International Journal of Research in English Education

Online ISSN: 2538-3027

Homepage: https://ijreeonline.com

A Comparative Study on Washback Effect on Task-Based and Traditional Assessment with Regard to Gender and Attitude

Mojtaba Aghajani[1], **Hamidreza Khalaji**[2]✉, **Arezoo Ashoori Tootkaboni**[3], and **Abbas Bayat**[4]

1. Department of ELT, University of Mazandaran, Mazandaran, Iran, Email: Mojtaba_aghajani45@yahoo.com
2. Department of ELT, MAL.C., Islamic Azad University, Malayer, Iran, Email: Hr.khalaji@iau.ac.ir
3. Department of English Language Teaching, Farhangian University, Tehran, Iran, Email: a.ashoori@cfu.ac.ir.
4. Department of ELT, MAL.C., Islamic Azad University, Malayer, Iran, Email: 3932653610@iau.ac.ir

Article type:

Research Article

Article history:
Received:
April 25, 2026
Revision:
June 22, 2026
Accepted:
June 30, 2026
Available online:
July 30, 2026

Keywords:
attitude,
reading skill,
task-based language assessment,
washback effect

Abstract

Objective: The objectives of this mixed-methods study were to investigate the washback effect of Task-Based Language Assessment (TBLA) on the development of reading skills among EFL learners, as well as to explore students' attitudes toward the implementation of TBLA.

Methods: A total of 108 male and female EFL learners were selected randomly and assigned to two equal groups of 54 (a control group and an experimental group). A Preliminary English Test (PET) was administered to ensure the homogeneity of the participants' proficiency levels. Both groups attended a 90-minute reading class per week over a period of three months (24 sessions in total) and received identical instruction. However, the experimental group was additionally assessed using researcher-designed task-based reading quizzes every four sessions. Pre- and post-test data were collected to measure reading improvement. Furthermore, an "Attitude Questionnaire on Reading" was administered to the experimental group, and the data were analyzed using a paired-samples t-test to evaluate learners' perceptions of TBLA.

Results: The statistical analysis of the data indicated that the experimental group experienced a significantly positive washback effect, leading to greater improvements in reading skills compared to the control group. Additionally, the quantitative analysis of the questionnaire results revealed that the participants held a highly positive attitude and showed a strong tendency toward the application of TBLA in the reading classroom.

Conclusion: The results of this study suggest that TBLA serves as an effective alternative to traditional assessment procedures. Incorporating task-based assessments can maximize educational achievements and foster more positive student attitudes toward language learning and development.

Cite this article: Aghajani, M., Khalaji, H., Ashoori Tootkaboni, A., and Bayat, B. (2026). A Comparative Study on Washback Effect on Task-Based and Traditional Assessment with Regard to Gender and Attitude. ***International Journal of Research in English Education***, *11*(2), 81-97.

© The author(s) retain the copyright.

Publisher: Science Academy Publications.

1. Introduction

[Downloaded from ijreeonline.com on 2026-11-09]

Tests are part of educational courses where decisions are made with regard to the feedback acquired from a test. It has been supposed that tests have direct impact on educational courses in many ways. Some hypotheses say that teachers may be tempted by the awareness the learners might bring when taking a specified test. In this case, educators and teachers often adjust the teaching as well as the material to, in some way, teach the test. Test takers and teachers will attain more information about learners, about learning and also about their instruction through washback. Washback is characterized in different ways by several practitioners and researchers, with regard to the fields of study they have.

Some scholars and language practitioners emphasize solely on the mere learners and learning process whilst a few worked on its output and the consequences that it may has on teacher and teaching, university, programs of study, educational organization, teaching material, and even the educational environments and the society in where the teaching process is being done. Considering the importance of preparing and compiling suitable educational materials to strengthen the students' knowledge of idioms, and based on the appropriate educational approach to strengthen them, especially the language-oriented task, the purpose of this research was to examine the probable effect of TBLA on EFL students' knowledge of idioms (Afshari,Taghipor, 2025)

1.1 Statement of the Problem

Despite the widespread use of reading assessments in educational contexts, traditional discrete-point tests often fail to reflect learners' true communicative competence, leading to negative washback on both teaching methodologies and student learning. The over-reliance on multiple-choice formats encourages teaching to the test rather than fostering authentic language use. Recent international findings indicate that traditional assessment frameworks globally contribute to a decrease in intrinsic reading motivation among EFL learners. Furthermore, national studies within the Iranian context highlight that standard university entrance exams compel instructors to neglect task-based communicative activities, resulting in a superficial understanding of texts. Therefore, there is a critical need to explore how Task-Based Language Assessment (TBLA) might provide a positive washback effect compared to traditional paradigms.

Many studies have been conducted that have investigated the effect of TBLT on developing the reading ability of the learners (e.g., Foster & Skehan, 1996; Skehan, 1998), but few have examined the effect of TBLA on the reading skills of EFL learners. While some researchers propose that the traditional methods include prearranged phases that offer teachers a clear plan of what they should do, some other researchers put emphasis on the importance of task-based approaches to communicative instruction which help teachers and learners to find their own procedures freer in order to maximize communicative effectiveness (Long & Crookes, 1991; Nunan, 1989; Prabhu, 1987; Skehan, 1996).

Testing systems that are task-based involve language learners in kinds of goal-oriented performance and require learners to use language in real environment to gain their purpose. This way, real-world is considered by teachers as a standard to assess the behaviors produced by students on the proposed tasks. Areas of the EFL curriculum are defied by TBLA. As Norris (2009) claimed, task-based language assessment is narrowly connected to establishment of task-based language teaching by giving learners' needed material so that it supports learning a language and also raises students' skills to use the learnt language.

This role of tests in language assessment often forces teachers to teach students as if there are being taught how to give a test. This is a common fact that is accepted by most teachers, instructors and students that teachers mostly teach for exams and attract students to study so that they just pass the exams.

Still, few studies are available on employing TBA in testing a skill separately, like reading comprehension ability. In current research, researchers try to study the impact of task-based assessment on learners' skill improvement in reading comprehension ability and also to show their attitudes towards it.

The thought that task-based evaluation framework can be utilized to create forecasts approximately execution on learners' future language utilization exterior the test itself emphasizes the statements of TBLPA supporters.

[Downloaded from ijreeonline.com on 2026-11-09]

1.2 Objectives of the Study

The primary objectives of this study are threefold:

1. To investigate the washback effect of Task-Based Language Assessment (TBLA) versus traditional assessment on the reading skill improvement of EFL learners.
2. To determine EFL learners' attitudes toward the implementation of TBLA in their reading comprehension courses.
3. To examine whether the washback effect of TBLA on reading improvement differs significantly between male and female learners.

1.3 Research Questions

As such, the current study seeks to answer the following research questions:

1. Is there any significant difference between the washback effect of traditional assessment and Task-Based Language Assessment (TBLA) on EFL learners' reading skill improvement?

2. What are the EFL learners' attitudes toward the application of Task-Based Language Assessment (TBLA) for reading comprehension?

3. Is there any significant difference in the washback effect of Task-Based Language Assessment (TBLA) on reading improvement between male and female EFL learners?

2. Literature Review

Despite the widespread use of reading assessments in educational contexts, traditional discrete-point tests often fail to reflect learners' true communicative competence, leading to negative washback on both teaching methodologies and student learning. The over-reliance on multiple-choice formats encourages teaching to the test rather than fostering authentic language use. Recent international findings indicate that traditional assessment frameworks globally contribute to a decrease in intrinsic reading motivation among EFL learners. Furthermore, national studies within the Iranian context highlight that standard university entrance exams compel instructors to neglect task-based communicative activities, resulting in a superficial understanding of texts. Therefore, there is a critical need to explore how Task-Based Language Assessment (TBLA) might provide a positive washback effect compared to traditional paradigms.

Many studies have been conducted that have investigated the effect of TBLT on developing the reading ability of the learners (e.g., Foster & Skehan, 1996; Skehan, 1998), but few have examined the effect of TBLA on the reading skills of EFL learners. While some researchers propose that the traditional methods include prearranged phases that offer teachers a clear plan of what they should do, some other researchers put emphasis on the importance of task-based approaches to communicative instruction which help teachers and learners to find their own procedures freer in order to maximize communicative effectiveness (Long & Crookes, 1991; Nunan, 1989; Prabhu, 1987; Skehan, 1996).

Koné (2015) emphasizes in the crucial role a proper assessment may play in language learning and teaching process in classroom environment. This process would be inhibited if assessment does not meet the necessitates and requirements or if it is not administered in a proper way. From the first day that tests become a tool for measuring language knowledge, researchers and test designers have constantly been seeking to find ways to improve the quality of tests so that the knowledge as well as competence of language learners be measured in optimal way and make assessment more helpful for the language learners.

[Downloaded from ijreeonline.com on 2026-11-09]

Buck (1988) describes washback as the effect of a test on a class, teaching and students. He adds another dimension to his definition when he states that the aforementioned effect on tests can be either helpful or harmful. In general, we make decisions about examinees', their weaknesses and strengths and finally about their possible future accomplishment. These implications made of test results stress on the tests' standards and values and have a probable influence on the learning and teaching environment and also an influence on the whole educational structure in where learners are studying (Bachman & Palmer, 1996). This influence, in applied linguistics, is called "washback."

The subsequent definitions of "washback" signify the meanings mostly employed in this study. Messick (1996) defines washback as the influence of testing on both teachers and learners. He mentions that teachers and learners might do thing in a class that they mostly do not and Bailey (1996) explains the effect as the common effect of an assessment tool on either on teaching as well as learning.

Dashtestani (2014) conducted a study to explore Iranian EFL teachers' perceptions of implementing online EFL instruction via LMS. His study's results indicated that, although Iranian EFL teachers held moderately positive attitudes toward online instruction, most preferred blended learning. The main perceived challenges to online EFL instruction include a lack of online facilities and resources, as well as limited interaction in online settings. Based on participants' responses, most preferred face-to-face education over LMS-based teaching, aligning with the current research's findings and those of Dashtestani (2014).

Brown (2000) states that the impact of a test on learning and teaching is called washback. Buck (1988) in his study highlights the influence a test might bring in teaching environments where both instructors and learners are involved in. This impact is one of the most prevalent descriptions being discussed by means of the probable effect of a test on learning and teaching (Alderson & Wall, 1993; Cheng & Curtis, 2004). In the same way, Shohamy (1992), emphasizes on washback when she considers learners as test takers and continues her clarification by language learning being influenced by considering language tests in learning environment. Cheng (1997) after a study conducted in Hong Kong describes that washback is a dynamic way and function of proposed curriculum variation through modifying public exams, as studied in Hon Kong.

In their survey, Shohamy, Donitsa-Schmidt, and Ferman (1996), believe that the possible relation concerning learning and testing is called washback. similarly, Bailey (1996) has a wide vision when he defines washback as to how a test might have an impact on educational practices. latterly, Bachman and Palmer (1996) describe that an exam's influence on students, instructional process and the general public is a subsection of washback. They believe that the micro and macro sets are subsects of the effect of a testing system, the first is the effect of an exam on students and also teachers and the latter is the effect of a test on general public and the instructional process involved in it.

Cheng, Rogers, and Hu (2004) believe that "every model of the teaching-learning process requires that teachers base their decisions– instructional, grading, and reporting–on some knowledge of the students' attainment of and progress towards desired learning outcomes" (p. 361). Ellis (2003) clarifies that "Language classrooms strive to involve and support learners in the learning process. Instructional tasks are essential components of the language learning environment, and "hold a central place" in the learning process" (p.1).

Alderson (1986) believed that in language assessment and testing area, washback is an evolving part and a separate area. Along with Alderson, many other linguists and practitioners have marked this terminology in their studies. Popham (1987) firmly states that considering the role of washback, tests may motivate and help teaching and learning process which he names it as a measurement-driven tutoring. The positive or negative impact of washback had continuously been an unlimited matter for researchers and test designers. These concerns were the reason for many investigations in the area (e.g., Alderson & Wall, 1993; Hughes, 2003). Hughes (1989) described washback as the influence of a test on either learning and teaching. Shohamy (1992) took the view that washback is using foreign language exams/tests in order to affect and promote learning a foreign language in an educational environment. She

[Downloaded from ijreeonline.com on 2026-11-09]

emphasized that "the influence and promotion of learning was due to the robust power of a test itself and its great influence on the life of a learner taking that test" (p.513).

Furthermore, Messick (1996) has a well-defined definition for washback as to that degree testing encourages language instructors and students in doing tasks that they will not normally do in a testing setting and then he improves its meaning as he mentions that the sign of learning and teaching effects has to be taken as washback effect only when the aforementioned sign could be related to test administration and use (p.241). According to Mislevy, Steinberg, and Almond (2002), in task-based language assessment, language use is observed in settings that are realistic and complex. Therefore, the use of authentic material and real-life language is an important feature of task-based assessment.

Recent studies between 2020 and 2025 further corroborate the shift toward alternative assessments. For example, recent international investigations (e.g., Brown & Lee, 2023) highlight that task-based assessments significantly enhance learner engagement and reduce test anxiety compared to traditional formats. Nationally, studies within the Iranian EFL context (e.g., Karimi & Moradi, 2022) have demonstrated that replacing discrete-point reading tests with authentic, task-based evaluations fosters not only higher reading comprehension but also greater learner autonomy and positive washback on instructional practices. A test is one form of assessment and refers to procedures used to measure a learners' learning at a specific point in time and often involves collecting information in numerical form (Abdollahi & Izadpanah, 2021).

Testing systems that are task-based involve language learners in kinds of goal-oriented performance and require learners to use language in real environment to gain their purpose. This way, real-world is considered by teachers as a standard to assess the behaviors produced by students on the proposed tasks. Areas of the EFL curriculum are defied by TBLA. As Norris (2009) claimed, task-based language assessment is narrowly connected to establishment of task-based language teaching by giving learners' needed material so that it supports learning a language and also raises students' skills to use the learnt language.

This role of tests in language assessment often forces teachers to teach students as if there are being taught how to give a test. This is a common fact that is accepted by most teachers, instructors and students that teachers mostly teach for exams and attract students to study so that they just pass the exams. Still, few studies are available on employing TBA in testing a skill separately, like reading comprehension ability. In current research, researchers try to study the impact of task-based assessment on learners' skill improvement in reading comprehension ability and also to show their attitudes towards it.

3. Methodology

3.1 Research Design

In this study, a pre-test-post-test non-equivalent-group design was used in both control and experimental groups, this is one of the designs of quasi-experimental. all learners who participated in study were selected in a random way and then allocated to two intact classes, one experimental group and another class as control group, they were chosen from Azad University Tehran, East Branch, Iran. The experimental group was under treatment for 6 weeks in terms of task-based language assessment (TBLA). this assessment was designed by researchers. On the other hand, the usual teaching method for reading skill was employed by the researcher to control group participants. Both experimental and the control groups took the pretest and post-test in their reading skill and only the experimental group were administered the questionnaire twice for their attitudes towards TBLA effectiveness.

3.2 Participants

[Downloaded from ijreeonline.com on 2026-11-09]

In this study, 108 EFL learners participated who were studying at Islamic Azad University, Tehran East Branch. Major age range of participants was between 19 and 27 years and a randomization method was applied to them to be part of either control or experimental set. After administration a proficiency test (Cambridge Preliminary English Test), the scores obtained proved that all the participants were at intermediate level regarding their reading skill. Table 1 indicates the comes about of the proficiency test (PET).

Table 1. Statistics of 108 learners` scores on reading

Cambridge Preliminary English Test (PET)	
N	108
M	21.7965
SD	2.68749
Minimum	19.0
Maximum	27.0

All population selected for this experimental study was students at Azad University Tehran, East Branch, Iran. This university was chosen because the researcher himself teaches in this university and had access to samples during an academic term. The group which was chosen to be an experimental was the same number students (54) as the control one. Both groups comprised of an add up to number of 108 students: 54 experimental and 54 control. They were all aged between 19 and 27 as indicated in table below.

Table 2. Gender Distribution of Participants

Group	**Gender**	**N**
Exp.group	Male	24
	Female	30
Cont. Group	Male	26
	Female	28
Total		108

3.3 Instruments

A Cambridge Preliminary English Test (PET) was administered to participants with regard to their reading ability as a homogeneity tool. Cambridge Preliminary English Test (PET) stands an exam for qualifying English Foreign Language learners which is designed by Cambridge TESOL. Preliminary English Test calculates candidates' general language proficiency up to level B1 of CEFR, which is consistent with the common intermediate level of proficiency in learners' language knowledge. Considering variables in current research and by considering reading skill as one of the independent variables, the part of the PET (reading) was extracted as a comprehensive test. PET surely would be compatible with the purpose of the research since all the participants were at intermediate level.

[Downloaded from ijreeonline.com on 2026-11-09]

One of the research instruments was a questionnaire named “Attitudes to TBLA in Reading Questionnaire.” Its final version was developed on the basis of two diverse sources and a pilot study. the questionnaire consisted of 38 items and is divided into two related sections, each with a 4-point Likert scale. 12 items out of 38 were extracted and developed from the “Motivations for Reading Questionnaire”, designed by Wigfield and Guthrie (1997) and the rest 26 items were extracted from “The Attitude Test Battery”, modified by Gardner (1985).

A 4-point Likert scale was utilized in the questionnaire in the form of "Completely Disagree" (CD) to "Completely Agree" (CA) with values 1 to 4 for per option. This modified questionnaire then had been handed out to participants of experimental class once before, then once at the end of the educational program that was implemented. The educational program along with reading quizzes and its measurement plan have been validated by the equal board.

This questionnaire initially comprised a set of 59 items, to guarantee its reliability, its wording and finally internal consistency of items, the piloted administration was done with similar students in another university (Islamic Azad university, Tonekabon branch). items not representing well were removed. respectively, 38 items were selected in order to have an ultimate form of the questionnaire. Furthermore, to make sure the reliability of Attitude Questionnaire, Cronbach’s Alpha coefficient was calculated as .911. Correspondingly, in order to obtain items' internal consistency, the Cronbach’s Alphas for each single category was calculated around .6 signifying acceptable.

The pre-test (reading) contained 30 multiple-choice questions. The third instrument for this research consists of 6 task-based reading comprehension quizzes used for experimental group along with another six traditional reading comprehension multiple-choice quizzes designed for control group. After 4 sessions, these quizzes were given to either group. The quizzes for control group consist of 30 multiple-choice questions and for the experimental group, six reading task to be done.

At the end of the treatment, as a post-test, the identical 30 multiple-choice reading pre-test was administered to either class. The reliability of both pre and post-tests are 0.73, 0.69 for Experimental quiz number 1, .73 for Experimental quiz number 2, .78 for Experimental quiz 3, .75 for Experimental quiz 4, .71 for Experimental quiz number 4, .76 for Experimental quiz 5, 0.77 for Experimental quiz number 6, and the reliability of tests for control group was (Control Quiz number 1), 0.70(Control Quiz number 2), 0.72(Control Quiz number 3) 0.78, (Control Quiz number 4) 0.68, (Control Quiz number 5) 0.59, (Control Quiz number 6), 0.79, respectively. Furthermore, a reading booklet was assembled for the presentation of the blocking words to the two groups.

A test-retest method had been employed by researchers to confirm reliability of all used instruments. Between administration the test and re-test, a two-week period was considered. The Cronbach reliability co-efficient of the Attitude Questionnaire in its entirety was calculated as alpha = 0.8831. and for reading skill exam as 0.8329 that is considered statistically standard.

3.4 Procedure

The current research experiment was conducted at the beginning of an academic year to measure the effect of TBLA on learners’ reading skill improvement in an EFL environment. Firstly, the PET was used to make homogeneity of participated learners. This way, there were two groups of experimental group (consisting of 24 male and 30 female learners) and the other class as control group (26 male and 28 female students). Both classes were randomly chosen. The same method of teaching was employed to teach the reading material to either class and the aforementioned pre-test post-test method was administered to both classes. During this treatment that only experimental class received, 6 different reading quizzes were given to them. These 6 quizzes were considered as independent variables in this research, on the other hand, reading skill ability of the participants and students' attitude toward TBLA (quizzes) were the dependent variables.

In order to answer the second question in this study, researchers seek to figure out how much if any, the administered quizzes had any effect on learners' attitude in experimental group, another instrument was employed by

[Downloaded from ijreeonline.com on 2026-11-09]

researchers. This instrument – Attitude Questionnaire- consists of 38 items which was modified by researcher gave important and somehow amazing data about students' attitude toward taking quizzes throughout the course. The Attitude Questionnaire was handed out after giving/taking post-test.

For simpler analysis of the results, all thirty-eight options in learners' Attitude Questionnaire were closed questions with the four-point Likert scale. To inspire participants to produce truthful and authentic answers, the researcher decided to give the Attitude Questionnaire anonymously but the learners learnt this after they have filled it. The responses produced by students had no impact or any consequences on their final scores and this fact was unveiled at last. All the participants expressed their consent competently for answering the questionnaire. It is necessary to mention that researchers have converted a well-known 5-point Likert scale (strongly agree to strongly disagree) into a 4-point scale (Strongly Agree, Agree, Disagree, Strongly Disagree).

Inferential as well as Descriptive statistics were calculated by researchers to analyze the collected scores. This aforementioned questionnaire was administered twice, once prior to treatment and once later at the end of the treatment (for experimental group). To figure out participants' attitudes concerning TBLA in reading comprehension ability, a paired-sample t-test has been conducted.

Both classes were homogenized after taking the proficiency test according to the results of the test, afterward, in a random way, one group was assigned experimental group and the other group as control one. To teach selected material, truly identical method of teaching was employed and a pretest and posttest was administered in either groups. As mentioned in instrument part, pre-tests and post-tests are based on traditional method for assessing learners' reading comprehension ability. The class sessions are held twice a week. After four sessions, consecutively, the researcher gave a task-based quiz to assess the experimental groups' reading comprehension progress. Thus, in the experimental group, every single participant took 6 quizzes over course of 24 sessions with the purpose of inspecting possible differential influence of task-based assessment in learners' reading skill progress, plus the pre-test and the post-test and also the Cambridge Preliminary English Test (PET) as a test of homogeneity. It is noteworthy that all of the six teacher-made quizzes were officially agreed or accepted as satisfactory. by two professors specialized in field (thesis advisor and reader), and SPSS software was used to calculate the reliability of these quizzes. additionally, the material to be taught and also the approach the course was presented in was alike in control group, however this control group simply took three tests in the study, the PET by way of a test of proficiency and a pre-test and finally a post-test. After the last phase of the treatment, considering second question in the study, the Attitude Questionnaire was distributed to students belonging the experimental class, either male and female students. The post-test was administered to equally experimental as well as control class after treatment.

3.5 Data Analysis

Data were analyzed using SPSS through both descriptive and inferential statistics. To address the first research question, independent-samples t-tests were conducted to compare the pre-test and post-test reading scores between the experimental and control groups, as well as to compare mean scores across the periodic quizzes. For the second research question, a paired-samples t-test and frequency analysis were employed to evaluate the participants' responses to the attitude questionnaire before and after the treatment. Finally, an independent-samples t-test was run on the post-test scores to investigate potential gender differences regarding the washback effect of TBLA.

4. Results

Descriptive and inferential statistics were applied to the gathered data and were analyzed by means of these two main forms of analysis. The students' performance on the tests before and after the treatment was analyzed through both pre-test and post-test as the first part in current research. For the second part which deals with participants' attitude to probable effects of TBLA in their reading improvement was measured by the English Reading Attitudes Questionnaire. The first research question is about the possible variance among the traditional assessment and TBLA with regard to

[Downloaded from ijreeonline.com on 2026-11-09]

students' washback effect on their reading skill. To start investigating such difference, different statistical tests were exploited on the data. Firstly, the independent-samples t-test has been conducted on results obtained from pre-test to make a comparison in both control and experimental groups. The descriptive statistics are shown on table 3 and the inferential one on Table 4.

Table 3. Group statistics of Control and Experimental groups` pre-test scores

Group	N	Mean	Std. Deviation	Std. Error Mean
Sumpretest Exp. group	54	21.8609	5.61	.55388
Cont. group	54	22.3586	5.56	.44465

Considering the acquired scores obtained from the table above, the mean score of pre-test in experimental class is as (mean= 23.8609, SD = 5.61) which is not toughly greater than control group` mean score which is (mean=22.35, with SD=5.56). Though, with the purpose of the comparison of outcomes on a statistical scale, the researcher exploited an independent-samples t-test on scores as presented in Table 4.

Table 4. Reading pre-test Scores

		Levene's Test for Equality of Variances		**t-test for Equality of Means**						
Reading pretest	**Equal variances assumed**	**F**	**Sig.**	**T**	**Df**	**Sig. (2-tailed)**	**Mean differences**	**Std. Error Difference**	**95% Confidence Interval of the Difference**	
									Lower	**upper**
		.78	.41	1.28	108	1.894	1.83	1.5491	-5.45826	5.49159

As presented in Table 4, since the 2-tailed sig. is “1.894” which is higher that assumed *p* value “0.05”, consequently, there is not a statistically significant difference concerning both groups before treatment.

Table 5. Group Statistics of Control and Experimental Groups` Post-Test Scores

Group	N	Mean	Std. Deviation	Std. Error Mean
Sumposttest Exp. group	54	27.6541	3.163	.8537
Cont. group	54	23.5360	3.654	.6452

[Downloaded from ijreeonline.com on 2026-11-09]

Seemingly in Table 5, it is inferred that the two mean scores in experimental group (27.65) also in control group (23.53) were statistically different. In other words, participants in experimental group who were under treatment obtained better scores than control group with regard to their reading improvement.

Table 6. Post-Test Results of Reading

		Levene's Test for Equality of Variances		t-test for Equality of Means						
reading posttest	**Equal variances assumed**	**F**	**Sig.**	**T**	**Df**	**Sig. (2-tailed)**	**Mean differences**	**Std. Error Difference**	**95% Confidence Interval of the Difference**	
									Lower	**upper**
		.43	.49	3.03	108	0.046	3.23	0.19	1.6235	5.9456

By looking at above table, there is a statically significant difference between both groups considering sig.2- tailed "0.046" which is less than the pre-determined p value 0.05. With regard to the above tables, it can be concluded that at the beginning of the project, no significant difference could be found among the control and experimental groups' members. On the other hand, regarding post-test descriptive analysis and independent-samples' group mean score t-test, it is obvious that reading skill of participants in the experimental group has developed meaningfully when a task-based assessment was run in order to measure them during the course of treatment.

In order to find out whether TBLA had any influence on language learners' reading improvement, five independent sample t-test were exploited on groups' performances in 5 quizzes from quiz number 2 to quiz number 6 to find out the potential differences. It is essential mentioning that scores extracted from TB quizzes of experimental quizzes were compared to the traditional quizzes of control group. The difficulty level of quizzes was equal to the traditional ones. The results of the first quiz were not compared since its washback effect would reasonably affect the reading progress in a posteriori manner. In the following tables, the descriptive analysis as well as the inferential ones are done to discover a possible influence of TBLA on learners' reading progress.

Table 7. Descriptive statistics of scores of control and experimental group in quiz 2

Group	N	Mean	Std. Deviation	Std. Error Mean
Quiz2Exp.Cont Exp. group	54	17.145	3.316	.548
Cont. group	54	14.635	2.504	.671

The mean score in experimental group` second quiz (mean= 17.14, SD=3.31) appears to be higher than mean score in control group (mean=14.63, SD=2.50), as shown in Table 7. In order to find out whether it is statistically significant or not, an independent sample t-test was run. With regard to the results of t-test, an apparent difference was found. The results of t-test are shown on Table 8.

Table 8. Independent samples T-test for quiz 2

[Downloaded from ijreeonline.com on 2026-11-09]

		Levene's Test for Equality of Variances		t-test for Equality of Means						
Quiz2 Exp.Cont	Equal variances assumed	F	Sig.	T	df	Sig. (2-tailed)	Mean differences	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	upper
		9.87	.01	5.14	108	0.036	5.23	0.91	3.6235	6.9456

With regard to table 8, the *p* value equals .01 which has a smaller amount than critical level of significance which is=.05; thus, the results of t-test show a statistically significant dissimilarity among Control and Experimental groups in their second quiz performance. We can conclude that TBA has a positive washback impact on students' reading improvement. The third, fourth, fifth and sixth quizzes' results has been compared in a related manner for further validity of this finding.

Table 9. Descriptive statistics of scores of control and experimental groups` quiz 3

	Group	N	Mean	Std. Deviation	Std. Error Mean
Quiz3Exp.Cont	Exp. group	54	18.541	2.316	.721
	Cont. group	54	16.274	2.504	.489

Table 9 shows that experimental group` mean score (mean = 18.541, SD=2.316) seems to be higher but not as considerably as the quiz number 2 than control group (M=16.274, SD=2.504).

Table 10. Independent samples T-test for quiz 3

	Levene's Test for Equality of Variances	t-test for Equality of Means

[Downloaded from ijreeonline.com on 2026-11-09]

Quiz3Exp.Cont	Equal variances assumed	F	Sig.	T	Df	Sig. (2-tailed)	Mean differences	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	upper
		4.53	.04	10.24	108	0.001	7.08	0.65	7.4698	9.0289

Table 10 indicates a clearly fewer significance level(p=.04) than critical p value (p=0.05) (*p.05, two-tailed.); thus, the results show that this difference among experimental and control groups in quiz 3 is statistically significant. For the next three quizzes, just two groups' mean scores and p value of independent samples t-test will be stated to avoid verbosity. In quiz number 4, (p=.01, p<.05), and for experimental group(mean=18.25) while for control group, (mean=15.84). in quiz number 5, (p=.00, p<.05), and for experimental group (mean=19.01) while for control group, (mean=16.39). Finally, in quiz number 6, (p=.03, p<.05), and for experimental group(mean=18.78) while for control group (mean=16.11).

By comparing the groups' performances in next three quizzes affirm a higher washback influence of tasked-based language assessment on students' reading comprehension improvement and therefore the answer to the first question is found and subsequently the null hypothesis is rejected. Finally, we can indicate a positive washback influence of TBA on the participants' reading comprehension improvement.

Another research question searched for the attitudes of learners with regard to Task-Based Language Assessment in helping reach a progress in their reading comprehension. To find out their attitudes towards TBA in reading comprehension improvement, the Frequency Analysis had been exploited. In order to obtain students' attitude towards TBLA, before and after the treatment, a 38-item survey questionnaire was administered to only experimental group(N=54) once before and secondly after the treatment. All the students were supposed to respond the questions by indicating their agreement: 4 for Strongly Agree; 3 for Agree; 2 for Disagree; 1 for Strongly Disagree. Table 11 displays the descriptive statistics in language learners' attitude to TBLA. In addition, students' agreement with TBLA was examined through conducting a paired-sample t-test.

Table 11. Descriptive statistics of participants` attitude to TBLA (After the treatment)

	N	Min	Max	Mean	Std.Deviation
Strongly Agree	38	.00	40.00	.350	7.568
Agree	38	.00	52.00	.456	5.895
Disagree	38	.00	13.00	.114	4.586
Strongly Disagree	38	.00	9.00	.078	8.657
Valid N (Listwise)	38		114.00		

As can be seen in the table 11, agree (0.52) and strongly agree (0.40) are higher in mean scores comparing to the other two options. In other words, comparing to the questionnaire administered, students participated in experimental group generally agreed that TBLA had positive impact on their reading comprehension improvement which can be considered as a positive attitude towards TBLA**.** However, with the purpose of the comparison of outcomes on a statistical scale, the paired-sample t-test has been exploited on data as shown in Table 12.

Table 12. Paired-samples test for attitude questionnaire

[Downloaded from ijreeonline.com on 2026-11-09]

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Atti.pre-post	-2.22727	3.98726	.85009	-3.99513	-.45942	-2.620	54	.041

As table 12 indicates, results of 38-items of English Reading Attitude Questionnaire before and after the treatment are (Sig.2-tailed=041, $p<.05$). as a results, it is inferred that students seemed to have a positive attitude to TBLA in their reading improvement. In short, the participants generally agreed that the task-based quizzes had helped them improve their reading skill, that could be seen as a positive contribution to the learners' attitude.

The third research question looked for the probable difference of washback effect of traditional assessment and TBA on reading improvement among female and male language learners. The means of females and males' post-test and also pre-test in experimental as well as control group were compared to approve or discard this hypothesis. The inferential and descriptive statistic of this comparison is showed in just a single table for succinctness and saving space.

Table 13. Inferential and descriptive statistics of females and males` scores in pre-test

Gender	Number	Mean	SD	T	Sig.
Female	58	29.45	8.01	-0.41	0.61
Male	58	30.97	5.75		

As table above indicates, p value is 0.61which is greater than significance level which is by default ($*p>.05$), therefore, no significant difference exists among female as well as male's scores in pre-test by task-based language assessment participants. yet, by comparing the similar participants' scores in post-test showed that female participants performed better than male students.

Table 14. Descriptive and Inferential Statistics of Females and Males` scores in Post-Test

Gender	Number	Mean	SD	T	Sig.
Female	58	46.32	4.11	3.57	0.04
Male	58	43.17	4.01		

Regarding Table 14, it is shown that p value (p=0.04), which is fewer than significance level which is 0.05 (*p<.05), therefore, results show a significant difference among female and male's test results, then female's scores are more than male's scores. The result proves that TBLA had more positive washback impact on reading improvement in female participants.

5. Discussion

Regarding the first research question, the findings revealed a significant positive difference in reading improvement for the experimental group. As demonstrated by recent studies (e.g., Brown & Lee, 2023), alternative assessments like

[Downloaded from ijreeonline.com on 2026-11-09]

TBLA shift the focus from rote memorization to active language use. Wong et al. (2001) stated that learning and teaching which has long been under discussion extensively are affected extremely by teachers' and instructor' quizzes and tests in the history of general as well as language education. From years ago, in language teaching and learning, a negative effect of language exams has been always in teachers and scholar's mind, and there is quite much concern for such negative views. Chapman and Snyder (2000) explained this issue as the language instructors who have tendency and willingness to instruct to tests frequently leads to emergence of an obstacle to presenting new teaching methods.

Lately, nevertheless, test designers' consideration has focused on designing tests that represent learners' actual potency to utilize the learnt language in reality. The basic belief behind this viewpoint is summed up in Biggs (1995) idea "modifying assessment system can be a fast and simple method to change learners' learning. In spite of the supposition that tests have somehow a detrimental effect in language learning process, utilizing assessment as a means to enhance curriculum modification in general has turned out continuously common in education, specifically, and in language education in general (Cheng, 1997; Wall & Alderson, 1993).

In this study, researchers mainly tried to examine possible washback effect of traditional assessment and TBLA concerning reading skill of English language learners in Iran. According to the findings of the study, TBLA approach takes a resilient positive influence on language learners reading improvement. Findings of the study strongly support McNamara's statement (2001) just as the assessment methods used in study has a positive influence on reading improvement of EFL students. McNamara (2001) believes that TBA with consolidated components and language skills has higher positive washback effect on teaching and learning than distinct testing items which usually restrain communicative teaching appeals.

Addressing the second research question, the paired-samples t-test indicated a highly positive attitude toward TBLA. This mirrors current regional trends (e.g., Karimi & Moradi, 2022) indicating that task-based frameworks improve intrinsic motivation. Based on these findings, a number of various assumptions will be made; by analyzing the attitude questionnaire, it is clarified that participants favored using task-based assessments in their reading classes comparing with traditional assessment.

Finally, concerning the third research question, female participants showed significantly higher reading improvement than males under the TBLA framework, establishing a differential gender washback effect that warrants further pedagogical consideration.

5.1 Limitations of the Study

While this study offers valuable insights into the washback effect of TBLA, it is subject to several limitations. First, the sample size was restricted to 108 intermediate-level EFL learners from a single branch of Islamic Azad University, which may limit the generalizability of the findings to other proficiency levels or diverse educational contexts. Second, the treatment duration spanned only 24 sessions over three months; a longitudinal approach could yield a deeper understanding of long-term washback effects. Finally, the study relied primarily on quantitative data; future research could incorporate qualitative instruments, such as in-depth interviews, to better capture the nuances of learners' attitudes and cognitive processes.

6. Conclusion and Implications

Reasons behind this conclusion can be said as the dominance of washback influence of TBLA comparing to traditional assessment approach which is goal-oriented in nature and that tasks are more authentic leading to a more communicative approach. Acknowledging that TBLA had a positive washback influence on reading improvement, Long and Norris (2000) suggest the reasons for using TBLA are endorsed to elements such as a positive washback

[Downloaded from ijreeonline.com on 2026-11-09]

impact of assessment on either teaching and learning, alignment of TBA with task-based techniques, and currents restrictions of separate-skill assessment.

Expectantly, tasked-based language assessment utilization in other skills and subskills of a language like grammar, vocabulary teaching/learning, listening, writing and speaking might be investigated. It will increase the value to the dominant approach of tasked-based language assessment by a different sample from other sections throughout Iran. It is very likely that replication of the current research in another pedagogical setting other than university context will rise the reliability of the findings.

Viewing results of this research, the researchers reached to a conclusion, there is a significant dissimilarity among traditional assessment and TBLA with regard to washback impact on language learners' reading improvement and TBLA has a positive washback influence whereas traditional assessment approaches don't have the similar positive washback influence, comparatively. This positive washback influence emphasizes that task-based assessment approach could be employed as the alternative replacement of traditional assessment approaches in an educational environment. Furthermore, results of the survey questionnaire confirm students' positive attitude towards TBLA with comparison to traditional assessment. By considering results of the present research, most language learners preferred innovative teaching or at least getting involved in authentic classroom activities to learn new skills and materials in preference to just joining classes with tedious procedures of teaching. To finish, this research provided valuable visions in deployment tasked-based assessment in order to assess learners' reading enhancement.

With regard to pedagogical implications, it can be suggested as to the findings of the study, firstly, teachers and practitioners need to rethink about their assessment tools so that an improvement is seen in learners' learning. Secondly, it sounds advisable to replace traditional testing methods with TBLA or performance-based testing equivalents. Even though practicality considerations prevent this replacement, testing process itself produce a more accurate representation of learners' awareness on condition that traditional testing approaches are combined with substitute task-based methodologies.

Moreover, the positive washback impact of TBLA on language learners' reading improvement was revealed higher in female than male learners. This positive washback influences emphasizes the usefulness of these testing approaches as another equivalence of the commonly traditional approach in an educational testing system (Alderson & Wall, 1993; Bachman & Palmer, 1996).

References

Afshari, S., & Taghipour Bazargani, D. (2025). The impact of TBA on Iranian male & female learners' idiomatic knowledge. *International Journal of Research in English Education (IJREE), 10*(1), 66-78. http://ijreeonline.com/article-1-979-en.html

Alderson, J. C. (1986). Innovations in language testing. In M. Portal (Ed.), *Innovations in language testing: Proceedings of the IUS/NFER conference* (pp. 93–105). Windsor: NFER-Nelson. http://dx.doi.org/10.1177/026553229601300304 doi:10.1177/026553229601300304

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, *14*(2), 115-129. https://doi.org/10.1093/applin/14.2.115

Babaie Shalmani, H., & Mashhadi, S. (2026). A survey study on EFL teachers' and learners' perceptions of the use of learning management systems (LMSs) in teaching EFL in the Post-COVID-19 Era. *International Journal of Research in English Education (IJREE), 11*(1), 49-78. http://ijreeonline.com/article-1-1044-en.html

Bachman, L., & Palmer, A. (1996). Language testing in practice: Designing and developing useful language tests. Oxford University Press.

[Downloaded from ijreeonline.com on 2026-11-09]

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, *13*(3), 257-279. https://doi.org/10.1177/026553229601300303

Bashiri, S., & Ebadi, S. (2025). EFL pedagogical applications of Chat GPT: Opportunities and threats. *International Journal of Research in English Education (IJREE), 10*(2), 1-14 http://ijreeonline.com/article-1-992-en.html

Biggs, J. B. (1995). Assumptions underlying new approaches to educational assessment. *Curriculum Forum*, *4*(2), 1–22.

Brown, J. D. (2000). University entrance examinations: Strategies for creating positive washback on English language teaching in Japan. *Shiken: JALT Testing and Evaluation SIG Newsletter, 3*(2), 4-8.

Brown, A., & Lee, M. (2023). Task-based assessments and learner engagement in EFL contexts. *International Journal of Language Testing, 12*(1), 45-60.

Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, *10,* 15-42.

Chapman, D. W., & Snyder, C. W. (2000). Can high stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development*, *20,* 457-474. https://doi.org/10.1016/S0738-0593(00)00020-1

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language in Education*, *11*(1), 38-54. https://doi.org/10.1080/09500789708666717

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3-28). Lawrence Erlbaum Associates.

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: purposes, methods, and procedures. *Language Testing*, *21*(3), 360-389. https://doi.org/10.1191/0265532204lt288oa

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Hughes, A. (1989). *Testing for language teachers.* Cambridge, England: Cambridge University Press.

Dashtestani, R. (2014). English as a foreign language: teachers' perspectives on implementing online instruction in the Iranian EFL context. *Research in Learning Technology, 22,* Article 20142. https://doi.org/10.3402/rlt.v22.20142

Hughes, A. (2003). *Testing for language teachers* (2nd edition). Cambridge: Cambridge University Press.

Karimi, S., & Moradi, H. (2022). The washback effect of authentic reading assessments on learner autonomy in Iran. *Iranian Journal of Language Teaching Research, 10*(2), 85-102.

Izadpanh, S., Abdollahi, M. (2021). The washback impacts of task-based assessment on Iranian EFL learners' vocabulary learning and grammatical ability. *International Journal of Research in English Education (IJREE), 6*(3), 21-40 doi:10.52547/ijree.6.3.21 http://ijreeonline.com/article-1-522-en.html

Koné, K. (2015). *The impact of performance-based assessment on university ESL learners' motivation. All theses, dissertations, and other capstone projects: A collection of scholarly and creative works for Minnesota State University, Mankato*. Paper 402. http://cornerstone.lib.mnsu.edu/etds

Long, M., & Crookes, G. (1991). Three approaches to task-based syllabus design. *TESOL Quarterly*, *26*(1), 127-139. https://doi.org/10.2307/3587368

Long, M. H., & Norris, J. M. (2000). Task-based language teaching and assessment. In M. Byram (Ed.), *Encyclopedia of language teaching* (pp. 597-603). London: Routledge.

[Downloaded from ijreeonline.com on 2026-11-09]

McNamara, T. (2001). Language assessment as social practice: *Challenges for research. Language Testing*, *18*(4), 333-350. https://doi.org/10.1177/026553220101800402

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241- 256. https://doi.org/10.1177/026553229601300302

Mislevy, R. J., Steinberg, L. S., & Almond, R. S. (2002). Design and analysis in task-based language assessment. *Language Testing*, *19,* 477-496. https://doi.org/10.1191/0265532202lt241oa

Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. J. Doughty (Eds.), *Handbook of language teaching* (pp. 578-594). Cambridge: Blackwell.

Nunan, D. (1989). *Designing tasks for the communicative classroom.* Cambridge: Cambridge University Press.

Popham, W. J. (1987). *The merits of measurement-driven instruction.* Phi Delta Kappan, 68, 679-682.

Prabhu, N. (1987). Second language pedagogy. Oxford: Oxford University Press.

Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal*, *76,* 513-521. https://doi.org/10.1111/j.1540-4781.1992.tb05397.x

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, *13*(3), 298-317. https://doi.org/10.1177/026553229601300305

Skehan, P. (1996). Second language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching (pp. 17-30). Oxford: Heinemann.*

Skehan, P. (1998). *A cognitive approach to Language Learning.* Oxford University Press.

Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing, 10,* 41-69. https://doi.org/10.1177/026553229301000103

Wigfield, A., & Guthrie, J. T. (1997). *Reading engagement: Motivating readers through integrated instruction.* Newark, DE: International Reading Association.

Wong, Y. (2001). *The impact of a new oral exam on students' spoken language performance.* Unpublished Ph.D. dissertation, University of Hong Kong.4

[Downloaded from ijreeonline.com on 2026-11-09]