[1]

* Correspondence:
zohrehzaretoofan@gmail.com
1. Department of English Language and Literature, Islamic Azad University, Ayatollah Amoli Branch, Amol, Mazandaran, Iran

Received: 24 February 2019
Revision: 26 April 2019
Accepted: 4 July 2019
Published online: 20 September 2019

Abstract

Assessment has been taken to demonstrate that learning is aligned with external standards which is almost related to students' goals in a curriculum in English language teaching and it plays an integral role in the success of language learning program that is following by so many EFL teachers. Consistency in scoring (mid-term and final exams) highly depends on the way of conducting paper assessment, validation of the process, reliability, experience of teachers, and different interpretation of assessors to make justified decision. Hence, the present study was an attempt to probe Iranian EFL learners' perceptions toward paper assessment in mid-term and final exams in a language institute. To this end, a total of 100 participants (50 males and 50 females) between ages of 15 to 28 at intermediate level, were selected based on Nelson Proficiency Test. Data were collected through scores of two sequential semesters and a Likert scale questionnaire. The findings of this study indicated that there is a direct positive relation between learners' viewpoints on paper assessment and their progress. Generally speaking, paper assessment in both formative and summative assessment would be a great progress among female and male English language learners.

Keywords: perception, mid-term exam, final exam, EFL learners, assessment

[DOI: 10.29252/ijree.4.3.21]

1. Introduction

Assessment is a key term in English language teaching research and it plays an integral role in the success of language learning programs that it is following by so many linguists, philosophers, scholars, EFL teachers, practitioners, and all staffs who are involved in English language teaching process. Experts in language assessment believe that raters and teachers can help the assessment reliability and facilitate language proficiency in students' performance (Bachman & Palmer, 2010; Hamp-Lyons, 2003).

If we have an accurate and clear perception toward paper assessment, it demands for its implications that we have interrelated factors which play a supportive role in English language teaching process. As identified by O'Malley and Valdez Pierce (1996 as cited in Rea-Dickins, 2001), we have certain aims in assessment that provides a clear understanding for

1. Placement of students,

2. Screening and identification of any issues,

3. Replacement or promotion,

4. Checking the rate of students' progress,

5. Appraisal of the existing educational program, and

6. Diagnosing probable liabilities of the program.

Assessment is one of the important issues in teaching a foreign language. This study is aimed to find learners' encouragement in learning process, how paper assessment due to their interests would lead to a better learning and if they were willing to pass such tests to evaluate what was their recognitions of taking exams (mid-term and final) during a semester of teaching English language, whether exams' questions make our learners prepared them for their future needs or not, and if exams were authentic to evaluate real statues of their progress in learning process.

Bachman (2007) provides a comprehensive summary of the strengths and weaknesses of three approaches to defining constructs in language assessment, with a focus on "the dialectic of abilities and contexts" (p. 41) (Purpura, 2016). In a "trait/ability-focused" approach, "context" refers to the methods for eliciting language performance, or in Bachman's (1990) words, "contextual features that determine the nature of language performance that is expected for a given test or test task" (p. 112). From a "task/context- focused" perspective, the context and the task to be performed in the context are inseparable. The construct to be measured is equivalent to "ability for use" (Bachman, 2007, p. 56). A strong form of the "task/context-focused" approach views construct as consisting of "abilities to accomplish particular tasks or task types" (Brown et al., 2002, p. 9, in Bachman, 2007, p. 56). In the third approach, the construct of language proficiency is defined from a social interactional perspective. Context is viewed as a separate dimension, and the focus has shifted to the interaction between the ability and context. The current study aimed to investigate Iranian EFL learners' perceptions toward paper assessment in mid-term and final exams in an English language institute.

1.1 Statement of the Problem

There are several issues that make it difficult for teachers to develop formative assessment practices. Firstly, formative assessment practice is complex (Vingsle, 2014), and using assessment information to plan subsequent instruction is especially difficult (Heritage, Kim, Vendlinski, & Herman, 2009; Schneider & Meyer, 2012). Secondly, external factors, such as accountability (Darling-Hammond & McCloskey, 2008; Klenowski, 2011; OECD, 2005) and the focus on examination and summative assessment (Bennett, 2011; Brown, Kennedy, Fok, Chan, & Yu, 2009; Wiliam, 2006) can impede implementation of formative assessment.

In the investigation, we will take a motivation perspective. The expectancy-value theory of achievement motivation (Wigfield & Eccles, 2000) will be used as an interpretive lens in the analysis of the teachers' reasons to fully engage in the professional development and in the implementation of a new formative classroom practice. Motivation is the driving force of human behavior, and the failures of ambitious professional development programs often seem to be related to teachers not sufficiently directing and sustaining their efforts towards program implementation (Bell et al., 2008; Randel et al., 2011; Schneider & Randel, 2010).

[DOI: 10.29252/ijree.4.3.21]

That is why we are going to consider paper assessment tool as a valuable perceptive from our learners' points of view toward their satisfactions. We need to keep data tracking in perspective, and giving weights to teachers' judgments, paper assessments at classroom level to provide enough detail to inform teaching and learning directly, without asking them is centrally impossible to flash new database for the illusion that creates those statements which have been tracked in a perspective.

1.2 Research Questions

Based on the objectives of the study, the following research questions were formulated:

1. Do mid-term and final exams affect EFL learners' perceptions to do better for the next level?

2. Is there any relationship between exams' marks and EFL learner's changing behavior for the next level?

3. Is there any dichotomous relationship between paper assessment and EFL learners' progress in an educational semester?

4. Are there any correlations among learners' perceptions and their progress?

1.3 Research Hypotheses

The following hypotheses were made to answer research questions:

1. Mid-term and final exams have the most effects on their progress to do better for the next level, and encourage to be prepared or not prepared.

2. There is a positive relationship between learners' exams mark and their changing behavior before and after exams to be prepared for the next step.

3. There is a negative dichotomous relationship between paper assessment and making progress in a learning process.

4. There is a highly positive correlation between learners' exams and their perceptions in two sequential terms.

2. Literature Review

Assessment literacy is an important issue in language teaching and trained teachers are more involved with assessment rather than untrained teachers. Lam (2015) suggested that assessment literacy in educational program was not enough and language assessment is needed to equip with pre-service teachers with assessment strategies. Tasagari (2014) analyzed online assessment tutorial materials and its usefulness to EFL teachers in the United States and Europe. Malone (2013) got online assessment from the perspective of language experts in the United States. All studies revealed various research methods toward the perception of assessment literacy and their needs. There is also another view of assessment in language teaching that is called static and dynamic assessment. In static assessment individuals are involving with no scaffolding in the parts of mediators or test takers, and it can be more practical and convenient than dynamic assessment (Lantolf & Poehner, 2010).

Many researchers now (Lantolf & Poehner, 2006; Ohta, 2000; Swain, 2000) argue that language learning and acquisition can be achieved through interaction and dynamic assessment helps students to become more involved in interaction and dynamic assessment helps learners to perform tasks by mediators or test takers (Gibbons, 2003; Lantolf & Poehner, 2004, 2006). Dynamic assessment also can be traced through Vygostky's view that stressed social environment is a facilitator in learning process (Karpov & Haywood, 1998; Kozulin & Garb, 2002). So mediation, zone of proximal development, contingency and scaffolding are cornerstones in dynamic assessment. Different studies carried out on traditional and static assessment in very convenient and practical rather than dynamic assessment which is more timing and spending energy to follow in language teaching process but dynamic assessment is more related to comprehending of the task and interaction.

When students are learning, they need to self-assess that how much effort they must take to be successful, when they are wrong which learning strategies work well for them. Accurate self-evaluation is going to enable students to see what they have mastered and identify what needs further work (Mcmillan & Hearn, 2008). Learners practically don't need to sugarcoat or exaggerate things about themselves in these papers. Assessment is a powerful force in student learning. From the students' perspective, only the most important activities in a subject are assessed (Kandlbilder, 2009). Vogt and Tsagari (2014) conclude that assessment procedures such as designing tests, giving grades, placing students in their corresponding levels, and awarding certificates are not fully developed skills in teacher participants

[DOI: 10.29252/ijree.4.3.21]

and most probably they are learned on the day to day practice. Basically, a self-assessment paper is something that sums up all sides and aspects of personality, while still giving off a rather positive image of learner.

Lopez Mendoza and Bernal Arandia (2009) added that trained teachers had more positive views towards assessment in comparison to the non-trained teachers. In their study, assessment was perceived among trained teachers as a tool to monitor learning, to communicate with the student, to align learning with teaching, and to empower students. Those with less training experience viewed assessment as mandate, as a summative process and a tool of power and control over students. The researchers concluded that teachers' previous assessment training experience may also have a role in their use and perceptions of language assessment.

Another important study in almost the same area of interest was conducted by Malone (2013). This study was focused on the content of a tutorial which was developed to promote foreign language instructors' knowledge about assessment basics. Feedback was collected from all study participants (44 language instructors and 30 language testers) in the U.S. Results of the study indicated a clear difference in the opinions of language testers and language educators regarding the requirements of the significant technical information about assessment. These findings have shed light on the fact that instructors' perceptions about assessment differ from their practical approach in the same field, which should be seriously considered by the designers of the assessment training programs while developing the tutorials or materials. The importance of enhancing the language assessment literacy among in-service teachers was investigated by Scarino (2013). She stated the importance of integrating the language assessment knowledge with understanding, culture and learning trends of a language. Moreover, a great stress was put forth on the test users and test developers to become aware of assessment procedures in order to infer and gauge their own assessment practical approach.

In this paper, we take up a suggestion by Perrenoud (1998) that any theory of formative assessment must be embedded within a wider theoretical field, specifically, within a theory of pedagogy. We propose a model whereby the design of educational activities and associated assessments is influenced by the theories of pedagogy, instruction and learning, and by the subject discipline, together with the wider context of education. We explore how teachers may develop productive relationships between the formative and summative functions of classroom assessment, so that their judgements may inform the formal external assessment of students, thus increasing the validity of those assessments. We also show how the model informs the development of theories that give appropriate weight to the role of assessment as part of pedagogy. The first issue in 1998 had important contributions from scholars such as Perrenoud (1998), Biggs (1998), Hattie and Jaeger (1998), Sadler (1998), Sebatane (1998), and Dwyer (1998). Their comments and reflections pushed the research field of assessment forward and offered important research themes on feedback, self-assessment, and formative assessment processes for years to come. It is a model that we need to continue using as we move forward. It opens up for important debates on controversial topics and reminds us of what has been achieved and where we still need to focus our research (Newton & Baird, 2016; Wiliam, 2017).

3. Methodology

3.1 Participants

A total of 100 EFL learners [male (*n*=50) and female (*n*=50)] studying English as a foreign language were recruited. The sample was connivance sampling procedure from Toofan Foreign Language Center in Mashad, Khorasan Rzavi province of Islamic Republic of Iran. Their age ranged from 15 to 28 years and the majority of them were 18.

3.2 Design of the Study

The nature of this study was purely descriptive and thus a survey method was used in this study. The reason is that based on the views of Ary, Jacobs, Razavieh, and Sorensen (2006), survey research designs are procedures in quantitative research in which "ask questions about peoples' beliefs, opinions, characteristics, and behavior of the population" (p.400).

3.3 Instruments

3.3.1 Nelson Proficiency Test (NPT)

Nelson Proficiency Test (NPT) was used at first in order to determine the language participants' proficiency level. With regard to homogeneity, participants were selected at the intermediate level. This instrument, in the form of multiple choice questions, consisted of 50 questions involving one cloze comprehension passage as well as vocabulary, grammar, and pronunciation sections.

[DOI: 10.29252/ijree.4.3.21]

3.3.2 Mid-term and Final Exams

Midterm and final exams in two sequential terms for comparing numerical data were taken from participants. In comparing two sets of scores in two sequential terms, we have been received low markers less than high markers to perceive their attention toward midterm and final exams. All mid-term and final exams in two sequential semesters were Oxford Standard Tests that were specifically made for American English file books at intermediate level. They evaluated learners' language proficiency in four skills such as reading, writing, listening, and vocabulary. All tests have different tasks to complete such as fill in the blanks, completion, true or false, checking comprehension questions, multiple choice items, answering the question, synonyms or opposites, and full writing paragraphs.

Cronbach's alpha reliability for each two sets of exams was calculated to understand whether it is reliable. As it can be seen in Table 1, comparing between mid-term1 and 2 among female learners revealed that there is a significant reliability which is 0.90 that shows both exams are reliable for assessing. In Table 2, reliability became 0.94 in their final exams in first and second term. For male learners (Table 3), it became 0.93 in their mid-term exams and it was 0.876 in their final exams as it is shown in Table 4, so two sets of exams are reliable.

Table 1. Reliability Statistics (F.Mid-term1 & F.Mid-term 2)

Cronbach's Alpha	N of Items
.903	2

Table 2. Reliability Statistics (F.final 1 & F.Final 2)

Cronbach's Alpha	N of Items
.946	2

Table 3. Reliability Statistics (Male learners in mid-term exams)

Cronbach's Alpha	N of Items
.932	2

Table 4. Reliability Statistics (Male learners in final exams)

Cronbach's Alpha	N of Items
.876	2

[DOI: 10.29252/ijree.4.3.21]

3.3.3 Language Learners' Satisfaction of Assessment

Language learners' satisfaction of assessment was developed by the researchers in the format of Likert scale from the weakest points to the strongest viewpoints. Learners were provided with this questionnaire to help us more about their satisfaction of paper exams in their own perceptions in that to what extent they are satisfied from their semester exams to progress gradually. It has been provided to mark it on paper by numbers (1-2-3-4-5) very weak, weak, medium, strong, and very strong of learners' ideas toward paper assessment. The questionnaire has gone through the validation processes. Its validity has been proved by two experts in the field. And its reliability determined in a pilot study which was conducted before the actual research. The questionnaire was scattered among 50 intermediate English language learners both male and female to respond. The obtained Cronbach's alpha was 0.70, denoting that the measure had a high degree of reliability.

Table 5. Reliability statistics for language learners' satisfaction of assessment

Cronbach's Alpha	No. of Items
.700	16

3.4 Procedure

100 participants at intermediate level were selected by NPT and given them a questionnaire before and after exams. They were 50 intermediate EFL female learners and 50 EFL male learners who were studying in an English language institute and their mid-term and final marks were gathered to compare their first semester marks with their second term in next level for considering EFL learners' progress numerically. Mid-term exam had 40 points out of 100 scores and final exam had 60 points out of 100 scores which the learners should totally get 100 scores in both mid-term and final exams. All kinds of questions were used in the formative and summative evaluations such as multiple choice, true or false, explanation, fill in the blanks, completing sentences, answering questions to evaluate their reading, writing and listening skills. The researchers had interview with learners to evaluate their speaking skill and added this score to their both mid-term and final exam. Then a questionnaire in the form of Likert scale was scattered to gather information about EFL learners' view towards their understanding in assessment and how much they agree or disagree with this particular statement that depends on what if the curve goes up or down, then providing open-ended questions to interview EFL learners to pick up their answers and suggestions to questions toward taking exams such as mid-term and final. At the end, the researchers analyzed the data to discuss if assessment would encourage EFL learners to make progress, what is more important is that how it would create a change in EFL learners' behavior toward paper assessment during a semester of teaching English language.

3.5 Data Analysis

In order to answer the research questions of the present study, the data were analyzed in the following way: First, as shown in Tables 6 and 7 below, the descriptive statistics were performed to see what participants do generally in the first and second term. Then, based on the research hypotheses, the researchers tabulated and the data were analyzed using SPSS statistical software, examining T-test, One way ANOVA, and Pearson product-moment correlation (Pearson r), which are the appropriate inferential statistics.

[DOI: 10.29252/ijree.4.3.21]

Table 6. Descriptive statistics in the first term

	N	Minimum	Maximum	Mean	Std. Deviation
Fmidterm1	50	22.00	40.00	33.2500	4.40576
Mmidterm1	50	17.00	40.00	33.2900	4.23071
Ffinal1	50	30.00	60.00	46.6300	9.95388
Mfinal1	50	30.00	58.00	46.5800	7.84113
Valid N (listwise)	50				

Table 7. Descriptive statistics in second term

	N	Minimum	Maximum	Mean	Std. Deviation
Fmidterm2	50	27.00	40.00	34.4700	3.61884
Mmidterm2	50	20.00	40.00	33.6600	4.06885
Ffinal2	50	30.00	60.00	48.2200	8.91293
Mfinal2	50	34.50	59.50	49.3500	6.56385
Valid N (listwise)	50				

Table 8 is the summary of data collection for the first and second midterm exams for both genders that shows the percentage of participants' cooperation in taking exams.

Table 8. Case processing summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Fmidterm2 * Fmidterm1	50	100.0%	0	.0%	50	100.0%
Mmidterm2 * Fmidterm1	50	100.0%	0	.0%	50	100.0%
Fmidterm2 * Mmidterm1	50	100.0%	0	.0%	50	100.0%
Mmidterm2 * Mmidterm1	50	100.0%	0	.0%	50	100.0%

4. Results

4.1 Do mid-term and final exams effect on EFL learners' perceptions to do better for the next level?

One sample T-test was performed among female and male learners for their first term of investigation. As it is shown, sig tailed is 0.0 so there is a positive relation between their mid-term and final exams progress. It means that taking

[DOI: 10.29252/ijree.4.3.21]

paper exams leads to positive results and learners' better perceptions in the next level. It promotes learners' knowledge so it approves the hypothesis of the first research question that mid-term and final exams have the most effects on learners' progress for the next level.

Table 9. Results of the one-sample t-test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Fmidterm1	53.365	49	.000	33.25000	31.9979	34.5021
Mmidterm1	55.640	49	.000	33.29000	32.0876	34.4924
Ffinal1	33.125	49	.000	46.63000	43.8011	49.4589
Mfinal1	42.005	49	.000	46.58000	44.3516	48.8084

Moreover, Paired sample T-test was carried out between female and male learners at their first term of examination as indicated in Tables below that show sig tailed is 0.965 which is so close to 1 and it is more than 0.05, so there is a negative relation to make them both gender compared each other. In Tables 10 and 11, it is shown that there is a positive relation between mid-term and final exams progress among female learners. The sig tailed is 0.0 and it is less than 0.05. It is also believed that learners make progress according to the level of their language proficiency. It would be better to take paper assessment for language learners because it could have been improved by learners' perceptions and it is essential to revise such tests to happen a reliable evaluation.

Table 10. Results of the Paired Samples Test (Between female and male learners at their first term)

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Fmidterm1 - Mmidterm1	-.04000	6.40379	.90563	-1.85994	1.77994	-.044	49	.965

In Table 11, it is particularly shown that the correlation between male midterm and final exams is 0.68 in the first term that caused a good relation between mid and final exams. In Table 12, sig tailed is 0.0. So there is a positive relation between their mid-term and final exam respectively. Learners' perception will be affected by mid and final exam as they make progress toward the next level.

[DOI: 10.29252/ijree.4.3.21]

Table 11. Results of the Paired Samples correlations (Between male midterm and final exams)

		N	Correlation	Sig.
Pair 1	Mmidterm1 & Mfinal1	50	.686	.000

Table 12. Results of the Paired Samples Test ((Between male midterm and final exams)

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Mmidterm 1 - Mfinal1	-1.32900E1	5.81789	.82277	-14.94343	-11.63657	-16.153	49	.000

The same statistics were also conducted for female learners to compare their mid-term and final exams relation. In Table 13 the sig tailed is 0.000 and this means that there is a positive relation between two sets of exam scores. As female learners have been taken mid and final exams, the range of perceptions toward paper assessment will change to do better during the term and final term. It is shown that learners have a positive view regarding their paper assessment in each term and they are gradually changed.

Table 13. Results of the Paired Samples Test (Comparison of female learners' mid-term and final exams)

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Fmidterm2 - Ffinal2	-1.37500E1	6.92102	.97878	-15.71693	-11.78307	-14.048	49	.000

In Table 14, the sig tailed is 00.1. It is less than 0.05 and there is a significant and positive relation in females' midterm exams both in their first and second term. Again we have a progress in female learners to take paper tests and they make progress. Generally female learners in both terms have been agreed to take paper assessment and they make progress to the next level. It is completely true and approved the first hypothesis which has been made to answer the first question of this research.

[DOI: 10.29252/ijree.4.3.21]

Table 14. Results of the Paired Samples Test (Relationship between females' midterm exams in the first and second term)

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Fmidterm1 - Fmidterm2	-1.22000	2.40145	.33962	-1.90248	-.53752	-3.592	49	.001

In Table 15, sig tailed is 0.172 that is more than 0.05 so it demonstrates that there is a negative relation between their first midterm and second midterm exams among male learners. They did not have any progress toward our hypothesis that we estimate to make them progress but it shows that their first term formative evaluation was not helpful to push them toward a better learning situation. So the hypothesis will be rejected and it is not approved for male learners to make progress in paper assessment in two sequential terms.

Table 15. Results of the Paired Samples Test (Relationship between the first midterm and second midterm exams among male learners)

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Mmidterm1 - Mmidterm2	-.37000	1.88661	.26681	-.90617	.16617	-1.387	49	.172

4.2 Is there any relationship between exams' marks and EFL learner's changing behavior for the next level?

In Table 16, sig tailed is 0.022 that is less than 0.05 percent and there is a positive significant relation between females' first term of their final exams and the second term final exams. At any consideration, female learners did better than male learners, female learners make progress for the next level and their behaviors have been changed contrary to the male learners. So the hypothesis of the second question of this research will be approved by female learners not male learners. Female learners are trying to make changes to do better for the next level and all paper tests at first and second terms are related to each other and there is a positive relation between first and second terms' exams. So it is accepted for female learners but there has not been any considerable change in male learners. Therefore, the hypothesis is rejected.

[DOI: 10.29252/ijree.4.3.21]

Table 16. Results of the Paired Samples Test (Relationship between females' first term and second term final exams)

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Ffinal1 - Ffinal2	-1.59000	4.76776	.67426	-2.94498	-.23502	-2.358	49	.022

4.3 Is there any dichotomous relationship between paper assessment and EFL learners' progress in an educational semester?

One way ANOVA was performed between and within groups of two sets of scores: mid-term and final exams for female learners in their second term. Table 17 indicates the sum of squares between and within groups of females' second term of their midterm and final exams that midterm1 is considered as predicator. It shows a dichotomous relation between paper assessment and learners' progress, so it approves the hypothesis to answer the third question of this research.

Table 17. Results of the One way ANOVA Fmidterm2 Ffinal2 by Mmidterm1

		Sum of Squares	df	Mean Square	F	Sig.
Fmidterm2	Between Groups	228.476	23	9.934	.625	.871
	Within Groups	413.229	26	15.893		
	Total	641.705	49			
Ffinal2	Between Groups	1374.451	23	59.759	.617	.877
	Within Groups	2518.129	26	96.851		
	Total	3892.580	49			

Table 18 shows the sum of squares between and within groups of female learners' second term in their midterm and final exams that their final exam in first term is considered as predicator. Sig tailed is .000 and .005 which is less than 0.05, so there is a positive relation between mid and final exams regarding female learners that have been acted between and within groups.

[DOI: 10.29252/ijree.4.3.21]

Table 18. Results of the One way NOVA Fmidterm2 Ffinal2 by Ffinal1

		Sum of Squares	df	Mean Square	F	Sig.
Fmidterm2	Between Groups	506.976	27	18.777	3.066	.005
	Within Groups	134.729	22	6.124		
	Total	641.705	49			
Ffinal2	Between Groups	3372.747	27	124.917	5.287	.000
	Within Groups	519.833	22	23.629		
	Total	3892.580	49			

Table 19, it reveals the sum of squares between and within groups of male learners' second term in their midterm and final exams that their midterm exam in first term is considered as predicator. Sig is .118 in final exams and is .000 in mid-term exam that shows that there is no direct positive relation between mid and final exams among male learners' progress because they have not been changed between and within groups. So the result shows that male learners disagreed to have paper assessment and it rejects the hypothesis to answer the third question of this research. It doesn't make any changes in male learners' behaviors to improve their scores. So they might disagree to have such assessment during the term or at the final section of the term. Therefore, the hypothesis "there is a negative dichotomous relationship between paper assessment and making progress in a learning process" is rejected for female learners and is supported for male learners.

Table 19. Results of the One way ANOVA Mmidterm2 Mfinal2 by Mmidterm1

		Sum of Squares	df	Mean Square	F	Sig.
Mmidterm2	Between Groups	720.116	23	31.309	8.935	.000
	Within Groups	91.104	26	3.504		
	Total	811.220	49			
Mfinal2	Between Groups	1242.496	23	54.022	1.617	.118
	Within Groups	868.629	26	33.409		
	Total	2111.125	49			

Table 20 represents the sum of squares between and within groups of male learners' second term in their midterm and final exams that their final exam in first term is considered as predicator.

[DOI: 10.29252/ijree.4.3.21]

Table 20. Results of the One way ANOVA Mmidterm2 Mfinal2 by Mfinal1

		Sum of Squares	df	Mean Square	F	Sig.
Mmidterm2	Between Groups	594.703	26	22.873	2.430	.017
	Within Groups	216.517	23	9.414		
	Total	811.220	49			
Mfinal2	Between Groups	1818.954	26	69.960	5.507	.000
	Within Groups	292.171	23	12.703		
	Total	2111.125	49			

4.4 Is there any correlations among learners' perceptions and their progress?

Table 21 indicates the correlations in the first term of evaluation for both genders and Table 22 shows the correlations in the second term of evaluation for both genders. It has been demonstrated that learners' final exams correlation was significant because in females' mid-term exam correlation was 1 and it has changed to 0.72 in their final exam. To some extent, it is concluded that it would be challengeable to take such assessment in class environment to make them progress. In females' view, it became less important to increase their scores in their final exams. It approves the hypothesis to answer the fourth question of this research that "there is a correlation between learners' perceptions and their progress in paper assessment." As it has been shown in Tables 17 and 18, there is a high range of correlations between the first and second term of exams among both genders. Therefore, the hypothesis is supported by the information in Tables 21 and 22 concerning a high correlation between learners' exams and their perceptions.

[DOI: 10.29252/ijree.4.3.21]

Table 21. Correlations in the first term of evaluation for both genders

		Fmidterm1	Mmidterm1	Mfinal1	Ffinal1
Fmidterm1	Pearson Correlation	1	-.099	.211	.721**
	Sig. (2-tailed)		.493	.141	.000
	N	50	50	50	50
Mmidterm1	Pearson Correlation	-.099	1	.686**	.028
	Sig. (2-tailed)	.493		.000	.846
	N	50	50	50	50
Mfinal1	Pearson Correlation	.211	.686**	1	.354*
	Sig. (2-tailed)	.141	.000		.012
	N	50	50	50	50
Ffinal1	Pearson Correlation	.721**	.028	.354*	1
	Sig. (2-tailed)	.000	.846	.012	
	N	50	50	50	50

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

[DOI: 10.29252/ijree.4.3.21]

Table 22. Correlations in the second term of evaluation for both genders

		Fmidterm2	Mmidterm2	Ffinal2	Mfinal2
Fmidterm2	Pearson Correlation	1	-.050	.692**	.130
	Sig. (2-tailed)		.729	.000	.370
	N	50	50	50	50
Mmidterm2	Pearson Correlation	-.050	1	.180	.509**
	Sig. (2-tailed)	.729		.210	.000
	N	50	50	50	50
Ffinal2	Pearson Correlation	.692**	.180	1	.280*
	Sig. (2-tailed)	.000	.210		.049
	N	50	50	50	50
Mfinal2	Pearson Correlation	.130	.509**	.280*	1
	Sig. (2-tailed)	.370	.000	.049	
	N	50	50	50	50

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

5. Discussion

The findings of this study are in accordance with the majority of the previous findings. Paper assessment can be a kind of dynamic assessment that will measure a comprehended task and it can be traced through Vygostky's view that stressed social environment is a facilitator in learning process (Karpov & Haywood, 1998; Kozulin & Garb, 2002). So mediation, zone of proximal development, contingency and scaffolding are cornerstones in dynamic assessment. Different studies carried out traditional and static assessment in very convenient and practical rather than dynamic assessment which more timing and spending energy to follow in language teaching process but dynamic assessment is more related to comprehending of the task and interaction.

We have several implications to establish that paper assessment is beneficial for language learners during and end of the semester of a learning process. Most of them should become aware of what was the final purpose of the task. Learners did not have the same level of recognition to make it happened truly, it is still under question that how we can rely on their attitudes in social environment only by providing such questions that were conducted by the teacher. We cannot be so sure that the reality of such questionnaire is applicable for learners because they should have made by trained teachers to understand the exact approach of answering to such questions that were almost due to their understanding and comprehension of English books to study for developing social environment that a foreign language occurs. Lopez Mendoza and Bernal Arandia (2009) added that trained teachers had more positive views towards assessment in comparison to the non-trained teachers. The findings were noticeable to follow up this procedure that looking at learners' perceptions was very helpful and it could give us feedback to underpin a new practical and dynamic assessment of tests by EFL teachers and specialists in testing. So all paper tests are important parts of a learning process for measuring learners' comprehension of the task and it should be prepared by a trained teacher in this field. That's why we are regarding learners' perceptions as a featuring feedback to make a plausible test.

[DOI: 10.29252/ijree.4.3.21]

Scarino (2013) stated the importance of integrating the language assessment knowledge with understanding, culture and learning trends of a language. Moreover, a great stress was put forth on the test users and test developers to become aware of assessment procedures in order to infer and gauge their own assessment practical approach. Testing all materials were so difficult to examine at the same time in each exam, it required energy and it was timing to plan such valid and reliable test to consider the purpose of learners' progress relating their cultures and trends of a language.

Various kinds of content for checking four skills were administered in learners' tests relating their theoretical parts of learning process that should be more related to their specific needs of analysis and would be more practical in a class as a social environment example according to Vygostkys' view. Therefore, needs analysis for providing a good test is essential for test takers and course designers. As we discussed in data analysis, most of the learners' perceptions in paper assessment were different in any statement, because they have been under their own recognition of assessment by getting different range of scores in their papers.

Vogt and Tsagari (2014) conclude that assessment procedures such as designing tests, giving grades, placing students in their corresponding levels, and awarding certificates are not fully developed skills in teacher participants and most probably they are learned on the day to day practice. Basically, a self-assessment paper is something that sums up all sides and aspects of personality, while still giving off a rather positive image of learner. It is used by potential or current employers and colleagues to create a character sketch about the learner, and know about learner's personality traits, strengths, and weaknesses. Having all this technical knowledge about learner makes other people more trustworthy, leads them to make better decisions about learner's placement in the office environment and enables them to foresee the type of a situation.

Even while being honest, one has to specifically take care about writing this type of a paper, because obviously it involves learner's self-image as well as how people generally perceive learner to be, and that is not something that should be taken lightly. Being honest in a self-assessment paper is a great thing, but you need to know when to say things and how exactly to say them. Learners practically don't need to sugarcoat or exaggerate things about themselves in these papers. We should be aware of learner's worth. We should be aware of learner's strengths and play onto them. Assessment is a powerful force in student learning. From the students' perspective, only the most important activities in a subject are assessed (Kandlbilder, 2009).

When students are learning, they need to self-assess that how much effort they must take to be successful, when they are wrong, and which learning strategies work well for them. Accurate self-evaluation is going to enable students to see what they have mastered and identify what needs further work. Students who experience success with moderately difficult and challenging tasks will attribute their success to ability and effort rather than to external attributions such as luck or help from other students. Making these internal attributions is, in turn, based on the ability of students to self-assess and self-evaluate. This knowledge helps students develop self-efficacy for future performance in similar tasks (Mcmillan & Hearn, 2008). As a result of the discussion, we need a dynamic assessment during and at the end of a course book term to supply all learners' goals via paper assessment. It can be done by self and peer assessment to locate their viewpoints of their weaknesses and strengths in a teaching semester period.

6. Implications

One of the implications of this study is about conducting psychometric evaluation of learners achieving pedagogical goals. Learners' recognition of assessment is different from teachers' viewpoints and it is basically divided in two main parts of assessment self and peer assessment. Instead of having paper assessment, teachers can use technology to peruse their learners for getting feedback of their own and classmates' assessment during a term of language learning program. The other implication toward assessment is teachers' own views that evaluation would happen through a numerical process. If there is absent of numerical assessment, it is crucially difficult to make decision to plan our judgments for the next term. This perspective makes us to have delusion toward another way of dynamic assessment aside from paper assessment. It is almost difficult or problematic for teachers to make a decision based on a qualitative assessment without any marks or numbers. Most teachers' judgments are made upon numerical assessment that are frequently used by paper.

7. Conclusion

We have come to this conclusion that paper assessment in both formative and summative assessment would be a great progress among female and male English language learners. It was considerable to have formative evaluation during a term such as mid-term exam and summative evaluation such as final exam which provide us all negative and positive

[DOI: 10.29252/ijree.4.3.21]

points toward assessment. We need to improve our paper assessment as it mentioned earlier through dynamic assessment to create more interactive questions that would be change each term by some test takers, and activate learners' cognitive and metacognitive strategies. Of course it takes time and we would need some practical and trained teachers to be more analytical and critical in this field to test such critical questions in four skills. Formative and summative evaluation are needed for four skills in each term for each learner to pass the present term and go to the next level. The finding revealed that paper assessment could manage learners to accomplish more learning issues to their views that it would become helpful to assist them in their learning process. The learners can construct their cognitive and metacognitive strategies of learning to become more successful than their previous term by paper assessment as they also have competitive relation with other peers to pose themselves apparently. Empirical research shows that assessment is a key term for providing the next curricular program and it will help them to be interactive and it is essential to train reflective teachers who are critical and responsible toward their jobs. Assessment, especially dynamic assessment makes them prepare for the next level and it gives us feedback to reach a potential way of teaching in our classrooms. Assessments based on situations relevant to students' own experiences can motivate them to give their best performances. We need to keep data tracking in perspective, giving weight to teachers' judgments, quizzes, tests, and assessments at classroom level without asking them to log them centrally. We want assessment to provide enough detail to inform teaching and learning directly – this is what drives student outcomes.

References

Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. (2006). *Introduction to research in education*. Florence. KY Thomson/Wadsworth.

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

Bachman, L. F. (1990). *Fundamental consideration in language testing.* Oxford University Press.

Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 41–71). Ontario, CA: University of Ottawa Press.

Bell, C., Steinberg, J., Wiliam, D., & Wylie, C. (2008). *Formative assessment and teacher achievement: Two years of implementation of the keeping learning on track program*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25, doi: 10.1080/0969594X.2010.513678

Biggs, J. (1998). Assessment and Classroom Learning: a role for summative assessment? *Assessment in Education: Principles, Policy & Practice, 5*(1), 103-110, doi: 10.1080/0969595980050106

Brown, J. D., Hudson, T., Norris, J., & Bonk, W. (2002). *An investigation of second language task-based performance assessments.* Technical report (University of Hawaii at Manoa. Second Language Teaching & Curriculum Center); #24.

Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice, 16*(3), 347-363, doi: 10.1080/09695940903319737

Darling Hammond, L., & McCloskey, L. (2008). Assessment for learning around the world what would it mean to be internationally competitive. *PDK International Journal, 90*(4), 263-272. doi: 10.1177/003172170809000407

Dwyer, A. M. (1998). The texture of tongues: Languages and power in China. In William Safran, ed., *Nationalism and Ethnic Politics 4.1/2: Special Issue.* Nationalism and Ethnoregional Identities in China. Frank Cass, pp. 68–85.

Gibbons, P. (2003). Mediating language learning: Teacher interaction with ESL students in a content based classroom. *TESOL Quarterly, 37*(2), 247-273.

Hamp-Lyons, L. (2003). Writing teachers as assessors of writing. In Kroll, B. (Ed.), *Exploring the dynamics of second language writing* (p.p. 162-189). New York, USA: Cambridge University Press.

[DOI: 10.29252/ijree.4.3.21]

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28*(3), 24–31. doi: 10.1111/j.1745-3992.2009.00151.x

Hattie, J. A., & Jaeger, R. (1998). Assessment and classroom learning. A deductive approach. *Assessment in Education, 5*(1), 111-122. doi: 10.1080/0969595980050107

Kandlbinder, P. (2009). Key concepts in postgraduate certificates in higher education teaching and learning in Australasia and the United Kingdom. *International Journal for Academic Development, 14*(1), 19-31, doi: 10.1080/13601440802659247

Karpov, U. V., & Heywood, H. C. (1998). Two ways to elaborate Vygotsky's concept of mediation: Implication for instruction. *American Psychologist, 53*(1), 27-36. doi: 10.1037//0003-066X.53.1.27

Klenowski, V. (2011). Assessment for learning in the accountability era: Queensland, Australia. *Studies in Education Evaluation Journal, 37*(1), 78-83. doi: 10.1016/j.stueduc.2011.03.003

Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension. *SAGE Journal, 23*(1), 112-127. https://doi.org/10.1177/0143034302023001733

Lam, R. (2015). Language assessment training in Hong Kong: implications for language assessment literacy. *Language Testing, 32*(2), 169-197. https://doi.org/10.1177/0265532214554321

Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the part into the future. *JAL, 1*(1), 49-72. doi: 10.1558/japl.1.1.49.55872

Lantolf, J. P., & Poehner, M. E. (2010). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research, 15*(1), 11-33. https://doi.org/10.1177/1362168810383328

Lantolf, J. P., & Poehner, M. E. (2011). *Dynamic assessment in the foreign language classroom: A teacher's guide.* University Park, PA: CALPER Publications.

Lopez Mendoza, A. A., & Bernal Arandia, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile: Issues Teach. Prof. Dev., 11*(2), 55-70.

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and, users. *Language Testing, 30*(3), 329-344. https://doi.org/10.1177/0265532213480129

Mcmillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons, 87*(1), 40-49.

Office of Economic Co-operation and Development. (2005). *Formative assessment: Improving learning in secondary classrooms.* Paris: OECD Publishing.

Ohta, A. S. (2000). Rethinking interaction in SLA: Developmentally appropriate assistance in the zone of proximal development and the acquisition of L2 grammar. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 51–78). Oxford, United Kingdom: Oxford University Press.

Perrenoud, P. (1998). L'évaluation des élèves. Paris/Bruxelles: De Boeck Université Collection Pédagogie en développement. *Revue des sciences de l'éducation*, *26*(3), 483-716.

Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal, 100*(1), 190-208. https://doi.org/10.1111/modl.12308

Randel, R. et al. (2011). The importance of the verbal shift handover report: A multi-site case study. *Int J Med Inform, 80*(11), 803-12. doi: 10.1016/j.ijmedinf.2011.08.006.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: identifying processes of classroom assessment. *Language Testing, 18*(4), 429-462. https://doi.org/10.1177/026553220101800407

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/BF00117714

[DOI: 10.29252/ijree.4.3.21]

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, *30*(3), 309-327. https://doi.org/10.1177/0265532213480128

Sebatane, E. M. (1998). Assessment and classroom learning: a response to Black & Wiliam. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 123-130. doi: 10.1080/0969595980050108

Schneider, M. C., & Meyer, J. P. (2012). Investigating the efficacy of a professional development program in formative classroom assessment in middle school English language arts and mathematics. *Journal of Multidisciplinary Evaluation*, *8*(17), 1–24.

Schneider, M. C., & Randel, B. (2010). Research on characteristics of effective professional development programs for enhancing educators' skills in formative assessment. In H. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp. 251−276). New York: Routledge.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford: Oxford University Press.

Vingsle, C. (2014). *Formative assessment: teacher knowledge and skills to make it happen* (Licentiate dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-91247

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: findings of a European study. *Language Assessment Quarterly, 11*(4), 374-402. doi:10.1080/15434303.2014.960046

Wigfield, A., & Eccles, J. S. (2000). Expectancy value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wiliam, D. (2006). *Does assessment hinder learning?* Speech delivered at the ETS Europe Breakfast Salon. Institute of Civil Engineer, UK, London.

Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester, (ed.), *Second handbook of mathematics teaching and learning* (pp.1053-1058). Greenwich CT: Information Age Publishing.

Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551-575. doi: 10.1080/0969594X.2018.1441807

[DOI: 10.29252/ijree.4.3.21]

Appendix

In the Name of God

Language learners' satisfaction of assessment

This questionnaire is going to find the answers about assessment satisfaction of language learners. We hope to have your cooperation in this field. It would be our pleasure to answer the questions honestly and thank you very much for spending your time. All information in this questionnaire will be used only in a research project, and there is no need to write your name. We really appreciate your kind cooperation in this field.

Before answering the questions, please complete the following chart. Personal characteristics:

Age	15-20	20-25	25-30	30-35	Above 35
Gender	Male		Female		
Education & major	Diploma	B.A student	B.A	M.A student	M.A
Years of learning English	Less than one year	1-2 years	2-3 years	3-4 years	Above 4 years

[DOI: 10.29252/ijree.4.3.21]

Dear responder, please put check mark in the box for your answers.

No	Questions	Very strong	Strong	Medium	Weak	Very weak
1	To what extent do you believe in mid-term and final tests are applicable for using English daily in your life?					
2	To what extent do you really think that you learn English more than previous term?					
3	To what extent do you do your English homework outside the classroom?					
4	To what extent do you care about your mid-term and final tests to spent time for studying your book?					
5	To what extent are you stressed about your exams?					
6	To what extent do you feel freedom that you have class without paper assessment?					
7	To what extent are satisfied from your exams during and after the term?					
8	To what extent are you willing to have classes without paper exams?					
9	To what extent are you trying to have more studies when you have exam?					
10	To what extent do you want to be evaluated by paper assessment?					
11	To what extent do you feel that paper assessment is required for going to the next level?					
12	To what extent does English effects on your learning satisfaction?					
13	To what extent do you satisfy from the method of teaching to pass or fail your exams?					
14	To what extent do your paper assessment effect on your satisfaction outside the classroom?					
15	To what extent do you feel that your education and major have effect on passing or failing mid-term and final exams?					
16	To what extent do you consider the importance of motivation to pass your exams successfully?					

[DOI: 10.29252/ijree.4.3.21]