

The Washback Effect, Assessment for Learning, and Assessment of Learning Among Cycle-2 Teachers in Muscat Governorate, Oman: An Explanatory Sequential Mixed-Methods Study

Ruqaiya Al-Baroomi¹, Prof. Hussain Alkharusi¹, and Dr. Yousif Abu-Shendi¹

Journal for Educators, Teachers and Trainers, Vol.17 (1)

<https://jett.labosfor.com/>

Date of reception: 5 July 2026

Date of revision: 10 August 2026

Date of acceptance: 30 September 2026

Ruqaiya Al-Baroomi¹, Prof. Hussain Alkharusi¹, and Dr. Yousif Abu-Shendi¹(2026). The Washback Effect, Assessment for Learning, and Assessment of Learning Among Cycle-2 Teachers in Muscat Governorate, Oman: An Explanatory Sequential Mixed-Methods Study. *Journal for Educators, Teachers and Trainers, Vol.17 (1) 147-171*



Journal for Educators, Teachers and Trainers, Vol.17 (1)

ISSN 1989-9572

<https://jett.labosfor.com/>

The Washback Effect, Assessment for Learning, and Assessment of Learning Among Cycle-2 Teachers in Muscat Governorate, Oman: An Explanatory Sequential Mixed-Methods Study

Ruqaiya Al-Baroomi¹, Prof. Hussain Alkharusi¹, and Dr. Yousif Abu-Shendi¹

¹Department of Psychology, College of Education, Sultan Qaboos University, Muscat, Oman

Abstract:

This study aimed to identify the degree to which elementary teachers in Muscat Governorate, Sultanate of Oman, practice washback-related assessment, assessment for learning, and assessment of learning, and to examine differences by gender, subject, and years of experience, as well as the relationships among these dimensions. A mixed-methods explanatory sequential design was used, with quantitative data from 379 teachers via questionnaire and qualitative data from focus group interviews with 12 teachers. Results showed high levels of practice in all three dimensions, no significant differences by gender or experience, and differences in washback due to gender-subject interaction. Positive correlations were found among the dimensions, with qualitative data confirming teachers' use of tests for planning and instructional improvement. Qualitative data reinforced the quantitative results, highlighting teachers' reliance on test results for planning, teaching decisions, and collaboration with colleagues. The study points to the importance of strengthening formative assessment applications and promoting positive washback practices within national assessment policies to improve classroom teaching quality and student learning.

Keywords: Teachers, Assessment, Washback effect, Assessment for learning, Assessment of learning, Elementary Teachers.

1. Introduction

Assessment plays an effective role in improving and developing the educational process. It is considered a complex process that is diagnostic, formative, and remedial, aiming at identifying the progress individuals make toward achieving the desired objectives. It reveals the actual extent of students' acquisition of skills, knowledge, attitudes, and values, enabling learners to employ knowledge and apply the necessary skills they have acquired in their lives (Jafal & Hammadnah, 2015).

Assessment is also concerned with essential changes in the learner's personality and the major objectives that educational systems seek to achieve. It is not limited to academic achievement alone; rather, it extends to identifying the inclinations and attitudes that learners develop during the educational process. Assessment practices extend beyond measuring academic achievement, as they may also influence psychological factors associated with learning success, such as students' self-concept, motivation, and engagement in learning (Baiduri, 2022).

Classroom assessment is considered a fundamental component of the educational process within the classroom. It is used to collect evidence about students' learning and their progress toward achieving educational objectives, and it is based on specific learning competencies. It includes multiple methods characterized by validity and reliability, and its results are used to enable students to reach their maximum potential and to improve the learning process and its outcomes (Al-Hakami, 2007).

Classroom assessment serves multiple purposes, including assessment of learning, which is the process aimed at determining the extent of success or failure in achieving the general objectives of the curriculum and identifying its strengths and weaknesses (Darandari, 2017). It also includes assessment for learning, which is a pre-planned process in which teachers use evidence obtained from assessment to adjust their ongoing instructional practices, or in which students use such evidence to modify their learning strategies (Popham, 2008).

The tools of classroom assessment vary according to teachers' objectives and include discussions, observations, assignments, and written tests (Korbaa, 2021). Tests are considered the most commonly used tools, as they are used to identify students' abilities and progress in the educational process (Qanooa & Fohamah, 2017).

There has been increasing interest in the washback effect of tests on teaching and learning, referred to as Test Washback, that is, the impact of tests on teaching and learning (Alderson & Banerjee, 2001). Alderson and Wall (1993) explained that tests influence what is taught to students, whereas Erfiani et al. (2018) indicated that such washback may be positive if it directs students toward deep and comprehensive learning, and negative if it drives teachers to narrow curricula or teach solely for the purpose of testing.

Asare (2002) also emphasized that measurement and assessment practices in Arab schools have shifted from being a means to becoming an end in themselves, which has led many teachers to focus on preparing students to pass tests that emphasize the recall of information. Popham (2005) further clarified that this reality poses a challenge for teachers, as their performance is evaluated based on students' scores, creating pressure from school administrations to increase pass rates and leading to teaching that is oriented exclusively toward testing.

Accordingly, the importance of studying test washback practices and their relationship with teachers' practices of assessment for learning and assessment of learning becomes evident. This study aims to shed light on the reality of assessment practices and to propose developmental suggestions that enhance positive test washback and its effective application in improving teaching and learning.

Despite the growing interest in classroom assessment practices, limited research has examined the interconnected relationship among washback effect, Assessment for learning, and Assessment of learning within the Omani educational context, particularly among Cycle-2 teachers across different subject areas. Furthermore, few studies have employed an explanatory sequential mixed-methods design to explore how these assessment practices interact within classroom settings. Therefore, the current study seeks to address this gap by investigating the relationship between washback effect, assessment for learning (AfL), assessment of learning (AoL) among Cycle-2 teachers in Muscat Governorate, Oman.

Although considerable efforts have been made to promote assessment for learning in Omani schools, field observations and supervisory reports indicate shortcomings in its implementation in the desired manner (Al-Saidi & Al-Sheedi, 2022). Recent studies have also recommended the need to equip teachers with modern assessment methods that support classroom practices (Haji, 2023). Moreover, Arabic research in the field of test washback remains limited, compared to foreign studies, which have often focused on language education (Bakir, 2019; Xie & Jia, 2025). Hence, there became a need to examine test washback practices and their relationship with assessment for learning and assessment of learning within the Omani context.

Therefore, this study aims to answer the following research questions:

1. What is the level of test washback practices among teachers of the second cycle of elementary education in Muscat Governorate, Sultanate of Oman?
2. Are there statistically significant differences in the level of test washback practices among teachers of the second cycle of elementary education in Muscat Governorate, Sultanate of Oman, attributable to gender, subject taught, and years of teaching experience?
3. What is the level of assessment for learning practices among teachers of the second cycle of elementary education in Muscat Governorate, Sultanate of Oman?
4. Are there statistically significant differences in the level of assessment for learning practices among teachers of the second cycle of elementary education

- in Muscat Governorate, Sultanate of Oman, attributable to gender, subject taught, and years of teaching experience?
5. What is the level of assessment of learning practices among teachers of the second cycle of elementary education in Muscat Governorate, Sultanate of Oman?
 6. Are there statistically significant differences in the level of assessment of learning practices among teachers of the second cycle of elementary education in Muscat Governorate, Sultanate of Oman, attributable to gender, subject taught, and years of teaching experience?
 7. What is the relationship between the level of test washback practices and both the level of assessment for learning practices and the level of assessment of learning practices among teachers of the second cycle of elementary education in Muscat Governorate, Sultanate of Oman?

2. Methodology

The study adopted a mixed-methods approach following an Explanatory Sequential Design, which begins with the collection of quantitative data through a questionnaire administered to a large sample of teachers, followed by qualitative data collection via focus group interviews to provide a deeper interpretation of the quantitative findings (Creswell & Clark, 2017). This approach aimed at determining the level of test washback, assessment for learning, and assessment of learning practices, as well as to understand their dimensions in light of teachers' experiences and perceptions.

2.1 Sampling

The study population consists of all second-cycle elementary education teachers in Muscat Governorate for the academic year 2023/2024, totaling 2,823 teachers distributed across 70 schools (Ministry of Education, 2023). The quantitative sample size was determined using the Cochran formula (Cochran, 1977) via Raosoft software at a significance level of (0.05) and a confidence level of (95%). The minimum sample size was (339) individuals, and (379) male and female teachers were selected using a cluster random sampling method from 12 schools (6 males and 6 females). The sample was distributed based on gender, specialization, and years of experience. Table 1 shows the distribution of the study sample according to the study variables.

Table 1. Distribution of the Study Sample According to Study Variables

Variable	Category	N
Gender	Male	187
	Female	192
Subject	Islamic Religion	45
	Arabic Language	65
	English Language	42

	Social Studies	40
	Science	90
	Mathematics	97
Years of Teaching Experience	(1-10) years	137
	(11-20) years	170
	More than 20 years	72

The qualitative sample was purposively selected from 12 teachers, representing the six main subject areas: Arabic, Islamic Religion, English, Mathematics, Science, and Social Studies. The selection also considered the variables of gender and teaching experience, categorized as 1–10 years, 11–20 years, and more than 20 years.

2.2 Instruments

2.2.1 Quantitative Instruments

A questionnaire was developed, consisting of three sections: Test Washback (20 items), Assessment for Learning (13 items) and Assessment of Learning (10 items). The items were prepared based on the theoretical literature and previous studies (Al-Damgh, 2010; Bakir, 2019; Al-Shaqsi & Al-Qara'an, 2018; Haji, 2023; Kong & Zhang, 2024). A five-point Likert scale was used for responses, ranging from 1 (Very Low) to 5 (Very High).

The questionnaire was reviewed by nine experts, including university professors and educational supervisors. The validity ratio among reviewers was calculated using Lawshe's formula. The results showed that the average agreement among the experts for the first section on test washback was 99.5%, for the second section on assessment for learning was 94.2%, and for the third section on assessment of learning was 89.2%. These agreement levels are considered acceptable for a panel of nine experts (Lawshe, 1975). To verify the consistency of the questionnaire items, the item-total correlation coefficients were calculated between the score of each item and the total score of the section, based on responses from the pilot sample (n = 60). Table 2 shows the results.

Table 2. Pearson Correlation Coefficients Between Each Item Score and the Total Score of the Test Washback Scale

Item Number	Correlation Coefficient	Item Number	Correlation Coefficient
1	0.68	11	0.48
2	0.66	12	0.58
3	0.74	13	0.56
4	0.59	14	0.80

5	0.76	15	0.81
6	0.60	16	0.71
7	0.67	17	0.70
8	0.69	18	0.70
9	0.64	19	0.55
10	0.39	20	0.85

All values are significant at the ($\alpha = 0.01$) level.

It is observed from Table 2 that the values of the correlation coefficients between the item scores and the total score of the test washback scale ranged between (0.39-0.85), and these correlations were statistically significant at the significance level of ($\alpha = 0.01$), which indicates that the test washback scale is internally consistent. Table 3 shows the Pearson correlation coefficients between each item score and the total score of the Assessment for Learning scale.

Table 3. Pearson Correlation Coefficients between Each Item Score and the Total Score of the Assessment for Learning Scale

Item Number	Correlation Coefficient
1	0.74
2	0.75
3	0.75
4	0.63
5	0.47
6	0.65
7	0.77
8	0.54
9	0.68
10	0.55
11	0.63
12	0.67
13	0.83

Values are significant at ($\alpha = 0.01$) level.

Table 3 shows that the correlation coefficients between the item scores and the total score of the Assessment for Learning scale ranged between (0.47-0.83), and these correlations were statistically significant at the ($\alpha = 0.01$) level, which indicates that the Assessment for Learning scale is internally consistent. Table 4 shows the Pearson correlation coefficients between the score of each item and the total score of the Assessment of Learning scale.

Table 4. Pearson Correlation Coefficients between Each Item Score and the Total Score of the Assessment of Learning Scale

Item Number	Correlation Coefficient
1	0.67
2	0.83
3	0.64
4	0.81
5	0.74
6	0.62
7	0.60
8	0.45
9	0.79
10	0.49

Values are significant at ($\alpha = 0.01$) level.

Table 4 shows that the correlation coefficients between the item scores and the total score of the Assessment of Learning scale ranged between (0.45-0.83), and these correlations were statistically significant at the ($\alpha = 0.01$) level, which indicates that the Assessment of Learning scale is internally consistent.

For reliability, the questionnaire was administered to the pilot sample. It was found that Cronbach's alpha coefficients reached 0.92 for Test Washback, 0.88 for Assessment for Learning and 0.85 for Assessment of Learning. These high coefficients indicate good internal consistency and reliability of the instrument.

2.2.2 Qualitative Instruments

Focus group interviews were conducted with 12 participants (6 males and 6 females), following a pre-prepared protocol linking interview questions to the quantitative analysis results. Each interview lasted 60–90 minutes and was audio-recorded after obtaining participants' consent.

The Interview reliability was ensured through:

1. Full recording and transcription of the interviews (Audit Trail).
2. Participant feedback (Member Check).
3. Peer review by experts in qualitative data analysis with an 80-100% agreement rate.

2.3 Data Analysis

1- Quantitative: SPSS software was used to generate means and standard deviations for descriptive questions. Three-way ANOVA was used to analyze differences by gender, subject, and years of experience, and Pearson's correlation coefficient was used to examine relationships between variables. Effect size was calculated according to Cohen standard (Cohen, 1988).

2- Qualitative: the data were analyzed using Thematic Analysis, beginning with initial coding and progressing to the extraction of major themes that explained the quantitative findings (Creswell & Clark, 2017).

3. Results and Discussion

3.1 Degree of Test Washback Practices Among Teachers

The arithmetic means and standard deviations of the sample's responses on all items of the Test Washback Practices Scale, and the scale as a whole, were calculated, as shown in Table 5.

Table 5. Arithmetic Means and Standard Deviations for Test Washback Practices Items

Item No.	Rank	Items	Arithmetic Mean	Standard Deviation	Degree of Practice
1	1	I align my lesson plans with the content and specifications of the tests.	4.43	0.72	Very High
4	2	I use past test questions to guide my teaching practices.	4.34	0.83	Very High
3	3	I focus my explanations more on topics or skills that are likely to appear on the test.	4.31	0.86	Very High
6	4	My classroom activities focus on improving students' test-taking skills.	4.31	0.78	Very High
7	5	I design classroom activities that simulate test specifications.	4.29	0.81	Very High

14	6	I vary my teaching methods based on students' levels of test readiness.	4.21	0.82	Very High
5	7	I prioritize teaching methods that directly prepare students for exams.	4.17	0.82	High
16	8	I collaborate with colleagues to share strategies for effective exam preparation.	4.17	0.93	High
13	9	I provide extra support to students who find difficulty with exam preparation.	4.13	0.83	High
15	10	I seek out resources and training programs that enhance my ability to prepare students for exams.	4.13	0.93	High
2	11	My teaching objectives are influenced by the course content that will be covered in the exams.	4.12	0.99	High
12	12	I adapt my teaching strategies to meet the diverse needs of students to ensure they perform well on exams.	4.10	0.90	High
18	13	I use exams as a tool to motivate students to more deeply engage with the subject content.	4.07	0.92	High
17	14	I provide tailored feedback to improve student performance on exams.	4.06	0.90	High
20	15	I adjust my teaching methods in response to changes in policies or guidelines related to exams.	3.98	0.97	High
8	16	I dedicate a significant part of class time to reviewing past exam papers.	3.52	0.98	High

19	17	I believe it is important to prepare students for exams, even if it means not addressing some of the course content.	3.46	1.23	High
9	18	I spend less time explaining topics that are not included in the exam.	3.21	1.10	Moderate
11	19	I avoid using new teaching techniques if they do not prepare students to perform well on the test.	3.02	1.25	Moderate
10	20	My teaching tends to emphasize memorization of facts rather than critical thinking.	2.84	1.23	Moderate
		Overall arithmetic mean	3.94	0.60	High

The results in Table 5 show that the arithmetic means for the washback practice items of the tests ranged between (2.84-4.43). Item (1), “I align my lesson plans with the content and specifications of the tests,” received the highest average score (4.43), followed by item (4), “I use past test questions to guide my teaching practices.” (4.34), and then item (6), “My classroom activities focus on improving students’ test-taking skills” (4.31). In contrast, item (9), “I spend less time explaining topics that are not included in the test,” received a moderate average score of (3.21), while item (10), “My teaching tends to emphasize memorization of facts rather than critical thinking,” came in last with an average score of (2.84). The overall average score for the axis was (3.94), indicating a high level of practice.

The qualitative data supported these findings, as teachers confirmed that tests guide lesson planning, activity design, and the prioritization of content. They explained that they use past test questions as training and teaching tools and employ classroom activities that simulate exam patterns. The teachers also highlighted that professional collaboration; such as analyzing test papers and sharing experiences, constitutes an important aspect of test washback. With respect to item 10, participants expressed their awareness of the risks of rote memorization and passive teaching, emphasizing their efforts to promote critical thinking and link lessons to students’ real-life experiences, thereby reducing the negative effects of testing. These findings indicate that test washback is an influential phenomenon in teaching practices, which can be positive when employed with pedagogical awareness (Hughes, 2020; Cheng, 2005; Al-Damgh, 2010; Fasih, 2022).

3.2 Differences in Washback Practices

A three-way analysis of variance was conducted, including the main effects and two-way interaction effects of the study variables (gender, subject area, and years of teaching experience) in the level of test washback practices among teachers. Prior to conducting the analysis, the following assumptions were examined:

- Independence: This assumption was satisfied, as the sample was selected using the cluster random sampling method.
- Normality of distribution: The values of skewness and kurtosis ranged between (-2) and (+2) (George & Mallery, 2001), which indicates an approximately normal distribution.
- Homogeneity of variance: This assumption was tested using Levene's Test for all dependent variables (test washback, assessment for learning, and assessment of learning). The F values were (1.81, 1.36, and 1.59), and they were not statistically significant at the level of $\alpha = 0.05$, which confirms the homogeneity of variances across groups.
- After verifying these assumptions, the means and standard deviations of the dependent variables were calculated according to the study variables, as presented in Table (6).

Table 6. Means and standard deviations of test washback practices according to the study variables

Variable	Category	N	Arithmetic Mean	Standard Deviation
Gender	Male	187	3.95	0.55
	Female	192	3.94	0.65
Subject	Islamic Religion	45	3.86	0.74
	Arabic Language	65	4.20	0.50
	English Language	42	3.87	0.43
	Social Studies	40	4.06	0.49
	Science	90	3.90	0.67
	Mathematics	97	3.83	0.59
Years of Teaching Experience	1-10 years	137	3.91	0.63
	11-20 years	170	4.00	0.56
	More than 20 years	72	3.88	0.64

Table 7 shows the results of the three-way ANOVA for the degree of washback of test practices according to the study variables.

Table 7. Results of the three-way ANOVA for the degree of washback of test practices according to the variables of study

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F Value	Probability Value	Effect Size (Eta, η^2)
Gender (A)	0.24	1	0.24	0.72	0.394	-
Subject (B)	5.24	5	1.05	3.15	0.009	0.04
Years of Experience (C)	0.08	2	0.04	0.15	0.881	-
Interaction (A) × (B)	6.95	5	1.39	4.17	0.001	0.06
Interaction (A) × (C)	0.20	2	0.10	0.31	0.735	-
Interaction (B) × (C)	4.20	10	0.42	1.26	0.251	-
Error	117.60	353	0.33			

The results in Table 7 show an interaction effect between gender and subject on the degree of test washback practices among second-cycle teachers in Muscat Governorate, with an effect size of ($\eta=0.062$), which is an average effect according to Cohen (1988).

Qualitative data supported these findings. Interviews with female teachers (especially in Islamic Religion and Arabic Language) revealed a focus on aligning activities with the exam document, while male teachers (especially in science subjects) emphasized connecting knowledge to everyday life contexts. As one science teacher said, “We do not teach only for the exam; we connect concepts to students’ lives, and this is what strengthens understanding more.” Previous studies also suggest that instructional and assessment practices may vary according to subject characteristics and learners’ educational needs, particularly in subject requiring specialized instructional support such as mathematics (Rasheed & Sunday, 2026). Accordingly, the results indicate that differences in washback practices are not directly attributable to gender or teaching experience; rather, they are influenced by the nature of the subject and by the interaction between gender and specialization. This reflects the practical nature of washback within the Omani educational context.

3.3 Level of Assessment for Learning (AfL) Practices

Results of Question (3): What is the level of assessment for learning practices among teachers of the second cycle of elementary education in Muscat Governorate, Sultanate of Oman?

To examine the level of assessment for learning practices among teachers, means and standard deviations of the respondents' answers to the items related to the AfL practices were calculated. Table 8 presents the results.

Table 8. Means and standard deviations of the items related to assessment for learning practices

Item No.	Rank	Items	Arithmetic mean	Standard Deviation	Degree Practice
13	1	I record student performance during and after the educational situation in a way that serves the learning process.	4.48	0.67	Very High
6	2	I provide feedback immediately after the assessment process to improve the learning process.	4.47	0.68	Very High
11	3	I record student performance assessments during the educational process.	4.45	0.70	Very High
12	4	I write encouraging phrases for excellent students on the short test paper.	4.42	0.84	Very High
2	5	I use the results of assessment instruments to make decisions that support the educational process.	4.39	0.70	Very High
3	6	I use the results of continuous assessment to improve methods I use in teaching.	4.37	0.75	Very High
1	7	I rely on continuous assessment to motivate students to learn.	4.35	0.75	Very High
7	8	I explain to students the objectives of the assessment process used during the lesson.	4.28	0.81	Very High
4	9	I give students the opportunity to discuss the test questions immediately after the assessment process to support learning.	4.26	0.84	Very high
9	10	I inform students of their learning weaknesses based on the used	4.17	0.85	High

		continuous assessment instruments.			
5	11	I give students pilot tests during the learning process that did not affect their final grade	4.10	0.96	High
8	12	I use a diagnostic test at the beginning of the year to determine students' cognitive level.	3.81	1.17	High
10	13	I give students the opportunity to correct their own assignments.	3.77	1.04	High
		The overall	4.26	0.55	Very high

Table 8 shows that the arithmetic mean for the assessment practices for learning items ranged between (3.77-4.48). Item (13), "I record student performance during and after the educational situation in a way that serves the learning process," ranked first (4.48), followed by item (6), "I provide feedback immediately after the assessment process to improve the learning process" (4.47), and then item (11), "I record student performance assessments" (4.45). Item (10), "I give students the opportunity to correct their own assignments," received the lowest average (3.77). The overall average was (4.26), indicating a very high level of practice.

The interviews supported these findings, as teachers' statements emphasized the importance of formative assessment and continuous monitoring of students' performance. A science teacher stressed that assessment during the lesson ensures that all students achieve understanding before moving on, while an Arabic language teacher believed that immediate feedback prevents the consolidation of errors. Some teachers also reported using performance records to monitor participation and achievement, even to varying degrees.

The interviews further indicated that assessment results are used to modify instruction. As a science teacher explained, "We change strategies and focus on improvement to raise students' achievement." Likewise, an Islamic religion teacher stated, "Sometimes I use games or innovative methods to motivate students." In contrast, some practices, such as self-assessment, were not sufficiently implemented, as reflected in Item (10). These findings are consistent with previous studies, such as (Saefurrohman & Balinas, 2016), which highlighted the role of assessment in diagnosis and feedback, and Alt (2018), which showed that teachers who encourage interaction tend to use formative assessment instruments more frequently. However, they contrast with Haji (2023), which indicated that Arabic language teachers in Madinah implemented assessment for learning at a low level. Based on these findings, it is evident that second-cycle teachers in Muscat Governorate practice assessment for learning at a very high level, with a clear awareness of its function as a tool for improving learning rather than merely judging performance.

3.4 Differences in Assessment for Learning (AfL)

A three-way ANOVA was used to examine differences in the degree of assessment for learning practices among second-cycle teachers attributable to gender, subject, and years of teaching experience. The analysis considered the main effects and two-way interactions of the study variables. Table 9 shows the means and standard deviations of assessment for learning practices according to the study variables. Table 10 shows results of the three-way ANOVA for the degree of assessment for learning practices according to the variables of gender, subject and years of teaching experience.

Table 9. Arithmetic Means and standard deviations of assessment for learning practices according to the study variables

Variable	Category	N	Arithmetic Mean	Standard Deviation
Gender	Male	187	4.21	0.57
	Female	192	4.30	0.52
Subject	Islamic Religion	45	4.21	0.64
	Arabic Language	65	4.34	0.58
	English Language	42	4.26	0.41
	Social Studies	40	4.22	0.63
	Science	90	4.33	0.51
	Mathematics	97	4.16	0.53
Years of Teaching Experience	1-10 years	137	4.18	0.62
	11-20 years	170	4.32	0.48
	More than 20 years	72	4.24	0.56

Table 10. Results of the three-way ANOVA for the degree of assessment for learning practices according to the variables of study

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F Value	Probability Value	Effect Size (Eta, η^2)

Gender (A)	2.16	1	2.16	7.32	0.007	0.02
Subject (B)	2.71	5	0.54	1.84	0.104	-
Years of Experience (C)	0.66	2	0.33	1.12	0.328	-
Interaction (A) × (B)	3.16	5	0.63	2.15	0.060	-
Interaction (A) × (C)	0.77	2	0.38	1.31	0.270	-
Interaction (B) × (C)	2.68	10	0.26	0.91	0.524	-
Error	103.92	353	0.29			

The results in Table 10 show that there was no statistically significant effect of the two-way interactions among gender, subject, and teaching experience on assessment for learning practices, as the p-values exceeded the significance level ($\alpha = 0.05$). In contrast, statistically significant differences were found for the gender variable in favor of female teachers, although the effect size was low ($\eta^2 = 0.02^2$). No significant differences were observed for the subject or teaching experience variables.

The qualitative data further reinforced these findings, as female teachers' statements revealed emphasized diversifying assessment methods (such as brainstorming, discussion, and the integration of higher-order thinking skills), while ensuring immediate feedback and regularly documenting students' performance. On the other hand, some male teachers' interviews revealed variability in implementation, with weaker daily documentation and greater reliance on final achievement results, despite the fact that there are initiatives such as "Little Teacher" or the use of technology. The lack of differences between academic disciplines and teaching experience was further confirmed by the qualitative data, as all participants expressed a shared awareness of the importance of assessment for learning and the use of feedback to improve educational practices.

These findings are consistent with the study by Asare (2021), which found that the only significant main effect of formative assessment practices was attributed to gender, with a small effect size. They also align with the findings of Al-Shaqsi and Al-Qara'an (2018), which confirmed the superiority of female teachers in implementing formative assessment practices. Accordingly, the results indicate that the differences in favor of females can be attributed to greater consistency and diversity in their practices, whereas the other variables did not show notable differences.

3.5 Level of Assessment of Learning (AoL) Practices

Means and standard deviations of the respondents' scores on the scale items and on the scale as a whole were calculated to describe the level of assessment of learning practices among second-cycle teachers in Muscat Governorate. Table 11 summarizes the results.

Table 11. Means and standard deviations of the level of assessment of learning practices among second-cycle elementary teachers

Item No.	Rank	Items	Arithmetic Mean	Standard Deviation	Degree of Practice
7	1	I encourage students to achieve the best possible grades in the final exams.	4.62	0.59	Very High
6	2	I adhere to the exam specifications table provided by the Ministry of Education before writing the questions.	4.57	0.68	Very High
4	3	I ensure that the final assessment criteria are clearly explained to students before the final exams.	4.24	0.76	Very High
9	4	I rely on final exam results to determine students' academic support needs for upcoming semesters.	4.14	0.90	High
2	5	I rely on final exam results to determine the extent to which educational objectives have been achieved.	4.11	0.85	High
1	6	I use final assessment results to group students according to their academic level.	4.10	0.93	High
10	7	I communicate with parents regarding their children's final assessment results.	3.95	1.03	High
3	8	I design comprehensive practice tests that cover the course content.	3.88	1.00	High
8	9	I consider the final assessment results as the basis for students' success or failure in a subject.	3.81	0.94	High
5	10	I compare the student's performance on the final	3.78	1.07	High

		exam to the group's performance level.			
		Total arithmetic mean	4.12	0.59	High

According to Table 11, the arithmetic means ranged between (3.78-4.62). Item (7), "I encourage students to achieve the best possible grades in the final exams," ranked first with an average of (4.62), followed by item (6), "I adhere to the exam specifications table," with an average of (4.57), and then item (4), "I ensure that the final assessment criteria are clearly explained to students," with an average of (4.24). In contrast, item (5), "I compare the student's performance on the final exam to the group's performance level," ranked last with an average of (3.78), while the overall average was (4.12), which indicates a high level of learning assessment practices among teachers.

Qualitative data reinforced these results; since teachers' reports showed a variety of motivational strategies for students, such as educational competitions and psychological encouragement, and their compliance with the specifications table as a guarantee of quality and institutional control. It also revealed a variety in the disclosure of assessment criteria; where some teachers explained them to students and parents, while others only provided the final grade. Other statements indicated the use of exam results for student redistribution or the development of remedial plans, despite challenges in communicating with parents.

These results confirm that assessment of learning practices are regularly implemented, with a clear emphasis on final achievement. This is consistent with the results of previous studies, such as Abolubbeh (2014), which showed that teachers tend to favor assessment of learning over assessment for learning, as well as with the findings of Al-Shaqsi and Al-Qara'an (2018), which indicated teachers' preference for assessment of learning, even at a moderate level. The results also align with the study by Saefurrohman and Balinas (2016), which demonstrated that teachers rely on assessment of learning instruments to determine final grades and to communicate with school administration and parents. Accordingly, it becomes evident that assessment of learning represents the dominant approach among second-cycle teachers in the Sultanate of Oman, accompanied by efforts to enhance its role in motivation and in improving educational decision-making, despite the need to further develop self-assessment practices and broaden family involvement.

3.6 Differences Assessment of Learning (AoL)

A three-way ANOVA was used to examine differences in the degree of assessment of learning practices among second-cycle teachers attributable to gender, subject, and years of teaching experience. The analysis considered the main effects and two-way interactions of the study variables. Table 12 shows the means and standard deviations of assessment of learning practices according to the study variables. Table 13 shows results of the three-way ANOVA for the degree of assessment of learning practices according to the variables of gender, subject and years of teaching experience..

Table 12. Means and standard deviations of the level of assessment of learning practices according to the study variables

Variable	Category	N	Arithmetic Mean	Standard Deviation
Gender	Male	187	4.10	0.59
	Female	192	4.14	0.59
Subject	Islamic Religion	45	4.06	0.67
	Arabic Language	65	4.14	0.64
	English Language	42	4.12	0.38
	Social Studies	40	4.02	0.72
	Science	90	4.27	0.52
	Mathematics	97	4.04	0.58
Years of Teaching Experience	1-10 years	137	4.06	0.64
	11-20 years	170	4.18	0.54
	More than 20 years	72	4.08	0.60

Table 13. Results of the three-way ANOVA for the degree of the assessment of learning practices according to the variables of study

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F Value	Probability Value
Gender (A)	0.44	1	0.45	1.32	0.25
Subject (B)	3.15	5	0.63	1.86	0.10
Years of Experience (C)	0.16	2	0.08	0.24	0.79
Interaction (A) × (B)	3.66	5	0.73	2.16	0.06
Interaction (A) × (C)	0.21	2	0.10	0.30	0.74
Interaction (B) × (C)	2.79	10	0.28	0.82	0.61

Error	119.81	353	0.34		
-------	--------	-----	------	--	--

The results shown in Table 13 indicate that there were no effects of two-way interactions among the variables. Moreover, the statistical analyses revealed no statistically significant differences in assessment of learning practices attributable to gender, subject, or years of experience.

The qualitative data supported this trend, since male and female teachers' statements indicated that all disciplines share a common understanding of assessment of learning as a tool for judging learning outcomes and directing them toward performance improvement, with a clear commitment to official test specifications and their review by the senior teacher and the supervisor. The interviews also revealed that final assessment results are used to identify students in need of support and to redistribute them across people, in addition to using them to enhance communication with parents, which in turn contributes to improving students' performance. Moreover, strategies for motivating high performance varied, including the use of competitions, rewards, and early psychological and cognitive preparation.

These results are consistent with the study by Al-Shaqsi and Al-Qara'an (2018), which showed that assessment of learning practices were at a moderate level, with no differences associated with personal variables. Accordingly, it becomes evident that assessment of learning practices among second-cycle teachers in Muscat are applied in a relatively consistent and systematic manner, with no substantial differences across gender, specializations, or years of experience, which reflects the impact of unified educational policies and the growing professional awareness of the importance of assessment in improving the quality of teaching and learning.

3.7 Correlations among Washback, Assessment for Learning, and Assessment of Learning

To know the nature of the relationships between the level of washback practices and both the level of assessment for learning practices and the level of assessment of learning practices, Pearson's correlation coefficient was used. Table 14 shows the results.

Table 14. Pearson correlation coefficients between washback practices and each of the following: the level of assessment for learning practices and the level of assessment of learning practices

Variable	Test Washback	Assessment for Learning	Assessment of Learning
Test Washback	-	0.57*	0.55*
Assessment for Learning	-	-	0.75*
Assessment of Learning	-	-	-

$p < 0.01$.

The results shown in Table 14 indicate the existence of a statistically significant positive correlation between washback practices and each of the following assessment for learning ($r = 0.57$) and assessment of learning ($r = 0.55$). The results also reveal a strong relationship between assessment for learning and assessment of learning ($r = 0.75$). These results indicate that washback practices serve as a catalyst for the integration of assessment practices, as they enhance the use of continuous feedback (i.e., assessment for learning) and improve the accuracy of summative judgments about learning outcomes (i.e., assessment of learning).

The qualitative data supported these results, as teachers reported that their adherence to test specifications was reflected in the design of lessons and classroom activities, and that the results of final examinations are used to identify areas of weakness, redistribute students, and coordinate with parents, reflecting a practical integration among different assessment instruments. These results are consistent with previous studies (Alderson and Wall, 2001; Fasih, 2022; & Suryanto et al., 2023), which emphasized the role of washback in redirecting teaching activities, motivating students, and improving formative and summative assessment practices.

4. Limitations of the Study

This study has several limitations that should be acknowledged. First, the data were primarily based on teachers' self-reports, which may have been influenced by social desirability bias. Second, the study was conducted within the context of Muscat Governorate, which may limit the generalizability of the findings to other educational settings. Third, although the instrument demonstrated acceptable validity and reliability indicators, additional validation procedures such as exploratory and confirmatory factor analyses were not conducted. Finally, the qualitative findings were based on a relatively small group of participants, which may not fully represent the experiences of all teachers.

5. Conclusions

Based on the aforementioned findings, the following conclusions were made:

1. Test washback is implemented at a high level, with compliance with the ministry specifications, the use of previous test questions, and a positive awareness of their role in improving instructional planning and professional collaboration.
2. Assessment for learning is practiced at a very high level, with a strong focus on feedback, classroom records, and teaching improvement, particularly among female teachers.
3. Assessment of learning is practiced at a high level, with the use of assessment results to support students and enhance communication with parents.
4. No statistically significant differences were found according to gender or years of experience, except for differences in favor of females in assessment for learning, in addition to an interaction effect between gender and subject in washback practices.

5. Statistically significant positive correlations were found between test washback and both assessment for learning and assessment of learning, reflecting the integration of assessment practices in classroom.

6. Recommendations

1. Develop training programs to familiarize teachers with the concept of washback and mechanisms for its positive application.
2. Encouraging the activation of assessment for learning practices in more diverse ways, especially self-assessment and transparency.
3. Developing effective communication channels with parents to enhance their participation in supporting students' learning.

7. Suggestions for Future Research

1. Carrying out studies on washback in specific subjects; such as science or Arabic language, to understand how its impact varies according to the nature of the subject.
2. Carrying out comparative research across different (first, second and third) educational stages to understand how assessment and practices evolve across stages.
3. Conducting quasi-experimental studies to measure the impact of training programs regarding washback on classroom practices.
4. Conducting qualitative studies to explore students' own perceptions of the impact of examinations on their learning.
5. Making research connecting between the level of applying washback and academic achievement on cumulative subjects.

References

- Abolubbeh, I. S. S. (2014). *Assessment of learning and assessment for learning practices among teachers of the first three grades in Jordan* (Unpublished master's thesis). Al-Hussein Bin Talal University, Ma'an, Jordan.
- Al-Damgh, K. A. (2010). Test washback and connectivity and their effects on educational programs. *Arabic for Non-Native Speakers*, 10, 1–30.
- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part I). *Language Teaching*, 34(4), 213–236. <https://doi.org/10.1017/S0261444800014464>
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistic*, 14, 115–129. <https://doi.org/10.1093/applin/14.2.115>
- Al-Hakami, A. (2007). *Educational evaluation and quality assurance in education*. In *Proceedings of the 14th Annual Meeting: Quality in General Education* (May 2007, pp. 809–829). Riyadh, Saudi Arabia: Saudi Educational and Psychological Association (JUSTIN), College of Education, King Saud University.
- Al-Saidi, H. M. S., & Al-Sheedi, K. J. K. (2022). Employing assessment for learning techniques by mathematics and science teachers in classroom settings in basic education

schools in the Sultanate of Oman. *Journal of Curriculum and Teaching Methodology*, 11(1), 25–42.

Al-Shaqsi, A. A., & Al-Qara'an, M. F. A. (2018). *The use of post-basic education school teachers of assessment for learning and assessment of learning practices in the Sultanate of Oman* (Unpublished master's thesis). Yarmouk University, Irbid, Jordan.

Alt, D. (2018). Teachers' practices in science learning environments and their use of formative and summative assessment tasks. *Learning Environments Research*, 21(3), 387–406. <https://doi.org/10.1007/s10984-018-9259-z>

Asare, A. (2002). The Reality of Achievement Tests in General and University Education and Their Impact on Educational Outcomes. *The Tenth Annual Meeting - Educational and Psychological Measurement and Assessment*, Riyadh, Saudi Association for Educational and Psychological Sciences and King Saud University, 314-321.

Asare, E. (2021). Basic school teachers' formative assessment practices: Influence of demographic variables. *International Journal of Current Approaches in Language, Education and Social Sciences*, 3(1), 57–68. <https://doi.org/10.35452/caless.2021.3>

Baiduri, B. (2022). Self-concept and learning achievement of mathematics education college student. *Journal of Psychological Perspective*, 4(2), 69–76. <https://doi.org/10.47679/jopp.424322022>

Bakir, S. (2019). Washback and assessment: Concept, effects, and dimensions. *Arabic*, 47, 213–236. <https://doi.org/10.33705/0114-000-047-010>

Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge University Press.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic Press.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage Publications.

Darandari, I. Z. A. (2017). Developing learning outcomes assessment plans in higher education institutions. *Bulletin of the Faculty of Arts*, 77(1), 13–72.

Erfiani, N., Musthafa, A., & Kurniawan, E. (2018). Washback effects of English summative test on teachers' instruction. *Indonesian EFL Journal*, 4(2), 13–23. <https://doi.org/10.15642/ijet2.2018.7.2.103-109>

Fasih, P. (2022). A critical literature review on the washback effect of task-based vocabulary assessment on Iranian EFL learners. *Frontiers in Education*, 7, Article 950203. <https://doi.org/10.3389/educ.2022.950203>

George, D., & Mallery, P. (2001). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.

Haji, K. M. O. (2023). Practices of assessment for learning among Arabic language female teachers in Al-Madinah Al-Munawwarah and the obstacles from their perspectives. *Journal of Educational and Psychological Sciences*, 16(3), 669–737.

Hughes, A. (2020). *Testing for language teachers* (2nd ed.). Cambridge University Press.

Jafal, F. N., & Hammadnah, I. M. (2015). *The extent to which upper basic stage mathematics teachers in Mafraq Governorate possess alternative assessment strategies and the obstacles to their implementation* (Unpublished master's thesis). Al al-Bayt University, Jordan.

Kong, F., & Zhang, Y. (2024). Investigating the washback of the Chinese Language Proficiency Test (HSK): From the perspective of CSL students. *SAGE Open*, 14 (1), 1–15. <https://doi.org/10.1177/21582440231224599>

Korbaa, A. (2021). Methods and principles of educational evaluation to achieve educational goals. *Al-Shamil Journal of Educational and Social Sciences*, 4(2), 48–57.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>

Ministry of Education (2023), *Annual educational statistics book*. Muscat, Oman: Ministry of Education.

Popham, W. J. (2005). *Classroom assessment: What teachers need to know* (4th ed.). Pearson/Allyn and Bacon.

Popham, W. J. (2008). *Transformative assessment*. Association for Supervision and Curriculum Development.

Qanooa, A. L., & Fohamah, K. (2017). The extent to which middle school teachers use the table of specifications in constructing achievement tests: A field study in El Oued middle schools. *Al-Siraj Journal of Education and Social Issues*, 1, 170–185.

Rasheed, H. T., & Sunday, E. M. (2026). Effects of Braille mathematical code instruction and concrete materials on mathematics performance in primary school learners with total blindness: A quasi-experimental study in Oyo State, Nigeria. *Nusantara Journal of Behavioral and Social Science*, 5(1), 37–46. <https://doi.org/10.47679/njbss.202614851>

Saefurrohman, S., & Balinas, E. (2016). English Teachers Classroom Assessment Practices. *International Journal of Evaluation and Research in Education*, 5(1), 82-92. <http://doi.org/10.11591/ijere.v5i1.4526>

Suryanto, J., Sorohiti, M., Isdiarto, I., Alya, B. C. P., & Zahra, A. (2023). The students' and the teachers' perception of the washback effect of the english progress test. *Research and Innovation in Social Science Education Journal*, 1(2), 55–62. <https://doi.org/10.30595/rissej.v1i2.126>

Xie, Q., & Jia, Q. (2025). Three decades of research on washback (1993–2023): A bibliometric study. *Language Testing in Asia*, 15, Article 23. <https://doi.org/10.1186/s40468-025-00357-w>