

Washback Effect of Task-Based Assessment on Iranian Intermediate EFL Learners' Writing Performance: A Case of Accuracy, Fluency and Complexity

Maryam Farahi*¹

Dr. Hamidreza Kargozari¹

¹Tabaran Institute of Higher Education, Mashhad, Iran

Abstract: The washback effect is a factor that describes how the use of a test can influence the teaching and learning process. However, task-based assessment specifically focuses on evaluating how effectively learners can apply their learning in real life. In addition, the writing performance of EFL learners' is a construct that involves various linguistic, and cognitive factors that affect the quality of the learners' output. Consequently, the present study investigated the washback effect of task-based assessment on the writing performance of Iranian EFL intermediate learners in terms of accuracy, fluency and complexity. The data was collected from a sample of 50 intermediate Iranian EFL learners through the pre-and post-writing tests. For investigating the homogeneity of the participants an OPT was taken before the treatment. Two groups of 50 participants, each control and experimental, took the same writing task as pre-test and post-test. The participants of the experimental group had some task-based writing assessments during the course. The courses for both groups were online. The results indicated that task-based assessment had no washback effect on students' writing performance in terms of, accuracy, fluency and complexity. The study's findings can guide teacher training programs by emphasizing the importance of developing pedagogical skills related to task-based assessment. Educators should be equipped with the knowledge and strategies to design, implement, and evaluate writing tasks that foster complexity, fluency, and accuracy in EFL learners' writing.

Keywords: Accuracy; Complexity; Fluency; Task-based Assessment; Washback.

Copyright © 2023 The Author(s): This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 (CC BY-NC 4.0) International License.

INTRODUCTION

TBLA (Task-Based Language Assessment)

Task-Based Language Assessment (TBLA) came into view as a reaction against the contemporary traditional classroom evaluation. TBLA is considered as a new trend in assessment which is commonly applied as a stimulator to reinforce learner participation in their evaluation, their interplay with other learners, instructors, parents, and their real community (Aschbacher, 1994). In other words, the authentic tasks are utilized to assess what a learner can do with language in a real or semi-real world, on top of that these tasks are designed to evaluate the behaviour of students in a natural context (Ellis, 2017).

TBLA is concerned with the language learners' ability to apply their language knowledge to accomplish and achieve the goal and purpose of the task, similar to how they use their first language to get things done in daily life. The tasks can range from simple daily things, i.e., ordering pizza in a restaurant and responding to emails, to more professional activities, such as listening to an academic lecture in class and taking notes. Thus, TBLA was developed to fill the gap of an appropriate assessment that measures what language learners can do in the target language. In other words, what distinguishes TBLA from other modes of assessment is that performance is in the construct of assessment in TBLA (Long & Norris, 2000).

Four main features characterize TBLA. First, it is a formative assessment; that is, it is an assessment undertaken as part of an instructional program to improve learning and teaching.

Second, it is a performance-referenced assessment; that is, it is an assessment that seeks to provide information about learners' abilities to use the

language in specific contexts, that is directed at assessing a particular performance of learners, and that seeks to ascertain whether learners can use the L2 to accomplish real target tasks.

Third, it is a direct assessment; that is, it is an assessment that involves a measurement of language abilities that involves tasks where the measure of the testee's performance is incorporated into the task itself, like information-transfer test tasks such as information-gap, opinion-gap, and reasoning-gap tasks. It must be noted, however, that direct assessment still involves some level of inferring because it is necessary to observe performance and then infer ability from that performance. Put differently, you can measure outcomes, but you are still left with inferring the ability that produced the outcome.

Fourth, it is an authentic assessment; that is, it is an assessment that involves either real-world language use (or as close as possible to this), or the kinds of language processing found in real-world language use, that is, the test task's characteristics must match those of the target-language task (Ellis, 2005).

Washback Effects

Washback effects (Alderson & Wall, 1993), the notion that testing and its structure influence teaching and learning, are commonplace in the education literature (Biggs, 1995; & Cheng, 1999). Washback can be perceived as either positive or negative, depending on whether it results in "good" teaching practices or undesirable teaching and learning outcomes (Holm & Kousholt, 2019). However, washback was a complex phenomenon, and it has been debated whether it was determined by the test itself or factors beyond it. In this study, the washback effect was measured by conducting three task-based assessment and their effects on the participants' writing performance in terms of accuracy, complexity and

fluency.

CAF of Writing Performance

The dependent variables in this study are complexity, accuracy, and fluency of written language. Mainly these three constructs are utilized as the dependent variables to measure differences concerning independent variables (Pallotti, 2014). The components above mentioned are so complex and multidimensional that there is not any general agreement among scholars to specify and operationalize them. In addition, Ellis (2005) proposes two approaches for evaluating accuracy, complexity, and fluency. Firstly, it is acquiring ratings. Secondly, it is calculating different discourse-based measures. They assert as if the first approach is preferable by language testers. This is the reason that ratings have a high face and assess student language pragmatically. Although, the second approach is mainly utilized by SLA scholars to measure various perspectives of language implementation precisely. Also, they believe that the particular scales of accuracy and complexity can be feasible for both spoken and written language however, fluency scales for written and spoken language contrast with each other.

Purpose

Many studies have investigated (CAF) Complexity, Accuracy and Fluency in the writing area, however, none so far have examined the three CAF measures jointly as an index of proficiency. This gap has led to our limited understanding of proficiency-related differences in the CAF in L2 writing. As Polio (2017) recommends, more research should be conducted to measure writing accuracy and fluency.

However, they must be examined and interpreted in accompany of complexity measures to obtain firmer conclusions (Foster & Wigglesworth, 2016; & Polio, 2017).

The gap in the literature and understanding the impact of task-based assessment on learners' writing performance is essential for developing effective language teaching methods and materials. If task-based assessment can enhance accuracy, fluency, and complexity in writing, this approach can be incorporated into language curricula and instructional practices to promote better writing skills among intermediate EFL learners in Iran.

Additionally, investigating the washback effect of task-based assessment on writing performance can provide valuable insights into the effectiveness of different assessment types and their impact on students' learning outcomes. This can lead to better-informed decisions about the design and implementation of language assessments that are both reliable and valid. Moreover, investigating the washback effect of assessment on writing performance can help identify areas of strength and weakness in learners' writing abilities.

This research aimed to investigate the impact of this assessment on the student's overall writing abilities. By examining the effect of task-based assessment on writing performance, researchers can better understand how this type of assessment affects student performance and can provide insights into how to improve EFL writing instruction in Iran. Additionally, the results of this research may be used to inform curriculum performance and assessment practices in other countries with similar EFL language education contexts. Ultimately, this research aimed to improve the quality of EFL education for Iranian intermediate students and potentially contribute to advancements in EFL education worldwide.

Research Questions

Based on the purposes of the study the following questions are set to be answered:

- **RQ-1:** Does task-based assessment have any washback effect on Iranian intermediate EFL learners' writing performance in terms of complexity?
- **RQ-2:** Does task-based assessment have any washback effect on Iranian intermediate EFL learners' writing performance in terms of accuracy?
- **RQ-3:** Does task-based assessment have any washback effect on Iranian intermediate EFL learners' writing performance in terms of fluency?

Research Hypotheses

The research Hypotheses were set:

H₀₁: Task-based assessment does not have any washback effect on Iranian intermediate EFL learners' writing performance in terms of complexity.

H₀₂: Task-based assessment does not have any washback effect on Iranian intermediate EFL learners' writing performance in terms of accuracy.

H₀₃: Task-based assessment does not have any washback effect on Iranian intermediate EFL learners' writing performance in terms of fluency.

REVIEW OF LITERATURE

Task-based Assessment

Ellis (2017) studied the task-based teaching methods, and the results showed the efficiency of task-based teaching methods on participants of the experimental group in contrast to conventional strategies in the control groups. Despite showing the good effect of TBLT along with TBLA in this study, the number of students was not enough to generalize the idea. In addition, there is no evidence to what extent this method helped the learners to learn new vocabulary. Moreover, there is a need to explore teachers' attitudes to have more accurate results.

Norris (2018) stated that over the past several decades, communicative performance tasks have come

to play a crucial role in language assessments on a variety of levels, from classroom-based tests to professional certifications, to large-scale language proficiency exams. However, the use of such tasks for assessment purposes remains contentious, and numerous language testing alternatives are available at potentially lower costs and degrees of effort. To facilitate decisions about when and why to adopt task-based designs for language assessment, I first outline the relationship between assessment designs and their intended uses and consequences. Then introduced two high-stakes examples of language assessment circumstances (job certification and admissions testing) that suggest a need for task-based designs, and reviewed the corresponding fit of several assessments currently in use for these purposes. In relation to these purposes, I also suggest some of the positive consequences of task-based designs for language learners, teachers, and society, and I point to the dangers of using assessments that do not incorporate communicative tasks or do so inappropriately. I conclude by highlighting other circumstances that call for task-based designs, and I suggest how advances in technology may help to address associated challenges.

In Hima *et al.* (2019) study, the results of the pre-observation or situation analysis have shown that the school faces a major problem from the aspect of teaching performance of English teachers. To overcome this problem, the project team proposed best practices in teaching English using the Task-Based Language Teaching (TBLT) approach as an innovative teaching approach in the context of English for Specific Purposes (ESP). Furthermore, the results of the preliminary study were used as a reference for the preparation of four learning meetings in each class.

Prepare for the teaching and learning process, starting from designing lesson plans including the ESP materials, media and evaluation instruments, 2. Implement the teaching and learning process by using TBLT, 3. Evaluating students. It is hoped that with the methods proposed in this program, the English teachers from the partner school can improve their teaching skills which then have a positive effect on improving the quality of the school in creating a more effective learning environment for students.

The washback effects of different test formats on the writing performance of students have always been of great importance. However, this area of research has not been fully touched upon by researchers of second language testing. Despite the importance of the issue, there is a dearth of empirical studies to unravel the effects of different types of tests on learning. To shed some light on the current issue, (Shirzadi & Amerian, 2020) intend to look into the washback effects of tests on students who are learning and using some special grammatical points in writing tasks. To fulfil this project, we made a set of questions

in three formats cloze, multiple-choice and metalinguistic on a grammatical form (i.e., present perfect and present perfect continuous) to use after each session of teaching (2 sessions of training) as an activity. The researchers devised and validated three tests on the target form; each test contained 20 questions and was in different formats cloze, multiple-choice or metalinguistic. Two focused writing tasks were implemented at the end of this two-session training. The results indicated that supporting teaching grammatical points with metalinguistic tests yields the highest positive washback on students' writing. Finally, some practical implications were suggested (Shirzadi & Amerian, 2020).

The study of Elahi *et al.* (2021) explored the effects of task-based language learning and integrating blended learning on the reading perception of Iranian EFL students. In their research, they just investigated the effect of Task-Based Instruction (TBI) on the development of grammar proficiency of EFL Learners. The participants in experimental groups were involved in TBI using different tasks, while the control group benefited from the traditional method. The results of the post-test indicated the high efficiency of the task-based teaching method in developing the grammar skills of EFL learners.

Noroozi & Taheri (2022) claimed that task-based language learning has been developed in response to the teacher-dominated, focus-on-forms methods such as Present, Practice, Produce (PPP). The body of the literature is replete with studies examining the learning efficacy of the PPP approach versus TBLT; however, these studies did not use assessment tasks in comparing these two methods. To this end, their study used an Assessment Task, a Grammaticality Judgment Test (GJT), and an Elicited Imitation Test (EIT) to compare the efficacy of PPP versus TBLT. Thirty-four lower-intermediate English language learners in Iran were randomly assigned to TBLT, PPP, and Control groups. Their study results indicated that only the TBLT group made substantial improvements in TBLA in the post-assessment, while the PPP and Control groups' performance did not significantly improve.

CAF of Writing

The application of integrated writing tasks in academic writing assessment is increasing and research on these tasks is growing. However, the use of individual distinct variables in students' performance in source-based writing is under-researched. Thus, Golparvar & Khafi (2021) purport to investigate the predictive contribution of L2 writing self-efficacy to the summary writing strategies used by EFL learners and their performance in a reading-to-write task. The participants of this study were 191 undergraduate university students, who answered an integrated writing task and completed questionnaires measuring L2 writing self-efficacy views and digest writing strategy

use. The outcomes of structural equation modelling (SEM) revealed that the three components of writing self-efficacy, linguistic, self-regulatory, and performance self-efficacy, significantly predicted summary writing performance. It was also found that linguistic self-efficacy predicted discourse synthesis and source use strategies, while self-regulatory and performance efficacy could only predict metacognitive strategies of planning and evaluation. The consequences are discussed and pedagogical conceptions are offered (Golparvar & Khafi, 2021).

Writing assessment literacy (WAL) has received research attention over the past few years. Tayebi *et al.* (2022) aimed at investigating the writing assessment knowledge of Iranian English language teachers along with their conceptions and practices of writing assessment based on (Crusan *et al.*, 2016) study to gain a better understanding of their current situation and to predict and accommodate for future writing assessment requirements. The study further aimed at examining how teachers' knowledge, conception, and practice of writing assessment are influenced by contextual and experiential factors. To accomplish this goal, a test of writing assessment knowledge together with an adapted version of a questionnaire developed by (Crusan *et al.*, 2016) for writing assessment conceptions and practice was distributed among 120 Iranian in-service teachers selected based on convenience sampling. The results of the study have shown inadequate levels of writing assessment knowledge for participating teachers. Concerning the conceptions of writing assessment, the majority of the participants valued innovative assessment methods like portfolio and self/peer assessment methods although, in practice, they rarely used these methods. The study revealed no impact of teaching experience and context on teachers' writing assessment knowledge and practice. The findings of this study contribute to our current understanding of WAL development and provide a more accurate picture of the writing assessment training needs of Iranian teachers and the development of more efficient teacher education courses (Tayebi *et al.*, 2022).

Leyi (2023) claimed that the use of lexical features in academic writing settings has sparked much interest in second language studies. However, prior research has dealt with productive lexicon primarily from complexity-oriented measures exhibited in low-stakes testing. From a CAF-based (complexity, accuracy and fluency) analysis approach, this study is aimed to explore lexical features in high-stakes tests that were extracted from a large corpus containing TOEFL iBT independent writings.

To this end, a pool of 727 and 275 sampled essays for Group Medium and High, respectively, and a total of 16 specific lexical items were selected to measure how learners' writing performance varied across writing proficiency. The results indicated that

though in general, learners showed significant differences between group means, the correlations between the presence of lexical devices and writing quality were found to be negligible in that only two indices (hyponymy, MTLTD) were revealed to correlate significantly with human-rated writing quality. Implications for second-language writing were also discussed in this study (Leyi, 2023).

Wash-Back Effect

According to Hung *et al.* (2019), washback refers to the influence of tests on learning and teaching. Several studies have revealed that tests affect teaching content, course design, and classroom practices. However, in Asian higher education contexts, little research has examined the washback of proficiency tests on English learning in comparison with the efforts on teaching. Thus, their study bridged this research gap by exploring the washback effects of a proficiency test on student learning in a campus-wide English curriculum, uncovering relationships between washback and learner characteristics such as major, gender, and proficiency level. A total of 694 students from engineering-, business-, and foreign language-related disciplines at a national university in Taiwan were surveyed. The results revealed that washback effects on personal image, learning motivation, emotion, and future job opportunities were especially salient. In addition, the relationship between washback and proficiency level was found to be statistically significant. However, male and female students did not differ statistically in washback nor was there a statistically significant difference in washback among different majors.

Lutfiana *et al.* (2020), in their study, investigated the teaching and learning of English through online classes and the washback effect of online testing during the COVID-19 pandemic. The results have indicated the effectiveness of WhatsApp groups, as the learning platform, for online teaching and learning. The positive washback effect motivated students to prepare more effectively for the following test, and their honesty was put to the test. The negative washback effect of the test was related to the time limit of the test, which caused the students to feel anxious. In addition, the students were not as concentrated and focused at home, as they were in a real classroom area. In the end, they could not achieve the desired score. The positive washback effect on the teacher was related to their improved ability to control the media during online instruction and prepare test items.

Besides, Pitoyo *et al.* (2020) explored the washback effect of quizzes on students' learning. To collect the data, the researcher used a questionnaire, observation, and in-depth interviews and analyzed qualitatively. The result of the study shows that students were motivated and wanted to learn more after doing integrated gamified tests with quizzes. As it was stated

by previous writers, the positive washback of a test can bring about benefits and encourage students to change their study methods.

In a study by Jamalifar *et al.* (2021), the authors attempted to explore the washback effect of the English Proficiency Test (EPT) on learners. They used an attitude questionnaire to gather data from 200 participants. The collected data showed the test's positive and negative washback impacts on learning materials. A negative washback effect led to narrowing down the curriculum and measuring Iranian Ph.D. candidates' proficiency based on grammar, vocabulary, and reading comprehension. In terms of the positive washback effect, the result showed the effectiveness of the syllabus but the levels of the textbooks needed to be improved to be aligned with the content.

METHOD

Research Design

The method which was adopted in this study was quantitative research in nature. The independent variable in this study was task-based assessment and the dependent variables were accuracy, fluency and complexity. Consequently, because the study used convenient sampling and it was non-random sampling, the study design is quasi-experimental. This design was utilized in the way that there were experimental and control groups based on non-random sampling. There was also a pretest along with a posttest for both groups and a treatment period. The research design included pretest-treatment-posttest.

Setting and Participants

Dornyei (2007) stated that the number of participants for experimental research and quasi-experimental sampling, sample size can be 50 participants. By the same token, this study which is quasi-experimental research consisted of 60 participants, including 15 males and 35 females at the intermediate level. Approximately, they were aged 10-45. The courses were held online. Consequently, the participants were from different cities in Iran. The students were chosen randomly for two groups, one experimental and one control group. The instructor was the researcher, and the classes were in the same conditions.

Instruments

This study examined the washback effect of task-based assessment (TBA) on the writing performance of Iranian intermediate EFL learners. Specifically, it focuses on the impact of TBA on accuracy, fluency and complexity. The study used the Oxford Placement Test and the Task-Based Assessment to measure students' writing performance. The aim was to provide insights into the effectiveness of TBA in enhancing writing skills for Iranian intermediate EFL learners, informing language educators and curriculum developers.

Oxford Placement Test (OPT)

The first instrument was an Oxford Placement Test from Oxford University Press (OUP), which was for non-native speakers of English, reporting at Pre-A1, A1, A2, B1, B2, C1, and C2 levels of the Common European Framework of Reference (CEFR). This test was divided into two parts: Part One (Questions 1 – 40) and Part Two (Questions 41 – 60).

According to the University of Cambridge Local Examination Syndicate (UCLES) in OPT, 24-47 correct answers illustrated the intermediate level. Consequently, to check the level of general language proficiency of the participants at the beginning of the study and find a homogenous sample, the OPT test was utilized.

Writing Pretest and Posttest

After OPT, one standard writing pretest which was the same for both groups, was given to evaluate students' writing performance in CAF. Then the researcher administered the tasks in the experimental group and traditional writings in the control group. At the end of treatment, the same writing posttest was administered in both groups to find out the effects of treatment on students' writing performance.

Tasks

After the pretest, every three sessions standard tasks, from Wolfe-Quintero *et al.* (1998); & (Polio *et al.*, 1998) model, were given to students as the other required instrument in this study. The aim was to determine whether these tasks had any effect on participants' writing performance. The writing tasks were given to both experimental and control groups. The participants of the experimental group had some task-based writing assessments during the course.

Procedures

There were two groups experimental and control groups and each group included 25 participants. After the OPT was administered, the participants were chosen randomly from two different groups. As already mentioned, according to the University of Cambridge Local Examination Syndicate (UCLES) in OPT, 24-47 correct answers illustrated the intermediate level. Subsequently, the participants were chosen from this average. Before the study, all participants in the experimental and control groups had been assessed for their general English expertise via OPT.

Thereupon, the measure of their proficiency in English language writing was examined by utilizing a standard pre-test managed by the instructor on the learners at the onset of the first session. The goal of the pretest was to certify both groups' homogeneity in their knowledge of writing before establishing the treatment, and later to utilize the pretesting data for later diagnosing the distinction with the posttest. Following the implementation of OPT, half of the students were

specified to the experimental group, and the other half of them were in the control group through their levels they should have equal conditions.

Calculating CAF Terms

The statistic of phrases in each T-unit had been used for complexity, the statistic of error-free T-units in each T-unit had been utilized for accuracy and the statistic of words in each T-unit had been utilized for fluency. These scales were highly reliable and valid and applied in various studies (Ellis, 2005; & Freeman, 2009).

PRELIMINARY ANALYSES

Descriptive Statistics of Oxford Quick Placement Test (OQPT)

Out of 60 participants, 50 of them who had scored between 24 and 47 according to the scoring guidelines of OQPT were chosen for this study. Ten students' scores were not used in data analysis because their scores were either below 24 or above 47. The descriptive statistics of participants' performance on OQPT is summarized in Table (1).

Table 1: Descriptive Statistics of Students' Performance on OQPT

Groups	N	Mean	SD	Minimum	Maximum
Experimental	25	33.92	7.75	24	47
Control	25	34.16	6.46	24	47

Note. OQPT= Oxford Quick Placement Test

RESULTS

Reliability

Cronbach's alpha procedure was performed to measure the reliability of the tasks. This test affirms the internal consistency of a measurement device used to assess various questionnaire characteristics. Cronbach's alpha coefficient is a numerical value between 0 and 1,

representing the correlation coefficient of data collected over time. The closer Cronbach's alpha coefficient is to 1, the higher the reliability of the variables would be. Additionally, a Cronbach's alpha coefficient greater than 0.7 illustrates the highly reliable tasks. Table (2), reports the results of the Cronbach's alpha test.

Table 2: Results of the Cronbach's alpha coefficient

Cronbach's alpha ($\alpha > 0.7$)		
	Control	Experimental
Fluency	0.83	0.81
Complexity	0.75	0.78
Accuracy	0.75	0.73

As shown in Table (2), Cronbach's alpha coefficient for the study's variables is greater than 0.7, signifying that each structure is sufficiently reliable.

At the beginning of the study, a pretest in writing in terms of complexity, accuracy, and fluency was taken from Iranian intermediate EFL learners in both control and experimental groups. The descriptive results of the pretest scores are shown in Table (3).

Analysis of the Results

Table 3: Descriptive Statistics of Control and Experimental Groups on Writing Pretest

Group Statistics					
	Groups	N	Mean	Std. Deviation	Std. Error Mean
Fluency	Experimental	25	12.59	8.60	1.72
	Control	25	12.88	5.36	1.07
Complexity	Experimental	25	1.92	1.16	0.23
	Control	25	2.09	1.34	0.27
Accuracy	Experimental	25	1.13	1.03	0.20
	Control	25	0.99	0.03	0.01

Table (4) shows that the pretest mean scores of the participants in the experimental and control groups for fluency were 12.59 and 12.88, complexity was 1.92 and 2.09, and accuracy was 1.13 and 0.99, respectively. It seems there was not a great difference between the control and experimental groups in their writing scores at the beginning of the study.

the participants in the experimental and control groups for fluency were 12.59 and 12.88, complexity was 1.92 and 2.09, and accuracy was 1.13 and 0.99, respectively. It seems there was not a great difference between the control and experimental groups in their writing scores at the beginning of the study. The results are summarized in Table (4).

Table (4) shows that the pretest mean scores of

Table 4: Results of Independent Samples T-test for Writing Pretest

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Fluency	1.84	0.181	-0.14	48	0.888	-0.28	2.02	-4.36192	3.78
Complexity	0.24	0.625	-0.47	48	0.636	-0.16	0.35	-0.88260	0.54
Accuracy	3.92	0.053	0.66	48	0.507	0.13	0.20	-0.27671	0.55

An independent samples t-test was run to compare the mean scores of the two groups to see if there was any significant difference between the groups regarding their writing pretest scores. Table (4) indicates that for fluency $t(48) = -.141$, $\alpha = .05$, $p = 0.888$, for complexity $t(48) = -.477$, $\alpha = .05$, $p = 0.636$, and accuracy $t(48) = -.668$, $\alpha = .05$, $p = 0.507$. This result suggests that the two groups were homogeneous in terms of their writing and no significant difference was observed

between groups in the pretest.

RQ1

To examine this question, whether the task-based assessment has any positive washback effect on Iranian intermediate EFL learners' writing performance in terms of complexity, a comparison of the means of pre and post-test within control and experimental groups by a paired samples t-test was performed.

Table 5: Independent Samples t-test between the Control and Experimental Groups in terms of complexity on Post-test

t-test for Equality of Means									
Levene's Test for Equality of Variances									
F	Sig.	t	df	Sig	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
							Lower	Upper	
5.00	0.030	0.43	48	0.669	0.06	0.14	-0.23	0.36	

As shown in Table (5), there was not a statistically significant difference between the experimental and control groups ($t(48) = 0.430$, $p > 0.05$) in their performance on the post-test. Although both groups did better in the pre-test, the control group outperformed the experimental group. Therefore, it can be concluded that treatment in the experimental group was not effective and applying the task-based assessment had no significant washback effect on Iranian intermediate EFL learners' writing performance

in terms of complexity.

RQ2

At the beginning of the study, a pretest in writing performance in terms of accuracy was taken from Iranian intermediate EFL learners in both control and experimental groups. This result suggested that the two groups were homogeneous in terms of their accuracy and no significant difference was observed between groups in the pretest.

Table 6: Independent Samples t-test between the Control and Experimental Groups in terms of Accuracy on Post-test

t-test for Equality of Means									
Levene's Test for Equality of Variances									
F	Sig.	t	df	Sig.	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
							Lower	Upper	
4.35	0.042	-1.00	48	0.322	-0.00	0.00	-0.02	0.00	

As shown in Table (6), there was not a statistically significant difference between the experimental and control groups ($t(48) = -1.000$, $p > 0.05$) in their performance on the post-test. Although both groups did better in the pre-test, the experimental group outperformed the control group. Therefore, it can be concluded that treatment in the experimental group was not effective and applying the task-based assessment had no significant effect on Iranian

intermediate EFL learners' writing performance in terms of accuracy.

RQ3

At the beginning of the study, a pretest in fluency in writing performance was taken from Iranian intermediate EFL learners in both control and experimental groups. This result suggested that the two groups were homogeneous regarding their fluency and

no significant difference was observed between groups in the pretest.

Table 7: Independent Samples t-test between the Control and Experimental Groups in terms of Fluency on Post-test

Levene's Test for Equality of Variances					t-test for Equality of Means				
F	Sig.	t	df	Sig.	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
							Lower		Upper
1.13	0.293	-0.08	48	0.930	-0.14	1.65	-3.47		3.17

As shown in Table (7), there was not a statistically significant difference between the experimental and control groups ($t(48) = -0.088$, $p > 0.05$) in their performance on the post-test. Although both groups did better in the pre-test, the control group outperformed the experimental group. Therefore, it can be concluded that treatment in the experimental group was not effective and applying the task-based assessment had no significant effect on Iranian intermediate EFL learners' writing performance in terms of fluency.

DISCUSSION

In Ellis' (2017) study on task-based teaching methods, the results showed the efficiency of task-based teaching methods on participants of the experimental group in contrast to conventional strategies in the control groups. Despite showing the good effect of TBLT along with TBLA in this study, the number of students was not enough to generalize the idea. In our study, the number of participants was enough. However, the results were not satisfying. They did not have any positive or negative effect on the washback task-based assessment.

Bulté & Housen (2012) complain that many L2 studies inadequately define or fail to define this term resulting in mixed and sometimes contradictory results. While freeing complexity from any theoretical assumptions about performance would allow for a more consistent application across studies, it also presents serious problems where frequency measures are concerned. According to this statement, due to the difficulties in measuring lexical frequency and grammatical structure of complexity of academic writing to investigate the better result of complexity, our research illustrates the clausal use which may not count as the complexity measurement.

Based on work by Norris & Ortega (2009), propose that syntactic complexity is manifested as sentence (or sentential) complexity, clausal complexity and phrasal complexity. Consequently, in our study, the clausal complexity did not demonstrate any influence on writing performance through the procedure that the researcher conducted.

Polio & Shea (2014) tested the reliability and

validity of each of these accuracy measures. Their findings suggested that not any one of these measures may be deemed more valid than the others. They cautioned researchers to consider both the strengths and weaknesses of these measures based on the context of their studies. Nonetheless, their findings showed that the ratio of error-free T-units to all T-units (EFT/T) along with the ratio of error-free clauses to all clauses (EFC/C) had high correlations with the different error measures.

Similarly, in the current study, the ratio of error-free T-units to all T-units had been conducted. However, the effect of Error Free T-units on all T-units (EFT/T) had been measured and the results indicated non-satisfying performance in learners' writing performance.

The study by Leyi (2023) focused on the use of lexical features in high-stakes tests, specifically TOEFL iBT independent writing samples. It aimed to explore how learners' writing performance varies across writing proficiency levels and the correlation between the presence of lexical devices and writing quality. The study selected a total of 16 specific lexical items to measure these aspects. However, our study focused on the impact of task-based assessments on writing accuracy, complexity and fluency. The study examined how task-based assessments influence the performance of writing skills among Iranian intermediate EFL learners, with a specific focus on accuracy, complexity and fluency.

Likewise, this study is related to second language writing, however, it has differed in specific focus. The study by Leyi (2023) examined the use of lexical features and its correlation with writing quality in high-stakes tests, while this study focuses on the impact of task-based assessment on writing accuracy among a specific group of learners. Also, in this study, the effects and influences of the two types of assessment were investigated.

The study by Leyi (2023) highlights the importance of considering specific lexical devices in high-stakes tests, while to compare in this study sheds light on the potential impact of task-based assessments on writing accuracy through the trusted models of measuring these features among Iranian intermediate

EFL learners.

Many studies (Sasaki 2000; Storch 2009) utilized two or more measures for evaluating writing fluency clearly which demonstrates the conceptual uncertainty they may have about it. Subsequently, in our study, the word counts in T-units represented the fluency of learners' writing performance.

Generally, fluency measurement is based on writing coherently, rapidly and correctly. As this aspect of fluency is defined it is shown that enumerating words in measuring may not give an exclusive result. However, the method that is used in this study for measuring fluency was the word count in T-units which illustrates no significant effect on Iranian intermediate EFL students' writing performance.

The study conducted by Tayebi *et al.* (2022) primarily focused on investigating the writing assessment knowledge, conceptions, and practices of Iranian English language teachers. The researchers aimed to understand the current situation of writing assessments among teachers and identify the factors influencing their assessment practices. The study employed a questionnaire and a test of writing assessment knowledge to collect data from in-service teachers.

On the other hand, in our study, the washback effect of task-based assessment on Iranian intermediate EFL learners' writing performance in terms of fluency would likely have involved gathering data directly from the learners. This could be done through pre- and post-writing assessments, where students' writing samples were collected before and after the implementation of task-based assessments. The samples were analyzed for fluency, which was measured through factors such as the use of cohesive devices, sentence complexity, and overall coherence.

In terms of the washback effect on learners' writing performance, the researchers aimed to understand how task-based assessment impacts learners' fluency in writing. The study on teacher assessment practices contributes to the broader understanding of how writing assessment practices influence students' writing performance. It highlights the importance of effective assessment practices in promoting fluency in writing, which could be further explored and investigated in the study of the washback effect on learners.

For instance, if the study on teacher assessment practices reveals a lack of utilization of innovative assessment methods like portfolios and self/peer assessment, it may suggest that the washback effect on students' fluency could be limited. On the other hand, if teachers were well-versed in effective assessment practices and incorporated them into their instruction, it could indicate a positive washback effect

on students' writing fluency.

Entirely, there is a valuable insight into the washback effect of task-based assessment on Iranian intermediate EFL learners' writing performance in terms of fluency. While Tayebi *et al.* (2022) focused on teacher assessment practices and our study on direct student assessments, there could be a complement in providing a holistic view of the factors influencing students' writing fluency and the effectiveness of task-based assessment in promoting it.

CONCLUSION

As mentioned, the purpose of this study was to investigate the washback effect of the task-based assessment on Iranian intermediate EFL learners' writing performance in three areas, complexity, accuracy and fluency. The students demonstrated their level of complexity in their writing, as they were required to express their ideas in a more elaborated and nuanced manner.

Regarding accuracy, the implementation of task-based assessment highlighted the importance of language accuracy and pushed the students to pay more attention to grammar, spelling, and punctuation. The incorporation of specific criteria for accuracy assessment provided clear guidelines for the learners to focus on correcting their linguistic errors. As a result, the students' writing performance did not show noticeable improvement in terms of grammatical accuracy and proper language use.

In terms of fluency, the tasks provided in the assessment encouraged the students to express their ideas in a more organized and coherent manner. The requirement of completing tasks within a given time limit pushed the learners to think quickly and write more fluently. In this study, task-based assessment did not show a contribution to enhancing the students' ability to organize their thoughts and effectively convey their ideas in writing.

This research provided evidence of the washback effect of task-based assessment on the writing performance of Iranian EFL intermediate learners in terms of complexity, accuracy, and fluency. The results indicated that task-based assessment is a valuable approach as the same as the traditional approaches in writing performance of language skills, as it encourages critical thinking, accuracy, and fluency in the students' writing. These findings have significant implications for language educators and curriculum designers, as they highlight the importance of incorporating task-based assessment in language learning contexts for students' overall writing proficiency.

REFERENCES

1. Safa, M. A. (2014). The washback effects of task-based assessment on the Iranian EFL learners' grammar development. *Procedia-Social and Behavioral Sciences*, 98, 90-99.
2. Ahmed, R. Z., & Bidin, S. J. B. (2016). The effect of task based language teaching on writing skills of EFL learners in Malaysia. *Open Journal of Modern Linguistics*, 6(03), 207.
3. Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language testing*, 13(3), 280-297.
4. Amrein-Beardsley, A. U. D. R. E. Y. (2009). The Unintended, Pernicious Consequences of " Staying the Course" on the United States' No Child Left Behind Policy. *International Journal of Education Policy and Leadership*, 4(6), 1-13.
5. Aschbacher, P. R. (1994). Helping educators to develop and use alternative assessments: Barriers and facilitators. *Educational Policy*, 8(2), 202-223.
6. Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational researcher*, 36(5), 258-267.
7. Bachman, L. F. (2001). Designing and developing useful language tests. *Experimenting with uncertainty: Essays in honour of Alan Davies*, 109-116.
8. Barrot, J. S., & Agdeppa, J. Y. (2021). Complexity, accuracy, and fluency as indices of college-level L2 writers' proficiency. *Assessing Writing*, 47, 100510.
9. Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language testing*, 13(3), 257-279.
10. Benson, S. D. (2016). Task-based language teaching: An empirical study of task transfer. *Language Teaching Research*, 20(3), 341-365.
11. Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2-20.
12. Biggs, J. B. (1995). Assumptions underlying new approaches to educational assessment: Implications for Hong Kong. *課程論壇*.
13. Birjandi, P., & Malmir, A. (2011). The effect of task-based approach on the Iranian advanced EFL learners' narrative vs. expository writing. *Iranian Journal of Applied Language Studies*, 1(2), 1-26.
14. Brown, J. D. (2004). Performance assessment: Existing literature and directions for research. *Second Language Study*, 22, 91-139.
15. Bruton, D. L., & Kirby, D. R. (1987). Research in the classroom: Written fluency: Didn't we do that last year?. *The English Journal*, 76(7), 89-92.
16. Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT journal*, 10(1), 15-42.
17. Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, 32, 21.
18. Chapman, D. W., & Snyder Jr, C. W. (2000). Can high stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development*, 20(6), 457-474.
19. Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and teacher education*, 15(3), 253-271.
20. Cheng, L. (2003). Looking at the impact of a public examination change on secondary classroom teaching: A Hong Kong case study. *The Journal of Classroom Interaction*, 1-10.
21. Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34(2), 213-238.
22. Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66-79.
23. Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in educational evaluation*, 45, 62-69.
24. De Wolf, I. F., & Janssens, F. J. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of education*, 33(3), 379-396.
25. Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School accountability, postsecondary attainment, and earnings. *Review of Economics and Statistics*, 98(5), 848-862.
26. Dolnicar, S., & Chapple, A. (2015). The readability of articles in tourism journals. *Annals of Tourism Research*, 52, 161-166.
27. Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
28. Elahi, A., & Heidari, D. M. (2021). The Impact of integrating blended learning with task-based language learning on reading comprehension of Iranian EFL learners. *Journal of Language Horizons*, 5(1), 125.
29. Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in second language acquisition*, 27(2), 305-352.
30. Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied linguistics*, 30(4), 474-509.
31. Ellis, R. (2017). Position paper: Moving task-based language teaching forward. *Language Teaching*, 50(4), 507-526.
32. Evans, N. W., Hartshorn, K. J., Cox, T. L., & De Jel, T. M. (2014). Measuring written linguistic accuracy with weighted clause ratios: A question of

- validity. *Journal of Second Language Writing*, 24, 33-50.
33. Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *Tesol Quarterly*, 28(2), 414-420.
34. Foster, P., & Tavakoli, P. (2009). Lexical diversity and lexical selection: A comparison of native and non-native speaker performance. *Language Learning*, 59(4), 866-896.
35. Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics*, 36, 98-116.
36. Golparvar, S. E., & Khafi, A. (2021). The role of L2 writing self-efficacy in integrated writing strategy use and performance. *Assessing Writing*, 47, 100504.
37. Ha, N. T. T. (2019). A literature review of washback effects of assessment on language learning. *HO CHI MINH CITY OPEN UNIVERSITY JOURNAL OF SCIENCE-SOCIAL SCIENCES*, 9(2), 3-16.
38. Nidhomil, H. A., & Saputro, T. H. (2019). The Implementation of Task-Based Language Teaching (TBLT) and English for Specific Purposes (ESP) Context in SMK Muhammadiyah 1 Malang. *ELITE JOURNAL*, 1(2), 111-122.
39. Holm, L., & Kousholt, K. B. (2019). Beyond washback effect: A multi-disciplinary approach exploring how testing becomes part of everyday school life focused on the construction of pupils' cleverness. *Annual Review of Critical Psychology*, 16, 917-952.
40. Hunt, K. W. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3.
41. Hung, S. T. A., & Huang, H. T. D. (2019). Standardized proficiency tests in a campus-wide English curriculum: A washback study. *Language Testing in Asia*, 9, 1-17.
42. Jamalifar, G., Salehi, H., Tabatabaei, O., & Jafarigohar, M. (2021). Washback effect of the English proficiency test (EPT) on PhD candidates' language learning strategies. *Journal of Language and Translation*, 11(3), 179-191.
43. Jennings, J. L., & Bearak, J. M. (2014). "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher*, 43(8), 381-389.
44. Tayyebi, M., Abbasabady, M. M., & Abbassian, G. R. (2022). Examining classroom writing assessment literacy: A focus on in-service EFL teachers in Iran. *Language Testing in Asia*, 12(1), 12.
45. Hakim, P. K., & Srisudarso, M. (2020). A washback study on portfolio assessment. *ELT in Focus*, 3(1), 9-14.
46. Kermad, A., & Kang, O. (2019). Effect of classroom assessment stakes on English language learners' oral performance. *Tesol Journal*, 10(2), e00392.
47. Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied linguistics*, 18(2), 141-165.
48. Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*, 30(4), 579-589.
49. Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, 16(3), 307-322.
50. Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168-182.
51. Qian, L. (2023). Use of lexical features in high-stakes tests: Evidence from the perspectives of complexity, accuracy and fluency. *Assessing Writing*, 57, 100758., 57, 1075-2935.
52. Lisa A. R., & Zemach, D. (2005). Academic writing from paragraph to essay. *Macmillan Publishers Limited*, 4, 25-33.
53. Long, M. (2014). Second language acquisition and task-based language teaching. *Studies in Second Language Acquisition*, 37(3), 594-595.
54. Long, M., & Norris, J. (2000). Task-Based teaching and assessment. *Encyclopedia of language teaching*, 1, 597-603.
55. Lutfiana, L., Tono, S., & Mahmuda, A. (2020). Overseas students' language and culture barriers towards acquiring academic progress: A study of Thai undergraduate students. *International Journal of Current Science and Multidisciplinary Research*, 3(4).
56. Martel, J. (2019). Washback of ACTFL's Integrated Performance Assessment in an Intensive Summer Language Program at the Tertiary Level. *Language Education & Assessment*, 2(2), 57-69.
57. McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and linguistic computing*, 15(3), 323-338.
58. McKinley, J., & Thompson, G. (2018). Washback effect in teaching English as an international language. *The TESOL encyclopedia of English language teaching*, 1-12.
59. Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 5-19.
60. Messick, S. (1996). Validity and washback in language testing. *Language testing*, 13(3), 241-256.
61. Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation1. *Journal of linguistics*, 21(2), 339-384.
62. Neary-Sundquist, C. A. (2017). Syntactic complexity at multiple proficiency levels of L2 German speech. *International Journal of Applied Linguistics*, 27(1), 242-262.

63. Noroozi, M., & Taheri, S. (2022). Task-based language assessment: a compatible approach to assess the efficacy of task-based language teaching vs. present, practice, produce. *Cogent Education*, 9(1), 2105775.
64. Norris, J. M. (2016). Current uses for task-based language assessment. *Annual Review of Applied Linguistics*, 36, 230-244.
65. Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied linguistics*, 30(4), 555-578.
66. NORRIS, J. M. (2018). Task-based language assessment aligning designs with intended uses and consequences. *JLTA Journal*, 21, 3-20.
67. O'Donnell, R. C. (1974). Syntactic differences between speech and writing. *American speech*, 49(1/2), 102-110.
68. O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language testing*, 12(2), 217-237.
69. Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of second language writing*, 19(4), 218-233.
70. Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492-518.
71. Ortega, L. (2012). Epilogue: Exploring L2 writing-SLA interfaces. *Journal of Second Language Writing*, 21(4), 404-415.
72. Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, 31(1), 117-134.
73. Palupiningsih, A., & Kusumastiti, W. (2020). WASHBACK OF BROADCAST PROJECT-BASED ASSESSMENT FOR TOURISM 4.0 ERA ON STUDENTS'LEARNING. SAGA: *Journal of English Language Teaching and Applied Linguistics*, 1(1), 59-66.
74. Pitoyo, M. D., & Asib, A. (2020). Gamification-Based Assessment: The Washback Effect of Quizizz on Students' Learning in Higher Education. *International Journal of Language Education*, 4(1), 1-10.
75. Polio, C., & Fleck, C. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7(1), 43-68.
76. Polio, C. (2017). Second language writing development: A research agenda. *Language Teaching*, 50(2), 261-275.
77. Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of second language writing*, 26, 10-27.
78. Pouladian, N., Bagheri, M. S., & Sadighi, F. (2017). An analysis of errors in writing skill of adult Iranian EFL learners preparing for the IELTS. *International Journal of English Linguistics*, 7(3), 85-96.
79. Reynolds, D. W. (2005). Linguistic correlates of second language literacy development: Evidence from middle-grade learner essays. *Journal of Second Language Writing*, 14(1), 19-45.
80. Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. Routledge.
81. Rottensteiner, S. (2010). Structure, function and readability of new textbooks in relation to comprehension. *Procedia-Social and Behavioral Sciences*, 2(2), 3892-3898.
82. Salehi, H., & Yunus, M. M. (2012). The Washback Effect Of The Iranian Universities Entrance Exam: Teachers' Insights. *GEMA Online Journal of Language Studies*, 12(2).
83. Salma, N., & Prastikawati, E. F. (2021). Performance-based assessment in the English learning process: Washback and barriers. *Getsempena English Education Journal*, 8(1), 164-176.
84. SARIGÖZ, İ. H., & Fişne, F. N. (2019). Integration of task-based language assessment into young learner classrooms. *Eurasian Journal of Applied Linguistics*, 5(3), 403-422.
85. Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of second language writing*, 9(3), 259-291.
86. Sawyer, A. G., Laran, J., & Xu, J. (2008). The readability of marketing journals: Are award-winning articles better written?. *Journal of Marketing*, 72(1), 108-117.
87. Shirzadi, D., & Amerian, M. (2020). Washback effects of multiple-choice, cloze and metalinguistic tests on EFL students writing. *Indonesian Journal of Applied Linguistics*, 9(3), 536-544.
88. Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal*, 76(4), 513-521.
89. Skehan, P. (1992). *Strategies in second language acquisition* (Vol. 1, pp. 178-208). Thames Valley University Working Papers in English Language Teaching.
90. Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, 17(1), 38-62.
91. Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics*, 30(4), 510-532.
92. Skehan, P., & Foster, P. (2001). Cognition and tasks. *Cognition and Second Language Instruction*, 36(1), 181-203.
93. Snellings, P., Van Gelderen, A., & De Glopper, K. (2004). Validating a test of second language written lexical retrieval: A new measure of fluency in written language production. *Language*

- Testing, 21(2), 174-201.
- Sumera, A., Burua, A., & Navamoney, M. D. (2015). Exploring the effect of backwash in first-year medical learners and comparison with their academic performances. *Procedia-Social and Behavioral Sciences*, 174, 491-495.
94. Tayyebi, M., Abbasabady, M. M., & Abbasian, G. R. (2022). Examining classroom writing assessment literacy: A focus on in-service EFL teachers in Iran. *Language Testing in Asia*, 12(1).
95. University of Cambridge Local Examination Syndicate (UCLES),(2023). Cambridge Assessment Archives & Heritage, GBR/2086/UCLES.
96. Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). Second language development in writing: Measures of fluency, accuracy & complexity. *Studies in Second Language Acquisition*, 23(3), 423-425.
97. Yang, W. (2014). Mapping the relationships among the cognitive complexity of independent writing tasks, L2 writing quality, and complexity, accuracy and fluency of L2 writing. *Applied Linguistics and English as a Second Language Dissertations*, 1, 80-86.
98. Zenker, F., & Kyle, K., (2021). Investigating minimum text lengths for lexical diversity, *Indices. Assessment Writing*, 47, 100-505.