

## Contributions of Cloze Test Varieties and Procedures in Reading Instruction: A Washback Study

Gholam-Reza Abbasian<sup>1</sup>, Kowsar Ghasemi<sup>2\*</sup>

- 
1. Assistant Professor, Dept. of English, Imam Ali University & South Tehran Branch, Islamic Azad University, Tehran, Iran. Email: gabbasian@yahoo.com
  2. Dept. of English, South Tehran Branch, Islamic Azad University, Tehran, Iran.
- \* Corresponding Author's Email: kosar.ghasemi@yahoo.com
- 

**Abstract** – Aside from language ability, different factors including the way (e.g., test method/format) testee's ability is measured can have an influence on a learner's performance. Among various test methods, cloze test has received momentum in the field such that it is rendered through a number of varieties and procedures assumed to affect the testee's learning and performance. To shed an empirical light on the issue, the present study investigated the effectiveness of its varieties and procedures (i.e., fixed-ratio and rational multiple choice cloze test and cloze-elide test) in developing reading comprehension ability. To do so, 113 Iranian Pre-university students were chosen and screened through PET to make sure of their homogeneous nature; among them only 80 participants were ultimately organized into four groups. Each group was provided with a 10-session alteration of integrative treatment of testing-teaching through which they experienced rendering reading instruction through various varieties and procedures of cloze test. The results indicated that there were no significant differences among the average performances of four groups. However, the rational multiple choice group outperformed the other group. So, it was revealed that the integration of cloze test with its varieties and procedures affects learners' performance and creates positive washback to some extent; however, the study leaves the room open for further investigations.

**Keywords:** cloze test, cloze-elide, cloze multiple-choice, reading comprehension, washback

---

### I. INTRODUCTION

The endemic areas of language testing in human lives, employment, economic policy, particularly in language learning and curriculum development make language testing as a controversial topic (Fulcher, 2010). This debate is based on the fact that language testing is of importance to teachers since it provides them with some useful educational feedback. Actually, tests are given to learners to probe their weaknesses and strengths in language learning and to dispose learners to learn (McNamara, 2005). Contrary to this idea, it seems that assessing English language proficiency has been subservient in many educational systems all around the world, including Iran to the extent that passing a test is more commonplace than learning language. Most pupils study and take a test merely to pass it, and then they forget what they have learnt. Whereas, one of the axioms of language testing is the

notion of the test effect on the teaching and individual's learning. Thus, considering the fact that testing can influence teaching (Alderson & Wall, 1993) follows on learners' development in language skills, has attracted many researchers' attention in order to propose the most effective test method/s to assess language proficiency to promote teaching and learning processes.

Contrary to the diversity of language testing methods/forms, selecting a proper test method, and as Palupiningsi (2011) indicates, assessing students' language competency in specific language skill has been a big hurdle for stakeholders. He holds that "such problems require careful planning in terms of the alternative solution. In order to arrive at the best solution for the most proper test, it is necessary to understand what kind of testing can be appropriately applied in a particular language skill" (p. 2). In the same vein, Heaton (1991) claimed that testing can improve teaching and boost students' learning.

In its turn, both testing and teaching are bound up with each other such that test washback may virtually make it possible to change teaching. Hence, testing is an inbuilt component of both teaching and learning processes and it would likely have a great influence on teaching, known as washback or backwash. On the test method/form, Brown (2004) proposed some testing techniques such as "reading aloud, written response, multiple-choice, picture-cued items, matching test, editing, gap filling test, cloze test, c-test, cloze-elide test, short-answer test, ordering test, and summarizing test" (p. 190). Similarly, Heaton (1991) classifies some testing techniques namely: "word matching, picture and sentence matching, true/false reading test, multiple-choice items, completion test, rearranging items, cloze procedure, and open-ended and miscellaneous items for reading test" (p.105), while Alderson's classification (2000, p.202) concludes "multiple-choice, cloze test, gap-filling test, matching, ordering, editing, cloze-elide, short-answer, free-recall, summary, gapped summary, information-transfer" (cited in Lu, 2006, p. 16).

Additionally, Brown (2004) pointed out that precise assessment of language skills especially reading comprehension is a compelling need which may be met in many various forms such as multiple choice items, short answer items, cloze tasks, selective deletion gap filling, c-tests, cloze elide, true/false, filling in the blanks and cloze test, etc. Meanwhile, Brown (1980) indicated that, among them, cloze test enjoying an integrative nature has received prime importance in testing language skills in general and in testing reading skill in particular.

### **A. Cloze Test**

Initially, the word cloze has been driven from *closure* in Gestalt theory coined by Wilson Taylor in 1953. Cloze test is in the form of a reading passage consisting of a portion of text with certain words removed, devised to measure students' ability, skill and knowledge or achievement. As such, it is used to determine how well the students are in supplying the words deleted by tapping into their background schemata and making them to think critically about the missing words to reconstruct the mutilated passage (cited in Hadley, G. & Naaykens, J., 1999. p. 64). Lu (2006, p.15 as cited from Rye, 1982) also explains that "cloze

procedure is essentially a cognitive task. The reader has to reason and construct suggestions to fill the gap on the basis of the evidence derived from the context. The completion of meaning, based on understanding and reasoning, is a cognitive task". In line with Rye (1982), Farhady (2006, p. 35) claims that "according to the theory of expectancy grammar, the more proficient the reader is, the better he would decide on the missing words" (p. 3). In fact, cloze tests can be utilized as a valid measurement to identify students' knowledge and understanding of the reading process, as well as assessing the extent of students' vocabularies and knowledge of a text, and also persuading students to think critically and analytically about the text and content (cited in Hadley & Naaykens, 1997, p.66).

### **B. Cloze Test Varieties and Procedures and Testing Reading Skill**

Following the emergence of the integrative approach and the popularity of the cloze test that focuses mostly on improving critical thinking of learners, it has been proven that it can be as a valuable device to test reading comprehension. Conrad (1970), Oller (1973), Porter (1976) and some researchers concluded that these types of test are quite reliable and valuable to the fields of language learning in general and language testing in particular (cited in Darwesh, 2010, p. 109). Additionally, Culhane (1970) suggested that both teachers and learners would derive great benefit from the cloze test as a device to improve reading skill. Also, cloze test has been adapted and modified in some forms and procedures. By former it is meant the general physical appearance such as 1) *modified cloze test*, 2) *standard cloze test*, 3) *c-test*, 4) *multiple-choice*, and 5) *cloze- elide test*, while the latter refers to the fixed - ratio, rational and random deletion procedures (Brown, 2004). Farhady and Keramati (1996) claimed that both features (varieties and procedures) affect testees' performance differently. In this regards, Porter (1978) and Alderson (1983) posited that "much of the discrepancy in performance on cloze test may be attributed to the deletion procedures. In fact, the deletion rate may play a significant role in the results of cloze test" (cited in Farhady, 2006 p. 272).

### **C. Washback**

The term washback refers to a way public examinations influence learning and teaching. So, this concept can be reckoned as the connection between language testing and teaching. In fact, it deals with both the impact of the testing on the teaching and on the educational system or even the society (Harwood, 2007). By washback, Porter (1985) meant the effect of testing on teaching. Bailey (2007, p. 6) cited from Berry (1994, p. 31) defined it as "effect a test has on classroom practice". Cheng (2005) considered it as a neutral term since it may refer to both the intended positive effects and the unintended negative effects (cited in Harwood, 2007, p. 16).

Given the literature, both cloze test and washback have been extensively studied; however, the former has rarely been addressed in terms of the consequences of its varieties and procedures and the latter has not been investigated first in terms of the test form in general and in relation to the varieties and procedures of cloze test in particular. In a bid to

address these issues empirically, one macro research question was raised which was specifically followed up through four micro questions as follows:

#### **D. Research Questions**

##### **Macro research question:**

Q1. Is there any relationship between cloze test varieties and procedures in developing EFL learners' reading performance?

##### **Micro research questions:**

Q2. Is the integration of cloze–elide test with the fixed-ratio deletion procedure more effective than that of cloze-elide test with rational deletion procedure and multiple choice with the rational and fixed- ratio deletion procedures?

Q3. Is the integration of cloze–elide test with the rational deletion procedure more effective than that of multiple choice form with fixed-ratio and rational deletion procedures?

Q4. Is the integration of multiple choice form with fixed-ratio more effective than that of multiple choice form with the rational deletion procedures?

Q5. Does the distinct integration of cloze test varieties and procedures have any significant effects on developing EFL learner's reading comprehension ability?

#### **E. Research Hypotheses**

H0.1. There is no relationship between cloze test varieties and procedures in developing EFL learners' reading performance.

H0.2. The integration of cloze–elide test with the fixed-ratio deletion procedure is not more effective than that of cloze-elide test with rational deletion procedure and multiple choice form with the rational and fixed- ratio deletion procedures.

H0.3. The integration of cloze–elide test with the rational deletion procedure is not more effective than that of multiple choice form with fixed-ratio and rational deletion procedures.

H0.4. The integration of the multiple choice form with fixed-ratio is not more effective than that of multiple choice form with the rational deletion procedures.

H0.5. The distinct integration of cloze test varieties and procedures has no significant effects on developing the EFL learner's reading comprehension ability.

## **II. METHOD**

### **A. Participants**

The participants of the study were 80 homogeneous Iranian female EFL learners aged 17-18 who were selected based on cluster sampling out of an original sample of 100 from a secondary high school in Tehran.

## **B. Instrumentation and Validation**

Four instruments were employed to gather the intended data: (1) Preliminary English Test (PET) for selecting a homogeneous sample, (2) A Researcher-made Reading Comprehension test administered as a Diagnostic (i.e., pretest), and (3) 10 different varieties of cloze tests with the same text difficulty incorporating the target varieties and procedures: (i.e, multiple-choice and close-elide test which are based on two procedures including fixed-ratio and rational deletion procedures). (4) Another 10-item Researcher-made Reading Comprehension multiple choice test used as an achievement test (i.e., posttest)

## **C. Procedure**

At the outset of the experiment, the learners' scores on the PET and the Researcher-Made Reading Comprehension test were collected. After obtaining the scores, the participants were grouped into four experimental groups based on their language level. Then, all four groups received instructions on reading, but multiple cloze tests of various varieties and procedures were incorporated into the treatment. It means that at the end of each session, each experimental group received its special cloze varieties and procedures. In fact, throughout the treatment they were exposed to an "integrative instruction"; alternative integration of teaching and testing.

Henceforth, to the first group, cloze-elide test was administered through fixed ratio deletion procedure. Moreover, cloze-elide test based on rational deletion procedure was given to the second group. Based on the fixed-ratio procedure, a multiple choice cloze test was administered to the third group. Eventually, the fourth group attempted a multiple choice cloze test with the rational deletion procedure. After all these administrations, all groups received the same achievement test on reading skill. Based on the performances of the participants, the results obtained from cloze passages constructed from the different texts using two different cloze tests and procedures: cloze-elide test and multiple choice cloze test with the combination of two different procedures including, fixed-ratio deletion and rational deletion were statistically analyzed.

## **D. Reliability Indices**

The KR-21 reliability indices for the PET, pretest and posttest of reading comprehension and ten quizzes are displayed in Table 1. The moderate and low reliability indices for the reading comprehension tests were due to the fact that these tests had rather few items while the sample size is a crucial factor for having a high reliability.

**Table 1: KR-21 Reliability Indices**

	N	Mean	Variance	KR21
PET	80	73.78	156.683	0.89
Pretest	80	15.41	5.056	0.32
Posttest	80	16.46	5.543	0.50
t1	80	14.12	10.718	0.64
t2	80	14.00	9.544	0.59
t3	80	14.05	11.314	0.66
t4	80	14.96	9.328	0.63
t5	80	15.06	9.806	0.65
t6	80	14.90	7.939	0.55
t7	80	15.34	7.872	0.57
t8	80	15.35	7.547	0.55
t9	80	15.78	7.392	0.58
t10	80	15.75	7.937	0.61

### E. Construct Validity

Two separate factor analyses were run to probe the underlying constructs of the PET, pretest, posttest of reading comprehension and the ten quizzes. The results of the factor analysis using a varimax rotation method indicated that the PET, pretest and posttest of reading comprehension measured two distinct traits which accounted for 80.35 percent of the total variance. As displayed in Table 2, the pretest and posttest of reading comprehension loaded on the first factor which can be labeled as “reading ability” factor and the PET test load on the second factor which can be called “general language proficiency” factor.

**Table 2: Rotated Component Matrix**

	Component	
	1	2
Pretest	.837	
Posttest	.835	
PET		.992

### F. Construct Validity of Ten Quizzes

The results of the factor analysis using a varimax rotation method indicated that the ten quizzes measured a single trait which accounted for 58.43 percent of the total variance

(Table 3). Table 3 displays the factor loadings of the ten quizzes under the only extracted factor. These results suggested that the quizzes measured the same trait across ten sessions.

**Table 3: Component Matrix**

	Component
	1
t3	.841
t9	.820
t1	.787
t5	.773
t6	.767
t10	.760
t2	.741
t4	.733
t7	.724
t8	.686

### III. RESULTS AND DATA ANALYSIS

Having checked the data in terms of normality assumptions, the researchers employed parametric statistical analyses. So, mainly t-tests, ANOVA and MANOVA were run.

#### A. Testing Assumptions

The first assumption [tables are not reported due to space restrictions], checked through data skewness and kurtosis, showed neither ratio was over the absolute value of 1.96; indicating normal distribution of the data. The respective tables are not included here due to space restrictions.

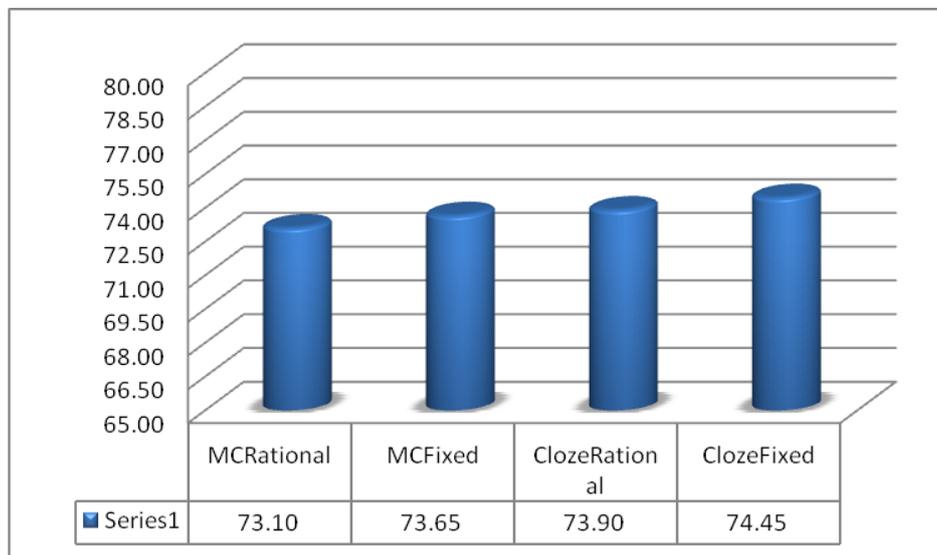
#### B. The PET (General Language Proficiency)

The parametric one-way ANOVA was run to compare the fixed and rational cloze-elide and multiple choice (MC) groups' means on the PET in order to prove that they enjoyed the same level of general language proficiency prior to the main study. The assumption of homogeneity of variances was met since (Levene's  $F(3, 76)$  sig.  $.98 = .039$ ,  $P > .05$ ). The descriptive statistics for the four groups on the PET test are as follows: The rational MC ( $M = 73.10$ ,  $SD = 12.90$ ), fixed MC ( $M = 73.65$ ,  $SD = 12.93$ ), rationale cloze-elide ( $M = 73.90$ ,  $SD = 12.59$ ) and fixed cloze-elide ( $M = 74.45$ ,  $SD = 12.57$ ) the groups showed almost the same means on the PET.

Based on the results displayed in Table 4 ( $F(3, 76) = .039, P > .05, \omega^2 = .037$ , representing a weak effect size), it can be concluded that there were not significant differences among the means of the four groups on the PET. Thus, the groups proved to be homogeneous in terms of their general language proficiency, too.

**Table 4: One-Way ANOVA, PET by Groups**

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	18.850	3	6.283	.039	.990
Within Groups	12359.100	76	162.620		
Total	12377.950	79			



**Figure 1: PET by Groups**

### C. Pretest of Reading Comprehension

A one-way ANOVA was also run to compare the four groups' means on the pretest of reading comprehension in order to prove that they enjoyed the same level of reading ability prior to the main study. Similarly, the assumption of homogeneity of variances was met (Levene's  $F(3, 76) = 2.54, P > .05$ ) (Table 5).

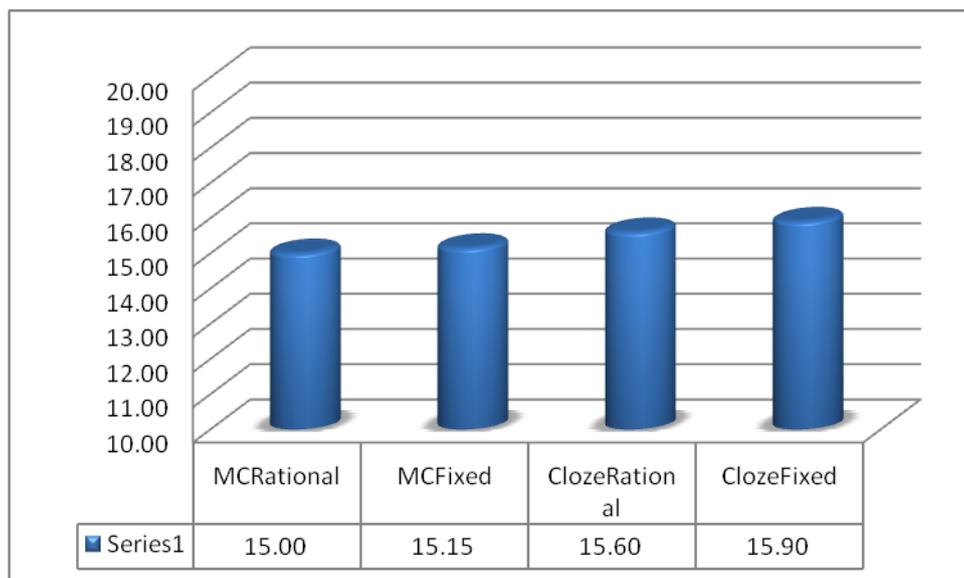
**Table 5: Levene's Test of Equality of Error Variances**

F	df1	df2	Sig.
2.544	3	76	.062

The descriptive statistics for the four groups on the pretest of reading comprehension are as follows: The rational MC (M = 15, SD = 1.62), fixed MC (M = 15.15, SD = 2.58), rationale cloze-elide (M = 15.60, SD = 2.81) and fixed cloze-elide (M = 15.90, SD = 1.80) the groups showed almost the same means on the pretest of reading comprehension. Based on the results displayed in Table 6 ( $F(3, 76) = .66, P > .05, \omega^2 = .013$ , representing a weak effect size), it can be concluded that there were not significant differences among the means of the four groups on the Pretest of Reading Comprehension. Thus, it can be claimed that they were homogeneous in terms of their reading ability prior to the main study.

**Table 6: One-Way ANOVA, Pretest of Reading Comprehension by Groups**

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	10.238	3	3.413	.666	.575
Within Groups	389.150	76	5.120		
Total	399.388	79			



**Figure 2: Pretest of Reading Comprehension Test by Groups**

#### D. Investigation of the Research Questions

##### Macro Research Question:

A one-way ANOVA plus planned (a-priori) contrasts were run to probe the research questions posed in this study. Before discussing the results, it should be mentioned that the assumption of homogeneity of variances was met (Levene's  $F(3, 76) = 2.49, P > .05$ ) (Table 7).

**Table 7: Levene's Test of Equality of Error Variances**

F	df1	df2	Sig.
2.495	3	76	.066

The descriptive statistics for the four groups on the posttest of reading comprehension are as follows: The rational MC (M = 18, SD = 2.10) had the highest mean on the posttest of reading comprehension. This was followed by fixed MC (M = 17.35, SD = 1.92), rationale cloze-elide (M = 15.35, SD = 2.47) and fixed cloze-elide (M = 15.15, SD = 1.53) groups.

Based on the results displayed in Table 8 ( $F(3, 76) = 9.81, P < .05, \omega^2 = .24$  representing a large effect size), it can be concluded that there were significant differences among the means of the four groups on the posttest of Reading Comprehension. Thus, it can be claimed that the major null-hypothesis as there was not any relationship between cloze test varieties and procedures in developing EFL learners' reading performance **was rejected**.

**Table 8: One-Way ANOVA, Posttest of Reading Comprehension by Groups**

	Sum Squares	of Df	Mean Square	F	Sig.
Between Groups	122.238	3	40.746	9.810	.000
Within Groups	315.650	76	4.153		
Total	437.887	79			

Although the F-value of 9.81 indicated significant differences between the means of the four groups on the posttest of reading comprehension, the planned (a-priori) contrasts were run in order to probe the three minor research questions. The three planned contrasts were compared:

- 1: cloze-elide fixed deletion with the other three groups;
- 2: cloze-elide rational deletion with fixed and rational MC groups, and
- 3: fixed MC with rational MC.

Based on the results displayed in Table 9, it can be concluded that:

A: The combination of the following three groups: cloze-elide rational and fixed and rational MC groups (M = 16.90) significantly outperformed the cloze-elide fixed deletion group (M = 15.15) on the posttest of reading comprehension. They mean difference (MD = 1.75;  $t = 3.32, p < .05$ ) indicated that the cloze-elide fixed deletion group had significantly lower mean than the other three groups. Thus, the first minor null-hypothesis as the integration of cloze-elide test with the fixed-ratio deletion procedure was not more effective than that of

cloze-elide test with rational deletion procedure and multiple choice with the rational and fixed-ratio deletion procedures **was not rejected**. The combination of fixed and rational MC groups outperformed the cloze-elide fixed deletion group on the posttest of reading comprehension.

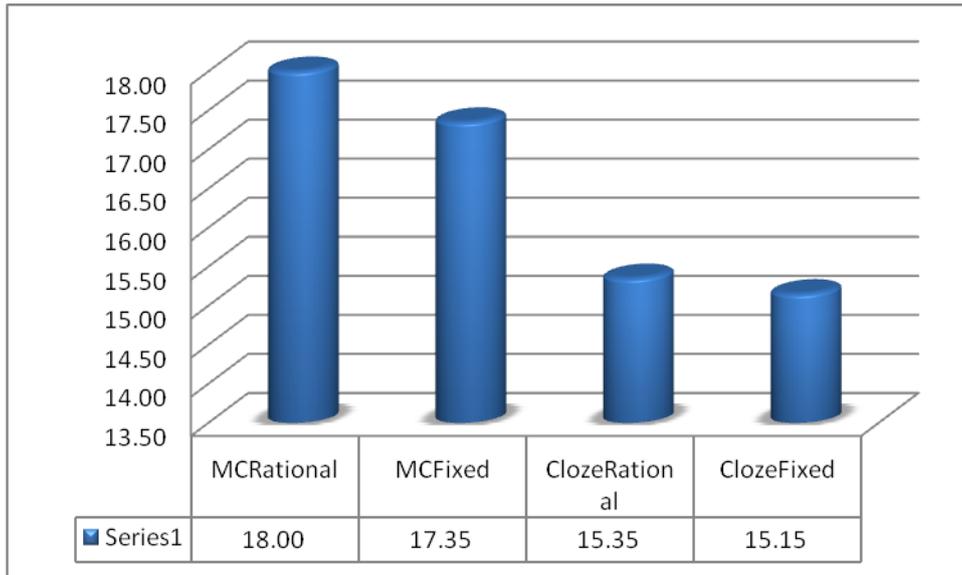
**Table 9: Contrast Tests**

	Contrast	Value of Contrast	Std. Error	T	df	Sig. (2-tailed)
Assume	G4 vs. G1+G2+G3	1.75	.526	3.323	76	.001
equal	G3 vs. G2+G1	2.33	.558	4.166	76	.000
variances	G1 vs G2	.65	.644	1.009	76	.316
Does not	1	1.75	.443	3.944	46.049	.000
assume equal	2	2.33	.639	3.638	31.918	.001
variances	3	.65	.638	1.019	37.714	.315

Note. Since the assumption of homogeneity of variances was met (Table 4.8) the results of “assume equal variances” were reported.

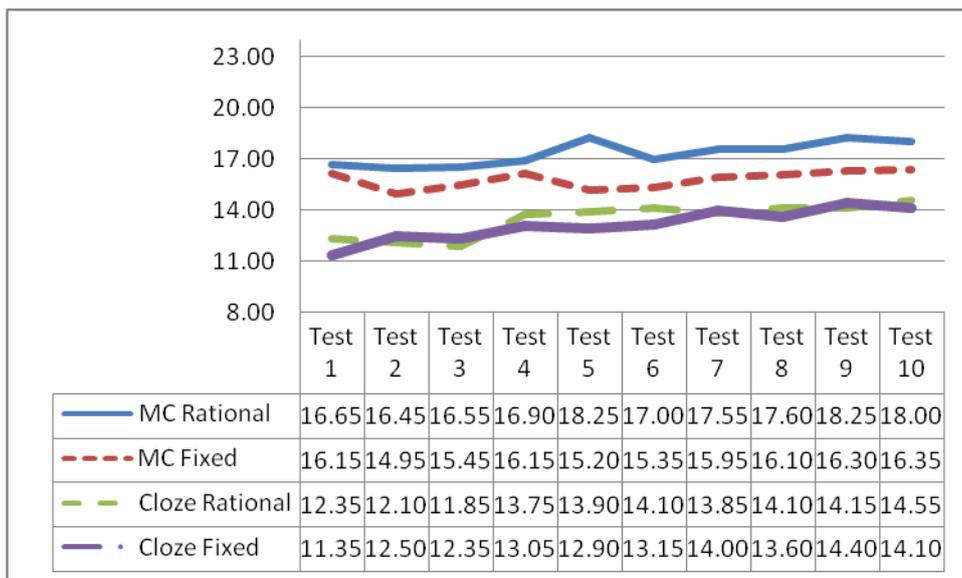
B: The combination of the fixed and rational MC groups ( $M = 17.68$ ) significantly outperformed the cloze-elide rational deletion group ( $M = 15.35$ ) on the posttest of reading comprehension. The mean differences ( $MD = 2.33$ ;  $t = 4.16$ ,  $p < .05$ ) indicated that the cloze-elide fixed deletion group had significantly lower mean than the other two groups. Thus, the second minor null-hypothesis as the integration of cloze–elide test with the rational deletion procedure was not more effective than that of multiple choice form with fixed-ratio and rational deletion procedures **was not rejected**. The combination of fixed and rational MC groups outperformed the cloze-elide rational deletion group on the posttest of reading comprehension.

C: There was not any significant difference between fixed MC group ( $M = 17.35$ ) and rational MC ( $M = 18$ ) group on the posttest of reading comprehension ( $MD = .65$ ;  $t = 1$ ,  $p > .05$ ). Thus, the third null-hypothesis **was not rejected**.



**Figure 3: Posttest of Reading Comprehension Test by Groups**

In order to investigate the minor research question one addressing if the distinct integration of cloze test varieties and procedures have any significant effects on developing the EFL learner’s reading comprehension ability, the subjects took ten quizzes before the protest of reading comprehension. Line Graph 4 displays the means for the four groups on the ten tests. The rational MC group had the highest means on all sessions. This was followed by fixed MC which took the second passion across the ten tests. Despite, the first few sessions, the rational cloze-elide group showed higher means than the fixed cloze-elide group.



**Line Graph 4: Ten Tests by Groups**

All four groups – disregarding minor deviations from their increasing trend – showed an increase in their means across the ten sessions of which the ones for the rational MC and rational cloze-elide are more prominent. Four separate repeated measures ANOVA were run to investigate the development of reading ability of each group across the ten sessions.

**E. Rational Multiple Choice Group**

The results of repeated measures ANOVA ( $F(9, 11) = 2.80, p > .05, \text{partial } \eta^2 = .69$  representing a moderate effect size) (Table 10) indicated that there were not any significant differences among the means of the rational MC groups across ten sessions since neither the F-values nor those of the sig met the criteria of significance.

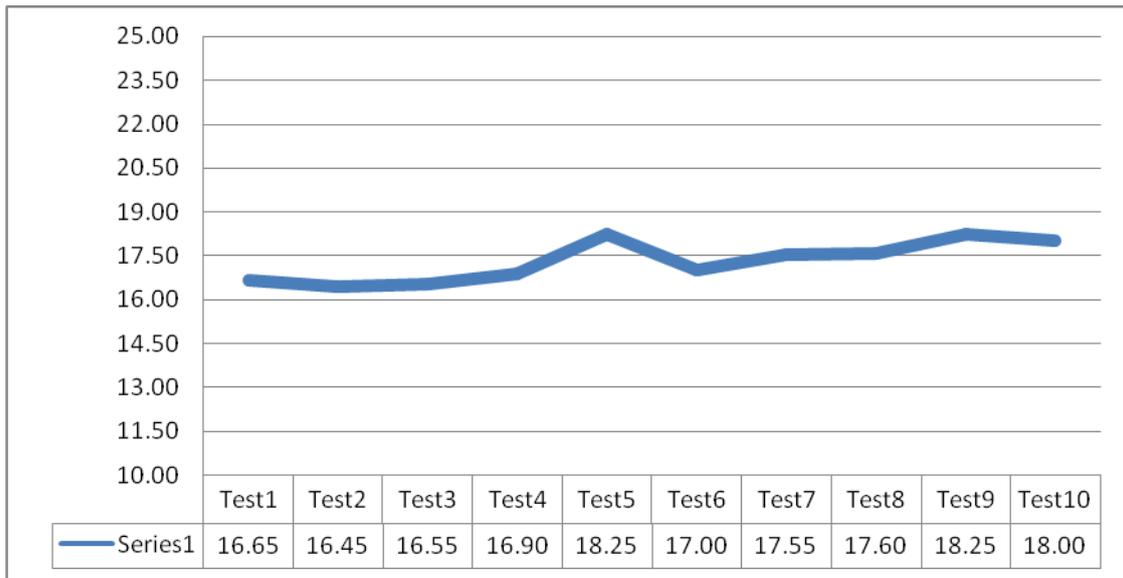
**Table 10: Multivariate Tests; MC Rational Group’s Ten Quizzes**

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Quizzes	Pillai's Trace	.697	2.809	9	11	.055	.697
	Wilks' Lambda	.303	2.809	9	11	.055	.697
	Hotelling's Trace	2.298	2.809	9	11	.055	.697
	Roy's Largest Root	2.298	2.809	9	11	.055	.697

Table 11 and Line Graph 5 display the increasing trends of means of the MC rational group.

**Table 11: Descriptive Statistics; MC Rational Group’s Ten Quizzes**

Quizzes	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	16.650	.412	15.787	17.513
2	16.450	.444	15.520	17.380
3	16.550	.559	15.379	17.721
4	16.900	.778	15.272	18.528
5	18.250	.403	17.406	19.094
6	17.000	.543	15.864	18.136
7	17.550	.618	16.257	18.843
8	17.600	.483	16.588	18.612
9	18.250	.481	17.244	19.256
10	18.000	.447	17.064	18.936



**Line Graph 5: MC Rational Groups' Ten Quizzes**

**F. Fixed Multiple Choice Group**

Contrary to rational multiple-choice groups, the results of repeated measures ANOVA ( $F(9, 11) = 3.66, p < .05, \text{partial } \eta^2 = .75$ , representing a moderate effect size) (Table 12) indicated that there were significant differences between the means of the fixed MC groups across the ten sessions since all sig values are lower than the set p-value (i.e., .05).

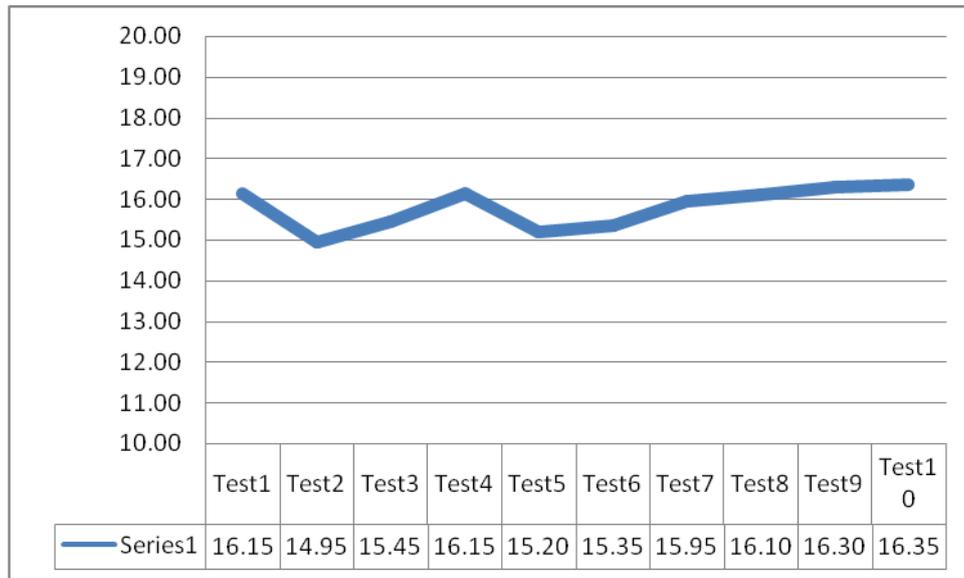
**Table 12: Multivariate Tests; MC Fixed Group's Ten Quizzes**

Effect	Va Lue	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Quizzes	Pillai's Trace	.750	3.665	9	11	.023	.750
	Wilks' Lambda	.250	3.665	9	11	.023	.750
	Hotelling's Trace	2.998	3.665	9	11	.023	.750
	Roy's Largest Root	2.998	3.665	9	11	.023	.750

As displayed through Table 13 and Line Graph 6, it can be concluded that the MC fixed group has fallen of means on the second and fifth sessions beyond which the means showed an increasing pattern.

**Table 13: Descriptive Statistics; MC Fixed Group's Ten Quizzes**

Quizzes	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	16.150	.504	15.095	17.205
2	14.950	.701	13.482	16.418
3	15.450	.694	13.998	16.902
4	16.150	.449	15.211	17.089
5	15.200	.663	13.812	16.588
6	15.350	.582	14.133	16.567
7	15.950	.510	14.882	17.018
8	16.100	.542	14.965	17.235
9	16.300	.612	15.020	17.580
10	16.350	.741	14.799	17.901



**Line Graph 6: MC Fixed Groups' Ten Quizzes**

### G. Rational Cloze-Elide Group

The results of repeated measures ANOVA ( $F(9, 11) = 3.40, p < .05, \text{partial } \eta^2 = .73$  representing a moderate effect size) (Table 14) indicated that there were significant differences between the means of the rational cloze-elide groups across ten sessions because all criteria values (i.e., sig and F) cooperatively are in favor of rejecting the hypothesis.

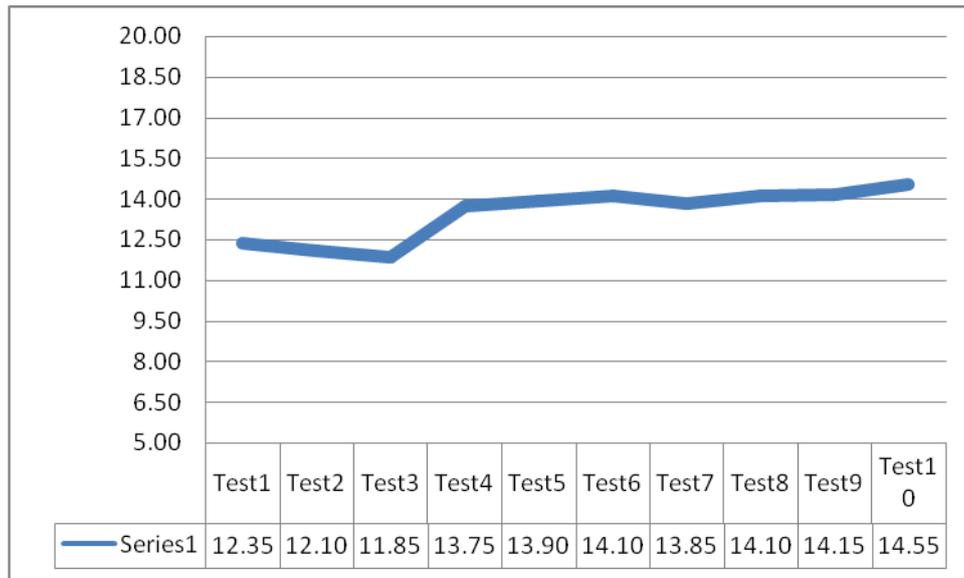
**Table 14: Multivariate Tests; Cloze-Elide Rational Group's Ten Quizzes**

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Quizzes	Pillai's Trace	.736	3.408	9	11	.030	.736
	Wilks' Lambda	.264	3.408	9	11	.030	.736
	Hotelling's Trace	2.789	3.408	9	11	.030	.736
	Roy's Largest Root	2.789	3.408	9	11	.030	.736

Table 15 and Line Graph 7 display an increasing trend of means of the cloze-elide rational group despite the fact that they show a certain decline on the third and seventh sessions.

**Table 15: Descriptive Statistics; Cloze-Elide Rational Group's Ten Quizzes**

Quizzes	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	12.350	.646	10.998	13.702
2	12.100	.593	10.858	13.342
3	11.850	.654	10.481	13.219
4	13.750	.632	12.427	15.073
5	13.900	.589	12.667	15.133
6	14.100	.507	13.038	15.162
7	13.850	.544	12.711	14.989
8	14.100	.528	12.996	15.204
9	14.150	.525	13.052	15.248
10	14.550	.578	13.340	15.760



**Line Graph 7: Cloze-Elide Rational Groups' Ten Quizzes**

### H. Fixed Cloze-Elide Group

The results of repeated measures ANOVA ( $F(9, 11) = 4.68, p < .05, \text{partial } \eta^2 = .75$ , representing a moderate effect size) (Table 16) indicated that there were significant differences between the means of the fixed cloze-elide groups across ten sessions as the sig and F values prove.

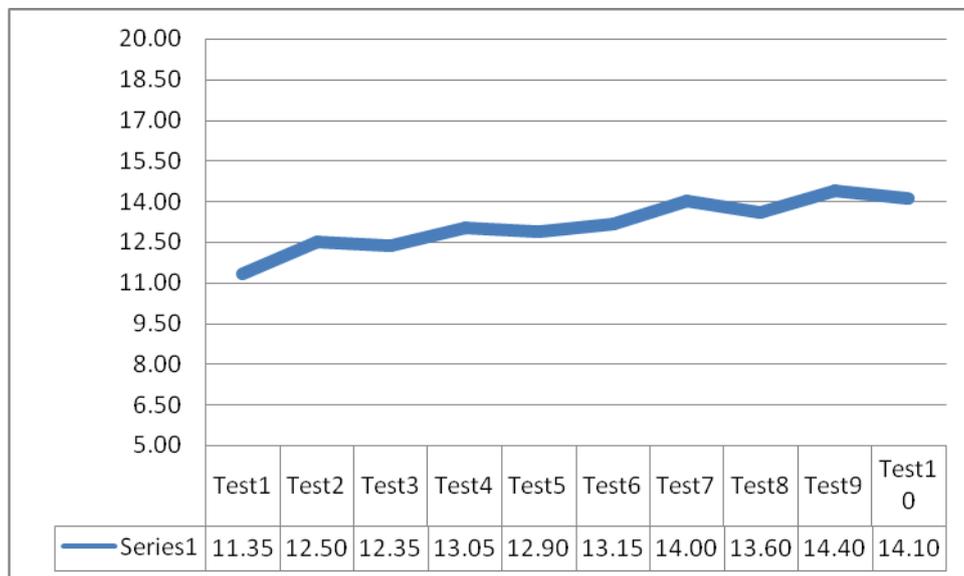
**Table 16: Multivariate Tests; Cloze-Elide Fixed Group's Ten Quizzes**

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Quizzes	Pillai's Trace	.793	4.681	9	11	.010	.793
	Wilks' Lambda	.207	4.681	9	11	.010	.793
	Hotelling's Trace	3.830	4.681	9	11	.010	.793
	Roy's Largest Root	3.830	4.681	9	11	.010	.793

Table 17 and Line Graph 8 show, it can be concluded that the cloze-elide fixed group has fallen off means on the third, fifth, eighth and tenth sessions.

**Table 17: Descriptive Statistics; Cloze Fixed Group's Ten Quizzes**

Quizzes	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	11.350	.514	10.273	12.427
2	12.500	.521	11.410	13.590
3	12.350	.539	11.221	13.479
4	13.050	.438	12.133	13.967
5	12.900	.486	11.883	13.917
6	13.150	.568	11.961	14.339
7	14.000	.459	13.040	14.960
8	13.600	.472	12.611	14.589
9	14.400	.285	13.804	14.996
10	14.100	.250	13.576	14.624



**Line Graph 8: Cloze-Elide Fixed Groups' Ten Quizzes**

Based on the results which indicated one non-significant and three significant but moderate differences among the means and an inspection of the line graphs which did not show any steady up-ward pattern for any groups, it can be claimed that the fifth null-hypothesis **was not fully rejected**.

#### **IV. DISCUSSION AND CONCLUSIONS**

This study explored the washback of cloze test varieties and procedures on the performance of the EFL learners as a measure of reading comprehension skill. As the data revealed in table 4 ( $F(3, 76) = .039, P > .05, \omega^2 = .037$ ), the means of the four groups on the PET were almost the same. So, the result indicated that the four groups were homogeneous in terms of their general language proficiency. In addition, based on the four groups means on the pretest of reading comprehension presented in table 6 ( $F(3, 76) = .66, P > .05, \omega^2 = .013$ ), the F ratio showed that they were homogeneous in terms of their reading ability. Then, posttest scores were analyzed to investigate the washback of rational and fixed cloze-elide and MC tests on the performance of the EFL learners after ten sessions. The data obtained in table 8 ( $F(3, 76) = 9.81, P > .05, \omega^2 = .024$  representing a large effect size), showed that there were significant differences among the means of the four groups on the posttest of reading comprehension. Therefore, based on the data obtained from this study, it can be claimed that in developing EFL learners' reading ability, not only cloze test varieties but also its procedures can have an effective influence.

By the same token, many researchers such as Taylor (1953), Bachman (1985), Oller and Conrad (1971), Manning (1978) and Gellert and Elbro (2013) supported the claim that cloze test highly correlates with standard tests of reading comprehension. Therefore, it can be used as a valid and reliable measurement tool. Moreover, the results of Sadeghi's (2008) study showed a significant correlation between cloze test and reading section of IELTS, so it can be concluded that both tests measure the same trait. Therefore, according to Shohamy (1983) and Oller (1973), cloze test appears to be a valid tool to measure reading comprehension (cited in Sadeghi, 2008. p.119). In line with Sadeghi (2008), Tabatabaei and Mirzaei (2014) explored a high correlation among participants' scores on c-test and standard cloze test against IELTS as a criterion measure. However, testees' scores on c - tests showed that c-test was a more valid test of reading comprehension than standardized cloze test. Totally, it can be claimed that in measuring reading comprehension ability, cloze tests have proved as a valid measurement device because of correlation with IELTS and other Standardized tests.

Indeed, given the findings of the first minor null hypothesis, it can be claimed that the following three groups; cloze-elide rational and fixed and rational MC groups ( $M = 16.90$ ) significantly outperformed the cloze-elide fixed deletion group ( $M = 15.15$ ) on the posttest of reading comprehension. Their means difference ( $MD = 1.75; t = 3.32, p < .05$ ) indicated that the cloze-elide fixed deletion group had significantly the lowest mean of the other three groups. Thus, the integration of cloze-elide test with fixed-ratio deletion procedure was not more effective than that of cloze-elide test with rational deletion procedure and multiple choice with the rational and fixed- ratio deletion procedures.

Accordingly, the findings of the present study are in line with those of Rostamzadeh (2010) who investigated cloze test varieties and c-test to compare which test forms have superiority to show individuals' performance on reading test. The reported finding revealed that rational deletion cloze test has the highest effect among the other formats. These results lend support to the findings of Bachman (1985) which supported rational cloze test as more

appropriate for measuring reading comprehension ability since it addressed discourse knowledge and syntax.

Furthermore, the result interpreted from the second and third minor null hypotheses showed that there were not any significant difference between fixed MC group ( $M = 17.35$ ) and rational MC ( $M = 18$ ) group on the posttest of reading comprehension ( $MD = .65$ ;  $t = 1$ ,  $p > .05$ ). So, it can be concluded that fixed and rational MC groups outperformed the cloze-elide rational deletion group on the posttest of reading comprehension. These results proved that rational MC cloze test ( $M = 18$ ,  $SD = 2.10$ ) had the highest mean on the posttest of reading comprehension. Moreover, Palupiningsih's (2011) study, in line with this research, has shown that participants in cloze-elide had a poor performance compared to the standardized cloze test. On the other hand, Bachman's (1985) study showed that learners' performance on fixed ratio compared with the rational cloze test was different, as fixed ratio participants found it hard to answer. Additionally, as cited from Chen (2008) the obtained results are in line with some findings such as those by Abraham and Chapelle (1992), Hinofotis (1987) and Yamashita (2003).

Eventually, the result as to the fourth minor null hypothesis showed that four cloze test groups' mean scores increased across the ten sessions over those of the rational MC and rational cloze-elide was more prominent. The rational MC group had the highest means on all sessions, which was followed by fixed MC. Without considering the first few sessions, the rational cloze-elide group showed higher means than the fixed cloze-elide group. Thus, it can be concluded that individuals outperformed on rational and fixed MC cloze tests, which can be interpreted as significant differences among the means of the fixed cloze-elide groups across ten sessions but cloze-elide fixed group had fallen off means on the third, fifth, eighth and tenth sessions. In addition, according to the reported results by Ajideh and Sattarpour (2014) who investigated the relatedness of cloze-elide test, multiple-choice cloze test, and c-test as measures of reading comprehension, there were not any significant difference among the performance of the testees' scores on the MC cloze test, cloze-elide test and c-test. Although the testees' performances on MC cloze test and c-test were approximately equal, they had poor performance in cloze-elide. So, it can be concluded that MC cloze test entails in much influence as to enhancing learners' reading comprehension ability.

In a nutshell, the findings revealed that different test formats affect learners' performance. Moreover, employing different test methods has positive washback on both learning and teaching process. So, different test methods and varieties of cloze test can contribute to the developing learners' reading ability. Nevertheless, the fixed-ratio group showed poor performance on the cloze test because of the students' unfamiliarity with cloze-elide test format, particularly that of fixed-ratio deletion procedure, and more importantly having a nodding acquaintance with semantic and syntactic features of words. On the other hand, MC test is regarded as a kind of proper and learner-friendly test type in Iranian schooling system in terms of all dimensions of test practicality including: easy to answer, easy to score, easy to construct, etc. Therefore, a proper test-based instruction on how to do different test formats should be supplemented into the classroom activities by the teachers. Ultimately, this study evidently introduces an "integrated instruction"; testing-teaching cycle,

as an effective, practical and pedagogical mechanism as to developing reading ability, but leaves the room open and warranted for further and similar studies on the other three remaining skills.

## REFERENCES

- Alderson, J.C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129. Retrieved from [http://www.comp.dit.ie/dgordon/Courses/ILT/ILT0005/ Does washbackExist.pdf](http://www.comp.dit.ie/dgordon/Courses/ILT/ILT0005/Does_washbackExist.pdf)
- Alderson, J. C. (2000). *Assessing Reading*. New York: Cambridge American Children. Unpublished doctoral dissertation, University of A Language Arts, *Journal of Michigan Volume 8 Issue 1 Historical*. Retrieved From <http://www.jstor.org/discover/10.237/3586211?sid=21105057523741&uid=70&uid=2129&uid=4>
- Ajideh, P., & Sattarpour, S.(2014). Investigating the Relatedness of Cloze-Elide Test, Multiple-Choice Cloze Test, and C-test as Measures of Reading Comprehension. *Journal of English Language Teaching and Learning No. 13, 2014*. Retrieved from [http://elt.tabrizu.ac.ir/pdf \\_2171\\_bd20aeca0612bdf0737a9765b747471d.html](http://elt.tabrizu.ac.ir/pdf/_2171_bd20aeca0612bdf0737a9765b747471d.html)
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: OUP. New York, ISBN 0194370038.Tehran: Rahnama, LB3060.
- Brown, H. D. (2004). *Language Assessment Principle and Classroom Practices* California, Los Angeles.Practices White Plains, NY: Pearson Education.
- Chapelle, C. A., & Abraham, R. G. (1990). Cloze method: What difference does it make? *Language Testing*, 7 (2), 121-146.retrieved from [ttp://ltj.sagepub.com/content/7/2/121](http://ltj.sagepub.com/content/7/2/121).
- Chen, W.Y.E. (2008). The Relationship between Rational Cloze Test and the Discourse Structure Test.Retrieved from [http://ir.lib.ntnu.edu.tw/retrieve/51106/metadata\\_02\\_02\\_s\\_05\\_0076.pdf](http://ir.lib.ntnu.edu.tw/retrieve/51106/metadata_02_02_s_05_0076.pdf).
- Cheng, L. (1997). How does wash back influence teaching? Implications for Hong Kong. *Language and Education*, 11, 38-54.Retrieved from [http://www.clal.org.cn/personal/testing/washback/cheng1997\(washback\).pdf](http://www.clal.org.cn/personal/testing/washback/cheng1997(washback).pdf).
- Culhane, J. W. (1970). *Cloze procedures and comprehension*. *Reading Research*, 23(5), 410-413, 464. Retrieved from [10.2307/20196335sid= 21105058007141 &uid=70 &uid=2129&uid=4&uid=2](http://www.jstor.org/discover/10.2307/20196335?sid=21105058007141&uid=70&uid=2129&uid=4&uid=2).
- Darwesh, A. A. J., (2010). Cloze Tests: An Integrative Approach. Retrieved from <http://www.icsei.net/icsei2011/Full%20Papers/0196.pdf>
- Farhady, H. (2006). *Twenty-five years of living with applied linguistics: Collection of Articles*. Tehran: Rahnam Publication.
- Farhady, H. & Keramti, M.N. (1996). A Text-driven Method for the Deletion Procedure in Cloze Passage. *Language Testing*, 13 (2). Retrieved from Farhady, H. (2006). *Twenty-five years of living with applied linguistics, Collection of Articles*. Tehran: Rahnama Publications.

- Fulcher, G. (2010). *Practical Language Testing*. London. Hodder Education/Routledge. ISBN-10: 0340984481; 342 pages + xiv.
- Gellert, A.S., & Elbro, C (2013). Cloze Tests May be Quick, But Are They Dirty? Development and Preliminary Validation of a Cloze Test of Reading Comprehension. *Journal of Psychoeducational Assessment*, 31(1) 16-28 © 2013 SAGE Publications Reprints and permission: sagepub.com/journals Permissions.navDOI: 10.1177/0734282912451971. Retrieved from <http://jpa.sagepub.com>.
- Hadley, G. & Naaykens, J. (1999). Testing the Test: Comparing SEMAC and Exact Word Scoring on the Selective Deletion Cloze. Retrieved from <http://www.nuis.ac.jp/~hadley/publication/kortesol/Hadley-Naaykens-KOTESOL.pdf>
- Harwood, Ch. (2007). Washback and the Cambridge ESOL Key English Test Speaking Component: a study from Japan a study from Japan. University of Leicester MA Applied Linguistics & TESOL Dissertation. Retrieved from [https://www.academia.edu/322863/Washback\\_and\\_the\\_Cambridge\\_ESOL\\_Key\\_English\\_Test\\_Speaking\\_Component\\_a\\_study\\_from\\_Japan](https://www.academia.edu/322863/Washback_and_the_Cambridge_ESOL_Key_English_Test_Speaking_Component_a_study_from_Japan).
- Heaton, J. B. (1991). *Writing English Language Tests. Writing English Language Tests*. Hong Kong:Longman. New York Hebrew University Retrieved from <http://www.scribd.com/doc/135763874/J-B-Heaton-Writing-English-Language-Tests-Lon-BookFi-org#scribd>
- Lu, G. (2006). Cloze Test and Reading Strategies in English Language. University of the Western Cape.Retrieved from [http://etd.uwc.ac.za/xmlui/bitstream/handle/11394/1676/Lu\\_MED\\_2006.pdf?sequence=1](http://etd.uwc.ac.za/xmlui/bitstream/handle/11394/1676/Lu_MED_2006.pdf?sequence=1).
- McNamara, T. (2005). Language Testing. Oxford Introduction to language study. Challenges for Research. *Language Testing*, 18(4), 333-349.
- Manning, M. (1978). A Comparison among Measures of Reading Achievement with Low Income Black Third Grade Students. TM 850 474. Retrieved from <http://files.eric.ed.gov/fulltext/ED261074.pdf>.
- Palupiningsih, A. (2010, 2011). Testing Reading Comprehension using Approach. PALGRAVE, *assessing reading* Article 6 1-1-1992 Early English Language Arts. Retrieved from [http://eprints.uny.ac.id/1059/1/Angesti\\_Palupiningsih\\_06202244139.pdf](http://eprints.uny.ac.id/1059/1/Angesti_Palupiningsih_06202244139.pdf)
- Rostamzadeh, A. (2010). *Cloze Test and C-test*. Zaban-Va-Adab-NO26-Allameh Tabataba'i University.
- Sadeghi, k. (2008). Cloze validation against IELTS Reading Paper: Doubts on Correlational Validation. *Journal of English Language Teaching and Learning* Year 53 No. 217. Reterived from [http://www.sid.ir/en/vewssid/j\\_pdf/1323201021706.pdf](http://www.sid.ir/en/vewssid/j_pdf/1323201021706.pdf).
- Sadeghi, K. (2008). Measuring Reading Comprehension: The Judgmental Validity of Cloze Procedure. Urmia University. *IJAL*, Vol. 11, No. 2, September 2008.
- Tabatabaei.O., & Mirzaei. E. (2014). Correlational Validation of Cloze Test and C - Test against IELTS.Journal of Educational and Social Research /MCSER Publishing, Rome-Italy. Vol. 4 No.1 January 2014.