

Please cite this paper as follows:

Sobouti, F., Gharagozloo, N., & Rahimi, A. H. (2025). Investigating ESP Instructors' Knowledge of Assessment Literacy Components and Instructional Practices in the Iranian Academic Context. *International Journal of Foreign Language Teaching and Research*, 13 (53), 11-34.

Research Paper

Investigating ESP Instructors' Knowledge of Assessment Literacy Components and Instructional Practices in the Iranian Academic Context

Fatemeh Sobouti¹, Neda Gharagozloo^{2*}, Amir Hossein Rahimi³

¹Ph.D. Candidate, Department of English Translation, Varamin- Pishva Branch, Islamic Azad University, Varamin, Iran

fatemeh.sobouti@iasbs.ac.ir

²Assistant Professor, Department of English Translation, Varamin- Pishva Branch, Islamic Azad University, Varamin, Iran

Neda.Gharagozloo@iau.ac.ir

³Assistant Professor, Department of English Language, Yadegar-e-Imam Khomeini (RAH) Shahre Rey Branch, Islamic Azad University, Tehran, Iran
rahimi.amirh@gmail.com

Received: November 14, 2024

Accepted: December 16, 2024

Abstract

In the Iranian academic setting, the current investigation aimed to identify the assessment literacy (AL) components that experienced and novice English for Specific Purposes (ESP) teachers with and without TEFL (teaching English as a foreign language) backgrounds comprehended. We designed an explanatory sequential mixed-methods study, and selected 100 PhD male and female ESP educators from various branches of Islamic Azad University via criterion sampling in order to complete the teacher assessment literacy scale (TALS). The standard deviation as well as mean of the TALS were computed. The mean scores of the TEFL/non-TEFL and novice/experienced teachers on the seven TALS components were then independently compared using two multivariate analyses of variances (MANOVA). After administering the scale to 100 ESP teachers and analyzing the results, 20 of the instructors were chosen via convenience sampling for observation during the study's qualitative phase. Afterwards, the qualitative information was collected through the use of the observation notes. Employing axial/open coding and content analysis of themes, all of the notes were classified and examined. The results showed that in terms of knowledge of assessment literacy components, instructors with TEFL backgrounds understand much better than teachers without TEFL backgrounds. Similarly, the means of all TALS components were greater for inexperienced teachers than for experienced ones. Furthermore, the results of the qualitative phase demonstrated that the novice TEFL teachers performed AL components more effectively than their non-TEFL counterparts as well as the experienced TEFL/non-TEFL instructors.

Keywords: Assessment literacy, Components of assessment literacy, Novice and experienced teachers, TEFL/non-TEFL background

بررسی میزان آگاهی مربیان ESP از مؤلفه‌های سواد ارزشیابی و شیوه‌های آموزشی در زمینه تحصیلی ایران در محیط دانشگاهی ایران، پژوهش حاضر با هدف شناسایی مؤلفه‌های سواد ارزیابی (AL) که معلمان انگلیسی با تجربه و مبتدی برای مقاصد خاص (ESP) با و بدون پیش‌زمینه TEFL (تدریس انگلیسی به عنوان یک زبان خارجی) آن را درک می‌کنند، انجام شد. ما یک مطالعه ترکیبی متوالی توضیحی را طراحی کردیم و ۱۰۰ نفر از مربیان دکتری ESP مرد و زن از واحدهای مختلف دانشگاه آزاد اسلامی را از طریق نمونه‌گیری معیاری به منظور تکمیل مقیاس سواد ارزیابی معلم (TALS) انتخاب کردیم. انحراف معیار و همچنین میانگین TALS محاسبه شد. سپس میانگین نمرات معلمان TEFL غیر TEFL و تازه کار/با تجربه در هفت مؤلفه TALS به طور مستقل با استفاده از دو تحلیل واریانس چند متغیره (MANOVA) مقایسه شد. پس از اجرای مقیاس بر روی ۱۰۰ معلم ESP و تجزیه و تحلیل نتایج، ۲۰ نفر از مربیان به روش نمونه‌گیری در دسترس برای مشاهده در مرحله کیفی مطالعه انتخاب شدند. سپس با استفاده از یادداشت‌های مشاهده‌ای، اطلاعات کیفی جمع‌آوری شد. با استفاده از کدگذاری محوری/باز و تحلیل محتوای مضامین، تمامی یادداشت‌ها طبقه‌بندی و بررسی شدند. نتایج نشان داد که از نظر آگاهی از مؤلفه‌های سواد ارزشیابی، مربیان با پیشینه TEFL بسیار بهتر از معلمان بدون پیش‌زمینه TEFL درک می‌کنند. به طور مشابه، میانگین تمام اجزای TALS برای معلمان بی‌تجربه بیشتر از معلمان با تجربه بود. علاوه بر این، نتایج مرحله کیفی نشان داد که معلمان مبتدی TEFL مؤلفه‌های AL را مؤثرتر از هم‌تایان غیر TEFL خود و همچنین مربیان با تجربه / TEFL غیر TEFL انجام می‌دهند.

واژه‌های کلیدی: سواد ارزشیابی، مؤلفه‌های سواد ارزشیابی، معلمان مبتدی و باتجربه، پیشینه تفل/غیر تفل

Introduction

Assessment literacy is critical to promote student achievement, student learning, and teacher instruction (Deluca & Klinger, 2010; Zhang, et al., 2021). Assessment literacy was firstly originated from Stiggins' (1991) work. He believes that educators who are proficient in assessment literacy understand the why, how, and what of assessment, and how to avoid the possible problems in assessing the learners. Assessment literacy also helps teachers to understand the negative consequences of inaccurate and poor assessment (Stiggin, 1991). Furthermore, assessment literacy is claimed to be the key to efficient teaching (Popham, 2014; Yan & Fan, 2020).

The importance of this study is in its pedagogical benefits for EFL (English as a foreign language) instructors and learners and the theoretical contribution to the second language research. Besides, the results of this research are thought to offer a thorough grasp of the nature of assessment and how it relates to TEFL/non-TEFL backgrounds. The empirical findings of the current research might affect language instructors, language educators, and materials developers. Besides, ESP instructors might use the findings to increase their knowledge of assessment and create an atmosphere of cooperation and coordination in their classrooms through employing assessment.

The past literature has demonstrated that Iranian EFL teachers have inadequate knowledge of assessment literacy (Ashraf & Zolfaghari, 2018; Farhady & Tavassoli, 2018; Mellati & Khademi, 2018; Razavipour et al., 2011). For example, Razavipour et al. (2011) who studied the interaction of test wash back and AL among Iranian EFL teachers proposed that "despite having limited assessment knowledge, instructors nonetheless adapt their English instruction and evaluation to meet the requirements of external exams" (p.156).

Therefore, by comparing the assessment literacy components known by ESP teachers with and without TEFL backgrounds, the present research project was designed to look at some previously unexplored areas of teacher assessment literacy. As far as the researchers are aware, very few Iranian studies have ever looked into whether ESP instructors' experience and educational backgrounds have any effect on their knowledge of assessment literacy components. Thus, in order to bridge this gap in the body of literature, the researchers conducted the present study. Therefore, the following questions are posed:

Q1. Is there any statistically significant difference between ESP instructors with TEFL and non-TEFL backgrounds in terms of their knowledge of assessment literacy components?

Q2. Is there any statistically significant difference between novice and experienced ESP instructors concerning their knowledge of assessment literacy components?

Q3. How do novice and experienced ESP instructors with TEFL and non-TEFL backgrounds display their knowledge of assessment literacy components in practice?

Literature Review

The language assessment literacy (LAL) is regarded as a necessary prerequisite for EFL teachers (Ashraf & Zolfaghari, 2018; Zamani & Ahangari, 2016). Deficiency in LAL, as Stobart (2008) argued, makes difficulties for the EFL instructors when they schedule their lessons. Similarly, the teachers who do not engage in assessments are usually a little severe and create a culture of competing rather than collaborating in their classrooms; something which surely leaves negative effects on the students' learning and second language development (Ellis, 2008).

The qualities of an effective English language instructor in the Iranian setting were studied by Babai Shishavan and Sadeghi (2009). They argued that assessment literacy can be considered as one of the significant features of EFL teachers. Likewise, Zamani and Ahangari (2016) accounted teacher assessment literacy (TAL) as a major teacher eligibility in the EFL domain.



Moreover, to raise student accomplishment, it is essential to comprehend and implement effective classroom assessments (Marzano, 2000). Bayat and Rezaei (2015) also asserted that “since the quality of applied assessment is directly related to the quality of instruction, one of a teacher's most significant duties is to assess pupils” (p. 139).

It is worth mentioning that, LAL has been measured through comparing different instructors' AL, for example, novice and experienced teachers (Tajeddin et al., 2018). Furthermore, Azadi (2018) examined the notional elements of TAL among instructors of ESP in Iran. Moreover, Mohammadi (2020) contrasted AL knowledge of ESP instructors with and without TEFL backgrounds.

The importance of second language assessment literacy is emphasized by Falsgraf (2005). He asserted that “it assists educators in understanding, evaluating, and using data of students' performance to improve instruction” (p. 6). Besides, in order to accomplish their learning goals, teachers can use the most reliable as well as effective tools when they are assessment literate. (Siegel & Wissehr, 2011). Ashraf and Zolfaghari (2018) believe that teachers' assessment literacy can be considered as an important link, which makes connection between student achievement and assessment quality. For this reason, in order for instructors to be effective, they need to possess the proper level of assessment literacy. Mellati and Khademi (2018) insist that language educators consider teachers' assessment awareness in their teacher education programs. Moreover, AL is necessary for EFL teachers to identify the issues in their classrooms and improve in their work by implementing changes. (Hajizadeh & Salahshour, 2014; Scarino, 2013).

Farhady and Tavassoli (2018) stated that current improvements in the field of education force teachers “to be informed of and utilize efficient teaching and evaluation techniques to enhance learning” (p. 45). Taylor (2013) signified that both ESP and EFL instructors should know different types of assessment to be more exact in the process of instruction as the mediators of teaching-learning process.

Assessment literacy knowledge also helps the ESP instructors to convey to test stakeholders the fundamentals, theories, and procedures of language testing while the ESP testing is regarded in a large scale (Popham, 2014). However, as Arani et al. (2012) argued, testing is more absorbing than assessment in the Iranian educational context. Malone (2013) suggested that the agreement needed between testers and users contains the fundamentals of assessment literacy, presenting the crucial elements of AL for the ESP educators. Then, he highlighted how ESP instructors with more expertise in AL may more effectively fill in the gaps and make analogies between the users and the testers.

Theoretical Framework

Mertler (2009) asserts that assessing how well pupils perform is one of the most crucial responsibilities of instructors because it has a big impact on everything they accomplish. Spolsky (1995, as cited in Jafarpour, 2003, p. 59) stated that “if a teacher is qualified to instruct in a language, then that person is qualified to assess the pupils as well.”. Instructors' appropriate level of assessment knowledge is also emphasized in order for appropriate evaluation of students (Inbar-Lourie, 2013). Theoretically speaking, instructors should be master in assessment notions such as developing, administering, and reporting test (Bachman & Palmer, 2010). Modern ideas about assessment recommend that EFL teachers should get mastery over the assessment literacy as a factor of their professional improvement (Atay, 2008).

The present study theoretically is based on the coalescence of notions such as agentic theory (Kögler, 2012), reasoned action theory (Yzer, 2013), planned theory of behaviors (Ajzen, 2020), and social cognitive theory (Bandura, 2005).

Agentic theory of Kögler (2012) declares that individual teachers are powerful agencies affecting their learners while dealing with their own performances. Indeed, instructors are



responsible for the effectiveness and progress of students in relation to assessing of learning and assessing for learning. Thus, in EFL classrooms, teachers are essential in introducing assessment concepts and improving assessment quality.

Reasoned action proposed by Yzer (2013) argues that teachers' individual viewpoints and attitudes on second language instruction, acquisition, and assessment could affect classroom practices so that it can cause success or failure in classroom assessment. It's also important to note that Sussman and Gifford (2019) described instructors' ideas as information derived from their daily environment. In this regard, teachers' perspectives towards education and learning, their previous educational backgrounds, and in-service training projects would affect such attitudes (Ramnarain & Hlatwayo, 2018). Furthermore, according to Dasgupta (2013), teachers' attitudes and views have a big impact on how they do L2 assessments. According to Ajzen's (2020) theory of planned behavior, teachers' actions within the classroom serve as a further means of demonstrating TAL, in addition to their knowledge, beliefs, and strategies. Social cognitive theory of Bandura (2005) deals with the outcomes of actions. In this regard, the most critical part of this theory is self-efficacy. Indeed, people assess their actions regarding the cognitive, affective, decisional, and motivational processes (Gotch & French, 2013). Therefore, what makes the instructors alter their assessment techniques or not is the notion of self-efficacy. Similarly, instructors' prosperity in the procedure of assessing students' performances is controlled by their strategies, world-views, beliefs, and self-efficacy (Bandura, 2005). Thus, regarding social cognitive theory, different aspects of social and cognitive issues would affect assessment. In other words, assessment is not restricted to only assessing the students' achievement. Therefore, in order to have a thriving assessment process, various aspects of teachers and learners' behaviors, different experiences of instructors and students regarding assessment, its outcomes on students' improvement, and its impact on students' life should be taken into consideration (Bandura, 2005). Accordingly, teachers' expertise over theoretical and operational notions of assessment is of prime importance (Zwozdiak-Myers, 2012).

Methods

Participants

In the current study, by using criterion sampling, 100 ESP teachers were chosen for the current study. The criteria included the educational backgrounds of the instructors (only PhD TEFL/non-TEFL) and their level of experience as teachers (novice/experienced). The quantity of participants satisfies the criteria established by Krejcie and Morgan (1970) for calculating sample size in research endeavors. Actually, there were 137 (N=137) ESP teachers working at various Azad university branches in the provinces of Zanjan and Tehran throughout the study's conduct. According to Krejcie and Morgan's sample size table, there should have been at least 100 participants (n=100). The ages of the participants ranged from 30 to 50. Additionally, convenience sampling was used to pick 20 individuals (10 with TEFL backgrounds and 10 without) for a follow-up classroom observation. Ethics was considered in selecting the participants through making them informed of the purposes and ensured of anonymity and confidentiality of data.

Instrumentation

To obtain more precise findings, the following three tools were used in accordance with the theoretical underpinnings of TAL:

- The Experienced and Novice Teacher Questionnaire,
- The Teacher Assessment Literacy Scale, and
- The observation notes.



The Novice and Experienced Teacher Questionnaire was created by Rodríguez and McKay (2010). Its revised version regarding the local and cultural notions, which had been already applied in the Iranian context (Baniali, 2018; Eezami, 2016), was the first instrument in the present study (Appendix A). This questionnaire is able to rather indicate the teachers' level of experience. It includes 12 questions on a five-point Likert scale (little, a little, to some extent, much, and very much). The original version of the questionnaire indicates the Cronbach's alpha reliability index of .72. By using factor analysis, Rodríguez and McKay (2010) confirmed the questionnaire's construct validity. Studies by Eezami (2016) and Baniali (2018) have revealed $\alpha=0.76$ and $\alpha=0.71$ respectively for the reliability of the modified Iranian version of this scale. 12 was the lowest possible score, while 60 was the highest. According to Rodriguez and McKay (2010), the cut score ranged from 30 to 36. The statement suggests that teachers classified as novices if their score was less than 30, and experienced instructors were identified if their score was higher than 36. The teachers with scores between 30 and 36 were excluded in order to define the precise scoring method (Rodríguez & McKay, 2010, p. 3).

The second tool (Appendix B) was the Iranian variant of the Teacher Assessment Literacy Scale (Azadi, 2018) which Mertler (2009) created its original version. There are two sections to the scale. The 35 items in the first section address each of the seven components that teachers must be aware of and use when assessing students' language proficiency. The scale includes items that examine general assessment concepts, such as how various activities are applied to inform learners of their assessment outcomes. The other questions check knowledge of classroom assessment and standardized testing. The following seven standards are suggested for instructors to comprehend and be able to perform (Mertler, 2009, as cited in Sobouti et al. 2023):

1. Choosing Appropriate Assessment Methods
2. Developing Appropriate Assessment Methods
3. Administering, Scoring, and Interpreting the Results of Assessments
4. Using Assessment Results to Make Decisions
5. Developing Valid Grading Procedures
6. Communicating Assessment Results
7. Recognizing Unethical or Illegal Practices

The second part comprises of questions regarding instructors' backgrounds as classroom teachers. In accordance with Cronbach's alpha ($\alpha=0.73$), "the Persian variant of the scale has a dependability value of 0.73" (P. 63), Azadi (2018) asserted. He also said that "expert judgment validity has verified the validity of the teacher's scale" (p. 68).

Table 1

Statistics of Reliability; Assessment Literacy and its Elements

	Cronbach's Alpha	N of Items
Selecting Appropriate Assessment Methods	0.73	5
Developing Appropriate Assessment Methods	0.80	5
Administering, Scoring, Interpreting Results	0.74	5
Using Assessment Results to Make Decisions	0.81	5
Developing Valid Grading Procedure	0.76	5
Communicating Assessment Results	0.78	5
Recognizing Illegal or Unethical Practices	0.74	5
Total	0.76	35

The third instrument was classroom observation notes taken by the researchers. In order to confirm the findings of the quantitative part of the study, the researchers applied observation. The

recorded sessions were reviewed by five specialists, and the notes made during in-class observations were compared to the information obtained from the audiotapes. The degree of agreement between the audiotape review and the notes taken throughout the observation sessions proved to be a confirmation of the reliability of the observation notes.

Design

A mixed-methods sequential explanatory design incorporating qualitative as well as quantitative approaches was used in the present investigation (Creswell & Clark, 2017). Sequential explanatory design entails gathering and analyzing quantitative data in the first phase, then gathering and analyzing qualitative data in the second phase depending on the first phase's outcomes (Creswell, 2009). In addition, the results of the quantitative part are used to identify the participants for the follow-up observation (Creswell, 2009). Furthermore, the incorporation of both research methods can counterbalance their shortcomings. According to Mackey and Gass (2016), "the benefit of triangulation lies in the fact that it promotes the validity and reliability of the data and minimizes interviewer or observer bias" (p. 182). In the current investigation, triangulation was used to analyze the convergence findings from several ways to obtain richer data. The observation notes and scale were employed in this process.

Procedure

At first, using criterion sampling, 100 PhD teachers, both TEFL and non-TEFL, who were teaching ESP at various Islamic Azad University branches in the provinces of Zanzan and Tehran were selected. Following that, the participants were given the Persian version of the Novice and Experienced Teacher Questionnaire (Baniali, 2018) and the Teacher Assessment Literacy Scale (Azadi, 2018). The survey was to be completed by the teachers outside of class, sealed, and returned to the researchers within a week after being received. The participants' identities and replies would be kept confidential, the researchers assured. The quantitative data which was obtained from scales and questionnaires was analyzed by SPSS version 25.

Of the 100 ESP teachers who participated in the study's quantitative phase, 20 individuals (20 percent of the total ESP teachers; in line with Bachman & Palmer, 2010) were selected through convenience sampling for observation. In the next step, following the approval and cooperation of the authorities and participants, the researchers divided the 20 ESP teachers into the following four categories. (Sobouti et al., 2023):

- five experienced TEFL teachers,
- five experienced non-TEFL teachers,
- five novice TEFL teachers, and
- five novice non-TEFL teachers.

In the end, 20 classes belonged to the 20 ESP teachers were observed by the researchers. Indeed, the researchers observed the instructors' assessment abilities in realistic situations with respect to the study's focus. In order to make the observation notes more reliable and to show more accurate view of how things are going in the instructors' assessment procedures, each class was observed three 90-minute sessions (Dörnyei, 2007).

Data Analysis

In the current investigation, data was gathered and examined in both quantitative and qualitative ways. To compare the means of TEFL/non-TEFL and novice/experienced ESP educators' knowledge on the seven standards of teacher assessment literacy, during the study's quantitative phase, a multivariate analysis of variances (MANOVA) was carried out.

In addition to the quantitative analysis of data, the information obtained from the observation notes was first classified into groups, open codes, and axial codes prior to being analyzed qualitatively.

Results

Q1: Is there any statistically significant difference between ESP instructors with TEFL and non-TEFL backgrounds in terms of their knowledge of assessment literacy components?

A multivariate analysis of variances (MANOVA) was performed to probe the first quantitative research question.

Table 2

Testing Normality of Data; Elements of Assessment Literacy Knowledge by Background

		N	Skewness		Kurtosis			
			Statistic	Std. Error	Ratio	Statistic	Std. Error	Ratio
TEFL	Choosing Methods	50	-0.09	.337	-0.27	-0.567	0.662	-0.86
	Developing Methods	50	-0.23	.337	-0.70	-1.066	0.662	-1.61
	Administration	50	0.06	.337	0.20	-0.943	0.662	-1.42
	Make Decisions	50	-0.38	.337	-1.14	-0.778	0.662	-1.18
	Grading Procedure	50	-0.48	.337	-1.45	0.790	0.662	1.19
	Communication	50	-0.16	.337	-0.49	-0.449	0.662	-0.68
	Recognizing Unethical	50	-0.54	.337	-1.61	-0.614	0.662	-0.93
Non-TEFL	Choosing Methods	50	0.14	.337	0.43	-0.529	0.662	-0.80
	Developing Methods	50	-0.286	.337	-0.85	-1.160	0.662	-1.75
	Administration	50	-0.281	.337	-0.83	-0.279	0.662	-0.42
	Make Decisions	50	-0.247	.337	-0.73	-0.270	0.662	-0.41
	Grading Procedure	50	-0.363	.337	-1.08	0.029	0.662	0.04
	Communication	50	-0.467	.337	-1.39	-0.644	0.662	-0.97
	Recognizing Unethical	50	-0.233	.337	-0.69	-0.288	0.662	-0.44

First, as shown in Table 2, the ratios of skewness and kurtosis over their standard errors were less than +/- 1.96, indicating that the assumption of normality was maintained.

Table 3

Box's Test of Equality of Covariance Matrices; Components of Assessment Literacy Knowledge by Background

Box's M	79.984
F	2.639
df1	28
df2	33465.844
Sig.	.000

Next, as displayed in Table 3, the presumption of homogeneity of covariance was not maintained (Box's $M = 79.98$, $p = .000$). As noted by Field (2018, p. 885), "Once sample sizes are identical, this test can be disregarded since particular MANOVA test statistics are resistant to breaches of this assumption".



Table 4

Levene's Test of Equality of Error Variances; Components of Assessment Literacy Knowledge by Background

		Levene Statistic	df1	df2	Sig.
Choosing Appropriate Assessment Methods	Based on Mean	6.490	1	98	.012
	Based on Median	4.443	1	98	.038
	Based on Median and with adjusted df	4.443	1	88.025	.038
	Based on trimmed mean	6.542	1	98	.012
Developing Appropriate Assessment Methods	Based on Mean	7.275	1	98	.008
	Based on Median	6.740	1	98	.011
	Based on Median and with adjusted df	6.740	1	89.913	.011
	Based on trimmed mean	7.354	1	98	.008
Administering, Scoring, and Interpreting the Results of Assessments	Based on Mean	6.674	1	98	.011
	Based on Median	5.582	1	98	.020
	Based on Median and with adjusted df	5.582	1	95.440	.020
	Based on trimmed mean	6.758	1	98	.011
Using Assessment Results to Make Decisions	Based on Mean	9.199	1	98	.003
	Based on Median	5.197	1	98	.025
	Based on Median and with adjusted df	5.197	1	91.005	.025
	Based on trimmed mean	8.667	1	98	.004
Developing Valid Grading Procedure	Based on Mean	.373	1	98	.543
	Based on Median	.297	1	98	.587
	Based on Median and with adjusted df	.297	1	91.768	.587
	Based on trimmed mean	.267	1	98	.606
Communicating Assessment Results	Based on Mean	1.817	1	98	.181
	Based on Median	1.419	1	98	.236
	Based on Median and with adjusted df	1.419	1	97.664	.236
	Based on trimmed mean	1.888	1	98	.173
Recognizing Unethical or Illegal Practices	Based on Mean	4.874	1	98	.030
	Based on Median	3.632	1	98	.060
	Based on Median and with adjusted df	3.632	1	88.454	.060
	Based on trimmed mean	4.666	1	98	.033

Then, the Levene's test of homogeneity of variances was run (Table 4). The outcomes revealed that the assumption of homogeneity of variances was met for developing valid grading procedures ($F(1, 98) = .297, p = .587$), communicating results ($F(1, 98) = 1.41, p = .236$), and recognizing unethical and illegal practices ($F(1, 98) = 3.63, p = .090$). However, the assumption was violated for choosing appropriate assessment methods ($F(1, 98) = 4.44, p = .038$),



developing appropriate assessment methods ($F(1, 98) = 6.74, p = .011$), administrating, scoring, and interpreting results ($F(1, 98) = 5.58, p = .020$) and utilizing assessment outcomes to make decisions ($F(1, 98) = 5.19, p = .025$). In order to conquer the issue of heterogeneity of variances, Tabachnick and Fidell (2014) suggested the reduction of the alpha level to .025 or .01. Therefore, the researchers discussed the results of MANOVA at $\alpha = .01$ level.

Table 5*Multivariate Tests; Components of Assessment Literacy Knowledge by Background*

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.976	536.055	7	92	.000	.976
	Wilks' Lambda	.024	536.055	7	92	.000	.976
	Hotelling's Trace	40.787	536.055	7	92	.000	.976
	Roy's Largest Root	40.787	536.055	7	92	.000	.976
Background	Pillai's Trace	.510	13.676	7	92	.000	.510
	Wilks' Lambda	.490	13.676	7	92	.000	.510
	Hotelling's Trace	1.041	13.676	7	92	.000	.510
	Roy's Largest Root	1.041	13.676	7	92	.000	.510

And finally, the findings of the MANOVA are shown in Table 5. Considering these outcomes ($F(7, 92) = 13.67, p = .000 < .01$, Partial eta squared = 0.510 indicating a significant effect size), the means of the TEFL and non-TEFL teachers on the seven components of assessment literacy knowledge showed significant differences.

Table 6*Descriptive Statistics; Components of Assessment Literacy Knowledge by Background*

Dependent Variable	Background	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Choosing Appropriate Assessment Methods	TEFL	16.980	.367	16.251	17.709
	Non-TEFL	13.220	.367	12.491	13.949
Developing Appropriate Assessment Methods	TEFL	18.340	.406	17.535	19.145
	Non-TEFL	13.960	.406	13.155	14.765
Administering, Grading, and Interpreting the Outcomes	TEFL	16.320	.363	15.600	17.040
	Non-TEFL	13.560	.363	11.840	14.280
Using Assessment Outcomes to Decide	TEFL	15.320	.350	14.626	16.014
	Non-TEFL	11.280	.350	10.586	11.974
Developing Valid Grading Procedure	TEFL	16.880	.376	16.135	17.625
	Non-TEFL	13.080	.376	12.335	13.825
Communicating Assessment Results	TEFL	16.240	.375	15.496	16.984
	Non-TEFL	12.880	.375	12.136	13.624
Recognizing Unethical or Illegal Practices	TEFL	15.240	.336	14.573	15.907
	Non-TEFL	10.780	.336	10.113	11.447

Besides, Table 6 displays the mean scores for TEFL as well as non-TEFL ESP educators on the elements of assessment literacy. The findings demonstrated that TEFL teachers outperformed

their non-TEFL colleagues teaching ESP in terms of mean scores on all assessment literacy components.

Table 7

Tests of Between-Subjects Effects: Components of A L Knowledge by Background

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Background	Choice	353.440	1	353.440	52.357	.000	.348
	Develop	479.610	1	479.610	58.232	.000	.373
	Admins	353.440	1	353.440	53.684	.000	.354
	Decision	408.040	1	408.040	66.762	.000	.405
	Grading	361.000	1	361.000	51.201	.000	.343
	Communicating	282.240	1	282.240	40.179	.000	.291
	Ethic	497.290	1	497.290	88.016	.000	.473
Error	Choice	661.560	98	6.751			
	Develop	807.140	98	8.236			
	Admins	645.200	98	6.584			
	Decision	598.960	98	6.112			
	Grading	690.960	98	7.051			
	Communicating	688.400	98	7.024			
	Ethic	553.700	98	5.650			
Total	Choice	23816.000	100				
	Develop	27369.000	100				
	Admins	21850.000	100				
	Decision	18696.000	100				
	Grading	23492.000	100				
	Communicating	22170.000	100				
	Ethic	17977.000	100				

It is possible to conclude, based on the data shown in Tables 6 and 7, that:

A. The TEFL instructors ($M = 16.98$) possessed a noticeably greater mean score on choosing appropriate assessment methods ($F = 52.35$, $p = .000$, partial eta squared = .348 displaying a significant effect size) compared to the non-TEFL instructors ($M = 13.22$).

B. The TEFL instructors ($M = 18.34$) possessed a noticeably greater mean score on developing appropriate assessment methods ($F = 30.25$, $p = .000$, partial eta squared = .373 displaying a significant effect size) compared to the non-TEFL instructors ($M = 13.96$).

C. The TEFL instructors ($M = 16.32$) possessed a noticeably greater mean score on administering, Scoring, and interpreting results ($F = 53.68$, $p = .000$, partial eta squared = .354 displaying a significant effect size) compared to the non-TEFL instructors ($M = 12.56$).

D. The TEFL instructors ($M = 15.32$) possessed a noticeably greater mean score on using assessment results to make decisions ($F = 66.76$, $p = .000$, partial eta squared = .405 displaying a significant effect size) compared to the non-TEFL instructors ($M = 11.28$).



E. The TEFL instructors ($M = 16.88$) possessed a noticeably greater mean score on developing

valid grading procedures ($F = 51.20$, $p = .000$, partial eta squared = .343 displaying a significant effect size) compared to the non-TEFL instructors ($M = 13.08$).

F. The TEFL instructors ($M = 16.24$) possessed a noticeably greater mean score on communicating results ($F = 40.17$, $p = .000$, partial eta squared = .291 displaying a significant effect size) compared to the non-TEFL instructors ($M = 12.88$).

G. The TEFL instructors ($M = 15.24$) possessed a noticeably greater mean score on considering

unethical and illegal practices ($F = 88.01$, $p = .000$, partial eta squared = .473 displaying an average effect size) compared to the non-TEFL instructors ($M = 10.78$).

In sum, the outcomes of the quantitative data analysis showed that, compared to the ones with non TEFL backgrounds, instructors with TEFL backgrounds enjoyed higher knowledge regarding different components of AL. The major issue in this respect was that the ESP teachers with TEFL credentials knew much about standards 1, 2, 3, 5, and 6. But they knew little about standard 4 and 7. It was also discovered that, although having little knowledge of assessment literacy, teachers without TEFL backgrounds were aware of its components. The mean scores for standards 1, 2, 3, and 5 were high among the teachers who were not trained in TEFL. On the other hand, they knew less about standards 4 and 6. Standard 7 was the least awareness-raising level for this group.

Q². Is there any statistically significant difference between novice and experienced ESP instructors concerning their knowledge of assessment literacy components?

To investigate the second research question, a multivariate analysis of variances (MANOVA) was performed to compare the means of the novice and experienced ESP teachers on the seven areas of teacher assessment literacy knowledge.

Table 8

Testing Normality of Data; Components of Assessment Literacy Knowledge by Teaching Experience

		N	Skewness			Kurtosis		
			Statistic	Std. Error	Ratio	Statistic	Std. Error	Ratio
Novice	Choosing Methods	55	-.059	.322	-0.18	-.595	.634	-0.94
	Developing Methods	55	-.145	.322	-0.45	-.884	.634	-1.39
	Administration	55	.043	.322	0.13	-.617	.634	-0.97
	Make Decisions	55	-.307	.322	-0.95	-.823	.634	-1.30
	Grading Procedure	55	-.313	.322	-0.97	-.094	.634	-0.15
	Communication	55	-.044	.322	-0.14	-.508	.634	-0.80
	Recognizing Unethical	55	-.459	.322	-1.43	-.804	.634	-1.27
Experienced	Choosing Methods	45	-.226	.322	-0.70	-.714	.634	-1.13
	Developing Methods	45	.172	.354	0.49	-.586	.695	-0.84



Administration	45	.452	.354	1.28	.432	.695	0.62
Make Decisions	45	.558	.354	1.58	1.308	.695	1.88
Grading Procedure	45	-.136	.354	-0.38	-.359	.695	-0.52
Communication	45	-.436	.354	-1.23	-.088	.695	-0.13
Recognizing Unethical	45	-.333	.354	-0.94	-.624	.695	-0.90

Remarkably, the assumption of normality was maintained; Table 8 shows that the ratios of kurtosis and skewness over their standard errors were less than +/- 1.96.

Table 9

Box's Test of Equality of Covariance Matrices; Components of Assessment Literacy Knowledge by Teaching Experience

Box's M	29.573
F	.975
df1	28
df2	30778.014
Sig.	.503

With Box's M = 29.57, $p = .503 > .001$, the homogeneity of covariance matrices (Table 9) was maintained. Field (2018) pointed out that the Box's test ought to be presented at .001 levels.

Table 10

Levene's Test of Equality of Error Variances; Components of Assessment Literacy Knowledge by Teaching Experience

		Levene Statistic	df1	df2	Sig.
Choosing Appropriate Assessment Methods	Based on Mean	7.664	1	98	.007
	Based on Median	6.950	1	98	.010
	Based on Median and with adjusted df	6.950	1	93.488	.010
	Based on trimmed mean	7.745	1	98	.006
Developing Appropriate Assessment Methods	Based on Mean	3.277	1	98	.073
	Based on Median	2.938	1	98	.090
	Based on Median and with adjusted df	2.938	1	96.918	.090
	Based on trimmed mean	3.302	1	98	.072
Administering, Grading, and Interpreting the Outcomes of Assessments	Based on Mean	3.732	1	98	.056
	Based on Median	3.573	1	98	.062
	Based on Median and with adjusted df	3.573	1	97.781	.062
	Based on trimmed mean	3.519	1	98	.064
Using Assessment Results to Make Decisions	Based on Mean	13.472	1	98	.000
	Based on Median	7.487	1	98	.007



Developing Valid Grading Procedure	Based on Median and with adjusted df	7.487	1	87.219	.008
	Based on trimmed mean	13.265	1	98	.000
	Based on Mean	1.739	1	98	.190
	Based on Median	1.264	1	98	.264
Communicating Assessment Results	Based on Median and with adjusted df	1.264	1	90.999	.264
	Based on trimmed mean	1.645	1	98	.203
	Based on Mean	1.691	1	98	.197
	Based on Median	1.822	1	98	.180
Recognizing Unethical or Illegal Practices	Based on Median and with adjusted df	1.822	1	97.469	.180
	Based on trimmed mean	1.715	1	98	.193
	Based on Mean	9.726	1	98	.002
	Based on Median	8.581	1	98	.004
	Based on Median and with adjusted df	8.581	1	92.393	.004
	Based on trimmed mean	9.169	1	98	.003

The findings of the Levene's test of homogeneity of variances are shown in Table 10. The findings indicated that except for choosing appropriate assessment methods ($F(1, 98) = 6.95, p = .010$), applying assessment outcomes to make decisions ($F(1, 98) = 7.48, p = .007$), and recognizing unethical and illegal practices ($F(1, 98) = 8.58, p = .004$), the presumption of variance homogeneity was confirmed for developing appropriate assessment methods ($F(1, 98) = 2.93, p = .090$), administering, grading, and interpreting outcomes ($F(1, 98) = 3.57, p = .062$), developing valid scoring process ($F(1, 98) = 1.26, p = .264$), and conveying outcomes ($F(1, 98) = 1.82, p = .180$).

Table 11
Multivariate Tests; Components of Assessment Literacy Knowledge by Teaching Experience

Effect	Value	F	Hypothesis	df	Error	df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.970	424.1187	92			.000	.970
	Wilks' Lambda	.030	424.1187	92			.000	.970
	Hotelling's Trace	32.2704	424.1187	92			.000	.970
	Roy's Largest Root	32.2704	424.1187	92			.000	.970
Experience	Pillai's Trace	.378	7.982	7	92		.000	.378
	Wilks' Lambda	.622	7.982	7	92		.000	.378
	Hotelling's Trace	.607	7.982	7	92		.000	.378
	Roy's Largest Root	.607	7.982	7	92		.000	.378

The outcomes of the MANOVA are shown in Table 11. Considering the seven dimensions of assessment literacy knowledge, there were significant differences between the means of novice and experienced teachers ($F(7, 92) = 7.98, p = .000 < .01$, Partial eta squared = .378 suggesting a strong effect size).

Table 12*Descriptive Statistics; Components of Assessment Literacy Knowledge by Teaching Experience*

Dependent Variable	Teaching Experience	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Choosing Assessment Methods	Appropriate Novice	16.600	.371	15.865	17.335
	Experienced	13.267	.410	12.454	14.080
Developing Assessment Methods	Appropriate Novice	17.600	.437	16.733	18.467
	Experienced	14.378	.483	13.419	15.337
Administering, Grading, and Interpreting the Outcomes of Assessments	Appropriate Novice	13.022	.435	12.159	13.885
	Experienced	15.600	.393	14.819	16.381
Using Assessment Outcomes to Decide	Appropriate Novice	11.511	.411	10.695	12.327

Table 13*Tests of Between-Subjects Effects; Components of Assessment Literacy Knowledge by Teaching Experience*

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Partial Eta Squared						
Experience	Choice	275.000	1	275.000	36.419	.000
	Develop	256.972	1	256.972	24.455	.000
	Admins	164.462	1	164.462	19.321	.000
	Decision	261.828	1	261.828	34.434	.000
	Grading	215.903	1	215.903	25.308	.000
	Communicating	210.620	1	210.620	27.158	.000
	Ethic	368.109	1	368.109	52.827	.000
Error	Choice	740.000	98	7.551		
	Develop	1029.778	98	10.508		
	Admins	834.178	98	8.512		
	Decision	745.172	98	7.604		
	Grading	836.057	98	8.531		
	Communicating	760.020	98	7.755		
	Ethic	682.881	98	6.968		
Total	Choice	23816.000	100			
	Develop	27369.000	100			
	Admins	21850.000	100			
	Decision	18696.000	100			
	Grading	23492.000	100			
	Communicating	22170.000	100			
	Ethic	17977.000	100			



Drawing from the data shown in Tables 12 and 13, it can be asserted that:

A. The novice instructors ($M = 16.60$) possessed a significantly greater average score on choosing appropriate assessment methods ($F = 36.41$, $p = .000$, partial eta squared = .271 indicating a significant effect size) compared to the experienced instructors ($M = 13.26$).

B. The novice instructors ($M = 17.60$) possessed a significantly greater average score on developing appropriate assessment methods ($F = 24.45$, $p = .000$, partial eta squared = .200 indicating a significant effect size) compared to the experienced instructors ($M = 14.37$).

C. The experienced instructors ($M = 15.60$) possessed a significantly greater average score on administering, scoring, and interpreting results ($F = 19.32$, $p = .000$, partial eta squared = .165 indicating a significant effect size) compared to the novice instructors ($M = 13.02$).

D. The experienced instructors ($M = 14.76$) possessed a significantly greater average score on applying outcomes of assessment to make choices ($F = 34.43$, $p = .000$, partial eta squared = .260 indicating a significant effect size) compared to the novice instructors ($M = 11.51$).

E. The novice instructors ($M = 16.30$) possessed a significantly greater average score on developing valid grading procedures ($F = 25.30$, $p = .000$, partial eta squared = .205 indicating a significant effect size) compared to the experienced instructors ($M = 13.35$).

F. The novice instructors ($M = 15.87$) possessed a significantly greater average score on communicating results ($F = 27.15$, $p = .000$, partial eta squared = .217 indicating a significant effect size) compared to the experienced instructors ($M = 12.95$).

G. The novice instructors ($M = 14.74$) possessed a significantly greater average score on considering unethical and illegal practices ($F = 52.82$, $p = .000$, partial eta squared = .350 indicating a significant effect size) compared to the experienced instructors ($M = 10.88$).

Q3. *How do novice and experienced ESP instructors with TEFL and non-TEFL backgrounds display their knowledge of assessment literacy components in practice?*

Three observations were made of 20 classrooms during the study's qualitative phase. Some notes were taken based on the researchers' inferences. Table 14 presents the data derived from the notes in the realm of themes and open/axial codes.

Table 14
Themes and Codes Derived out of the Observation Notes

Open Codes	Themes	Axial Codes
Choosing appropriate assessment methods	TEFL	They used various oral and written tests.
	Experienced	They aimed at testing communicative aspects of language.
	TEFL	They were more accustomed to assessment components.
	Novice	They could select appropriate assessment methods. They focused on tasks as well as tests.
	Non-TEFL	They could select rather appropriate assessment methods.
	Experienced	They mainly used the traditional testing activities. They relied on vocabulary and grammar tests.
	Non-TEFL	They couldn't choose many suitable assessment techniques.
	Novice	They focused on grammar and ESP vocabulary tests. They valued testing more than assessment.
		They helped the learners deal with tasks in the assessment.
	TEFL	They focused on recognition tests and tasks.
	Experienced	They used reliable teacher made and standardized tests.

Developing Appropriate Assessment Methods	TEFL Novice	They developed different test and task types. They focused on performance tasks. They used teacher made and standardized tests. They considered reliability and content validity of the tests they used. They could develop various tasks such as listing, ordering, describing, explaining, comparing, and contrasting.
	Non-TEFL Experienced	They did not design tasks. They used previously developed tests. They focused on recognition tests.
	Non-TEFL Novice	They relied on recognition tests of vocabulary. They focused on teacher-made tests of reading.
Administering, Grading, Interpreting Outcomes and the of Assessments	TEFL Experienced	They administered various tests regularly. They scored the papers very meticulously. They compared the scores with each other. They decoded scores as a norm-referenced concept.
	TEFL Novice	They applied tasks as summative activities. They relied on the learners' growth when interpreting scores. They decoded scores as a criterion-referenced concept.
	Non-TEFL Experienced	They administered the tests and not assessment tasks. They interpreted test results according to teacher feedback.
	Non-TEFL Novice	They once in a while administered quizzes and tests. They interpreted test results according to teacher feedback.
	TEFL Experienced	They relied on the test findings in designing new lessons. They depended on the test findings in designing homework. They considered results of tests to help the weaker learners.
Using Assessment Outcomes to Decide	TEFL Novice	They focused on the required information of their learners prior to teaching any new lessons. They planned their instructions based on the criterion-referenced concepts. They considered the outcomes of tests to determine the learners' weaknesses and strengths.
	Non-TEFL Experienced	They determined pass/fail of the students through test results. They relied on the test outcomes as a formative concept.
	Non-TEFL Novice	They relied on the test findings to decide about pass/fail of the students. They guided the students' learning techniques using test results.
	TEFL Experienced	They relied on norms.
	TEFL Novice	They developed criteria in scoring tasks.
Creating Reliable Grading Guidelines	Non-TEFL Experienced	They used recognition exams as the basis for their grading.
	Non-TEFL Novice	They didn't employ any particular scoring system.
	TEFL Experienced	They only announced the scores.
	TEFL	- They announced the results. - They discussed the answers with the learners.



Communicating Assessment Outcomes	Novice	• They informed the students of their misunderstandings.
	Non-TEFL	• They just announced the results.
	Experienced	
	Non-TEFL	• They only revealed the scores.
Recognizing Unethical or Illegal Practices	Novice	
	TEFL	• They announced the scores in public without considering the learner's privacy.
	Experienced	
	TEFL	• They had private conversations with students concerning their grades.
	Novice	
	Non-TEFL	• They announced the scores in public.
	Experienced	
	Non-TEFL	• They discussed grades with students in private.
	Novice	

The important aspects of the notes were as follows:

- Novice and experienced TEFL teachers selected more suitable assessment techniques compared to their non-TEFL colleagues.
- Novice TEFL teachers were the most competent and informed in creating tests and tasks, and using suitable assessment techniques.
- Experienced TEFL teachers outperformed their non-TEFL counterparts in terms of test administration issues.
- Experienced TEFL teachers had more success than non-TEFL instructors in implementing assessment findings when making choices.
 - Novice TEFL instructors gave careful consideration to the students' prior knowledge before beginning a new lesson.
- Novice TEFL instructors took into account issues such as grading procedures, sharing assessment outcomes, and ethical considerations for second language assessment.

In detail, with respect to the standard 1 of TALS, the data analysis of the observation notes displayed that compared to their non-TEFL colleagues, TEFL teachers, whether novice or experienced, could select more suitable techniques for assessment. For example, teacher No. 3 (TEFL/experienced) used various oral and written tests and tasks as assessment methods. Furthermore, he made use of both teacher-made tests and reliable standardized tests. While, teacher No. 14 (non-TEFL/experienced) could select quite suitable techniques for assessment. Notably, he concentrated on the conventional vocabulary and grammatical examination tasks. This is in line with what the study's quantitative phase found which shows that the participants with TEFL backgrounds ($M = 16.98$) and novice instructors ($M = 16.60$) had a significantly higher mean on choosing appropriate assessment methods than their non-TEFL ($M = 13.22$) and experienced ($M = 13.26$) counterparts.

Regarding the standard 2 of TALS, the data analysis of the observation notes showed that novice TEFL instructors were the most informed and diligent ones in terms of creating suitable assessment techniques. Moreover, novice TEFL instructors were aware of the significance of the standardized tests and the students' prior achievements in creating new tests. For example, teacher No. 8 (TEFL/novice) concentrated on the content validity and reliability of the tests she utilized in her class. She could create a variety of assessments as well as performance tasks, such listing, ranking, contrasting, explaining, and comparing. Consistent with the results of the observation notes, the TEFL instructors ($M = 18.34$) and the novice instructors ($M = 17.60$) had a



significantly higher mean on development of appropriate assessment methods than the non-TEFL ($M = 13.96$) and the experienced ($M = 14.37$) instructors.

Regarding the standard 3 of TALS, the data analysis of the observation notes showed that TEFL experienced teacher possessed a discriminating characteristic rather than their non-TEFL colleagues in terms of administering and grading tests as well as interpreting assessment results in the ESP classrooms. For example, teacher No. 2 (TEFL/experienced) performed exams on a regular basis. In addition, he assigned assignments as summative evaluations and meticulously graded the papers. Ultimately, he made a comparison between the results and utilized a norm-referenced interpretation. Consistent with the observation's findings, the TEFL instructors ($M = 16.32$) and the experienced instructors ($M = 15.60$) had a significantly higher mean on administering, scoring and interpreting results than the non-TEFL ($M = 12.56$) and the novice ($M = 13.02$) instructors.

Regarding standard 4 of TALS, the data analysis of the observation notes showed that experienced TEFL instructors outperformed their non-TEFL colleagues in terms of taking assessment findings into account while making choices. As an illustration, teacher No. 3 (TEFL/experienced) employed the test findings to design new lessons. Moreover, he created homework based on the tests' findings. Additionally, he used the testing's diagnostic power to direct the weaker students in light of their exam results. But before teaching any new sections, novice TEFL teachers (like teacher No. 7) devoted close attention to their pupils' prior knowledge. In accordance with the observation's findings, The TEFL instructors ($M = 15.32$) and the novice ones ($M = 14.76$) had a significantly greater average score on using assessment outcomes in decision making rather than the non-TEFL instructors ($M = 11.28$) and the experienced ones ($M = 11.51$).

Regarding standards 5, 6, and 7 of TALS, the data analysis of the observation notes showed that neither the novice non-TEFL instructors nor the experienced teachers in both groups acted satisfactorily. TEFL novice teachers were the only group that could relatively take into account such assessment literacy factors. For instance, in relation to standard 5, instructor No. 8 (TEFL/novice) created scoring guidelines and evaluated students' advancement in a range of skill areas. Similar to standard 6, instructor No. 9 (TEFL/novice) revealed the outcomes, talked about the answers with the students, and spotlighted the malpractices. Comparably, in reference to standard 7, instructor No. 10 (TEFL/novice) had a private conversation with the students on their scores. The findings of the quantitative data analysis support the observational findings, showing that the TEFL teachers' mean was noticeably higher on creating valid scoring procedures ($M = 16.88$), communicating outcomes ($M = 16.24$), and considering unethical practices ($M = 15.24$) than the non-TEFL instructors on developing valid scoring procedures ($M = 13.08$), communicating results ($M = 12.88$), and considering unethical practices ($M = 10.78$).

To put in a nutshell, the findings from the observation of the 20 ESP teachers who participated in the study's qualitative phase showed that compared to non-TEFL teachers, TEFL instructors had a better understanding of the components of assessment literacy. As a matter of fact, the results of the observations showed that the inexperienced TEFL instructors understood assessment literacy and its value in the context of teaching and evaluating ESP courses more fully. Additionally, they were amended with regard to selecting the most effective methods of assessment, creating tests, making decisions based on the findings of assessments, test validation, declaring assessment results, and the ethics in assessment.

However, in terms of other components of assessment (i.e., administration, scoring, and interpreting results) experienced TEFL instructors exhibited exact points as opposed to their non-TEFL colleagues and inexperienced TEFL/non-TEFL teachers.



Discussion

Through a mixed methods study, the current investigation aimed to identify the assessment literacy components that experienced and novice ESP teachers with TEFL / non-TEFL backgrounds knew about in the Iranian academic setting. The results of the quantitatively analyzed data firstly showed that ESP instructors with TEFL backgrounds are statistically significant in terms of their knowledge of assessment literacy components as opposed to non-TEFL instructors. The reason might lie in the sensitivity proposed on teaching assessment notions and principles to the TEFL students in the teacher training programs in the Iranian context (Farhady & Tavassoli, 2018; Mohammadi, 2020). In detail, although both groups of instructors with/without TEFL backgrounds had higher means on standards 1, 2, 3, and 5, they varied in the level of awareness about standard 6 of TALS. Both groups of instructors had relatively low means on standards 4 and 7.

The results of this investigation, in terms TEFL/non-TEFL ESP instructors' knowledge of AL components, are consistent with some of the recent investigations regarding the highest mean performance in the scale. For example, Jalilzadeh et al. (2022) in an intensive semi-structured examination with 20 EFL teachers identified that TEFL instructors were better at understanding assessment literacy components compared to non-TEFL instructors. Likewise, Plake and Impara (1993) used TALS and investigated some in-service instructors' assessment literacy. In their study, in line with the current investigation, standard 1 represented the greatest mean score on the scale. In another related study, Mertler (2003) discovered that standard 1 was where both in-service and pre-service instructors performed well. However, although standard 7 had the lowest mean in the current study, standard 6 and standard 5 had the lowest performance in Mertler's (2003) and Plake and Impara's (1993) investigations, respectively. Moreover, the present study findings in terms of components of AL is against the findings of some other studies such as Barnes et al. (2015) and Scarino (2013). As in those studies, standard 7 and 4 were more valued than those of the present study. In this respect, Barnes et al. (2015) stated that Social and cultural values, as well as the social policies, shape conceptions about teacher assessment and how such concepts are structured. In other words, teachers' priorities for AL components reflect expectations in both their macro- (such as their country/culture) and micro- (such as their department/university) settings (Chan & Luo, 2020). In fact, it is simpler to justify cross-cultural differences in teachers' assessment ideas when one is aware of the teachers' larger national assessment system. Furthermore, according to Remesal (2007), even educators from similar circumstances with similar socio-political norms have different perspectives on the idea and purpose of assessment. Altogether, a number of studies point to the necessity of more investigation on the structure as well as features of assessment both within and between cultures.

Likewise, in the second place, the quantitative findings of the study revealed that the novice instructors possessed a significantly greater mean on all the components of assessment literacy than their experienced counterparts. The reason might lie in the nature of novelty of classroom activities for the instructors who have just started teaching and are enthusiastic enough to apply their achievement in the L2 classroom and pursue what they have in mind as active people in the domain of education (Fathi & Derakhshan, 2019). Likewise, the novice instructors have not experienced burn out issues (Fathi & Saeedian, 2020) and are mostly interested in absorbing the students (Skaalvik & Skaalvik, 2017). Moreover, since novice teachers are more motivated and encouraged to reach professional development, they try to enhance their own assessment literacy. Additionally, they seek to promote their reputation in competitive working environment. Accordingly, they work more on their assessment literacy. Last but not least, they attempt not to lag behind their experienced colleagues. All these lead to higher levels of assessment literacy among them. This is in accordance with the findings of other identical international and Iranian investigations (Gareis & Grant, 2015; Jalilzadeh et al., 2022). For example, Gareis and Grant



(2015) who studied on assessment literacy for teacher candidates argued that AL could be absorbing to the teachers who have not reached the poisoning levels such as burnout and discouragement. In another related investigation in the Iranian academic setting, Jalilzadeh et al. (2022) showed that, compared to their experienced counterparts, novice EFL instructors had a greater understanding of the components of assessment literacy and the significance of AL in teacher preparation programs.

The analysis of the researchers' observation notes throughout the study's qualitative phase generally revealed that novice instructors with TEFL backgrounds outperformed their non-TEFL colleagues as well as their experienced TEFL/non-TEFL counterparts. This confirms the findings from this study's quantitative phase. Besides, the results of the study's qualitative phase are consistent with previous research on EFL instructors' AL practices in the Iranian academic setting. (Ashraf & Zolfaghari, 2018; Firoozi et al., 2019; Hajizadeh & Salahshour, 2014; Sobouti et al., 2023). They all confirmed that instructors with TEFL backgrounds significantly outperformed their non-TEFL counterparts in terms of using AL components during instructional hours. Moreover, they confirm the success of novice ESP instructors in applying AL principles in their classroom practices.

In this respect, the results of several worldwide investigations on the AL practices of EFL instructors support the current study's conclusions (Jeong, 2013; Lam, 2019; Looney et al., 2018; Xu & Brown, 2016). For example, Jeong (2013) argued that TEFL and non-TEFL teachers understand AL differently, as TEFL ones are more familiar with the notions and constructs to be assessed. Furthermore, according to Xu and Brown's (2016) study on teacher assessment literacy in practice in the Chinese context and Mertler's (1999) descriptive study of Ohio teachers' classroom assessment practices, newly graduate instructors are more likely than experienced teachers to apply their accomplishments in the L2 classroom.

However, the finding of the preset study in terms of the success of novice ESP instructors in applying AL principles in their classroom practices is inconsistent with some of the earlier research, which mainly showed that higher qualified teachers use more pertinent AL components and techniques in L2 classrooms. In this respect, Tajeddin et al. (2018) argued that "the findings revealed greater consistency in the experienced teachers' assessment literacy for speaking, despite a small difference in perceptions between novice and experienced instructors" (p. 57).

Conclusion

Focused on the ESP instructors with varying levels of experience and different educational backgrounds in Iranian academic settings, the current study aimed to demystify their understanding of assessment literacy components. Initially, it was discovered that teachers with TEFL credentials performed noticeably better in TALS than did teachers without TEFL backgrounds. Second, it was shown that in TALS, inexperienced teachers with TEFL backgrounds appeared noticeably better than both their non-TEFL peers and the experienced TEFL/non-TEFL teachers. Thirdly, findings from observation notes showed that inexperienced teachers with TEFL backgrounds practiced assessment elements better than their experienced and non-TEFL colleagues. Therefore, since AL plays a vital role in instructors' teaching (Ashraf & Zolfaghari, 2018) and learners' learning (Ellis, 2008), it can be concluded that it is necessary to hire ESP teachers with TEFL credentials rather than those without. The TALS results also showed that Iranian ESP teachers who participated in the study, regardless of their expertise and experience, were found to be inadequately informed regarding assessment literacy especially on decision making and ethical issues. Although the TEFL oriented instructors of ESP in this study had taken courses in language assessment and testing, those courses don't cover everything ESP



educators should know. Therefore, in order to compensate for the problems of low level of assessment literacy, constant in-service training programs on assessment literacy is needed.

References

- Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. *Human Behavior and Emerging Technologies*, 2(4), 314-324.
- Arani, A. M., Kakia, M. L., & Karimi, M. V. (2012). Assessment in education in Iran. *Assessment*, 9(2), 101-110
- Ashraf, H., & Zolfaghari, S. (2018). EFL teachers' assessment literacy and their reflective teaching. *International Journal of Instruction*, 11(1), 425-436.
- Atay, D. (2008). Teacher research for professional development. *ELT journal*, 62(2), 139-147.
- Azadi, A. (2018). *A study on the conceptual factors of teacher assessment literacy among ESP instructors*. Unpublished master's thesis, Islamic Azad University, Electronic Branch, Tehran, Iran.
- Babai Shishavan, H., & Sadeghi, K. (2009). Characteristics of effective English language teacher as perceived by Iranian teachers and learners of English, *Iranian Journal of Language Teaching Research*, 1(2), 130-132.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: OUP.
- Bandura, A. (2005). The evolution of social cognitive theory. *Great minds in management*, 6(2), 9-35.
- Baniali, S. (2018). *A study on Iranian experienced and novice EFL teachers' belief and practice in teaching vocabulary*. North Tehran Branch, Islamic Azad University. Tehran, Iran.
- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. *International handbook of research on teachers' beliefs*, 284-300.
- Bayat, K., & Rezaei, A. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education*, 3(1), 139-146.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). Psychometric analysis of an assessment literacy instrument: Applicability to pre-service teachers. In *Annual meeting of the mid-western educational research association*, Columbus, OH.
- Chan, C. K., & Luo, J. (2020). An exploratory study on teacher assessment literacy: Do novice university teachers know how to assess students' written reflection?. *Teachers and Teaching*, 26(2), 214-228.
- Creswell, J. W. (2009). *Research design qualitative and quantitative and mixed methods approaches* (3rd Ed.). California: Sage.
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.
- Dasgupta, N. (2013). Implicit attitudes and beliefs adapt to situations: A decade of research on the malleability of implicit prejudice, stereotypes, and the self-concept. *Advances in experimental social psychology*, 47(4), 233-279.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- Dörnyei, Z. (2007). Creating a motivating classroom environment. In *International handbook of English language teaching* (pp. 719-731). Springer, Boston, MA.
- Eezami, R. (2016). *A study on work engagement and fulfillment of basic psychological needs among novice and experienced EFL teachers in the Iranian institutes*. Kharazmi University, Tehran, Iran
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: OUP.



- Falsgraf, C. (2005). Why a national assessment summit? New visions in action. *National Assessment Summit*, 3, 6-9.
- Farhady, H., & Tavassoli, K. (2018). Developing a language assessment knowledge test for EFL teachers: A data-driven approach. *Iranian Journal of Language Teaching Research*, 6(3), 79-94.
- Fathi, J., & Derakhshan, A. (2019). Teacher self-efficacy and emotional regulation as predictors of teaching stress: An investigation of Iranian English language teachers. *Teaching English Language*, 13(2), 117-143.
- Fathi, J., & Saeedian, A. (2020). A structural model of teacher self-efficacy, resilience, and burnout among Iranian EFL teachers. *Iranian Journal of English for Academic Purposes*, 9(2), 14-28.
- Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(1), 2-14.
- Field, A. (2018). *Discovering statistics using IBM SPSS, statistics for statistics*. (5th ed.). London: SAGE Publications.
- Gareis, C. R., & Grant, L. W. (2015). Assessment literacy for teacher candidates: A focused approach. *Teacher Educators' Journal*, 20(3), 4-21.
- Gotch, C. M., and French, B. F. (2013). Elementary teachers' knowledge and self-efficacy for measurement concepts. *The Teacher Educator*, 48(1), 46-57.
- Hajizadeh, N., & Salahshour, N. (2014). Characteristics of effective EFL instructors: Language teachers' perceptions versus learners' perceptions. *International Journal of Applied Linguistics and English Literature*, 3(1), 202-214.
- Inbar-Lourie, O. (2013). Language assessment literacy: What are the ingredients?. *Language Testing*, 30(3), 301-307.
- Jafarpour, A. (2003). Is the test constructor a facet? *Language Testing*, 20(1), 57-87.
- Jalilzadeh, K., Alavi, M., & Siyyari, M. (2022). Comparing language assessment literacy and the challenges of Iranian EFL teachers: TEFL vs non-TEFL background. *Language and Translation*, 12(4), 177-196.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers?. *Language Testing*, 30(3), 345-362.
- Kögler, H. H. (2012). Agency and the other: On the intersubjective roots of self-identity. *New Ideas in Psychology*, 30(1), 47-64.
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and psychological measurement*, 30(3), 607-610.
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81(2), 78-89.
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Re-conceptualizing the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442-467.
- Mackey, A., & Gass, S. M. (2016). *Second language research: Methodology and design* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344.
- Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.



- Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education*, 43(6), 1-18.
- Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120, 285-296.
- Mertler, C. A. (2003). *Classroom assessment literacy inventory*. (Adapted from the Teacher Assessment Literacy Questionnaire (1993), by Barbara S. Plake & James C. Impara, University of Nebraska-Lincoln, in cooperation with The National Council on Measurement in Education & the W.K. Kellogg Foundation.
- Mertler, C. A. (2005). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(2), 76-92.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(1), 101-113.
- Mohammadi, A. (2020). *A mixed-methods study on the teacher assessment literacy of ELT instructors versus content instructors Islamic Azad University*. Unpublished master's thesis, University of Qom, Iran.
- Plake, B. S., & Impara, J. C. (1993). *Teacher assessment literacy questionnaire*. Nebraska-Lincoln: The National Council on Measurement in Education & the W.K. Kellogg Foundation.
- Popham, W. J. (2014). *Classroom assessment: What teachers need to know* (7th ed.). Boston: Pearson Education.
- Ramnarain, U., and Hlatwayo, M. (2018). Teacher beliefs and attitudes about inquiry-based learning in a rural school district in South Africa. *South African Journal of Education*, 38(1), 1-10.
- Razavipour, K., Riazi, A., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching*, 4(1), 156-161.
- Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam (2005). *Curriculum Journal*, 18, 27-38.
- Rodríguez, A. G., & McKay, S. (2010). Professional development for experienced teachers working with adult English language learners. CAELA Network Brief. *Center for Adult English Language Acquisition*. Retrieved from www.cal.org/caelanetwork
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Australia Language Testing*, 30(3) 309-327.
- Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Pre-service teachers' assessment literacy. *Journal of Science Teacher Education*, 22(4), 371-391 Retrieved from: <http://dx.doi.org/10.1007/s10972-011-9231-6>.
- Skaalvik, M., & Skaalvik, S., (2017). Motivated for teaching? Associations with school goal, teacher self-efficacy, job satisfaction and emotional exhaustion. *Teaching and Teacher Education*, 67(2), 152-160.
- Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Oxon: Routledge.
- Sussman, R., & Gifford, R. (2019). Causality in the theory of planned behavior. *Personality and Social Psychology Bulletin*, 45(6), 920-933
- Tabachnick, B.G., & Fidell, L.S. (2014). *Using multivariate statistics* (6th ed.). Pearson Inc.



- Tajeddin, Z., Alemi, M., & Yasaei, H. (2018). Classroom assessment literacy for speaking: Exploring novice and experienced English language teachers' knowledge and practice. *Iranian Journal of Language Teaching Research*, 6(3), 57-77.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412.
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.
- Yan, X., & Fan, J. (2020). Am I qualified to be a language tester? Understanding the development of assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219-246.
- Yzer, M. C. (2013). Reasoned action theory. The SAGE handbook of persuasion. *Developments in Theory and Practice*, 2(2), 120-136.
- Zamani, R., & Ahangari, S. (2016). Characteristics of an effective English language teacher (EELT) as perceived by learners of English. *International Journal of Foreign Language Teaching and Research*, 4(14), 69-88.
- Zhang, C., Yan, X., & Wang, J. (2021). EFL teachers' online assessment practices during the COVID-19 pandemic: Changes and mediating factors. *Asia-Pacific Edu Res*, 30(6), 499-507.
- Zwozdiak-Myers, P. (2012). *The teacher's reflective practice handbook: How to engage effectively in professional development and build a portfolio of practice*. Routledge.



© 2025 by the authors. Licensee International Journal of Foreign Language Teaching and Research, Najafabad Iran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY NC 4.0 license). (<http://creativecommons.org/licenses/by-nc/4.0/>).

