A Study on the Assessment Literacy of Novice and Experienced ESP Instructors with TEFL and non-TEFL Backgrounds in Iranian Academic Context

Faatemeh Sobouti [1]

PhD Candidtae, Department of English Translation,

Varamin- Pishva Branch, Islamic Azad University, Varamin, Iran.

Neda Gharagozloo (Corresponding Author) [2]

Assistant Professor, Department of English Translation,

Varamin- Pishva Branch, Islamic Azad University, Varamin, Iran.

Amirhosein Rahimi [3]

Assistant Professor, Department of English Language, Yadegar-e-Imam

Khomeini (RAH) Shahre Rey Branch, Islamic Azad University, Tehran, Iran.

Abstract

The present study was an attempt to investigate assessment literacy among ESP instructors in the Iranian academic context. A mixed-methods sequential explanatory study was designed, and 100 novice/experienced ESP instructors with TEFL/non-TEFL backgrounds at the Ph.D. level were selected through criterion sampling to complete a questionnaire and a scale. Then, from among the mentioned participants, 20 were selected for classroom observations and in-depth, semi-structured interviews based on convenience sampling. First, the quantitative data were collected using observation checklists and analyzed through running Chi-square to compare the assessment practices of novice/experienced instructors with TEFL/non-TEFL backgrounds. Next, the qualitative data were collected through taking notes and applying interviews. All interviews were transcribed, categorized, and analyzed through content analysis of themes and open/axial coding. The findings indicated that the instructors with TEFL backgrounds significantly practiced assessment methods more than the ones with non-TEFL backgrounds. Furthermore, it was found that novice instructors significantly practiced assessment methods more than the experienced ones. Likewise, it was identified that novice TEFL instructors perceived assessment components better than their non TEFL or experienced counterparts. The findings of this study can be useful for executive administrators in the ministry of higher education, university instructors and students.

Keywords

assessment literacy, components of assessment literacy, novice teachers, experienced teachers, TEFL/non-TEFL background

1. Introduction

Teachers have crucial effects on learners' learning and achievement. As Darling-Hammond (2000) states, the most critical factor which affects learners' learning is the instructor. Moreover, "good teaching is impossible in the absence of good assessment" (Eckhout et al., 2005, p. 3) as Stiggins (1999) argues "the quality of instruction in any classroom turns on the quality of the

[1] fatemeh.sobouti@iasbs.ac.ir

[2] n.gh@iauvaramin.ac.ir

[3] rahimi.amirh@gmail.com

assessments used there" (p. 20). Likewise, Deluca and Klinger (2010) state that Assessment Literacy (AL) is of great importance to promote learner achievement and teacher instruction.

Since EFL instructors need assessment literacy to pinpoint the problems of their classes and move toward betterment in their practice via applying modifications (Scarino, 2013), language educators contemplate instructors' assessment awareness in teacher education programs (Mellati & Khademi, 2018; Tavassoli & Farhady, 2018). Similarly, in large scale tests, assessment literacy knowledge assists English instructors to communicate the principles of language testing, as well as theory and practice, to test stakeholders (Popham, 2014).

Moreover, Jeong (2013) states that the definition of AL varies for non-language testers and language testers. Language educators also need to know what a consensual definition for AL is and how it is to be specified for groups with a variety of demands and levels of participation in language assessment practices. Therefore, the Language Assessment Literacy (LAL) that classroom teachers need varies from the LAL that a professional testing scholar may require; the demands are much higher for the latter (Malone, 2013).

Furthermore, Teacher Assessment Literacy (TAL) is considered as a critical qualification for instructors in EFL (English as a Foreign Language) domain (Ashraf & Zolfaghari, 2018; Bayat & Rezaei, 2015; Zamani & Ahangari, 2016). It is even argued that the development of Teacher Assessment Identity (TAI) is an outcome of performing various assessment practices (Estaji & Ghiasvand, 2021). Deficiency in AL can cause problems for the EFL instructors in their lesson planning (Stobart, 2008). Likewise, the instructors with less concern for assessment in their classes are probably a bit strict and bring about an atmosphere of competition rather than cooperation for their students; this affects both second language development and the learners' learning negatively (Ellis, 2008). AL is also claimed to play a critical role in effective teaching (Inbar-Lourie, 2013; Popham, 2014). It is argued that in the Iranian educational system testing has been more captivating than assessment (Arani et al., 2012).

Second language assessment literacy is also of paramount significance (Scarino, 2013), as Siegel and Wissehr (2011) have put forward, it helps teachers to adopt the most effective and reliable instruments to meet their learning objectives. Moreover, Tavassoli and Farhady (2018) argued that recent advances in education urge teachers "to be aware of and apply effective procedures of instruction and assessment to enhance learning" (p. 45). This requires teacher education programs to provide opportunities for teachers to improve their knowledge in different areas, including language assessment knowledge and practice, to meet the new challenges. Therefore, English as a Foreign Language (EFL) teachers also need assessment literacy to find the problems of their classes and move toward betterment in their work though running modifications (Hajizadeh & Salahshour, 2014; Scarino, 2013). Taylor (2013) signified that both English for Specific Purpose (ESP) and EFL instructors should be equipped with various forms of assessment to be more to the point in the process of teaching as this highly affect their own function as the moderators of teaching-learning process. Referring to the significant features of AL for the ESP instructors, Malone (2013) has emphasized that ESP teachers with higher AL knowledge can better fill the gaps and contrasts between the testers and the users.

It is worth mentioning that the significance of this study lies in its theoretical contribution to the second language research and the pedagogical implications for second language teachers and learners. The present study contributes to the growing body of research focusing on assessment literacy. Moreover, this study intends to help teachers get familiar with literacy assessment and assessment types which can affect students' learning, outcome, and teaching instruction. The results of this study will gain significance because language assessment literacy helps teachers "to understand, analyze and apply information on student performance to improve instruction" (Falsgraf, 2005, p. 6). Besides, familiarity with different assessment types allows teachers to adopt the most effective and appropriate instruments to meet the learners' needs and to accomplish learning objectives (Siegel & Wissehr, 2011).

Taking into consideration the aforementioned points, this study intends to investigate teacher assessment literacy practice of novice and experienced ESP instructors with TEFL and non-TEFL backgrounds at Islamic Azad University. In this regard, the perception of teacher assessment literacy by both novice and experienced ESP instructors enjoying TEFL and non-TEFL backgrounds was taken into consideration. Hence, the following questions are posed in the present study.

Is there any statistically significant difference between the assessment practices of ESP instructors with TEFL and non-TEFL backgrounds?

Is there any statistically significant difference between the assessment practices of experienced and novice ESP instructors?

How do experienced and novice ESP instructors with TEFL and non-TEFL backgrounds perceive teacher assessment literacy?

2. Review of Literature

What is fundamental in the processes of learning and teaching is assessment. It can affect the quality of both the teachers' teaching and the learners' learning. In this regard, a collection of research findings indicate that instructors struggle to incorporate instruction with assessment in association with contemporary assessment principles and theories (Amiri & Birjandi, 2015; Bachman & Palmer, 2010; Lam, 2019). In other words, a tension is created while on the one hand, there is a global move towards the employing of innovative assessment, such as criterion-referenced and performance-based assessments (Tao, 2014), and on the other hand, there is a strong preference on the side of teachers to use traditional assessment methods like objective tests.

To investigate the reason for this issue, different aspects of assessment literacy (AL) have been measured through comparing different instructors' AL (Edwards, 2017; Tajeddin et al., 2018). For instance, Azadi (2018) investigated the conceptual elements of TAL among Iranian ESP teachers, or Mohammadi (2020) compared AL knowledge of ESP teachers with TEFL and non-TEFL backgrounds. Assessment literacy practices of teachers in the classroom have also been

investigated (Crusan et al., 2016; Lam, 2019). Moreover, some standards have been suggested for such practices (McTighe & O'Connor, 2005; Pilcher, 2001). Some investigations have also noticed assessment literacy practices among EFL teachers' instructional strategies (Hajizadeh & Salahshour, 2014; Razavipour et al., 2011). Meanwhile, research results have revealed that instructors' assessment literacy practices do not necessarily illustrate their assessment literacy knowledge (Brookhart, 2011; Fard & Tabatabaei, 2018; Jannati, 2015; Lam, 2019). In this regard, Jannati (2015), who investigated EFL teachers' language assessment literacy reported that EFL teachers' perceptions of LAL have not been displayed in their practices.

In this regard, previous studies have indicated that Iranian EFL instructors undergo low levels of assessment literacy knowledge (Ashraf & Zolfaghari, 2018; Fard & Tabatabaei, 2018; Mellati & Khademi, 2018; Tavassoli & Farhady, 2018). Furthermore, since the concept is relatively new in the field, ESP teachers need to get acquainted with the concepts pertained to AL. To date little attention has been paid to demystifying teacher assessment literacy of ESP instructors with TEFL and Non-TEFL backgrounds at Iranian universities. It seems that many instructors lack assessment literacy and are not familiar with appropriate assessment practices, though the advantages of assessment literacy are recognized (Wang, Kao, & Lin, 2010). In order to fill this niche in literature, some researchers (e.g., Burry-Stock & Frazier, 2008; Mertler, 2005) have studied the multiple facets of teacher assessment literacy, regarding what it implies, how it works, and how teacher training programs may improve teacher assessment literacy (TAL).

Similarly, the current study tries to investigate some untapped aspects of teacher assessment literacy by comparing the assessment literacy of novice and experienced ESP teachers with TEFL and non-TEFL backgrounds. To the best knowledge of the researchers, few studies in Iran (Ashraf & Zolfaghari, 2018; Azadi, 2018; Birjandi & Tamjid, 2012; Fard & Tabatabaei, 2018; Hajizadeh & Salahshour, 2014; Mohammadi, 2020; Tavassoli & Farhady, 2018) have ever investigated how ESP teachers differ in terms of their familiarity with assessment practices. Nor have previous studies paid enough attention to the likely effect of ESP instructors' educational background on their literacy assessment perceptions on the basis of literature review in Iranian context. Thus, this gap in the literature and the paucity of research in this area in the Iranian context, and teachers' lack of familiarity with assessment practices in the same context are the impetus for the researchers to conduct the present study with the hope to contribute to the body of literature and help teachers to become more familiar with assessment practices.

3. Theoretical Framework

Among the most considerable responsibilities of classroom teachers is assessment of student performance because it has a great impact on whatever teachers do (Mertler, 2009). Spolsky (1995) stated, "In the event that a teacher is acceptable in teaching a language, the individual in question is acceptable in evaluating the students too" (as cited in Jafarpour, 2003, p. 59). It is also declared that teachers should have an appropriate level of assessment knowledge in order to evaluate learners appropriately (Inbar-Lourie, 2013). Theoretically speaking, teachers need get mastery over testing and assessment concepts such as test development, test administration, and test report (Bachman & Palmer, 2010). Modern ideas concerning assessment urge the assessment literacy of EFL teachers as a component of teachers' professional development (Atay, 2008).

With respect to theoretical notions, the present study relies on the combination of ideas presented in agentic theory (Bandura, 2008; Kögler, 2012), planned theory of behaviors (Ajzen, 2020; Sussman & Gilfford, 2019), theory of reasoned action (Hale et al., 2002; Yzer, 2013, 2017), and social cognitive theory (Bandura, 1986, 1997, 2005; Schunk, 2012). Agentic theory as proposed by Kögler (2012) notifies that individual teachers could be considered as vital and powerful agencies or persons dealing with their own performances while affecting their learners. In fact, teachers are paying contribution to the success and effectiveness of learners both in terms of assessment for learning and assessment of learning. Hence, teachers play a significant role in presenting ideals of assessment and increasing assessment qualities in the second language classroom.

Within reasoned action (Yzer, 2017) and planned behaviors theories (Ajzen, 2020), personal beliefs or attitudes of the individual teachers have been argued as a good predictor of their behaviors/ performances. That is, personal beliefs or attitudes of teachers can powerfully influence the ways in which they implement their classroom assessment. Sussman and Gifford (2019) defined personal beliefs/attitudes as the representation of information that individuals hold about any object, thing and other people surrounding them. Personal beliefs could be broadly defined as propositional attitudes, which refer to the attitudes of individuals toward a proposition about any object, thing and other people (Ramnarain & Hlatswayo, 2018).

Reasoned action proposed by Yzer (2013) and expanded (Yzer, 2017) deals with the fact that teachers' personal beliefs and attitudes could leave impacts on classroom practices and this action is so significant that it can lead classroom assessment to success or failure. Assessing L2 development would be affected by thoughts and perspectives of the teachers toward teaching, learning and assessing the second language. On the other hand, Ajzen (2020) presented idea of planned behaviors in terms of his own planned behavior theory emphasizing that teacher assessment literacy is manifested not only through the beliefs, knowledge, and strategies teachers claim, but also through the practices they follow in reality in the classroom context. It is also worth mentioning that Sussman and Gifford (2019) defined personal beliefs as information individuals present about the subject matter under investigation which in itself comes from the environment in which they are living. Such attitudes would be affected by the previous educational background, the in-service training projects, as well as teachers' perspective towards education and learning in general, and assessment in particular (Ramnarain & Hlatswayo, 2018).

Dasgupta (2013) commented that teachers' personal beliefs and attitudes have significant effect on their judgment, decision-making, and practices in L2 assessment. Accordingly, teachers' personal beliefs and perceptions about L2 assessment could effectively and significantly impact their assessment practices.

Within social cognitive theory (Bandura, 1986, 1997), self-efficacy has been argued to influence every phase of the individuals' personal evolution by regulating their behaviors /performances through their cognitive, motivational, affective and decisional processes (Gotch & French, 2013). Self-efficacy determines whether individual teachers think of altering their assessment implementation, and whether they have the motivation and perseverance needed to succeed with their assessment implementation.

Bandura's social cognitive theory (1986) deals with the consequences of actions. Hence, self-efficacy is considered as one of the most significant parts of this theory. In fact, as Gotch and French (2013) argue, people evaluate themselves, their own behaviors and performances, and what they do with respect to the cognitive, affective, motivational, and decisional processes. This is the notion of self-efficacy which makes the teachers change their assessment methods or not. Likewise, teachers' success in the process of assessing learners' performances is bound to their self-efficacy, world-view, beliefs, and the strategies they develop in this respect (Bandura, 1997). Under the scope of social cognitive theory, assessment is not just limited to assessing the learners' achievement but also various aspects of cognitive and social issues will affect this endeavor. Accordingly, a successful assessment process, takes different aspects of teacher and learner behaviors into consideration (Bandura, 2005) and relies on various experiences of teachers and learners with respect to assessment as a process, its consequences on learners' development and its effect of learners' life. In this regard, teachers' mastery over assessment notions, both theoretically and operationally, are of prime significance (Zwozdiak-Myers, 2012). That is why teachers' perceptions and practices altogether form one's assessment literacy.

4. Method

4.1. Participants

The participants of this study were 100 ESP instructors selected on the basis of criterion sampling including; instructors' educational backgrounds (TEFL/non-TEFL) and their teaching experience (novice/experienced). This number is in line with the criterion pertained to determining sample size for research activities (Krejcie & Morgan, 1970). In fact, when the study was being carried out, the population of ESP teachers at different branches of Azad universities in Zanjan and Tehran was 137 (N = 137) and based on the sample size table of Krejcie and Morgan, the number of participants should be 103 at least (n = 103). The participants' age ranged from 30 to 50. Furthermore, 20 participants (10 with TEFL backgrounds and 10 with non-TEFL backgrounds) were selected based on the convenience sampling for a follow up classroom observation and interview.

4.2. Instrumentation

It is worth mentioning that the present article is derived from a multiphase study conducted by the researchers which consists of two parts examining the ESP instructors' assessment literacy knowledge and practice respectively. The present article reports the second part (i.e., the ESP instructors' assessment practice). To achieve more accurate results for the whole project, five instruments were employed based on the theoretical principles of Teacher Assessment Literacy in the literature. The instruments were: *Novice and Experienced Teacher Questionnaire, Teacher Assessment Literacy Scale, classroom observation scheme, classroom observation notes,* and *a semi-structured interview with teachers.* The first two instruments were utilized directly in the first phase, and indirectly in the second phase of the whole project. That is, in order to select 20 participants of the second phase of the research (i.e., the focus of the present article), the researchers relied on the results obtained from the questionnaire and scale administered to 100 participants in the first phase. Then, they applied observation and interview to complete the research.

The first instrument was Novice and Experienced Teacher Questionnaire which was initially developed and confirmed by Rodríguez and McKay (2010). Its modified version considering the cultural and local notions of the Iranian context (Baniali, 2018; Eezami, 2016) was applied in the present study. Measuring the EFL teachers' experience as novice or experienced, the questionnaire incorporates a five-point Likert scale (little, a little, to some extent, much, and very much questions) with 12 items. The Cronbnach's alpha reliability index of the original version of the questionnaire indicates a good reliability ($\alpha = .72$). The construct validity of the questionnaire has been confirmed through factor analysis (Rodríguez & McKay, 2010). The reliability indices for this scale in the modified version used in the Iranian EFL studies have been reported as $\alpha = 0.76$ (Eezami, 2016) and $\alpha = 0.71$(Baniali, 2018). The minimum and maximum scores were 12 and 60, respectively. The cut score was 30 to 36 (Rodríguez & McKay, 2010). It implies that, in the present study, "the instructors who scored below 30 were specified as novice and those who scored above 36 were labeled as experienced. In order to specify the clear cut scoring system, the teachers with scores between 30 and 36 were left out" (Rodríguez & McKay, 2010, p.3).

The second instrument adopted to measure teacher assessment literacy in this study was the localized version of the *Teacher Assessment Literacy Scale* (Azadi, 2018) which was initially developed by Mertler (2009). This scale had been, in its own turn, adapted from the teacher assessment literacy questionnaire (Plake & Impara, 1993) developed by the University of Nebraska-Lincoln in cooperation with The National Council on Measurement in Education. The scale includes two parts. Part one consists of 35 items regarding the seven standards for teacher competence in the educational assessment of students. Some of the items in the scale measure general notions related to assessment and testing, such as the use of different assessment activities for assigning learners'

grades and communicating the results of assessment to learners and their parents; the remaining items refer to knowledge of standardized testing and classroom assessment. As Mertler (2009) argued, regarding classroom assessment, the following seven standards are good to guide what teachers should know and be able to do:

Choosing Appropriate Assessment Methods

Developing Appropriate Assessment Methods

Administering, Scoring, and Interpreting the Results of Assessments

Using Assessment Results to Make Decisions

Developing Valid Grading Procedures

Communicating Assessment Results

Recognizing Unethical or Illegal Practices

Part two encompasses items related to teachers' backgrounds as a classroom teacher.

The instrument enjoyed a reliability of 0.54 in the first study related to teachers' assessment literacy conducted by Plake and Impara (1993) and administered to 555 teachers in the United States (54). A higher level of reliability (0.74) was reported by Campbell, Murphy, and Holt (2002) in a similar study undertaken to 220 undergraduate pre-service teachers passing a course in measurement and test. Moreover, Azadi (2018) reported that "the Persian version of the scale enjoyed the reliability index of 0.73 based on Cronbach alpha ($\alpha = 0.73$)" (p. 63). Furthermore, he showed that "the validity of the teacher's scale has been confirmed through expert judgment validity" (p. 68).

Table 1
Reliability Statistics; Assessment Literacy and its Components

	Cronbach's Alpha	N of Items
Choosing 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 Unethical or Illegal Practices	0.74	5
Total	0.76	35

The third and fourth instruments were the observation checklist (Appendix A) and the researchers' notes. In line with Creswell and Clark (2017), who specified that checklists should be developed based on the related literature (p. 240), the observation checklist was developed by the researchers aimed at covering all seven components of assessment literacy based on a thorough review of the related literature on assessment literacy. The checklist

consisted of two parts: The first part included demographic features of the teachers who were observed, and the second part consisted of a five-point Likert scale (little, a little, into some extent, much, and very much) included 28 items covering the seven standards of teacher assessment literacy. In order to determine the reliability index of the observation checklist, the researchers asked five Ph. D. holders of TEFL to evaluate the checklist based on content-related evidence. In an attempt to calculate the reliability of the instrument, the researchers targeted the internal consistency of the ratings. Based on the results, it was concluded that there were significant agreements between the raters. The content validity of the observation checklist was gained through expert judgment validity (Mackey & Gass, 2016). Meanwhile, five experts analyzed the recorded sessions together and discussed the contents of the audiotapes against what had been written down during class observations. The degree of compatibility between the notes taken during the observation sessions and the review of the audiotapes was a reconfirmation for the reliability of observation notes.

The fifth instrument was an in-depth semi-structured interview (Appendix B). The questions for the interview were developed by the researchers based on the study objectives and concepts mentioned in the components of the Teacher Assessment Literacy Scale which cover all seven standards of assessment literacy. Then, the interview items were reexamined by two experts who were TEFL Ph. D. holders and familiar with assessment literacy arguments to ensure the appropriateness of content and language. Hence, the interview guide's content and construct validities were confirmed through expert judgment validity criteria. To ensure the reliability of the interview the researchers relied on the inter-rater reliability or intra-rater reliability indices (Creswell & Clark, 2017).

4.3. Design

The current study employed a mixed-methods sequential explanatory design based on Creswell and Clark (2017), including both quantitative and qualitative approaches with respect to the data collection and data analysis phases of the study. The justification for pursuing a mixed-methods approach is that both quantitative and qualitative research methods by themselves cannot be adequate for demonstrating and explaining the complexity of language studies (Creswell & Clark, 2017). In addition, the integration of both research methods can compensate for their shortcomings. As for timing, it is sequential. That is, first, the quantitative data is collected and analyzed. Then, in the second phase, by considering the results obtained in the first phase, the qualitative data is collected and analyzed (Creswell, 2009). As Mackay and Gass (2016) noted, "the value of triangulation is that it reduces observer or interviewer bias and enhances the validity and reliability (accuracy) of the information" (p. 182). In the present study triangulation was presented through employing, questionnaire, scheme, observation, observation notes, and interview. Interviewing the participants (TEFL and Non-TEFL ESP instructors) was done in order to get ensured of the accuracy of what they have presented in the questionnaire and scheme regarding the study variables. Also, classroom observations enriched the data which was collected. In other words, the triangulation technique was applied to examine the convergence of evidence from different methods to achieve richer data.

4.4. Procedure

At first, questionnaires and scales, namely Novice and Experienced Teacher Questionnaire (Rodríguez & McKay, 2010) and Teacher Assessment Literacy Scale (Mertler, 2005), were distributed through criterion sampling, among 100 ESP instructors holding master's degree or Ph. D. with TEFL or non-TEFL backgrounds at various branches of Islamic Azad University in Tehran and Zanjan provinces where ESP courses were being held. As it was requested by the researchers, the instructors completed the survey during non-instructional times, enclosed and gave them back to the researchers within 1 week of receipt. The researchers guaranteed that the participants' responses as well as their identity would be kept confidential. Then, the questionnaire and the scale were scored and SPSS version 25 was used in order to analyze the quantitative data.

Having been administered the questionnaire and the scale among 100 ESP instructors and conducted data analysis, in the qualitative phase of the study, 20 out of 100 ESP instructors were selected on the basis of convenience sampling for classroom observations and an in-depth semi-structured interview, provided that those selected instructors had given their consent for further cooperation.

In the next step, after obtaining the permission and consent of the participants and authorities, the researchers conducted classroom observations of twenty classes belonged to twenty ESP teachers. Five classes belonged to experienced teachers with a TEFL background, and five other classes belonged to experienced teachers with a non-TEFL background. Likewise, five classes belonged to novice teachers with a TEFL background, and five others belonged to novice ones with a non-TEFL background. Each class was observed for three sessions, and each session took about 90 minutes. This was done to increase the reliability of observation scheme as one single observation might not have provided the researchers with a sound understanding of the real classroom practices, but three sessions of observation (Dörnyei, 2007) would represent a status quo of the classroom assessment practices by the teachers. Indeed, observations allowed the researchers to investigate the teachers' assessment in natural settings regarding the focus of the present study (i.e., surveying teacher assessment literacy among English teachers with TEFL and Non-EFL backgrounds).

And finally, 20 observed instructors were interviewed (15-30 minutes). The interviewees were singled out on the basis of the outcomes obtained from the analysis of quantitative data ensued from the questionnaire and scale. The interview questions were designed by the researchers. Likewise, the interviews were recorded on a Digital Voice Recorder (DVR) and were transcribed, categorized, and analyzed for the purpose of the study.

4.5. Data Analysis

To answer the research questions of the study, data was analysed both quantitatively and qualitatively. In quantitative phase, since the sample size for this part of the research was limited (n = 20), analysis of Chi-square (crosstabs) was run to compare the TEFL/non-TEFL and novice/experienced ESP instructors' practices of assessment literacy.

Apart from utilizing an observation check list and quantitative data analysis of chi-square mentioned above, the results ensued from the notes taken by the researchers during observation, along with data obtained from interviews, were qualitatively analysed.

It should be noted that firstly, all observation notes and interviews were transcribed, summarized, categorized, and analyzed. Then, the themes and codes were extracted, and finally, frequencies, percentages, and standard residuals as descriptive units were used for an accurate and better interpretation of the obtained results.

5. Results

Q1. Is there any statistically significant difference between assessment practices of ESP instructors with TEFL and non-TEFL backgrounds?

Table 2								
Frequencies, Percentages and Std. Residuals; Assessment Practice by Background								
			Choices					Total
			Little	A little	To some extent	Much	Very much	
	TEFL	Count	0	1	58	124	97	280
		%	0.0%	0.4%	20.7%	44.3%	34.6%	100.0%
		Std. Residual	-4.4	-4.2	-.4	1.5	3.0	
	Non-TEFL	Count	38	38	65	93	46	280
		%	13.6%	13.6%	23.2%	33.2%	16.4%	100.0%
		Std. Residual	4.4	4.2	.4	-1.5	-3.0	
Total		Count	38	39	123	217	143	560
		%	6.8%	7.0%	22.0%	38.8%	25.5%	100.0%

Since the sample size for this part of the research was limited (n = 20), analysis of Chi-square (crosstabs) was run to compare the TEFL and non-TEFL instructors' assessment practices. Table 2 displays the frequencies, percentages and standardized residuals (Std. Residuals) for the TEFL and non-TEFL ESP instructors' practice of assessment methods. It should be noted that any Std. Residual higher than +/- 1.96, indicate that the related option was selected significantly more/less than what was expected. The results indicated that non-TEFL background instructors selected "little" (Std. Residual = 4.4 > 1.96) and "a little" (Std. Residual = 4.2 > 1.96) altogether more than TEFL background instructors who selected "much" (Std. Residual = 1.5 > 1.96)

and "very much" (Std. Residual = 3 > 1.96). Based on these results, it can be asserted that instructors with a TEFL background significantly practiced assessment methods more than the instructors with non-TEFL backgrounds.

Table 3
Chi-Square Tests; Assessment Practice by Background

	Value	Df	Asymptotic Significance (2-sided)
Pearson Chi-Square	96.118[a]	4	.000
Likelihood Ratio	120.881	4	.000
Linear-by-Linear Association	84.941	1	.000
N of Valid Cases	560		
Cramer's V	.414		.000
b. 0 cells (00.0%) have expected count less than 5. The minimum expected count is 19.			

The results of analysis of chi-square in Table 3 (χ^2 (4) = 96.11, p = .000, Cramer's V = .414 representing a moderate to large effect size) indicated that TEFL instructors significantly practiced assessment methods more than non-TEFL instructors.

Q2. Is there any statistically significant difference between assessment practices of experienced and novice ESP instructors?

Table 4
Frequencies, Percentages and Std. Residuals; Assessment Practice by Teaching Experience

		Choices					Total
		Little	A little	To some extent	Much	Very much	
Experienced	Count	38	38	43	88	73	280
	%	13.6%	13.6%	15.4%	31.4%	26.1%	100.0%
	Std. Residual	4.4	4.2	-2.4	-2.0	.2	
Novice	Count	0	1	80	129	70	280
	%	0.0%	0.4%	28.6%	46.1%	25.0%	100.0%
	Std. Residual	-4.4	-4.2	2.4	2.0	-.2	
Total	Count	38	39	123	217	143	560
	%	6.8%	7.0%	22.0%	38.8%	25.5%	100.0%

To answer this question, the researchers relied on the frequencies, percentages and std. residuals of assessment practices of experienced and novice ESP instructors. Table 4 displays the frequencies, percentages and standardized residuals (Std. Residuals) for the novice and experienced instructors' practice of assessment methods. The results indicated that experienced instructors selected "little" (Std. Residual = 4.4 > 1.96) and "a little" (Std. Residual = 4.2 > 1.96) altogether more than novice instructors who selected "much" (Std. Residual = 2 > 1.96) and "very much" (Std. Residual = 2 > 1.96). Based

on these results, it can be claimed that novice instructors significantly practiced assessment methods and components more than experienced instructors.

Table 5			
Chi-Square Tests; Assessment Practice by Teaching Experience			
	Value	Df	Asymptotic Significance (2-sided)
Pearson Chi-Square	92.042a	4	.000
Likelihood Ratio	116.604	4	.000
Linear-by-Linear Association	30.745	1	.000
N of Valid Cases	560		
Cramer's V	.405		.000
b. 0 cells (00.0%) have expected count less than 5. The minimum expected count is 19.			

The results of analysis of chi-square in Table 5 (χ^2 (4) = 92.04, p = .000, Cramer's V = .405 representing a moderate to large effect size) indicated that the novice instructors significantly practiced assessment methods and components more than the experienced instructors.

Twenty classes were observed three times. Apart from using a checklist to know if the ESP teachers were in line with the norms of assessment literacy, the researchers took some notes on the basis of their inferences. The data ensued from the checklist were calculated and reported in the quantitative part above. The data ensued from notes the researchers took in the qualitative part of the classroom observations with respect to the differences between assessment practices of novice and experienced ESP instructors with and without TEFL backgrounds are presented in Table 6 in terms of themes, open codes, and axial codes.

Table 6		
Themes and Codes Derived out of the Notes Taken in the Observations		
Open Codes	Themes	Axial Codes
• Choosing appropriate assessment methods	TEFL Experienced	• Used different written and oral test types as appropriate assessment methods. • Focused on testing communication aspects of language.
	TEFL Novice	• Could choose appropriate assessment methods. • Relied on tests as well as tasks. • Develope various tasks in line with the course purpose. • More familiar with assessment issues.

	Non-TEFL Experienced	• Could choose moderately appropriate assessment methods. • Mainly relied on the traditional testing activities. • Focused on testing grammar and vocabulary.
	Non-TEFL Novice	• Could choose little appropriate assessment methods. • Relied on testing ESP vocabulary. • Focused on testing grammar. • Showed that testing was more valued than assessment.
• Developing Appropriate Assessment Methods	TEFL Experienced	• Could help the learners recognize different test forms and deal with tasks in the assessment process. • Relied on recognition tasks and tests. • Made use of reliable standardized tests as well as instructor made tests.
	TEFL Novice	• Developed different test forms and task types. • Enjoyed creativity in designing tasks. • Relied on performance tasks. • Made use of reliable standardized tests as well as instructor made tests. • Considered content validity in their work and in some cases focused on reliability of the tests they use. • They were also able to develop different testing and performance tasks such as comparing, contrasting, criticizing, identifying, describing, explaining, specifying, listing, ordering, matching, selecting, defining, recalling, and restating.
	Non-TEFL Experienced	• Used previously developed tests. • Did not design tasks. • Developed recognition tests.
	Non-TEFL Novice	• Relied on testing vocabulary through recognition tests. • Focused on testing reading tests through teacher-made tests.
• Administering, Scoring, and	TEFL Experienced	• Administered the test regularly. • Administered tasks as a summative activity. • Scored the papers very meticulously.

Interpreting the Results of Assessments		• Compared the scores together. • Interpreted scores as a norm – referenced notion.
	TEFL Novice	• Administered tasks both in the process of learning and teaching and as a summative activity. • Focused on learners' development in interpreting scores. • Interpreted scores as a criterion reference notion.
	Non-TEFL Experienced	• Administered the tests and not assessment tasks. • Relied on teacher feedback in the test result interpretation.
	Non-TEFL Novice	• Once in a while administered tests and quizzes. • Relied on teacher feedback in interpreting the test results.
• Using Assessment Results to Make Decisions	TEFL Experienced	• Used the test results to design the new lesson. • Relied on the test results to design homework. • Used results of quizzes in the failing and passing of the course for the students. • Relied on the diagnostic perspective of testing and guided the weak students based in their test results.
	TEFL Novice	• Paid special attention to the prerequisite knowledge of their students before teaching the new unit. • They were the only group who were relatively aware of the concepts such as Norm-referenced and Criterion-referenced and could use the pertained information to plan for their instructions. • Used the test results to diagnose the weaknesses and strengths of the students.
	Non-TEFL Experienced	• Used the test results to pass or fail the students. • Used the test results as a formative assessment notion.
	Non-TEFL Novice	• Used test results to guide the learners in their learning strategies.

		• Used test results to decide up on the learners' passing or failing the course.
• Developing valid grading procedures	TEFL Experienced	• Did not use any specific grading procedure, but the psychometric testing views. • Relied on norms.
	TEFL Novice	• Relied on performance tasks. • Developed criterions in scoring tasks. • Valued leaners' development in different skills.
	Non-TEFL Experienced	• Relied on grading based on recognition tests.
	Non-TEFL Novice	• Did not use any specific grading strategy.
• Communicating Assessment Results	TEFL Experienced	• Communicate the test results through announcing the results.
	TEFL Novice	• Not only announced the results but asked the students to check their views with the teacher. • Discussed the answer with the leaners. • Highlighted the performance misconducts. and tried to make the learners aware of their misinterpretations.
	Non-TEFL Experienced	• Only announced the scores.
	Non-TEFL Novice	• Only announced the scores.
• Recognizing Unethical or Illegal Practices	TEFL Experienced	• Dis not consider learner's privacy. • Announce the scores in public.
	TEFL Novice	• Talked to students about their grades in private.
	Non-TEFL Experienced	• No special positive or negative function.
	Non-TEFL Novice	• Talked to students about their grades in private.

The crucial points of the notes the researchers took in the qualitative part of the classroom observations with respect to the differences between assessment practices of novice and experienced English instructors with and without a TEFL background go as follows:

Novice and experienced teachers with a TEFL background chose more appropriate assessment methods in comparison to their non-TEFL background counterparts.

Novice teachers with a TEFL background were the most knowledgeable and hardworking in terms of designing tasks, developing tests, and applying appropriate assessment methods.

Experienced TEFL teachers behaved noticeably comparing their non-TEFL colleagues regarding test administration factors in the EFL classroom.

Experienced teachers with a TEFL background were more successful than non-TEFL teachers in applying assessment results to make decisions.

Novice TEFL teachers paid special attention to the prerequisite knowledge of their students before teaching a new section.

Novice teachers with a TEFL background considered assessment literacy components like grading methods, communicating the assessment results, and ethical aspects of L2 assessment.

Consequently, it can be stated that novice instructors with a TEFL background more significantly practiced assessment components and methods in comparison to the experienced instructors with or without a TEFL background.

Q3. How do experienced and novice ESP instructors with TEFL and non-TEFL background perceive teacher assessment literacy?

To answer the qualitative research question of the study in the second phase, a thematic analysis of the interviews was conducted. The analysis of the qualitative data is a process which brings order to the data via arranging content into categories, patterns, and descriptive units. It should be noted that firstly, all interviews were transcribed, summarized, categorized, and analyzed. Thematic analysis through open coding showed that for each question there exist some themes. Likewise, axial coding revealed that for each theme, one can find some supporting notions. Hence, the themes and codes were extracted, and finally, descriptive units such as frequency, parentages, and standard residuals were used for an accurate and better interpretation of the obtained results. Table 7 below best shows the category of the main themes (groups), open codes (general views), and axial codes (specific notions) based on the interview data.

Table 7

Themes and Codes Derived out of the Interviews

Open Codes	Themes	Axial Codes	Frequency	Percent
1.Choosing appropriate assessment methods	TEFL Experienced	• I try to use appropriate assessment methods such as essay type items, oral group work.	4	%40
		• I like to use standardized tests. But they are not available for the courses we teach. That is why instructors mainly use teacher-made tests or the work books other instructors have written	4	%40
		• I try to test my students in line with the instructional objectives	4	%40
	TEFL Novice	• I try to use real life tasks, and creative performance tasks such as role playing.	5	%50
		• I think in the crowded classes we have to make use of testing, not assessment.	5	%50
		• In fact, we mainly uses tests and sometimes we employ oral exams.	5	%50
		• I ask the students to deliver mini-talks in the classroom and ask their peers to assess them.	5	%50
		• I would like to use dynamic assessment in my class, however, I do not have enough time. On the other hand, the final exams are in the multiple-choice form and testing oriented	5	%50
		• I try to test my students in line with the instructional objectives	3	%30
		• we mainly uses tests	3	%30
		• there is no difference between tests and tasks	4	%40

	Non-TEFL Experienced	• I use both oral and written quizzes as I think they are appropriate assessment methods	3	%30
	Non-TEFL Novice	• we mainly uses tests	2	%20
		• I use both oral and written quizzes as I think they are appropriate assessment methods	3	%30
		• learners should be able to translate ESP texts	4	%40
2.Developing Appropriate Assessment Methods	TEFL Experienced	• I try to develop tests which enjoy content validity	5	%50
		• I look at the student's past records and test scores to plan my own teaching and developing tests	4	%40
		• I develop tests based on the instructional manual of the ministry of education. So, I include different testing and performance tasks such as comparing, contrasting, identifying, specifying, and listing in my exams		
	TEFL Novice	• I develop appropriate assessment methods such as oral and written exams in line with instructional objectives	5	%50
	Non-TEFL Experienced	• I try to develop tests which enjoy content validity	5	%40
		• I sometimes make tests, but I mainly use the tests in the books available	3	%30
		• I look at the student's past records and test scores to plan my own teaching and developing tests	3	%30
		• I develop tests based on the instructional manual of the ministry of education. So, I include different testing and performance tasks such as comparing, contrasting, identifying, specifying, and listing in my exams	3	%30

	Non-TEFL Novice	• I never make tests. I use the tests in the books available	4	%40
3.Administering, Scoring, and Interpreting the Results of Assessments	TEFL Experienced	• I take exams every other session and score their performance by the help of the learners, themselves. We discuss the main points of the test in the same session	5	%50
		• I administer written tests every month or at the end of each lesson and score the students' papers on time. Meanwhile, I use oral exams every session	4	%40
		• I let the students check their problems with me and be fully informed of the nature of their scores	3	%30
		• I interpret the results of assessments in the class, especially if use the standardized tests issued by the ministry of education	4	%40
		• I ask the students to present their ideas about the difficulty level of the exam, test, or task they have taken	3	%30
	TEFL Novice	• I interpret the results of assessments in the class, especially if use the standardized tests issued by the ministry of education	4	%40
		• I let the students check their problems with me and be fully informed of the nature of their scores	5	%50
		• I ask the students to present their ideas about the difficulty level of the exam, test, or task they have taken	5	%50
	Non-TEFL Experienced	• I take exams every other session and score their performance by the help of the learners, themselves. We discuss the main points of the test in the same session	3	%30

		• I administer written tests every month or at the end of each lesson and score the students' papers on time. Meanwhile, I use oral exams every session	4	%40
	Non-TEFL Novice	• I ask the students to present their ideas about the difficulty level of the exam, test, or task they have taken	3	%30
4.Using Assessment Results to Make Decisions	TEFL Experienced	• I use the exam results to check for the prerequisite knowledge of my students before I start the new lesson	4	%40
		• I consider the general understanding of the previous lesson by the students (I rely on a norm-referenced view) to plan for the future instruction	4	%40
		• I make the learners familiar with the final exam samples which are standardized tests by the ministry of education	4	%40
	TEFL Novice	• I use the exam results to check for the prerequisite knowledge of my students before I start the new lesson	5	%50
		• I usually decide up on the type of homework based on the previous performance of the learners	5	%50
	Non-TEFL Experienced	• I consider the general understanding of the previous lesson by the students (I rely on a norm-referenced view) to plan for the future instruction	4	%40
		• I make the learners familiar with the final exam samples	3	%30

	Non-TEFL Novice	• Usually, I use the test results to pass fail the students in their ESP course	4	%40
5.Developing valid grading procedures	TEFL Experienced	• I try to use a valid grading procedure and focus on different times with varying grades • I explain to the students exactly how the grade has been determined	3	%30
		• I makes use of oral responses to questions asked in class of each student over the course as an assessment method showing a reliable student-performance	4	%40
		• I give a lot of tests to motive the students study harder and gain the best result possible	5	%50
	TEFL Novice	• I try to use a valid grading procedure and focus on different times with varying grades	5	%50
		• I makes use of oral responses to questions asked in class of each student over the course as an assessment method showing a reliable student-performance	4	%40
	Non-TEFL Experienced	• I explain to the students exactly how the grade has been determined • I give a lot of tests to motive the students study harder and gain the best result possible	3	%30
	Non-TEFL Novice	• I explain to the students exactly how the grade has been determined	4	%40
	TEFL Experienced	• I talk with the students about their performance in the class as well as in the exams	5	%50
		• I provide students with samples of their work	4	%40

6.Communicating Assessment Results		• I record the assessment results of the learners in the class portfolio every session	5	%50
	TEFL Novice	• I try to find the source of the weak learner's problems and use different methods to help him/her	4	%40
		• I focus on the learners' scores in my own class and discuss the case of weak students with my colleagues	5	%50
		• I consider the results of both formal and informal assessment types while reporting the students' performance	5	%50
		• I am really sensitive to the development of learners	4	%40
		• I record the assessment results of the learners in the class portfolio every session	5	%50
	Non-TEFL Experienced	• I record the assessment results of the learners in the class portfolio every session	3	%30
		• I am really sensitive to the development of learners	4	%40
	Non-TEFL Novice	• I record the assessment results of the learners in the class portfolio every session	3	%30
		• I am really sensitive to the development of learners	3	%30
	TEFL Experienced	• I will never let students see the other students' graded tests, so no violation of the students' right of privacy is allowed in my class	4	%40
		• I always assess the learners based on the concepts covered in the	4	%40

7.Recognizing Unethical or Illegal Practices		lesson, though I may add a pinch of creativity to it		
	TEFL Novice	• I recognize unethical and illegal practices in my assessment and do not differentiate between the learners while assessing them	5	%50
		• I do not announce the scores in public unless they are high to encourage the learners	4	%40
		• I do not belittle the weak students in front of their classmates	5	%50
		• I always assess the learners based on the concepts covered in the lesson, though I may add a pinch of creativity to it	4	%40
	Non-TEFL Experienced	• I will never let students see the other students' graded tests, so no violation of the students' right of privacy is allowed in my class	3	%30
	Non-TEFL Novice	• I recognize unethical and illegal practices in my assessment and do not differentiate between the learners while assessing them	4	%40
		• I do not announce the scores in public unless they are high to encourage the learners	3	%30
		• I do not belittle the weak students in front of their classmates	4	%40

To sum up, the results of the interview with the 20 ESP instructors taking part in the study revealed that instructors with a TEFL background had a more clarified and accurate understanding about components of assessment literacy compared to the non-TEFL instructors. In fact, the interview results revealed that novice instructors with a TEFL background were more aware of assessment literacy notions and the significance of AL knowledge in teaching and testing ESP courses. Moreover, they were more updated in terms of selecting the best assessment method, developing test types, decision making based on assessment results, test validation, communicating assessment results, and ethical issues.

However, in terms of other assessment components such as test administration, scoring procedure, and interpreting assessment results, experienced instructors with a TEFL background presented to-the-point concepts indicating that such instructors with a TEFL background could perceive the aforementioned assessment literacy components better than their non-TEFL counterparts as well as novice instructors with or without a TEFL background.

In the present study, based on the data analysis of observation notes regarding standard 1 of TAL, instructors with a TEFL background (both novice and experienced ones) could choose more appropriate assessment methods compared to their colleagues without a TEFL background (both novice and experienced ones). For example, teacher 5 (Linda) (TEFL/experienced) used different written and oral test types, assessment methods, and task types in the assessment process. Furthermore, she could help the learners recognize different test forms and deal with tasks in the assessment process. She made use of reliable standardized tests as well as instructor made tests. She also considered instructional objectives in the assessment methods she developed. While, teacher 12 (Shaahrokh) (non-TEFL/experienced) could choose moderately appropriate assessment methods. He mainly relied on the traditional testing activities, and focused on testing grammar and vocabulary.

In line with the results of observation, the interviewees with TEFL background (both novice and experienced ones) mentioned different ways of choosing appropriate assessment methods. For example, teacher 5 (Linda) (TEFL/experienced) argued:

> *"I try to use appropriate assessment methods such as essay type items, oral group work. I like to use standardized tests. But they are not available for the courses we teach. That is why instructors mainly use teacher-made tests or the work books other instructors have written. I try to test my students in line with the instructional objectives."*

While, teacher 12 (Shaahrokh) (non-TEFL/experienced) stated, "*I mainly use tests. There is no difference between tests and tasks. I use both oral and written quizzes as I think they are appropriate assessment methods.*"

Based on the data analysis of observation notes regarding standard 2 of TAL, novice instructors with a TEFL background were the most hardworking and knowledgeable ones in terms of developing tests, designing tasks, and developing appropriate assessment methods. Moreover, novice instructors with a TEFL background realized the significance of standardized tests and the previous performance of the learners in developing new tests. They were also able to develop different testing and performance tasks. For example, in this regard, teacher 6 (Sahar) (TEFL/novice) considered content validity in her work and in some cases focused on the reliability of the tests she used. She was also able to develop different testing and performance tasks such as comparing, contrasting, criticizing, identifying, describing, explaining, specifying, listing, ordering, matching, selecting, defining, recalling, and restating.

In accordance with the outcomes of observation, all of the novice interviewees with TEFL backgrounds were meticulous about the development of appropriate assessment methods. For example, teacher 6 (Sahar) (TEFL/novice) stated that *"I develop appropriate assessment methods such as oral and written exams in line with instructional objectives. I try to develop tests which enjoy content validity."*

Based on the data analysis of observation notes regarding standard 3 of TAL, experienced TEFL instructors had a more noticeable behavior compared to their non-TEFL counterparts in terms of test administration factors, scoring procedure, and interpreting assessment results in the ESP classrooms. This was evident with the experienced instructors and could be considered as their discriminating characteristic. For example, teacher 4 (Ehsaan) (TEFL/experienced) administered the test regularly. He also administered tasks as a summative activity and scored the papers very meticulously. Finally, he compared the scores together and interpreted them as a norm –referenced notion.

In line with the outcomes of observation, most of the experienced interviewees with TEFL background emphasized the administering, scoring, and interpreting the results of assessments. For example, teacher 4 (Ehsaan) (TEFL/experienced) argued:

> *"I administer written tests every month or at the end of each lesson and score the students' papers on time. I let the students check their problems with me and be fully informed of the nature of their scores. I interpret the results of assessments in the class, especially if use the standardized tests issued by the ministry of education. I ask the students to present their ideas about the difficulty level of the exam, test, or task they have taken."*

Based on the data analysis of observation notes regarding standard 4 of TAL, experienced instructors with a TEFL background were more successful than their non-TEFL counterparts in terms of utilizing assessment results to make decisions. However, novice TEFL instructors paid special attention to the prerequisite knowledge of their students before teaching the new unit. They were also the only group who were relatively aware of the concepts such as Norm-referenced and Criterion-referenced and could use the pertained information to plan for their instructions. For example, teacher 3 (Ali) (TEFL/experienced) used the test results to design the new lesson. He also made use of the results of quizzes in the failing and passing of the course for the students. Furthermore, he relied on the test results to design homework. He also relied on the diagnostic perspective of testing and guided the weak students based on their test results. However, teacher 7 (Alirezaa) (TEFL/novice) paid special attention to the prerequisite knowledge of his students before teaching the new unit. He also used the test results to diagnose the weaknesses and strengths of the students in different aspects of the course.

In harmony with the results of observation, most of the experienced interviewees with TEFL background stressed on using assessment results to make decisions. For example, teacher 3 (Ali) (TEFL/experienced) said:

> *"I use the exam results to check for the prerequisite knowledge of my students before I start the new lesson. I consider the general understanding of the previous lesson by the students (I mean I rely on a norm-referenced view) to plan for the future instruction. I make the learners familiar with the final exam samples which are standardized tests by the ministry of education."*

However, teacher 7 (Alirezaa) (TEFL/novice) stated that *"I use the exam results to check for the prerequisite knowledge of my students before I start the new lesson. I usually decide up on the type of homework based on the previous performance of the learners."*

Based on the data analysis regarding standards 5, 6, and 7, in terms of grading methods, communicating the assessment results, and ethical aspects of L2 assessment, neither the experienced instructors of both groups, nor the novice instructors without a TEFL background behaved appropriately. The only group who moderately could take such assessment literacy components into consideration was the novice instructors with a TEFL background. For example, regarding standard 5, teacher 8 (Zeinab) (TEFL/novice) relied on performance tasks. She developed criteria in scoring tasks and valued learners' development in different skills. Likewise, regarding standard 6, teacher 9 (Mehraan) (TEFL/novice) not only announced the results but asked the students to check their views with the teacher. He also discussed the answer with the learners, and highlighted the performance misconducts. Besides, he tried to make the learners aware of their misinterpretations. Similarly, with regard to standard 7, teacher 10 (Amir) (TEFL/novice) talked to students about their grades in private.

In accordance with the results of observation, most of the novice interviewees with TEFL backgrounds took assessment literacy components such as grading methods, communicating the assessment results and ethical aspects of L2 assessment into consideration. For example, regarding standard 5 on developing valid grading procedures, teacher 8 (Zeinab) (TEFL/novice) believed:

> *"I try to use a valid grading procedure and focus on different times with varying grades. I make use of oral responses to questions asked in the class of each student over the course as an assessment method showing a reliable student-performance."*

Likewise, regarding standard 6 on communicating assessment results, teacher 9 (Mehraan) (TEFL/novice) stated:

> "*I try to find the source of the weak learner's problems and use different methods to help him/her. I focus on the learners' scores in my own class and discuss the case of weak students with my colleagues. I am really sensitive to the development of learners.*"

Similarly, with regard to standard 7 on recognizing unethical or illegal practices, teacher 10 (Amir) (TEFL/novice) argued:

> "*I recognize unethical and illegal practices in my assessment and do not differentiate between the learners while assessing them. I do not announce the scores in public unless they are high to encourage the learners. I do not belittle the weak students in front of their classmates. I always assess the learners based on the concepts covered in the lesson, though I may add a pinch of creativity to it.*"

To sum up, the results of the interview with the 20 ESP instructors taking part in the study revealed that instructors with a TEFL background had a more clarified understanding about components of assessment literacy more accurately compared to the non-TEFL instructors. In fact, the interview results revealed that novice instructors with a TEFL background were more aware of assessment literacy notions and the significance of AL knowledge in teaching and testing ESP courses.

Moreover, they were more updated in terms of devolving test types, technical assessment knowledge, test validation, ethical issues, and decision making based on assessment results. However, in terms of other assessment components such as test administration, selecting the best assessment instrument choice, and communicating assessment results, experienced instructors with a TEFL background presented to-the-point concepts indicating that such instructors with a TEFL background could perceive the aforementioned assessment literacy components better than their non-TEFL counterparts.

6. Discussion

Considering the importance of assessment literacy as the key to effective teaching, deficiency in AL may create problems for the EFL teachers in their lesson planning, teaching, and assessing their students. Previous studies have shown that Iranian EFL teachers suffer from low levels of assessment literacy knowledge. To the best knowledge of the researcher, few studies in Iran have ever investigated how ESP teachers differ in terms of their familiarity with assessment practices. Nor have previous studies paid enough attention to the likely effect of ESP instructors' educational background on their literacy assessment perceptions on the basis of literature review in Iranian academic context.

Therefore, the present study was an attempt to investigate assessment literacy of novice and experienced ESP instructors with TEFL and non-TEFL backgrounds in the Iranian academic context through a mixed methods study.

The quantitative data analysis revealed that in the first place, ESP instructors with a TEFL background significantly had a higher practice of assessment literacy than the non-TEFL instructors. The present finding is in line with the findings of some studies on AL practices and EFL teachers in the Iranian context (Ashraf & Zolfaghari, 2018; Azadi, 2018; Birjandi & Tamjid, 2012; Fard & Tabatabaei, 2018; Firoozi et al., 2019; Hajizadeh & Salahshour, 2014; Khodabakhshzadeh et al., 2018; Mohammadi, 2020; Razavipour et al., 2011; Tavassoli & Farhady, 2018; Watmani et al., 2020). In fact, the mentioned studies emphasized AL practices as EFL teachers' instructional strategies. For example, the present finding is in line with Azadi's (2018) study investigating the conceptual factors of TAL among Iranian ESP teachers, and in line with Mohammadi's (2020) study comparing AL knowledge of ESP instructors with TEFL and non-TEFL background. Likewise, the findings of some international studies on EFL teachers' AL practices (Brookhart, 2011; Burry-Stock & Frazier, 2008; Crusan et al., 2016; Ellis, 2008; Jeong, 2013; Looney et al., 2018; Lam, 2019; McTighe & O'Connor, 2005; Pilcher, 2001; Popham, 2014; Xu & Brown, 2016) also support the above mentioned finding of the present study arguing that, for an EFL instructor, mastery over assessment methods is a must. Similarly, the result of the present study is in concordance with Jeong's (2013) finding which revealed that since the notions and constructs to be testes is known for language testers, the definition of AL varies for language testers and non-language testers.

Secondly, it was shown that novice ESP instructors with a TEFL background applied AL components and principles in their classroom practices more significantly than their experienced colleagues with or without a TEFL background. Although, this is not in line with some of the previous findings which mostly revealed that experienced instructors practice more relevant AL methods and components in the L2 classroom (Edwards, 2017; Gatbonton, 2008; Mertler, 2005), it is in line with findings of a number of the other studies (Pilcher, 2001; Xu & Brown, 2016) which confirm that newly university graduate instructors are more willing to apply their achievements in the L2 classroom comparing to the experienced instructors. Hence, they devise in-service training and continuous education for the EFL teachers in terms of AL. Barnes, Fives, and Dacey (2015) found that teacher assessment notions and the structure of those notions are figured by cultural and social priorities as well as the policies in a society. This idea has been proved by other authors (Scarino, 2013). In fact, understanding the larger, national assessment context makes it possible to explain cross-cultural differences in teachers' concept of assessment. Also, Remesal (2007) identified that even instructors from similar contexts who have been exposed to the same socio-political expectations and dominations reflected mixed and different ideas about the intention of assessment. Altogether, these studies propose the need for additional research in order to investigate the nature and structure of notions about assessment across and within cultures.

Furthermore, the qualitative data analysis revealed that TEFL instructors understood constituents of assessment literacy more scrupulously than non-TEFL instructors. In fact, among TEFL instructors, novice teachers perceived some components of AL better than their experienced counterparts and were more updated in terms of technical assessment knowledge. In the present study, the result of the interview is in harmony with the findings of to some extent similar previous studies (Drajati et al., 2018; Hervey, 2015; Nazari et al., 2019). For example, in a similar investigation, Drajati et al. (2018) found that less experienced instructors implement a variety of technological instruments in their teaching. Moreover, Hervey (2015) confirmed that, in assessing students, novice teachers are more aware of new techniques and take more risks in applying technologies in their classrooms. Likewise, Nazari et al. (2019) ran a mixed methods study to investigate novice and experienced EFL teachers' different perceptions of Technological Pedagogical Content Knowledge (TPACK) and its effects on their professional development. They identified that in comparison with their content and pedagogical knowledge, novice EFL instructors had a rich knowledge of technology.

The results of the qualitative data analysis suggested that ESP instructors with non TEFL backgrounds did not perceive AL components well. This is in accordance with the finding of a research on teachers' perception of classroom assessment components, run by Mertler and Campbell (2005). They proved that teachers, in a lot of cases, did not have a good knowledge of assessment literacy principles.

The interview results in the present study showed that novice instructors with TEFL backgrounds were more aware of assessment literacy and its significance in instructor education. Studies have proved that usual classroom assessment courses in teacher preparation programs are not in harmony with teachers' real requirements for classroom practice (Schaefer & Lissitz, 1987). The focus has traditionally been on standardized testing (Schaefer & Lissitz, 1987), although this trend is changing. Only one course in assessment may not really be sufficient to deal with everything teachers need to know (Mertler, 2005).

The present study also takes support from the study done by Volante and Fazio (2007). They investigated assessment literacy knowledge and practice of teacher candidates during all four years of their program. Most participants proposed summative goals for assessment and only a few suggested formative goals. They preferred personal communication and observational techniques.

7. Conclusion

The present study was an attempt to demystify teacher assessment literacy among ESP novice and experienced instructors with TEFL and Non-TEFL backgrounds in the Iranian academic context. It was concluded that instructors with TEFL backgrounds significantly practiced assessment methods more than instructors with non-TEFL backgrounds. In addition, it was found that novice instructors with TEFL backgrounds significantly practiced assessment methods and components more than their non TEFL counterparts as well as the experienced instructors. Moreover, it was identified that novice instructors with TEFL backgrounds perceived assessment components more than their non TEFL or experienced counterparts.

Since AL literacy plays a critical role in effective training (Ellis, 2008) and the instructors' reflective teaching (Ashraf & Zolfaghari, 2018), employing ESP instructors with a TEFL background instead of a non-TEFL background is suggested. Also, as the present study findings revealed, the newly graduated novice instructors are more successful than their experienced colleagues in terms of AL practices (Mohammadi, 2020). The reason might root in the willingness of the recently graduated TEFL instructors in applying their achievements in the L2 classroom, and the issue of teacher burnout on the part of the experienced teachers (Fathi & Derakhshan, 2019; Sadeghi & Khezrlou, 2014).

After all, both interview and classroom observation results revealed that all ESP instructors taking part in the study, especially novice and experienced ones with a non-TEFL background, need in-service training courses in L2 assessment concepts. This is in line with Lam's (2015) proposal on language assessment training for ESL instructors in Hong Kong and its effect on the L2 students' development.

In general, irrespective of experience and background of Iranian ESP instructors who took part in the study, it was identified that they did not have a good understanding of not only standards of assessment literacy (Mertler, 2005), but also standards of teacher competencies (Plake & Impara; 1993). This is more alarming when we notice that around half of these ESP teachers enjoy a TEFL background which means, in their undergraduate and graduate levels, they have surely passed courses in language assessment and testing. Therefore, with regard to assessment literacy, in-service training courses are needed for teachers who are active in the Iranian ESP field including those participated in the study. Specifically, it is crucial that instructors notice that assessment is an indispensable part of education and is necessary for the learner's L2 development (McMillan, 2000). Teachers have also shown less interest in basic measurement principles and more in application of assessment processes (Pilcher, 2001). Then, hopefully, making assessment more applicable in ESP instructors' views of teaching, teachers of assessment courses can teach the principles of assessment to ESP teachers and show them the vital connection between instruction and assessment.

Moreover, the conventional focus of pre-service assessment courses has been more on standardized tests. Since ESP instructors cope with new trends in assessment such as competency-based assessment (Carraccio et al., 2016) and assessment pedagogies (Kayler & Weller, 2007), the content of pre-service training and in-service training programs on educational assessment need to focus more on substitute assessment methods.

In fact, instead of traditional tests and testing services, ESP instructors could employ new trends of assessment literacy in their classes in order to assess the learners' performances and this way facilitate learning for the learners. This way a less stressful situation based on the operational activities will be created and classroom interactions could be enriched and would help subsequent L2 and ESP development of the learners.

Furthermore, in the ELT domain, materials developers can apply the results of the present study and those of the similar ones in order to propose tasks in which learners' awareness toward learning is improved. Such tasks can help the learners move towards meaningful learning, self-correction, performance, and autonomy.

It is also worth mentioning that, the current study encountered some limitations during conducting the study. The present study was conducted in a relatively small scale (100 ESP instructors). Therefore, the results of the present study must be generalized cautiously. Furthermore, this study concentrated on the instructors' responses to scale regarding their assessment literacy. Instructors' responses might have been affected by different factors such as class content, teacher characteristics, and timing of the evaluation, and on the top of all, individual differences (Williams & Burden, 1997). Thus, further research is needed to investigate the following issues:

The same hypothesis can be formulated for Iranian ESP instructors teaching at different academic institutes. It is worth investigating whether providing the ESP instructors with instruction concerning AL has any positive and significant effects on the learners' ESP development.

Future studies might consider examining the residual effects of in-service training and instruction of the ESP instructors in terms of AL to explore whether and how long-term these effects actually could be. A semi-longitudinal study of the concept of AL on a specific group of instructors can reveal if this theory energizes retention of assessment literacy in the instructors' mentality or not.

In addition, the present study focused on the novice and experienced ESP instructors with TEFL and Non-TEFL backgrounds. Future studies may be needed to replicate the findings with general English university professors.

Further research is recommended to explore the role of AL instruction in developing Instructors' competency-based assessment and its relationship with instructors' autonomy, self-regulatory factors of teaching, and learner motivation.

References

Ajzen, I. "The theory of planned behavior: Frequently asked questions." *Human Behavior and Emerging Technologies,* Vol. 2, No. 4, 2020, 314-324.

Amiri, M., and Birjandi, P. "Reliability and Content Validity of a Comprehensive Discourse Completion Test of Interlanguage Pragmatics for

Academic Situations." *Sino-US English Teaching,* Vol. 12, No. 9, 2015, 654-659.

Arani, A. M., et al. "Assessment in Education in Iran." *Assessment*, Vol. 9, No. 2, 2012, 101-110.

Ashraf, H., and Zolfaghari, S. "EFL Teachers' Assessment Literacy and Their Reflective Teaching." *International Journal of Instruction,* Vol. 11, No. 1, 2018, 425-436.

Azadi, A. *A study on the Conceptual Factors of Teacher Assessment Literacy among ESP Instructors.* 2018. Islamic Azad University, Electronic Branch, Tehran, Iran Unpublished dissertation.

Atay, D. "Teacher Research for Professional Development." *ELT journal,* Vol. 62, No. 2, 2008, 139-147.

Bachman, L. F., and Palmer A. S. *Language Assessment in Practice.* Oxford: OUP, 2010.

Bandura, A. "Social foundations of thought and action." *Englewood Cliffs,* Vol. 4, No. 1,1986, 23-28.

Bandura, A. "Self-efficacy: The exercise of control." *New York: Freeman. Barefoot College. Available at: www. barefootcollege. org. Accessed on July* 31 (1997): 2011.

Bandura, A. "The evolution of social cognitive theory." *Great minds in management*, Vol. 6, No. 2, 2005, 9-35.

Bandura, A. "Toward an agentic theory of the self." *Advances in self research,* Vol. 3, No. 1, 2008, 15-49.

Baniali, S. *A study on Iranian Experienced and Novice EFL Teachers' Belief and Practice in Teaching Vocabulary*. 2018. Islamic Azad University, North Tehran Branch, Tehran, Iran, Ph.D. Dissertation.

Barnes, N., et al. *Teachers' Beliefs about Assessment: International Handbook of Research on Teachers' Beliefs.* New York: Routledge, 2015.

Bayat, K., and Rezaei, A. "Importance of Teachers' Assessment Literacy." *International Journal of English Language Education,* Vol. 3, No. 1, 2015, 139-146.

Birjandi, P., and Tamjid, N. H. "The role of Self-, Peer and Teacher Assessment in Promoting Iranian EFL Learners' Writing Performance." *Assessment & Evaluation in Higher Education,* Vol. 37, No. 5, 2012, 513-533.

Brookhart, S. M. "Educational assessment knowledge and skills for teachers." *Educational Measurement: Issues and Practices,* Vol. 30, No. 1, 2011, 3-12.

Burry-Stock, J. A., and Frazier, C. H. "Revision of the Assessment Practice Inventory Revised (APIR): A combined exploratory factor analysis and polytomous IRT approach." *American Educational Research Association,* Vol. 5, No. 2, 2008, 24-28.

Campbell, C., et al. "Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers." *Annual meeting of the mid-western educational research association,Vol. 9, No. 4,* 2002, 24-30.

Carraccio, C., et al. "Advancing competency-based medical education: a charter for clinician educators." *Academic Medicine,* Vol. 91, No. 5, 2016, 645-649.

Creswell, J. W., and Clark V. L. P. *Designing and conducting mixed methods research.* Sage publications, 2017.

Crusan, D., et al. "Writing Assessment Literacy: Surveying Second Language Teachers' Knowledge, Beliefs, and Practices." *Assessing Writing,* Vol. 28, No. 4, 2016, 43-56.

Darling-Hammond, L. "Teacher quality and student achievement: A review of state policy evidence." *Education Policy Analysis Archives,* Vol. 8, No. 1, 2000, 20-32.

Dasgupta, N. "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,* Vol. 47, No. 4, 2013, 233-279.

DeLuca, C., and Klinger, D. A. "Assessment Literacy Development: Identifying Gaps in Teacher Candidates' Learning." *Assessment in Education: Principles, Policy & Practice,* Vol. 17, No. 4, 2010, 419-438.

Dörnyei, Z. *Research methods in applied linguistics.* New York: Oxford University Press, 2007.

Drajati, N. A., et al. "Investigating English Language Teachers in Developing TPACK and Multimodal Literacy." *Indonesian Journal of Applied Linguistics,* Vol. 7, No. 3, 2018, 575-582.

Eckhout, T., et al. *A Method for Providing Assessment Training to In-service and Pre-service Teachers.* New York: Arnolds, 2005.

Edwards, F. "A Rubric to Track the Development of Secondary Pre-service and Novice Teachers' Summative Assessment Literacy." *Assessment in Education: Principles, Policy & Practice,* Vol. 24, No. 2, 2017, 205-227.

Eezami, R. *A study on work engagement and fulfillment of basic psychological needs among novice and experienced EFL teachers in the Iranian institutes.* 2016.

Kharazmi University, Tehran, Iran. Unpublished dissertation. Ellis, Rod. *The study of second language acquisition.* 2nd ed., Oxford: OUP, 2008.

Estaji, Masoomeh, and Farhad Ghiasvand. "Assessment Perceptions and Practices in Academic Domain: The Design and Validation of an Assessment Identity Questionnaire (TAIQ) for EFL Teachers." *International Journal of Language Testing,* Vol. 11, No. 1, 2021, 103-131.

Fard, Z., and Tabatabaei, O. "Investigating Assessment Literacy of EFL Teachers in Iran." *Journal of Applied Linguistics and Language Research*, Vol. 5, No. 3, 2018, 91-100.

Fathi, J., and Derakhshan, A. "Teacher self-efficacy and emotional regulation as predictors of teaching stress: An investigation of Iranian English language teachers." *Teaching English Language,* Vol. 13, No. 2, 2019, 117-143.

Firoozi, Tahereh, et al. "The Language Assessment Literacy Needs of Iranian EFL Teachers with a Focus on Reformed Assessment Policies." *Language Testing in Asia*, Vol. 9, No. 1, 2019, 2-14.

Gatbonton, Elizabeth. "Looking beyond teachers' classroom behaviour: Novice and experienced ESL teachers' pedagogical knowledge." *Language teaching research,* Vol. 12, No. 2, 2008, 161-182.

Gotch, C. M., and French, B. F. "Elementary teachers' knowledge and self-efficacy for measurement concepts." *The Teacher Educator,* Vol. 48, No.1, 2013, 46-57.

Hajizadeh, N., and Salahshour, N. "Characteristics of Effective EFL Instructors: Language Teachers' Perceptions Versus Learners' Perceptions." *International Journal of Applied Linguistics and English Literature*, Vol. 3, No. 1, 2014, 202-214.

Hale, J. L., Householder, B. J., and Greene, K. L. "The theory of reasoned action." *The persuasion handbook: Developments in theory and practice,* Vol. 14, No. 1, 2002, 259-286.

Hervey, L. G. "Between the notion and the act: Veteran teachers' TPACK and practice in 1: 1 settings." *Technological pedagogical content knowledge*. Vol. 5, No. 2, 2015. 165-189.

Inbar-Lourie, O. "Language Assessment Literacy: What Are the Ingredients?." *Language Testing*, Vol. 30, No. 3, 2013, 301-307.

Jafarpur, A. "Is the test constructor a facet?." *Language Testing,* Vol. 20, No.1, 2003, 57-87.

Jannati, S. "ELT Teachers' Language Assessment Literacy: Perceptions and Practices". *The International Journal of Research in Teacher Education*, Vol. 6, No. 2, 2015, 26-37.

Jeong, H. "Defining assessment literacy: Is it different for language testers and non-language testers?." *Language Testing,* Vol. 30, No. 3, 2013, 345-362.

Kayler, M, and Weller, K. "Pedagogy, self-assessment, and online discussion groups." *Journal of Educational Technology & Society,* Vol. 10, No.1, 2007, 136-147.

Khodabakhshzadeh, H., Kafi, Z., and Hosseinnia, M. "Investigating EFL teachers' conceptions and literacy of formative assessment: Constructing and validating an inventory." *International Journal of Instruction,* Vol. 11, No. 1, 2018, 139-152.

Kögler, H. H. Agency and the other: On the intersubjective roots of self-identity. *New Ideas in Psychology*, Vol. 30, No. 1, 2012, 47-64.

Krejcie, R. V., and Morgan, D. W. "Determining sample size for research Activities." *Educational and psychological measurement*, Vol. 30, No. 3, 1970, 607-610.

Lam, R. "Language assessment training in Hong Kong: Implications for language assessment literacy." *Language Testing*, Vol. 32, No. 2, 2015, 169-197.

Lam, R. "Teacher Assessment Literacy: Surveying Knowledge, Conceptions and Practices of Classroom-based Writing Assessment in Hong Kong". *System*, Vol. 81, 2019, 78-89.

Looney, A., Cumming, J., van Der Kleij, F., and Harris, K. "Re-conceptualizing the role of teachers as assessors: Teacher assessment identity." *Assessment in Education: Principles, Policy & Practice*, Vol. 25, No. 5, 2018, 442-467.

Mackey, A., and Gass, S. M. *Second language research: Methodology and design*. 2nd ed., Mahwah, NJ: Lawrence Erlbaum Associates, 2016.

Malone, M. E. "The Essentials of Assessment Literacy: Contrasts between Testers and Users". *Language Testing*, Vol. 30, No. 3, 2013, 329-344.

McMillan, J. H. "Fundamental assessment principles for teachers and school administrators." *Practical Assessment, Research & Evaluation*, Vol. 7, No. 8, 2000, 1-9.

McTighe, J., and O'Connor, K. "Seven Practices for Effective Teaching." *Educational Leadership,* Vol. 63, No. 3, 2005, 10-17.

Mellati, M., and Khademi, M. "Exploring Teachers' Assessment Literacy: Impact on Learners' Writing Achievements and Implications for Teacher

Development." *Australian Journal of Teacher Education*, Vol. 43, No. 6, 2018, 1-18.

Mertler, C. A. "Secondary Teachers' Assessment Literacy: Does Classroom

Experience Make a Difference?." *American Secondary Education*, Vol. 33, No. 2, 2005, 76-92.

Mertler, C. A. "Teachers' Assessment Knowledge and Their Perceptions of the Impact of Classroom Assessment Professional Development." *Improving Schools,* Vol. 12, No. 1, 2009, 101-113.

Mertler, Craig A., and Cynthia Campbell. "Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the Assessment Literacy Inventory." *Online Submission* (2005).

Mohammadi, A. *A mixed-methods study on the teacher assessment literacy of ELT instructors versus content instructors Islamic Azad University*. 2020. University of Qom, Iran, Unpublished dissertation.

Nazari, Naghmeh, et al. "Evaluating Novice and Experienced EFL Teachers' Perceived TPACK for Their Professional Development." *Cogent Education*, Vol. 6, No. 1, 2019, 1-26.

Plake, B. S., and Impara, J. C. "Teacher Assessment Literacy: What Do Teachers Know about Assessment?." *Handbook of classroom assessment*. Vol. 3, No. 1, 1996. 53-68.

Pilcher, J. K. "The standards and integrating instructional and assessment practices." *ERIC,* Vol. 2, No. 1, 2001, 451-190.

Popham, W. J. *Classroom Assessment: What Teachers Need to Know.* 7th ed., Boston: Pearson Education, 2014.

Ramnarain, U., and Hlatswayo, M. "Teacher beliefs and attitudes about inquiry-based learning in a rural school district in South Africa." *South African Journal of Education*, Vol. 38, No. 1, 2018, 1-10.

Razavipour, K, et al. "On the interaction of Test Washback and Teacher Assessment Literacy: The Case of Iranian EFL Secondary School Teachers." *English Language Teaching*, Vol. 4, No. 1, 2011, 156-161.

Remesal, A. "Educational Reform and Primary and Secondary Teachers' Conceptions of Assessment: The Spanish Instance, Building upon Black and Wiliam." *The Curriculum Journal*, Vol. 18, No. 1, 2007, 27-38.

Rodríguez, A. G., and McKay, S. "Professional Development for Experienced Teachers Working with Adult English Language Learners. CAELA Network Brief." *Center for Adult English Language Acquisition*, Vol. 4, No. 2, 2010, 1-13.

Sadeghi, K., and Khezrlou, S. "Burnout among English language teachers in Iran: Do socio-demographic characteristics matter?." *Procedia-Social and Behavioral Sciences*, Vol. 98, No. 2, 2014, 1590-1598.

Scarino, A. "Language Assessment Literacy as Self-awareness: Understanding the Role of Interpretation in Assessment and in Teacher Learning." *Australia Language Testing,* Vol 30, No. 3, 2013, 309-327.

Schaefer, W. D., and Lissits, R. W. "Measurement Training for School Personnel: Recommendations and Reality." *Journal of Teacher Education*, Vol. 38, No. 3, 1987, 57-63.

Schunk, D. H. (2012). *Social cognitive theory.* In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA handbooks in psychology®. APA educational psychology handbook, Vol. 1. Theories, constructs, and critical issues* (p. 101–123). American Psychological Association. https://doi.org/10.1037/13273-005.

Siegel, M. A., and Wissehr, C. "Preparing for the plunge: Pre-service teachers' assessment literacy." *Journal of Science Teacher Education,* Vol. 22, No. 4, 2011, 371-391.

Spolsky, B. *Measured Words*. Oxford: Oxford University Press, 1995.

Smith, L. F., and Galvin, R. "Toward assessment readiness: An embedded approach in primary initial teacher education." *Assessment Matters,* Vol. 7, No. 2, 2014, 39-63.

Stiggins, R. J. "Are You Assessment Literate?." *The High School Journal,* Vol. 6, No. 5, 1999, 20-23.

Stobart, G. *Testing times: The uses and abuses of assessment*. Oxon: Routledge, 2008.

Sussman, R., and Gifford, R. "Causality in the theory of planned behavior." *Personality and Social Psychology Bulletin, Vol.* 45, No. 6, 2019, 920-933.

Tajeddin, Z., et al. "Classroom Assessment Literacy for Speaking: Exploring Novice and Experienced English Language Teachers' Knowledge and Practice." *Iranian Journal of Language Teaching Research*, Vol. 6, No. 3, 2018, 57-77.

Tao, N. *Development and validation of classroom assessment literacy scales: English as a Foreign Language (EFL) teachers in a Cambodian Higher Education Setting.* 2014. Victoria University, Australia, Unpublished dissertation.

Tavassoli, K., and Farhady, H. "Assessment Knowledge Needs of EFL Teachers." *Teaching English Language*, Vol. 12, No. 2, 2018, 45-65.

Taylor, L. "Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections." *Language Testing,* Vol. 30, No. 3, 2013, 403-412.

Volante, L., and Fazio, X. "Exploring Teacher Candidates' Assessment Literacy: Implications for Teacher Education Reform and Professional Development." *Canadian Journal of Education*, Vol. 30, No. 3, 2007, 749-770.

Wang, J., et al. "Pre-service Teachers' Initial Conceptions about Assessment of Science Learning: The Coherence with Their Views of Learning Science." *Teaching and Teacher Education,* Vol. 26, No. 3, 2010, 522-529.

Watmani, R., Asadollahfam, H., and Behin, B. "Demystifying language assessment literacy among high school teachers of English as a foreign language in Iran: Implications for teacher education reforms." *International Journal of Language Testing,* Vol. 10, No. 2, 2020, 129-144.

Williams, M., and Burden, R. "Students' developing conceptions of themselves as language learners." *The Modern Language Journal,* Vol. 83, No. 2, 1999, 193-201.

Xu, Y., and Brown, G. "Teacher Assessment Literacy in Practice:Reconceptualization." *Teaching and Teacher Education,* Vol. 58, 2016, 149-162.

Yzer, M. C. "Reasoned action theory. The SAGE handbook of persuasion." *Developments in Theory and Practice,* Vol. 2, No. 2, 2013, 120-136.

Yzer, M. "Theory of reasoned action and theory of planned behavior." *The international encyclopedia of media effects,* Vol. 2, No. 3, 2017, 1-7.

Zamani, R., and Ahangari, S. "Characteristics of an Effective English Language Teacher (EELT) as Perceived by Learners of English." *International Journal of Foreign Language Teaching and Research,* Vol. 4, No. 14, 2016, 69-88.

Zwozdiak-Myers, P. *The teacher's reflective practice handbook: How to engage effectively in professional development and build a portfolio of practice.* Routledge, 2012.

Appendix A

Classroom Observation Checklist

Part A: Ethnographic Features

1. Education: M.A………………… Ph.D.…………………………
2. Gender: Male / Female
3. Age: …………………….
4. Work Experience:………….Years
5. Taken part in in-service training for assessment …………..Yes/ No

Part B: The Teacher's Classroom Behavior

(Little=1), (A little=2), (Into some extent=3), (Much=4), (Very Much=5)

Standards	No.	Description of observed behavior	1	2	3	4	5
1. Choosing appropriate assessment methods	1	The teacher employs appropriate assessment methods such as essay type items, oral group work, real life tasks, and creative performance tasks.					
	2	S/he makes use of reliable standardized tests as well as teacher made tests.					
	3	The teacher uses appropriate assessment methods such as oral and written quizzes.					
	4	The teacher uses tasks aiming at assessing the students' knowledge in line with instructional objectives.					
	5	The teacher develops appropriate assessment methods such as oral and					

		written exams in line with instructional objectives.					
2.　Developing appropriate assessment methods	6	The teacher tries to use valid tests and match an outline of the instructional content to the content of the actual assessment.					
	7	The teacher looks at the student's records and standardized test scores to see which topics the students had not performed well on previously.					
	8	The teacher develops different testing and performance tasks such as comparing, contrasting, criticizing, identifying, specifying, listing, ordering, matching, selecting, defining, recalling, and restating.					
3.Administering, scoring, and interpreting the	9	The teacher administers the test or performativity tasks and scores the students' papers or their performance at the same session.					

results of assessments	10	The teacher interprets the results of assessments in the class.					
	11	Students are allowed to check their problems with the teachers and be fully informed of the nature of their scores.					
	12	Student know that a high percentage of the instructional content of the unit is covered by the test or the task.					
4. Using assessment results to make decisions	13	The teacher checks for prerequisite knowledge in her/his students before she begins the new unit.					
	14	The teacher builds up the new lesson based on the previous performance of the learners.					
	15	The teacher uses both Norm-referenced and Criterion-referenced information to plan for instruction.					
	16	The teacher informs the students of the nature of the final tests which are standardized tests by the ministry of education.					

	17	The teacher develops valid grading procedures as s/he explains exactly how the grade was determined.					
5. Developing valid grading procedures	18	The teacher makes use of oral responses to questions asked in class of each student over the course of the unit to gain the most reliable student-performance.					
	19	The teacher shows the parents samples of the student's work.					
	20	The teacher permits students to redo their assignments several times if they need more opportunities to meet his/her standards for grades.					
	21	The teacher reports the assessment results to the school board every session.					
6.Communicating assessment results	22	The teacher focuses on the learners' score in his own class, while discussing the case of weak students with the school counselor.					

	23	The teacher is sensitive to the development of leaners.					
	24	The teacher considers the results of both formal and informal assessment types while reporting the students' performance.					
7.Recognizing unethical or illegal practices	25	The teacher recognizes unethical or illegal practices as s/he does not announce the scores in public.					
	26	The teacher does not belittle the weak students in front of their classmates.					
	27	Students cannot see the other students' graded tests, so no violation of the students' right of privacy is allowed.					
	28	The teacher makes use of planned instructions to focus on the concepts covered in the test.					

Appendix B
Interview Guide
• How do you choose an appropriate assessment method?
• Do you usually develop tests, design tasks, and design interviews to assess your students? Explain the process to me, please.
• How do you administer, score, and interpret the results of assessments?
• Do you ever use assessment results to make decisions? How?
• How do you develop valid grading procedures?
• How do you communicate assessment results?
• Do you ever recognize unethical or illegal practices? How?
• Have you taken part in any in-service training courses considering assessment issues recently? Do you feel it is needed?
• Is there anything significant concerning testing and assessment you would like to mention?