Anatolian Journal of Education
e-ISSN: 2547-9652

October 2026 ● Vol.11, No.2
www.e-aje.net
pp. 199-212

Teachers' Assessment Practices: Washback, Assessment for Learning, and Assessment of Learning Profiles

Ruqaiya Al Barumi
Department of Psychology, Sultan Qaboos University, Muscat, Oman, *ruqaya3165@edu.moe.om*

Bahia Al Rashdi
Department of Psychology, Sultan Qaboos University, Muscat, Oman, *s42056@student.squ.edu.om*

Hussain Alkharusi
Department of Psychology, Sultan Qaboos University, Muscat, Oman, *hussein5@squ.edu.om*

This study investigated teachers' assessment practices by identifying latent profiles based on assessment washback, assessment for learning (AFL), and assessment of learning (AoL), using a person-centered analytical approach. It also examined whether selected demographic variables predicted teachers' membership in these profiles. The participants were 376 teachers teaching Grades 5 to 10 in public schools in Muscat Governorate, Sultanate of Oman. Data were collected using a researcher-developed questionnaire consisting of 43 items measuring the three assessment dimensions. Latent Profile Analysis (LPA) was conducted using washback, AfL, and AoL as continuous indicators. The findings supported a three-profile solution representing low, moderate, and high levels of engagement in assessment practices; the profiles differed mainly in the overall level of engagement rather than in distinct forms of the three dimensions. Teachers in the high profile reported consistently high engagement across washback, assessment for learning, and assessment of learning, whereas teachers in the low profile reported minimal use of these practices. Results from the conditional LPA indicated that teaching specialization was associated with profile membership. Teachers of Arabic language and Science were more likely than teachers of Mathematics to belong to the high rather than the low profile, whereas gender and years of teaching experience were not significant predictors. These results highlight variation in teachers' overall level of engagement with assessment practices and illustrate the value of person-centered approaches for identifying subgroups of teachers. These findings suggest that professional development could move beyond unified training models toward differentiated approaches aligned with teachers' assessment practice profiles.

Keywords: washback, assessment for learning, assessment of learning, latent profile analysis, learning

INTRODUCTION

Assessment plays an effective role in improving and developing the educational process. Assessment is not limited to documenting learning outcomes, but it also plays a role in shaping the quality of teaching and supporting educational decision-making. Contemporary perspectives consider assessment as an evidence-based practice contributing to improving learning, adapting instruction, and enhancing accountability within educational systems (Brookhart, 2024). Assessment is no longer understood as a traditional measurement tool; rather, it is considered as an educational process aimed at identifying the progress made by individuals for achieving desired objectives. It reveals the actual extent of students' acquisition of skills, knowledge, and values through the educational process, enabling learners to utilize knowledge and apply the skills they have acquired in their lives (Jafal & Hamadneh, 2015).

Citation: Al Barumi, R., Al Rashdi, B., & Alkharusi, H. (2026). Teachers' assessment practices: Washback, assessment for learning, and assessment of learning profiles. *Anatolian Journal of Education*, *11*(2), 199-212. https://doi.org/10.29333/aje.2026.11213a

Classroom assessment achieves multiple objectives, including assessment of learning, which refers to the process aimed at determining the extent of success or failure in achieving the general objectives of the curriculum, along with identifying areas of strength and weakness within such objectives (Darandari, 2017). Assessment for learning is a pre-planned process in which teachers use evidence obtained from assessment to understand students' learning status in order to adjust their ongoing instructional practices, or in which students use such evidence to modify their current learning strategies (Popham, 2008). Contemporary research indicates that effective teaching is not reached by prioritizing one of these functions over the other; rather, it emerges from integrating them in a balanced and complementary manner within a unified and consistent framework (Black & Wiliam, 2009; Brookhart, 2024).

Closely associated with classroom assessment practices is the concept of Washback, also referred to as test washback, which denotes the effects of tests on teaching and learning (Alderson & Banerjee, 2001; Alderson & Wall, 1993). In other words, it concerns the extent to which tests shape the nature and direction of students' learning within the educational process and instructional practices. Recent studies show washback as a complex and context-sensitive phenomenon influenced by the design of assessment tools, teachers' beliefs, and institutional expectations (Allan, 2020). Nevertheless, test washback can be positive. Taylor (2005) argues that positive washback occurs when assessment contributes to the improvement of educational practices. For instance, speaking tests may encourage teachers to focus on developing speaking skills, which are often neglected in English language instruction. Indeed, tests can act as catalysts for change, prompting teachers to adjust their instructional approaches to ensure that students achieve better learning outcomes (Allan, 2020). Moreover, tests help teachers cover the curriculum comprehensively, thereby contributing to the completion of curricular content within the allocated timeframe (Alderson & Wall, 1993).

Although these three concepts are interrelated, they are not conceptually equivalent. Assessment for learning and assessment of learning represent two major forms of classroom assessment, using assessment to support ongoing learning and to evaluate learning outcomes, respectively, whereas washback represents the effects that assessment and tests in particular exert on teaching and learning (Alderson & Wall, 1993). What connects the three at the classroom level is the teacher who pursues these assessment purposes and responds to the influence of tests through observable assessment and instructional practices. Washback, in particular, is not an automatic effect of tests; it is mediated by teachers, who shape the type and intensity of its influence through their instructional and assessment decisions (Sprat, 2005). Accordingly, the present study does not examine washback as a property of tests, but as the extent to which teachers' reported practices reflect the influence of tests, alongside their reported use of assessment of learning and assessment for learning. Defining the washback dimension in this way places it on the same basis as assessment of learning and assessment for learning, with each dimension represented by reported classroom practices. This enables the three dimensions to be examined together as continuous indicators of teachers' assessment practices in the latent profile analysis. Examining the three dimensions together, as facets of teachers' classroom assessment practices, offers a more integrated understanding of how assessment operates in practice than treating any single dimension in isolation.

Parallel research on assessment for learning and assessment of learning reveals noticeable variation in how teachers implement assessment functions in practice. While some teachers emphasize summative assessment for accountability and reporting purposes, others integrate formative strategies to support learning and enhance teaching (Alt, 2018; Haji, 2023). This variation indicates that differences among teachers are not limited to their assessment knowledge; rather, they extend to how they interpret the purposes of assessment and how they respond to contextual pressures. Accordingly, it is not possible to assume homogeneity in classroom assessment practices among teachers or across different educational contexts.

Recent evidence further supports such variability in teachers' assessment practices. In Omani context, Alkharusi (2021) identified distinct purposes underlying teachers' uses of classroom assessment, illustrating the multiple functions that assessment may serve in instructional practice. Similarly, Magaji (2025) demonstrated variations in teachers' understanding and implementation of formative assessment practices and emphasized the role of assessment competence and literacy in shaping these practices. Together, these findings reinforce the view that teachers may differ in how they conceptualize and enact assessment, providing further justification for examining assessment practices from an individual-centered perspective.

Despite the growing research interest in washback, assessment for learning, and assessment of learning, these dimensions have often been examined in isolation, which has constrained a comprehensive understanding of how they interact within teachers' everyday practices or the extent to which they may be integrated into distinctive patterns of assessment behavior. In particular, how these three dimensions cooccur within individual teachers, and whether teachers can be distinguished into distinct subgroups on this basis, has rarely been examined directly, especially in Arab education context. Moreover, a substantial proportion of previous studies has relied on variable-centered analytical approaches that assume population homogeneity and overlook fundamental differences among teachers. This limitation is particularly evident in research carried out within Arab educational contexts, where studies employing individual-centered methodologies remain relatively limited.

Within this framework, recent methodological developments in educational psychology and measurement have highlighted the importance of individual-centered approaches; such as Latent Profile Analysis (LPA), in uncovering variability within educational practices. Rather than assuming homogeneity among teachers, this approach enables the identification of distinct of teachers who share similar patterns of beliefs, perceptions, and practices, whether those subgroups differ in their overall level of engagement, or in the structure of the dimensions. Empirical studies employing latent category analysis have demonstrated the feasibility of classifying teachers and learners into differentiated categories that vary in their cognitive, affective, and behavioral patterns (Chen et al., 2023; Kong et al., 2025; Lee, 2023; Quainoo et al., 2025). This challenges the assumptions of homogeneity underlying variable-centered analyses and contributes to a more nuanced understanding of these dimensions combine within individual teachers.

Although washback, assessment of learning and assessment for learning are conceptually distinct reflecting, respectively, the influence of tests on teaching, the formative use of assessment and its summative use, they may be practically related as teachers who engage actively in one form of assessment practice often engage in the others (Fulmer, Lee & Tan, 2015). Association of this kind at the level of variables, however, do not reveal how the three dimensions combine within individual teachers. A person- centered approach addresses this question directly by grouping teachers according to their standing on all three dimensions, making it possible to determine whether teachers are distinguished mainly by their overall level of engagement or by distinct structures in which some dimensions are emphasized more than others, a distinction that variable- centered analyses can not reveal (Spurk et al., 2020). Establishing the number, size and character of these subgroups, and identifying what predicts membership in them, therefore shows knowledge beyond that provided by the correlations among dimensions.

Within the Omani educational context, assessment practices are shaped by ongoing curriculum reforms alongside persistent test-oriented expectations, which impose diverse and overlapping demands on teachers' classroom assessment practices (Nasser, 2019). This coexistence of reform-driven, learning-oriented assessment and enduring high-stakes examination makes the Omani setting a theoretically informative case for examining how washback, assessment of learning, and assessment for learning combine in teachers' reported practices. Previous studies indicate that teachers do not constitute a homogeneous group in their assessment attitudes, knowledge, and practices, for example,

Alkharusi et al. (2014) found that teachers' assessment competence and practices have been found to vary with factors such as teaching experience, in-service training, and workload.

Based on this theoretical framework, this study adopts individual-centered approach to examine teachers' assessment practices by simultaneously considering washback, assessment for learning, and assessment of learning. The study aims at identifying latent categories of teachers' assessment practices and to investigate differences among these categories, in light of selected demographic and professional characteristics. By uncovering distinctive patterns of assessment practice, the study seeks to provide evidence-based insights that may contribute to the design of differentiated professional development programs responsive to the diversity of teachers' orientations and assessment-related needs.

Research Questions

1-What latent categories of assessment practices can be identified among teachers based on their responses to measures of washback, assessment for learning, and assessment of learning?

2-How do these latent categories differ according to teachers' gender, teaching specialization, and years of experience?

Study Objectives

This study aims at analyzing teachers' assessment practices by identifying latent categories that reflect variations in washback, assessment for learning, and assessment of learning, through the following objectives:

1-Identifying latent categories of assessment practices based on the interaction among the three dimensions.

2-Describing the characteristics of each category, in terms of the level of adoption of the three assessment dimensions.

3-Examining differences among the identified categories, according to teachers' demographic and professional characteristics (gender, teaching specialization, and years of experience).

4-Providing an interpretive framework to inform professional development programs in ways that align with the diversity of teachers' assessment practice patterns.

METHOD

Participants

The study population included all Cycle 2 (Grades 5- 10) teachers in the public schools of Muscat Governorate during the 2023l2024 academic year. It included 2823 teachers across 70 schools (Ministry of Education, 2023). The sample size was estimated using the Raosoft calculator, which applies Cochran's (1977) formula, with a 95% confidence level and a ±5% margin of error, indicating a minimum of 339 teachers (Daniel, 2012). Cluster random sampling was then used. Usable responses were obtained from 376 teachers.

The participants were 376 teachers teaching grades 5 to 10 in public schools in Muscat governorate in the Sultanate of Oman. There were 193 females (51.33%) and 183 males (48.67%). They were teaching mathematics (26.33%), science (23.67%), Arabic language (17.02%), Islamic education (11.44%), English language (11.17%), and social studies (10.37%). The teaching experience of the participants ranged from 1 to 35 years with mean of 13.76 years and a standard deviation of 8.09 years.

Instrument

A questionnaire was developed by the authors for this study. The questionnaire consisted of two parts. In the first part, the participants were requested to indicate their gender, teaching subject area, teaching grade level, and years of teaching experience. The second part consisted of 43 items divided into three sections. The first section contained 20 items assessing the degree to which teachers' instructional and assessment practices are influenced by tests (assessment washback; for example: "I use questions from previous tests to guide my instructional practices"). The second section contained 13 items assessing the degree to which teachers practice assessment for learning (e.g., I provide feedback immediately after the assessment in order to improve student learning). The third section contained 10 items assessing the degree to which teachers practice assessment of learning (e.g., I use the results of the final assessment to group students according to their achievement level). The items were developed based on previous literature (e.g., Al-Damgh, 2011; Al-Shaqsi, 2018; Bakir, 2019; Haji, 2023; Kong & Zhang, 2024). The questionnaire was developed and administered in Arabic, and an initial pool of 47 items was generated from this review.

Responses were obtained on a 5-point Likert scale ranging from 1 (very low) to 5 (very high) with a higher score reflecting a higher degree of practices. In other words, the average score for each section reflects the extent to which teachers report implementing the assessment practices measured by that section according to the following interpretation scale (Alkharusi, 2022): 1.00 – 1.80: Very low practice, 1.81 – 2.60: Low practice, 2.61 – 3.40: Moderate practice, 3.41 – 4.20: High practice, 4.21 – 5.00: Very high practice.

Content validity of the items was verified by nine specialists in educational psychology, curriculum and instruction, and educational assessment from Sultan Qaboos University and Ministry of Education in the Sultanate of Oman. They were asked to judge the clarity of wording as well as the appropriateness of each item and its relevance to the dimension being measured. The degree of expert agreement was quantified using Lawshe's (1975) content validity ratio and mean agreement was 99.5% for the washback section, 94.2% for the assessment for learning section, and 89.2% for the assessment of learning section, values that are acceptable for a panel of nine experts (Lawshe, 1975). Their feedback was used for refinement of the items. In light of this review, the initial pool of 47 items was reduced to the final 43. Internal consistency reliability was measured by Cronbach's alpha. Table 1 summarizes the score reliability of each section along with the range of the corrected item-total correlation.

Table 1
Score reliability of each dimension along with the range of the corrected item-total correlation

Dimension	Cronbach's alpha	Range of Corrected Item-Total Correlation
Assessment washback	0.92	0.40 – 0.73
Assessment for learning	0.89	0.43 – 0.70
Assessment of learning	0.86	0.46 – 0.69

Procedure

Ethical approval was obtained from Sultan Qaboos University College of Education to conduct the study. Then, permission was requested from Ministry of Education and school principals to collect data from the teachers. The participants were informed that a study is being conducted to investigate teachers' assessment practices. The teachers were also informed that they were not obligated to participate in the study, and that if they wished, their responses would remain anonymous and confidential. Those who wished to participate in the study were provided a cover letter and the questionnaire along with brief instructions about the information that was requested in the questionnaire, how to respond to the items, and where to find directions that were also included both

on the cover letter and the questionnaire. The participants took on average ten minutes to complete the questionnaire.

Data Analysis

A person-centered analytic approach was adopted to identify latent subgroups of teachers based on three continuous indicators: washback assessment, assessment for learning, and assessment of learning. Model selection was determined by a number of statistical criteria such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), sample size adjusted BIC (SABIC), entropy, and Lo-Mendell-Rubin test (LMR), Vuong-Lo- Mendell-Rubin likelihood test (VLMR), and bootstrap likelihood ratio test (BLRT) from the Likelihood Ratio Tests (LRTs). Each indicator was operationalized as the mean score across the items of the corresponding subscale, computed on the 5-point scale, with higher scores indicating greater reported engagement. The analysis was conducted in Mplus version 8.3 using robust maximum likelihood (MLR) estimation, with the three indicators entered as observed mean scores.

In a subsequent conditional model, profile membership was regressed on gender (coded 1=male, 2= female), years of teaching experience (a continuous variable), and teaching specialization. Teaching specialization was represented by five dummy variables, with mathematics (the largest group) as the reference subject. The low assessment practice profile served as the reference profile.

For the LRTs, a significant p-value ($p < .05$) indicates that the model with higher profiles fits the data significantly better than the model with fewer profiles (Nylund, Asparouhov & Muthén, 2007). Entropy values ≥ 0.80 were interpreted as evidence of high classification accuracy (Tein, Coxe & Cham, 2013), and lower AIC, BIC, and SABIC values indicated better fit (Peng et al., 2025). In addition, theoretical and practical considerations were also used to ensure the appropriateness and interpretability of the final model. Model selection emphasized parsimony, and it chose the simplest profile that was coherent and empirically distinct. Each latent profile was required to represent at least 5% of the sample, and to show average posterior classification probabilities of ≥ 0.70 (Alhadabi & Karpinski, 2020).

The adequacy of sample was confirmed. As Kline (2011), a sample of 376 teachers was sufficient, and it exceeded the recommended minimum (≥ 200) for stable model estimation and reliable fit indices.

FINDINGS

Descriptive statistics

Descriptive statistics were examined (see Table 2). First, data was screened for normality and outliers. Skewness and kurtosis values for the three variables washback assessment, assessment for learning and assessment of learning were within acceptable limits ± 3, indicating normal distributions (Kline, 2011). The outliers were checked using the z-score method. No extreme cases were detected as the values of z-scores of all data points were located between ± 2.58 (Field, 2018). Moreover, Pearson correlation coefficients between the study's variables were computed (see Table 3). The correlations among the three assessment indicators were statistically significant.

Table 2
Descriptive statistics for demographic and main study variables (N = 376)

Variable	M	SD	Min	Max	Skewness	Kurtosis
Experience (yrs)	13.76	8.09	1	35	–0.08	–0.98
Washback	3.97	0.59	2.33	5.00	–0.66	1.18
Assessment for Learning (AFL)	4.29	0.50	2.63	5.00	–0.71	0.14
Assessment of Learning (AOL)	4.15	0.54	2.46	5.00	–0.52	–0.22

Table 3
Pearson Correlation Coefficients between study variables (N = 376)

Variable	1	2	3	4	5	6
Gender	—					
Specialization	.03	—				
Experience (yrs)	–.20**	.11*	—			
Washback	–.03	–.13*	–.01	—		
Assessment for Learning (AFL)	.06	–.07	.06	.55**	—	
Assessment of Learning (AOL)	.01	–.03	.02	.52**	.70**	—

* p < .05, **p < .01

Unconditional Models Results

To answer research question 1, a sequence of unconditional LPA models with one to four profiles was estimated (Table 4). Model fit improved as profiles were added, and the four-profile model showed the lowest AIC, BIC, and SABIC values, the highest entropy, and significant Lo-Mendell-Rubin (LMR), Vuong-Lo-Mendell-Rubin (VLMR), and bootstrap likelihood ratio (BLRT) tests. However, the four-profile solution produced a very small profile containing only six teachers (1.6% of the sample), which fell below the pre-specified 5% minimum and was neither stable nor substantively interpretable. The three-profile model was therefore retained because of adequate class sizes, parsimony, and interpretability, together with its high classification accuracy (entropy = 0.803) and significant LMR, VLMR, and BLRT results. Average posterior class-membership probabilities for the three-profile solution were high (.94, .89, and .92 for the low, moderate, and high profiles, respectively), all exceeding the .70 criterion.

Table 4
Fit indices of unconditional LPA models with increasing number of profiles

Fit statistics	One-profile	Two-profile	Three-profile	Four-profile
AIC	1823.42	1532.14	1399.72	1350.48
BIC	1847.00	1571.44	1454.73	1421.21
SABIC	1827.96	1539.71	1410.31	1364.10
Entropy	—	0.740	0.803	0.852
LMR-LRT (*p*)	—	0.002	0.000	0.018
VLMR-LRT (*p*)	—	0.002	0.000	0.016
BLRT (*p*)	—	0.000	0.000	0.000
Profile size (% / n)	P1 = 100 % (376)	P1 = 35.9 % (135) P2 = 64.1 % (241)	P1 = 12.0 % (45) P2 = 46.5 % (175) P3 = 41.5 % (156)	P1 = 1.6 % (6) P2 = 13.3 % (50) P3 = 39.4 % (148) P4 = 45.7 % (172)

The final model included three distinct teacher groups:

- Profile 1 (Low assessment practice, 12.0%): teachers with the lowest means across all assessment indicators (Washback assessment = 3.35; AFL = 3.39; AOL = 3.28).

- Profile 2 (Moderate assessment practice, 46.5%): teachers with moderate levels of engagement (Washback assessment = 3.75; AFL = 4.17; AOL = 4.01).

- Profile 3 (High assessment practice, 41.5%): teachers with the highest scores across all assessment indicators (Washback assessment = 4.39; AFL = 4.69; AOL = 4.57).

Overall, the three-profile solution represents clear distinctions in intensity rather than patterns of assessment engagement, with all variables showing consistent increases across profiles.

The unstandardized parameters of the unconditional three-profile model are presented in Table 5. Teachers in High Assessment Practice Profile

(Profile 3) exhibited the highest mean score on all three indicators, followed by the Moderate Assessment Practice (Profile 2) and Low Assessment Practice (Profile 1) groups. All differences were statistically significant ($p < .001$).

Table 5
Unstandardized parameters (means and standard errors) of the unconditional three-profile model

Study Variable	Profile 1: Low Assessment Practice		Profile 2: Moderate Assessment Practice		Profile 3: High Assessment Practice	
	M	SE	M	SE	M	SE
Washback	3.354***	0.060	3.748***	0.049	4.387***	0.047
AFL	3.385***	0.076	4.166***	0.035	4.690***	0.033
AOL	3.284***	0.058	4.007***	0.048	4.570***	0.039

Conditional Latent Profile Analysis Results

Regarding research question 2, and to examine whether teachers' demographic characteristics predicted membership in the three profiles, gender, teaching specialization, and years of experience were entered as covariates in a conditional LPA. The low Assessment Profile (Profile 1) served as the reference group.

As shown in Table 6, teaching specialization predicted membership in the high assessment practice profile. Relative to teachers of Mathematics (the reference group), teachers of Arabic language (OR = 3.68, $p = .020$) and Science (OR = 3.86, $p = .030$) were significantly more likely to belong to the high profile than to the low profile. The remaining subjects were not statistically significant, and no covariate significantly distinguished the moderate profile from the low profile.

Years of teaching experience showed a positive but non-significant association with profile membership; more experienced teachers tended to belong to the moderate (B = 0.044, $p = .114$) or high (B = 0.050, $p = .060$) profile rather than the low profile, but this result did not reach statistical significance. Gender also did not significantly predict membership ($p > .05$), which indicates that male and female teachers were similarly distributed across the profiles. Given the cross-sectional design, these results should be interpreted as associations with the likelihood of profile membership rather than as causal effects.

Table 6
Multinomial logistic regression of profile membership on covariates, relative to the low assessment practice profile (Profile 1); for teaching specialization, Mathematics is the reference subject

Predictors	profile2 (Moderate Assessment Practice)			Profile 3 (High Assessment Practice)		
	B regression coefficient (SE)	Odds Ratio	P value	B regression coefficient (SE)	Odds Ratio	P value
Gender	0.280 (0.376)	1.32 [	.457	0.348 (0.366)	1.42	.342
Experience (years)	0.044 (0.028)	1.05	.114	0.050 (0.026)	1.05	.060
Islamic Studies	0.273 (0.612)	1.31	.656	0.836 (0.574)	2.31	.145
Arabic Language	0.374 (0.600)	1.45 [0.45, 4.71]	.533	1.303 (0.559)	3.68	.020*
English Language	1.547 (0.876)	4.70	.077	0.847 (0.880)	2.33	.336
Social Studies	0.434 (0.662)	1.54	.512	0.666 (0.660)	1.95	.313
Science	1.192 (0.637)	3.29	.061	1.351 (0.623)	3.86	.030*

*$p < .05$.

DISCUSSION, AND CONCLUSION

This study aimed to identify latent profiles of teachers' assessment practices across assessment washback, assessment for learning, and assessment of learning, and to examine whether gender, teaching specialization, and teaching experience were associated with profile membership. Three profiles emerged, differing primarily in the overall level of engagement with assessment practices, low moderate, and high, rather than in qualitative distinct configurations of the three dimensions. These findings have both theoretical and practical implications for understanding how teachers' assessment practices vary and how professional development might be differentiated.

The three profiles findings align with prior research that highlights variability in teachers' engagement practices. The study also sought to examine the influence of demographic variables in terms of gender, teaching subject area, and teaching experience on profile membership. The findings indicated three distinct profiles characterized as low, moderate, and high in terms of overall engagement with assessment practices. These findings offer both theoretical and practical implications for assessment literacy, teacher experience, and institutional support.

The emergence of three profiles aligns with prior research highlighting variability in teachers' engagement in assessment practices (Looney et al., 2018; Xu & Brown, 2016). Teacher in high profiles reported consistently high engagement across washback, assessment for learning, and assessment of learning. This profile suggests an effective integration of assessment throughout the entire teaching and learning process. Further, this accountability oriented practices reflects assessment as both learning-oriented and accountability-oriented. In other words, teachers in this profile tend to use assessment to improve student learning and, that they hold themselves accountable for the results of the assessment. This resonates with the argument raised by Black and Wiliam (1998) that assessment for learning can increase the standards when practiced in an accountable and effective integrated way during the instructional process. Empirical work likewise shows that well-designed formative feedback can improve learning outcomes (Cobbold & Wright, 2021).

Conversely, the low profile represents teachers whose implementation of assessment practices remains minimal or superficial. This may reflect limited knowledge of assessment, constrained institutional support or heavy pressure from summative test demands. For example, studies in EFL and high-stakes contexts have found that when teachers lack assessment literacy, the washback of tests tends to be negative, narrowing teaching and learning to test preparation (Rahman et al., 2021). Also, inadequate resources or restrictive assessment policies might contribute to minimal engagement

in assessment (Banta & Blaich, 2010). Furthermore, the pervasive pressure from summative test demands can lead teachers to focus primarily on test preparation rather than pedagogical goals (Gebril & Eid, 2017).

The moderate group represents a transition zone: teachers who engage in some assessment practices but inconsistently or less deeply. This variability suggests that the adoption of reflective and learning-oriented assessment strategies is not the same across the teaching population. Thus, the uneven engagement in assessment washback, assessment for learning, and assessment of learning practices of teachers in this category suggests a need for targeted interventions to enhance teacher assessment literacy. Research on teachers' conceptions and practices of classroom assessment supports this view. For example, Takele & Melese (2022) found that teachers employed mixed and different assessment types, focusing more on assessment of learning, which suggests a moderate and selective engagement with various assessment practices. Similarly, Kyaruzi et al. (2018) reported that teachers used assessment information for both formative and summative purposes, reflecting moderate engagement with assessment practices.

With respect to the covariates, years of teaching experience showed a positive but non-significant association with profile membership where more experienced teachers tended to belong to the moderate or high profile rather than the low profile, but this result did not reach significance once teaching specialization was taken into account. Although prior research suggests that experience can contribute to the development of assessment practices, particularly when accompanied by continuous professional learning (Amirian, 2025; Brown, Barnes, & Finefter-Rosenbluh, 2024), the present results indicate that, in this sample, experience alone did not significantly distinguish teachers' profiles.

Gender was not significantly associated with profile membership, which indicates that male and female teachers were similarly distributed across the profiles. Teaching specialization, however, was correlated with membership; teachers of Arabic language and Science were significantly more likely than teachers of Mathematics to belong to the high profile. This suggests that the subject a teacher teaches may be related to their reported assessment practices. These differences might reflect subject-specific assessment cultures, curricular demands, or professional development opportunities. This result differs from some previous reports of no specialization effect (Nazim, Alzubi, & Fakih, 2024).

In sum, the current study underscores that teachers' assessment practices are not the same but cluster into distinct profiles of engagement ranging from low to high across washback, assessment for learning and assessment of learning dimensions and this engagement was associated with teaching specialization more than with gender or experience. From a practical standpoint, these results suggest that professional support could be differentiated according to teachers' current profile of practice, rather than one-size-fits-all training. Investing in assessment literacy and promoting learning-oriented assessment may help support teachers' engagement.

Finally, several limitations warrant acknowledgement. First, as with any cross-sectional design, the associations reported here cannot support causal inference, and the profiles describe engagement levels at one time point rather than developmental stages. Longitudinal designs would help track how teachers move across profiles over time. Second, all measures were based on teachers' self-reports, which may not correspond fully to observed classroom practice. Third, the sample context (e.g., Muscat educational governorate) may limit generalizability to other educational governorates or cultural contexts. Fourth, while key demographic variables were included, other contextual variables (such as school culture, leadership support, assessment policy, class size) may further explain profile membership and deserve future exploration. Future research could employ longitudinal designs such as latent transition analysis to model how teachers move across profiles over time; incorporate classroom observations and qualitative interviews to examine how teachers in different profiles enact

assessment. It could also further validate the assessment-practice instrument, including its factor structure. Moreover, it could examine contextual variables such as school leadership and assessment policy, investigate the subject-specific differences observed here, and link assessment-practice profiles to student outcomes.

REFERENCES

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., and Wall, D. (1993). Does washback exist? Appl. *Linguist. 14,* 115–129. https://doi.org/10.1093/applin/14.2.115

Alhadabi, A., & Karpinski, A. C. (2020). Grit, self-efficacy, achievement orientation goals, and academic performance in University students. *International Journal of Adolescence and Youth*, *25*(1), 519–535. https://doi.org/10.1080/02673843.2019.1679202

Alkharusi, H. (2022). A descriptive analysis and interpretation of data from Likert scales in educational and psychological research. *Indian Journal of Psychology and Education, 12*(2), 13-16.

Alkharusi, H. (2021). English language teachers' uses of classroom assessment. *Anatolian Journal of Education, 6*(1), 103–108. https://doi.org/10.29333/aje.2021.618a

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2014). Educational assessment profile of teachers in the sultanate of oman. *International Education Studies*, *7*(5), p116. https://doi.org/10.5539/ies.v7n5p116

Allan, G. (2020). Tests as drivers of change in education: contextualizing washback, and the possibility of wash-forward. VNU Journal of Foreign Studies, 36(4).

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

Amirian, S. M. R. (2025). Understanding EFL teachers' formative assessment literacy: Insights from a mixed-methods study. *Language Testing in Asia*, *15*(1), 25. https://doi.org/10.1186/s40468-025-00363-y

Banta, T. W., & Blaich, C. (2010). Closing the assessment loop. *Change: The Magazine of Higher Learning*, *43*(1), 22–27. https://doi.org/10.1080/00091383.2011.538642

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability(Formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Brookhart, S. M. (2024). Educational Assessment Knowledge and Skills for Teachers Revisited. Education Sciences, 14(7), 751. https://doi.org/10.3390/educsci14070751

Brown, T. D., Barnes, M., & Finefter-Rosenbluh, I. (2024). Teacher perspectives and experiences of assessment literacy in Victorian junior secondary schools. *Australian Journal of Education*, *68*(1), 5–22. https://doi.org/10.1177/00049441231214022

Chen, J., Lin, C.-H., & Chen, G. (2023). Adolescents' self-regulated and affective learning, teacher support and digital reading literacy: A multilevel latent profile approach. *Computers & Education*, *205*. https://doi.org/10.1016/j.compedu.2023.104883

Cobbold, C., & Wright, L. (2021). Use of formative feedback to enhance summative performance. *Anatolian Journal of Education, 6*(1), 109–116. https://doi.org/10.29333/aje.2021.619a

Daniel, J. (2012). *Sampling essentials: Practical guidelines for making sampling choices*. SAGE Publications, Inc. https://doi.org/10.4135/9781452272047

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

Eric Atta Quainoo, Daniel Asamoah, David Baidoo-Anu, Juyan Ye, & Clara Akuamoah Boateng. (2025). Examining profiles of classroom assessment literacy among basic school teachers. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2558163

Field, A. (2018). Discovering statistics using IBM SPSS Statistics (5th ed.). SAGE Publications.

Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice*, *22*(4), 475–494. https://doi.org/10.1080/0969594X.2015.1017445

Gebril, A., & Eid, M. (2017). Test preparation beliefs and practices in a high-stakes context: A teacher's perspective. *Language Assessment Quarterly*, *14*(4), 360–379. https://doi.org/10.1080/15434303.2017.1353607

Haji, K. bint 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.

Heritage, M., & Harrison, C. (2019). *The power of assessment for learning: Twenty years of impact in UK & US classrooms*. Corwin.

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.

Kline R. (2023).Principles and practice of structural equation modeling (5th ed). Guilford Press.

Kong, S. C., Lai, M., Li, Y., Dickson Chan, T.-Y., & Travess Zhang, Y. (2025). A latent profile analysis of teachers' knowledge about and perceived usefulness of computational thinking and how teacher profiles relate to student achievement. *Computers & Education*, *232*. https://doi.org/10.1016/j.compedu.2025.105281

Kyaruzi, F., Strijbos, J.-W., Ufer, S., & Brown, G. T. L. (2018). Teacher AfL perceptions and feedback practices in mathematics education among secondary schools in Tanzania. *Studies in Educational Evaluation*, *59*, 1–9. https://doi.org/10.1016/j.stueduc.2018.01.004

Lawshe, C. H. (1975). A quantitative approach to content validity[1]. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lee, Jihyun. (2023). Latent class profiles of teacher use of digital tools in PISA 2018 data. *Education and Information Technologies*, 1–33. https://doi.org/10.1007/s10639-023-12305-3

Looney, A., Cumming, J., Van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 442–467. https://doi.org/10.1080/0969594X.2016.1268090

Magaji, A. (2025). Exploring trainee science teachers' assessment as learning (AaL) and assessment for learning (AfL) practices in a teacher training program. *International Journal of Instruction, 18*(1), 111–130. https://doi.org/10.29333/iji.2025.1817a

Nasser, R. (2019). Educational reform in oman: System and structural changes. In *Education Systems Around the World*. IntechOpen. https://doi.org/10.5772/intechopen.84913

Nazim, M., Alzubi, A. A. F., & Fakih, A.-H. (2024). Teachers' attitudes towards student-centered pedagogy and assessment practices: Instruction efficacy in perspective. *International Journal of Evaluation and Research in Education (IJERE)*, *13*(2), 1276–1286. https://doi.org/10.11591/ijere.v13i2.28139

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.1217167

Peng, N., Sun, Q., Li, Z., Xia, X., & Li, Y. (2025). The relationship between problematic mobile phone use and meaning in life among adolescents: A latent profile analyses. *BMC Psychology*, *13*(1), 127. https://doi.org/10.1186/s40359-025-02470-0

Popham, W. J. (2006). All about accountability / Needed: A dose of assessment literacy. *Educational Leadership, 63*(6), 82-83.

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

Rahman, K. A., Seraj, P. M. I., Hasan, M. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia, 11*, 12. https://doi.org/10.1186/s40468-021-00129-2

Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, *9*(1), 5–29. https://doi.org/10.1191/1362168805lr152oa

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*, 103445. https://doi.org/10.1016/j.jvb.2020.103445

Takele, M., & Melese, W. (2022). Primary school teachers' conceptions and practices of assessment and their Relationships. *Cogent Education*, *9*(1). https://doi-org.squ.idm.oclc.org/10.1080/2331186X.2022.2090185

Taylor, L. (2005). Washback and impact. ELT Journal, 59(2), 154-155.

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(4), 640–657. https://doi.org/10.1080/10705511.2013.824781

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010