Iranian EFL Teachers' Classroom Assessment Practices: Discrepancy between Theory and Practice

Mohammad Aliakbar[1], Ali Yasini[2], Setayesh Sadeghi[3]*

Abstract
Classroom assessment practices play an important role in increasing students' learning. These practices are also key to the success of curriculum reforms. In line with this importance, the current study aimed to explore Iranian EFL teachers' classroom assessment practices following the new curriculum reform introduced in 2011. It also examined whether there were differences in teachers' practices based on variables such as gender, teaching experience, and educational degree through qualitative and quantitative approaches. A total of 28 EFL teachers at public high schools from three provinces in Iran (Lorestan, Kermanshah, and Ilam) participated in the semi-structured interviews. After transcribing the interviews, recurrent themes were identified by thematic analysis. The results revealed that English teachers used primarily *assessment for learning* or *formative assessment*, yet their assessments were still exam-based. The frequency of assessment practices indicated that oral questioning during instruction (60.7%), test after each lesson or unit of lesson (50%), written exams (35.7%), and group activity (32.1%) were the most frequently used assessment practices among teachers. The findings revealed that there is a discrepancy between teachers' classroom assessment practices and the curriculum reform. Across gender, there was a significant difference between participants in terms of role play, assigning summaries of lessons, and playing audio files. Regarding teaching experience and academic degree, there was no significant difference between participants. The findings would have implications for teacher education programs and EFL teachers.

Keywords: Classroom assessment; EFL teachers; Formative assessment; Practices

1. Introduction

The ELT curriculum in Iran has undergone some significant changes in methodology since the 1970s. The last curriculum reform was presented in 2011. This reform was seen as a transition from traditional testing to communicative teaching and testing. It was embarked on and implemented as a result of the fact that students were unable to communicate in English after studying English for six years at state schools (Ganji et al., 2018)**.** In the new curriculum, the purpose of teaching English at schools is to help students develop their English communication skills (Aghagolzadeh & Davari, 2014); English textbooks have altered to align

---
[1] English Department, Ilam University, Email: m.aliakbari@ilam.ac.ir. ORCID: orcid.org/0000-0002-6189-5464
[2] Management Department, Ilam University, Email: a.yasini@ilam.ac.ir. ORCID: 0000-0002-8556-6497
[3] Corresponding author: English Department, Ilam University, Email: sadeghy.setayesh @gmail.com. ORCID: 0000-0002-2034-7867

with this goal and teachers are now expected to focus on all four language skills rather than simply focusing on reading skills (Atai & Mazlum, 2013). The new reform was implemented through the presentation of a new series of English textbooks entitled *Prospect* (for junior high school students) and *Vision* (for high school students), to raise Iranian students' ability to communicate in English.

Teaching and assessment are inseparable since teaching is always followed by an assessment to determine the student's understanding of the subject. Additionally, assessment can be employed as an instrument to determine if students have achieved the learning goals or if more intensive learning improvements are needed. Any change in curriculum always embraces changes in assessment, and success of the curriculum reform rests on assessment policies and classroom practices (Razavipour & Rezagah, 2018). The reformed Iranian assessment policies emphasize assessing communicating competence in all the four language skills areas. Furthermore, these policies which are echoed in national macro-level policy documents, e.g. National Curriculum Document (2013), mandate assessment standards.

The documents, as mentioned in Firoozi et al. (2019), highlight various forms of diagnostic, progressive, and formative assessments. Teachers are required to opt for alternative forms of assessment to assess student's learning progress, achievements, and quality of education. For instance, one of the standards says "assessment includes assessments as diagnostic, progressive, and formative that can be achieved through paper-and-pencil, oral, peer-assessment, performance appraisal, self-assessment, and informal observation. Decisions on student advancement and educational achievement are carried out by the assessments" (National Curriculum Document [NCD]), 2013, p.126) and " to report qualitatively and quantitatively on students' learning outcomes, all oral, essay, performance, written, observation, and portfolio assessment methods must be conducted" (NCD, p. 130).

Following the implementation of the Iranian curriculum reform, many studies were conducted to compare the new and the old series of the textbooks (e.g., Ahour & Golpour, 2013; Sardabi & Koosha, 2015). Nevertheless, a few studies have concentrated exclusively on classroom assessment practices. In order to bridge this disparity, the current study intends to explore whether the classroom assessment practices of teachers are aligned with the reform. More specifically the study aims to examine what classroom assessment practices Iranian EFL teachers use and exercise in their classrooms.

2. Review of Literature

2.1. English Language Assessment in Iran

After the 1979 Islamic Revolution, Iranian educational system underwent a wave of reforms. Radical changes in the curriculum occurred at all levels of education with the aim of Islamizing the system. The centralized Education Ministry holds complete control over the system including textbooks, assessment guidelines, and employment of staff (Paivandi, 2012). The changes were introduced to ELT materials and programs as well. The English textbooks were first revised according to GTM and ALM, focusing on grammar and reading with little attention paid to other language skills. Dissatisfied with the outcomes of the methods, Education Ministry developed an educational innovation with the goal of changing both textbook content and pedagogy (Atai & Mazlum, 2013). Formerly, reading and grammar were

the main goals of the curriculum, while in the new program, problem-solving and communication skills are the main goals of education (Aghagolzadeh & Davari, 2014).

In regard to English language assessment, educational assessment has traditionally focused on discrete points, aiming at testing the fragments of linguistic knowledge rather than the communication skills. Simply put, for all school teachers and students, the concept of a language test is often equivalent to a series of discrete-point, often multiple-choice, written exam items (Farhady & Hedayati, 2009). Although the evidence of negative washback effect of summative approach for measuring fragments of English language proficiency has been presented, it has still remained the dominant approach for decades (Riazi & Razavipour, 2011). However, the last reform in ELT in 2011 focused on assessing communication skills in the four domains of language proficiency. Recognizing the power of assessments to shape and detour programs and defeat reforms, policymakers are calling for a move from testing fragments of linguistic knowledge to assessing the communication skills and assessing learning (Farhady & Hedayati, 2009).

This approach contrasts with the country's decades-long tradition of language testing, which concentrated solely on reading skills, isolated vocabulary, and grammar. According to the requirements of the new program, all the four language proficiency domains must be assessed, and cumulative scores should be calculated to make the final pass/fail decision (Razavipour & Rezagah, 2018). In grades 9 and 10, teachers themselves develop, administer, and score exams. In the 11th grade final exams take the form of standardized tests that are developed by language testing experts and supervised by the Center for Evaluation and Monitoring (CEM).

2.2. Classroom Assessment

Assessment is considered integral to education. Classroom assessment encompasses all activities that learners and teachers perform to obtain information that can be utilized for diagnostic purposes to change learning and teaching (Black & William, 1998). The process of classroom assessment requires teachers to gather, analyze, and use evidence about student' learning in order to diagnose students' strengths and weaknesses, track their progress to achieve desired competence levels, evaluate them, and inform parents" (McMillan, 2013, p. 4). It also deals with improving not only student learning and motivation to learn, but also classroom teaching (Gronlund, 2006; Shavelson et al. 2008). In other words, classroom assessments practices are an impactful means of affecting both learning outcomes and motivation levels. Defined as a rigorous process, these assessments aim to reveal not only what is really learned, what the students know, but also how effective the teacher is (McMillan, 2013). Moreover, classroom assessments help bring forth comprehensive understanding of students' inclinations and depth of acquisition. They are also designed to support teachers' decision-making processes in teaching and learning (McMillan, 2015). According to Cheng et al. (2004), "any model of the teaching-learning process entails that teachers make decisions (concerning instruction, grading, and reporting) based on their knowledge of student achievement and progress toward desired learning outcomes". Thus, teachers continually evaluate students to support their progress in teaching and learning (Fulcher & Davidson, 2007).

Classroom assessment practices can be explicit, focusing on summative assessments, or implicit, focusing on formative assessments (Bachman & Palmer, 2010). Some researchers equate classroom assessment to formative assessment, while others regard it as inclusive of both formative and summative functions of classroom assessment (Brookhart, 2004). In the present paper, the notion of classroom assessment embraces both formative and summative assessments.

Summative assessment or *assessment of learning* (McMillan, 2015) often places focus on product and aims to monitor learning outcomes, for external accountability purposes (Dixson &Worrell, 2016). At the end of the unit or term, this assessment is conducted to ascertain whether the students have comprehended the subject matter taught (Ainsworth & Viegut, 2006). Summative formal assessments usually use traditional paper-and-pencil exams and are simply followed by a score without additional feedback. Lack of feedback leads to a lack of diagnostic information, and students are hardly aware of their weaknesses (Ainsworth & Viegut, 2006).

In contrast, formative assessment (FA) or *assessment for teaching* (McMillan (2015) has been widely acknowledged as a highly effective method for improving both teaching and learning (Alahmadi et al., 2019; Bachman & Palmer, 1996; Buyukkarci & Sahinkarakas, 2021; Cheng et al., 2004; Davison, 2004; Rea-Dickins & Gardner, 2000). The objective of the "assessment for learning" approach is to motivate the students, provide feedback, and identify areas need further learning (McMillan, 2015). Inspired by criterion-referenced measurement (Glaser, 1994) and formative evaluation (Scriven, 1991), formative assessment studies have gained increasing attention since the 1990s in the fields of general education and language assessment in particular (Wang, 2017).

Formative assessments involve a set of formal and informal assessment methods used by teachers during the process of learning to adjust learning and teaching activities for enhancing student achievement (Crooks, 2001). They are carried out via self-assessments, peer assessments, and informal anecdotal evidence gathering (McMillan, 2015). Similarly, Ellis (2003) argues that formative assessment can be classified into two types: planned and incidental assessments. The planned formative assessment involves the use of formal assessment instruments and procedures, including tests and quizzes. The incidental formative assessment is embedded in the daily instructional activities of the classroom. This latter assessment can be realized in two types of external and internal assessments. The external formative assessment comprises teacher and student reflections of performance either during the activity or upon its completion and the internal formative assessment occurs when teacher asks questions and offers immediate feedback for learners' performance during the completion of tasks. Indeed, the defining feature of formative assessment is feedback from the teacher (Ellis, 2003).

Considering that the classroom assessment practices applied by teachers reflect their educational processes, the type of assessment, frequency in which it is carried out, and the student feedback the assessment provides are important in terms of understanding how classroom assessment practices are progressing (Brookhart, 1997). For this reason, the study aimed to investigate how teachers are implementing their classroom assessment practices,

whether these assessment practices are being carried out in accordance with the current educational paradigms and the factors that affect these practices.

In EFL context, the construct and practices of formative assessment did not receive attention until 2000 though its effectiveness and help had long been acknowledged by both researchers and teachers (Cheng et al., 2004; Rea-Dickins, 2004). A number of studies have indicated that formative assessment is indispensable in ESL/EFL teaching and learning (Hazim Jawad, 2020; Ismail et al., 2022; Naka, 2023; Rea-Dickins & Gardner, 2000; Ozan & Kincal, 2018).

The literature on teachers' classroom assessment practices revealed that investigations of EFL teachers' assessment practices have been focused on teachers' beliefs about assessment practices (Davison, 2004; Saad et al., 2013), and teachers' assessment literacy (e.g., Coombe et al., 2012; Öz & Atay, 2017; Yamtim & Wongwanich, 2014). In the Iranian EFL context, the great majority of the published articles on language assessment have been on dynamic assessment (Karimi & Shafiee, 2014; Mardani & Tavakoli, 2011; Momeni & Nushi, 2022; Naeini & Duvali, 2012; Pishgadam et al., 2011). Furthermore, some studies have dealt with teachers' perspectives on assessment literacy (Alavi et al., 2022; Dehqan & Asadian Sorkhi, 2020; Jalilzadeh et al., 2022; Watmani, Asadollahfam, & Behin, 2020) and the assessment needs of Iranian EFL teachers (Firoozi et al., 2019; Razavipour & Rezagah, 2018). For instance, Firoozi et al. (2019) examined the needs of Iranian EFL teachers for language assessment literacy with an emphasis on reformed assessment policies. The findings revealed that teachers' current perceptions of language assessment must change to comply with the requirements of the noted reform. Moreover, teachers require mastery of both the knowledge and skills to assess language proficiency.

Although the literature indicated studies on language assessment in EFL classrooms, none of the studies alluded to what assessment practices are exercised in the classrooms. This gap motivated the authors of the present study to investigate what classroom assessment practices are implemented and whether they are matched to the current ELT curriculum reform. In addition, this study aims to broaden the current investigation on classroom assessment practices by exploring teachers' classroom assessment practices in relation to gender, teaching experience, and academic degrees in public high schools. To achieve this objective, the following questions guide the study:

1) What classroom assessment practices are implemented in Iranian EFL classrooms?
2) Do Iranian EFL teachers' assessment practices vary across gender, teaching experience, and academic degree?

3. Method

3.1. Design of the Study

The study adopted a qualitative approach. The study used semi-structured interviews to probe into EFL teachers' perceptions of their assessment practices in the classroom. Thus, the design of the study was exploratory. This was followed by quantitative investigations for the aim of examining the frequency calculations and potential significant differences between the participants with regard to gender, teaching experience and academic degree.

3.2. Participants

The study participants were chosen using the convenience sampling method. Convenience sampling involves a non-probability sampling, which focuses on the target population members who meet certain practical criteria. For instance, they are easy to access, available at a particular time, or willing to take part in the study (Dörnyei, 2007). A total of 28 EFL teachers (16 male and 12 female) from three provinces of Iran, namely Lorestan, Ilam, and Kermanshah participated in this study, within the age range of 37 to 55. Their teaching experiences varied between 5 and 32 years. They hold various degrees of Bachelor, Master, and Ph.D. Ten teachers had Bachelor degree, and 12 teachers were Master holders. Two of them held Ph.D. degree, and four Ph.D. candidates. The participants' demographic details are presented in Table 1.

Table 1
Demographic Details of Participants in the Study

Province	Gender	Age	Teaching experience	Educational degree
Ilam (14)	Female: 7 Male:7	37-52	5-32	Bachelor:5 Master:6 Doctorate:3
Lorestan (8)	Female: 4 Male:4	42-49	11-30	Bachelor:2 Master:5 Doctorate:1
Kermanshah (8)	Female:3 Male:5	37-55	6-29	Bachelor:2 Master:4 Doctorate:2

3.3. Instruments

The instruments used in this study included interview questions and teachers were interviewed using a semi-structured approach. For the respondents' convenience and the accuracy of the responses, the interviews were mainly conducted in Persian. Each interview lasted approximately 30 minutes. Sample interview questions include "Could you please mention some specific ways to ascertain students' understanding or confusion about English language learning?", and "Could you give some examples of how you assess your students?" .

To answer the second research question, i.e., "Do Iranian EFL teachers' assessment practices differ across gender, teaching experience, and academic degree?", descriptive statistics, including frequency and percentage, and non-paramedic tests, such as Mann-Whitney U and Kruskal Wallis, were conducted to examine possible significant differences between and among the participants in relation to gender, teaching experience and academic degree.

3.4. Data Collection and Analysis

Data were collected through one-shot interviews. After transcribing the interviews, recurrent themes were distinguished by inductive thematic analysis following the standards set

up by Braun and Clarke (2006). In so doing some steps were taken. The transcribed interviews were read several times to identify initial ideas. The responses were then coded as a small phrase or keyword representing a particular element. Afterwards, the collected codes were searched for potential themes. Themes were recognized based on the number of previously coded important elements which different respondents clearly stated in their responses to the questions. The themes were also checked to see if they work in relation to the coded extracts and the entire data set, creating a thematic map from the analysis. Finally, clear definitions and names were created for each theme (Braun & Clarke, 2006). After coding and quantitizing the data (Dörnyei, 2007), the basic themes were identified and their frequency of occurrences was counted and tabulated. The interviewees were assured of anonymity regarding their names, and abbreviations such as "Teacher 3" or "T3" were used instead.

The trustworthiness of the results was demonstrated through the researcher triangulation strategy and member checking to clarify and assess the accuracy of results (Miles et al., 2014). For researcher triangulation two of the researchers of the study read the interviews transcripts and key portions of transcripts were discussed throughout the process of reading. The findings were then compared and inter-rater reliability was calculated as percent agreement, being equivalent to 83. The concepts and categories together with the final categorization of teachers' perceptions on classroom assessment practices were given to some participants for confirmation, clarification, and any feasible adjustment via member checking.

As the data were categorical and the sample of the study was small, non-paramedic tests, such as Mann-Whitney U and Kruskal Wallis, were conducted to examine possible significant differences between and among the participants in relation to gender, teaching experience and academic degree.

4. Results and Discussion

4.1. First Research Question

To respond the first research question, "What classroom assessment practices are implemented in Iranian EFL classrooms?", thematic analysis of the recorded interview transcripts, through coding of transcriptions and categorizing the related codes, resulted in the emergence of two main themes that encapsulate some sub-themes. The main themes comprised formative assessment and summative assessment.

Table 2 represents the frequency and percentage calculations of teachers' assessment practices in their classes. It can be seen that most teachers have concentrated on formative assessment in their classroom assessment practices. As Table 2 illustrates, formative assessment included two sub-themes of planned formative assessment and incidental formative assessment. The planned formative assessment practices of participants were evident in six major activities: test after each lesson (50%), written exam (35.7%), oral test (25%), quiz (25%), multiple-choice test (17.9%), and dictation (10.7%). The incidental assessment practices comprise the seven categories of oral questioning (60.7%), group activity or discussion (32.1%), playing audio files (28.6%), making conversation and new sentences (25%), summary of the lesson (17.9%), oral production (14.3%), and role play (10.7%). There are only 7 instances where participants put emphasis on the summative assessment practices, as final exams.

Table 2

The Number and Percentage of the Assessment Practices and Test Types Adopted by EFL Teachers

Assessment Practices		Test types	Number of teachers Adopting the Test Type	Percentage of Teachers Adopting the Test Type
Formative assessment Practices	Planned Formative Assessments	1) Test after each lesson or unit of study	14	50.0
		2) Written Exam	10	35.7
		3) Oral test	7	25.0
		4) Quiz	7	25.0
		5) Multiple choice test	5	17.9
		6) Dictation	3	10.7
	Incidental Formative assessments	1) Oral questioning during instruction	17	60.7
		2) Group activity	9	32.1
		3) Playing audio files	8	28.6
		4) Making conversation and new sentences	7	25.0
		5) Asking for the summary of lesson	5	17.9
		6) Oral production (speaking)	4	14.3
		7) Role play	3	10.7
Summative assessment Practices (Final Exams)			7	25.0

With respect to summative assessment, participants expressed concern about the accountability aspect of the assessment to communicate the final test results with the parents and school management, as described by one of the teachers: " *for me, summative assessment is not a criterion as a reliable indicator of students' proficiency; I consider final exams as a means of generating a report for parents and administrators" (Teacher 1).* It implies that accountability objectives of assessment can dominate teachers' assessment beliefs, suggesting a dichotomy between the accountability-driven and improvement-focused objectives of assessment (Saad et al., 2013).

Thematic analysis of data found that assessment practices of teachers were associated with teachers' concerns about students' understanding and confusion. Particularly, teachers assess students to ascertain whether the students have understood or been confused about English language learning. Applying various questions, nearly all teachers attempt to check students' understanding and evaluate their knowledge about the materials being taught. A group of these questions serves as display questions where the answer is typically known by the teacher. Other questions are employed to verify if the students follow their instructions. One of the participants also stated that she ascertains students' understanding by asking an indirect

question*, "I ask questions during the lesson, not to see if they are paying attention or not ... no . The students certainly respond that 'we have learned teacher'. I ask an indirect question. A question about whether they have learned or not. We have questions and answers in classroom every session(T10)."*

The teachers mostly acknowledged that they had assessment in every session and after each instruction, as represented by participants, "*I can find out whether students have understood the lesson or not by asking some questions during the lesson. For example, I ask a few questions about the subject that I am teaching if they understand, they can answer. If they don't, they can't answer (T 27).* Another teacher stated, "*according to dynamic assessment every moment and every session we are assessing our learners just to make sure if they have learned or not"(T6).*

It is noteworthy that teachers posed a range of questions and assessments that were tailored to the skill or component being taught. For example, some teachers stated that *"when assessing reading skill, I ask oral questions about the text under reading and with respect to vocabulary and grammar I use making-sentence practice" (T4).* Other teachers stated*, "When assessing speaking skill, I present students illustrations to discuss or write about (Ts 20, 28), or when assessing listening I play audio recordings in the classroom (T 8) or students are asked to transcribe audio recordings" (T24).* It suggests that teachers attempt to align assessment methods with intended learning objectives.

Emphasizing the peer-assessment, one teacher described checking for understanding by asking students to answer their classmates' questions related to the topic is being taught, as it is reflected in his words, "*after presenting the lesson I ask students ' Do you have any questions?'. If they had a question, I would first ask the other students to answer the question if they couldn't answer I myself would answer. In this way, I want to make sure the student has understood. This helps the teacher to engage students in classroom activities and increase interaction between students (T 16).*

Another teacher noted that through doing textbook and workbook exercises they can assess students' understanding of the lesson. When students are able to do the exercises without teacher's reinstruction and repetition, it indicates that students have understood the subject matter. Moreover, one teacher assessed the understanding by giving students a task to do. The teacher maintained that *"I ask students to develop a close test using the reading text or conversations in the textbook. This is an interesting activity. It allows students to work on vocabulary. They remove key words from the text or conversation and bring them at the bottom as a multiple-choice or even an option"(T 20).*

Along with the above-mentioned ways teachers use to evaluate their students' understanding, the teachers also maintained that facial expressions and mental states of students help them to recognize students' understanding or confusion. For example, one of the teachers explained that " *teachers can identify the extent to which the students have learned the lesson from their facial expressions and answers to the questions asked during the teaching process (T19).* Another teacher expressed*, "according to our experiences from students' facial expressions and postures we find out if they are confused or have understood. The student keeps himself busy with his textbook or tries to ask irrelevant questions and inhibit teachers' instruction "(T1).* Facial expressions, serving as a ubiquitous form of non-verbal

communication, provide a valuable source of feedback that teachers can get from students (Sathik & Jonathan, 2013).

While most teachers mentioned that they use teacher-made tests to gain information about students' advancement and identify their strengths and weaknesses and give feedback to students as they develop through the course, there were still teachers who used tests to prepare their students for standardized exams, such as Konkur, which students will need to take in the future. Konkur, Iranian University Entrance Exam, is a high stake exam that is held annually to choose students for university education. This test takes a few years of preparation for students, and success or failure at passing the test can affect the test takers' social and personal lives (Foroozandeh & Forouzani, 2015).

Though few teachers declared that they do not allow to be influenced by the impact of Konkur on their teaching, majority of the teachers acknowledged that they adapt their practices to the high-stake test and use or introduce published Konkur-preparation books. The teachers believed that students who have to acquire large amounts of vocabulary and grammar for Konkur are not interested in language activities, such as communicative and oral activities, which do not appear to be related to the national and high-stake test. It is reflected in a teacher's comment, *"unfortunately, because of Konkur and demands of reading and grammar, communicative skills are just overlooked at the cost of receptive skills" (Teacher 20).* Another teacher expressed*, "konkur is very important for students, and as it is composed of vocabulary, grammar and reading comprehension, teachers are expected to strengthen students in these three areas"*. This can definitely produce a negative washback effect (Ahmadi Safa & Sheykholmoluki, 2023; Foroozandeh & Forouzani, 2015).

Furthermore, participants repeatedly complained that though the new program of ELT education, i.e., CLT, has emphasized the assessment of speaking and listening abilities, lack of appropriate infrastructures in schools has led to unsuccessful implementation of the oral assessments, exclusively listening ability. As the teachers pointed out, inadequate audio/visual equipment impede the enactment of reformed-based assessments. The majority of teachers stated that schools do not have the audio/visual equipment needed to assess the communication abilities of students.

In line with the above-mentioned dichotomy, while the teachers acknowledged that assessing students in group work or pair work could be contributory in communicative assessment, they articulated concerns regarding logistical barriers such as limited time and the huge number of students, which can impede the effective implementation of the cooperative assessment practices and alternative assessments including peer assessment and individualized feedback.

In addition to the above-mentioned challenges, some teachers pointed out that some English teachers lack proficiency in speaking or listening skills; accordingly, teachers are not able to expound what they themselves do not practice. Furthermore, the participants claimed that teachers are not well-educated about the oral performance assessment criteria, nor are they provided with sample tasks or explicit assessment specifications. They are told not to take discrete-point tests but are left free to choose what to do. All the teachers agreed that teachers must be provided with adequate training in this regard. This finding supports the results from

previous studies which indicate that teachers around the world lack adequate knowledge base in educational assessment (Riazi & Razavipour, 2011; Stiggins, 1991).

4.2.Second Research Question

Determining the frequencies of the items of subthemes provided the basis for answering the second research question, "Do Iranian EFL teachers' assessment practices differ across gender, teaching experience, and academic degree?", which meant to investigate the possible differences in teachers' assessment practices across variables such as teaching experience, academic degree, and gender.

Table 3
The Results of Inferential Tests for Teachers' Planned and Incidental Formative Assessment Practices across Teaching Experience, Academic Degree and Gender

	Formative assessment Practices	Kruskal Wallis tests				Mann-whitney U test	
		Teaching Experience		Academic Degree		Gender	
	Chi-square Tests	value	Sig	value	Sig	value	Sig
Formative Assessment Planned Practices	Test after each lesson or unit	.244	.885	1.125	.570	82	.453
	Witten exam	.872	.647	.700	.705	86	.576
	Oral test	.081	.960	.306	.858	82	.386
	Quiz	2.693	.260	.306	.858	68	.083
	Multiple choice test	3.818	.148	2.270	.321	84	.401
	Dictation	1.402	.496	.420	.811	92	.729
Incidental Formative Assessment Practices	Oral questioning	3.424	.180	.361	.835	72	.188
	Group activity	.986	.611	3.108	.211	66	.085
	Summary presentation	.981	.612	2.465	.292	66*	**.036**
	Making conversation or new sentences	2.693	.260	.949	.622	82	.386
	Oral production	1.199	.549	1.312	.519	92	.759
	Playing audio files	.378	.828	4.050	.132	60*	**.033**
	Role play	.898	.638	.420	.811	72*	**.038**

Table 3 represents the results of inferential tests for teachers' planned and incidental formative assessment practices across teaching experience, academic degree and gender. The findings indicated there were no significant difference for participants' teaching experience and academic degree with respect to their classroom assessment practices.

However, the results of Mann-whitney U test for teachers' planned and incidental formative assessment practices across gender showed that there is significant difference between male and female participants in terms of some of incidental formative assessment practices such as summary presentation (Asymp. Sig.=0.36), playing audio files (Asymp. Sig.= .033), and role play (Asymp. Sig.=038), while in other practices there were no differences. To compare male and females with respect to the assessment practices cross-tabulation was run. The findings are presented in Table 3 below.

Table 4

Cross-tabulation of Gender and Classroom Assessment Practices (Role Play, Assigning summary, Playing Audio Files*)*

Gender * Role Play Cross tabulation				
		role play		Total
		No	Yes	
Male	Count (% within gender)	16 (100.0%)	0 (.0%)	16
	% of Total	57.1%	.0%	57.1%
Female	Count (% within gender)	9 (75.0%)	3 (25.0%)	12
	% of Total	32.1%	10.7%	42.9%
Total		89.3%	10.7%	100.0%
Gender * Assigning Summary of the Lesson Cross-tabulation				
		assigning summary of the lesson		
		No	Yes	Total
Male	Count (% within gender)	11 (68.8%)	5 (31.2%)	16
	% of Total	39.2%	17.9%	57.1%
Female	Count (% within gender)	9 (75.0%)	3 (25.0%)	12
	% of Total	42.9%	.0%	42.9%
Total		82.1%	17.9%	100.0%
Gender * Playing Audio Files Cross-tabulation				
		playing audio files		
		No	Yes	Total
Male	Count (% within gender)	14 (87.5%)	2 (12.5%)	16
	% of Total	50.0%	7.1%	57.1%
Female	Count (% within gender)	6 (50.0%)	6 (50.0%)	12
	% of Total	21.4%	21.4%	42.9%
Total		71.4%	28.6%	100.0%

Table 4 indicates that out of 16 male participants no one exercised the role play assessment practice while 3 female participants (25%) performed the "role play" to assess students. Regarding "assigning summary of the lesson", male participants reported a greater likelihood to use the classroom assessment practice (31. 2 %). Also, female participants were more willing to play audio files (50%). The findings demonstrated that teachers' assessment practices could not be predicted from their years of teaching, and educational degree, but gender may influence the enactment of classroom assessment practices.

The results are partially consistent with findings from previous studies. In 2012, Pishghadam and Shayesteh examined the Iranian EFL teachers' conceptual assessment beliefs with respect to major, degree, age, gender, and experience, and found that there was no significant relationship between the perceptions of assessment of teachers and their level of experience. On the contrary, Öz (2014) examined teachers' preferences of assessment methods in the English as a foreign language (EFL) classroom and their Assessment for Learning (AFL) strategies in Turkey. The researcher found that there was a statistically significant difference among EFL teachers in regard to teaching experience. Likewise, Estaji and Fassihi (2016) revealed that no significant correlation existed between the teachers' enactment of formative assessment strategies and the gender and level of experience variables.

5. Conclusion

This study aimed to explore Iranian EFL teachers' classroom assessment practices at public high schools. Furthermore, it intended to realize whether there was a significant difference between teachers' assessment practices with respect to variables such as teaching experience, academic degree, and gender. To this aim, the data gleaned from interviews with 28 teachers were analyzed qualitatively and quantitatively.

The results from Kruskal Wallis indicated there was no significant main effect for participants' teaching experience and academic degree in relation to their assessment practices. This reflects that teachers' teaching experience and academic degree may not be predictive factors influencing classroom assessment practices. However, Mann-Whitney U tests revealed a significant difference between male and female participants with respect to three incidental formative assessment practices such as role play, summary presentation and playing audio files.

The findings from thematic analysis suggest that the EFL teachers exercise both summative and formative assessment practices to ascertain their students' understanding, with special focus on formative assessment practices. The frequency of assessment practices indicated that oral questioning during instruction (60.7%), test after each lesson or unit of lesson (50%), written exams (35.7%), and group activity (32.1%) were used more frequently than other practices. These practices are used for assessing students' presentations, correcting their errors, and giving feedback.

The findings indicated Iranian EFL teachers' willingness to utilize different classroom assessment practices and activities to leverage student understanding and measure how much learning has taken place. Teachers implement their classroom assessments as an essential part of their instruction and assess students learning to receive feedback on their understanding of the knowledge, skills and content being taught so that they can identify their students' weaknesses. Exercising various and multiple practices in assessing students ensure the high

quality of assessments and permit teachers to obtain sufficient, precise evidence of student learning (Nitko, 2001; Stiggins et al., 2006*)*.

Although the findings of the study revealed that EFL teachers implement formative assessment practices, but there is still no room for formative assessment such as peer assessment, and self-assessment which are considered the crucial components for monitoring language learning and teaching (Spiller, 2012). One feature of assessment for learning is that "the self- and peer-assessment process encourages students to repeatedly scrutinize the assessment criteria and articulate their impressions with the help of teacher feedback" (Chow & Leung, 2011, p. 142). Spiller (2012) reasoned that when students receive feedback from their peers they receive a wide range of ideas about their work to encourage development and improvement. By involving peers in the assessment, students can be more autonomous, responsible and proactive (Assessment Reform Group [ARG], 2002).

Indeed, the analysis of interviews proposes that the classroom assessment practices have almost stopped at a superficial level and have failed to foster students' communication skills. Teachers are still struggling to the new curriculum, and the reform seems to be too much for teachers to apply (Gholamnejad & Raeisi-Vanani, 2021; Razavipour & Rezagah, 2018). Definitely, although the official curriculum documents emphasize alternative assessment forms, assessment practices of teachers are still exam-based and rely on written exams and tests, not being matched with the assessment ideals advocated by the new ELT curriculum. This finding is in congruent with prior studies of the Iranian literature (Atai & Mazlum, 2013; Firoozi et al., 2019; Razavipour & Rezagah, 2018; Zolfaghari & Ahmadi, 2016).

The data analysis also revealed that contextual challenges present in public high schools pose obstacles to the effective execution of classroom assessment compatible with the new program reform. As participants pointed out, the contextual difficulties such as high-stake tests such as Konkur, limited class time, and large class size pose numerous problems. They stated that Konkur affects the enactment of communicative assessments. This finding supports prior studies (Foroozandeh & Forouzani, 2015; Ghorbani & Neissari, 2015; Jalilzadeh et al., 2022) which declared that the nationwide and large-scale exams, such as the Konkur in Iran, are commonly assumed to have an effect on teachers' instructing and students' learning. The participants mentioned that due to the size of the class and the limited classroom time they are unable to engage students in performing group and peer activities. They added as the class time is only three hours per week, when they are conducting speaking and listening assessments the time is getting over and all students may not be assessed. Thus, in the following sessions, they have to pursue the traditional approach of assessment to compensate for the time.

Participants also expressed that it is troublesome to conduct CLT assessments in large classes and hence they attempt to deliver written exams, oral tests, multiple-choice and short-answer tests on the lessons of the textbook and delegate students to learn grammar and vocabulary to control over the whole class. While education scholars advocate for language classes to have no more than 15 students (Holliday, 1996; Richards, 2001), allowing teacher to "listen what students say and correct their mistakes where necessary" (Holliday, 1996, p. 6), there are more than 25 students in Iranian language classes. Since the contextual difficulties such as limited class time and large class size have repeatedly reported in almost all the previous studies in the Iranian literature, it is highly recommended that policy-makers,

curriculum-developers, and administrators take into account these features in their decision-making processes.

Another aspect of the contextual problems can be related to teaching resources. To be able to do speaking and listening assessments, schools need to be provided with some infrastructures. For example, to assess listening comprehension, schools need a language laboratory equipped with essential amenities or a trustworthy audio device**.** As claimed by the teachers, insufficient audio/visual equipment impedes the implementation of the classroom assessments. According to most teachers, schools lack the appropriate audio/visual equipment required to assess students' communication abilities. Teachers need to play the listening audio recordings of the textbooks in the classroom, but unfortunately most schools do not have the audio-visual equipment such as language labs. These finding agree with the results reported in studies conducted by Firoozi et al. (2019), Gholamnejad and Raeisi-Vanani (2021), and Razavipour and Rezagah (2018).

Moreover, the interviews demonstrated that teachers require training both in professional and pedagogical expertise of classroom assessment since the development of oral proficiency tests requires innovation, resourcefulness, time, and enthusiastic practitioners (Razavipour & Rezagah, 2018). Teachers need to know how to meaningfully communicate in English as well as how to assess the construct of communicative competence to assess communicative language abilities of their students (Morrow, 2018). It suggests that teacher training courses that can improve teachers' speaking and listening abilities are crucial for teachers to conduct appropriate performance assessments. Furthermore, teachers require precise assessment criteria and specifications in performing classroom assessment practices. As mentioned in Firoozi et al. (2019), the proposed assessment criteria provided by Iranian Education Ministry, known as *Barombandi*, do not meet the needs of teachers in the practical application of the assessment guidelines. Lack of clear criteria are significant challenges encountered by teachers, which necessitate the need for further development of clear criteria**.**

A short glance at the data above is enough to recognize that teachers' classroom assessment practices do not fulfill the necessities of communicative language testing. The assessment practices offer little opportunities for significant interaction as they are primarily based on written tests, paper-and-pencil exams or quizzes, and memorized information from textbooks. This situation suggests that there is a discrepancy between teachers' classroom assessment practices and assessment guidelines in Official Documents of assessment since the assessment guidelines appear to be cutting-edge and look for to facilitate the assessment of communicative skills, and teachers are advised to utilize alternative assessments, such as self-assessment, portfolio, and informal observations. Confirming the results of the prior studies, the findings from the study acknowledge that the curriculum reform in terms of classroom assessment has not achieved its goals. It leads to nowhere if the required infrastructures are not provided and modified.

To reduce the discrepancy, English communication skills of teachers need to be enhanced through in-service or pre-service programs. When this is done, it is imperative that teachers undergo training to master the art of creating and utilizing rubrics to methodically assess, quantify, and record the English communication skills of their students (Xu & Brown, 2016). Likewise, English teachers need to have the professional and logistic resources to move

from traditional exams focusing on discrete point testing to more genuine, performance assessments.

The results of this study provide pedagogical implications for professional development. They will provide knowledge to teachers, educators and curriculum developers on the types of assessment practices to incorporate into teaching and learning, and also the need for assessment training that would provide teachers with the expertise and confidence to use a variety of communicative assessment practices. It is essential to train and update teachers in their understanding of English proficiency and assessment practices, which must be an ongoing and closely monitored process. In order to make teaching and assessing more engaging and interactive, it is also necessary to enhance the required infrastructure of English classrooms through the integration of technological equipment.

Based on the results obtained from the study, some suggestions for further research can be recommended. As the current study employed the interview as the main data collection instrument, in order to achieve a more comprehensive understanding, other studies can triangulate the results by observing teachers' classrooms assessment practices and/or investigating students' perspectives and expectations. In addition, further research can be conducted on the teachers' sources for assessments. In order to enhance the generalizability and trustworthiness of the results, it is recommended to carry out a large-scale and nationwide duplication of this study, involving a greater number of participants.

References
Aghagolzadeh, F., & Davari, H. (2014). Iranian critical ELT: A belated but growing intellectual shift in Iranian ELT community. *Journal for Critical Education Policy Studies (JCEPS)*, 12 (1), 391-410. Retrieved from www.Jceps.com/archieves/477.

Ahmadi Safa, M., & Sheykholmoluki, H. (2023). The impact of a nationwide high-stakes test from high school teachers and principals' perspectives: A qualitative study. *International Journal of Language Testing*, 13 (1). 104-132, https://doi.org/10.22034/ijlt.2022.352176.1179

Ahour, T., & Golpour, F (2013). Iranian new junior high school book (Prospect 1) weighted against material evaluation checklist from teachers' perspectives. *Journal of English Language Pedagogy and Practice*. 6(13), 16–35. Retrieved from http://jal.tabriz.iau.ir/article_523993.html

Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction assessment.* Thousand Oaks, CA: Corwin Press.

Alahmadi, N., Alrahaili, M., & Alshraideh, D. (2019). The impact of the formative assessment in speaking test on Saudi students' performance. *Arab World English Journal*, *10*(1), 259–270. http://dx.doi.org/10.24093/awej/vol10no1.22

Alavi, S.Y., Rezvani, R., & Yazdani, S. (2022). A gap analysis between EFL teachers' assessment literacy in Iranian public schools and their pre-service assessment curriculum. *Teaching English as a Second Language Quarterly (TESLQ)*, 41(4), 59-86. https://doi.org/10.22099/tesl.2022.42432.3066

Assessment Reform Group (ARG) (2002). *Research-based principles to guide classroom practice*. University of Cambridge, School of Education.

Atai, M. R., & Mazlum, F. (2013). English language teaching curriculum in Iran: Planning and practice. *The Curriculum Journal,* 24(3), 389–411. https://doi.org/10.1080/09585176.2012.744327

Bachman, L. F., & Palmer, A. S. (1996/2010*). Language testing in practice*. Oxford University Press.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. Granada Learning.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3 (2), 77-101. http://dx.doi.org/10.1191/1478088706qp063oa.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education,* 10 (2), 161–180. http://dx.doi.org/10.1207/s15324818ame1002_4.

Brookhart, S.M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record.*, 106 (3), 429-458. http://dx.doi.org/10.1177/016146810410600302.

Buyukkarci, K., & Sahinkarakas, S. (2021). The impact of formative assessment on students' assessment preferences. *The Reading Matrix: An International Online Journal*, *21*(1), 142–161. Retrieved *f*rom https://www.readingmatrix.com/files/24-o03594m3.pdf.

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360-389. http://dx.doi.org/10.1191/0265532204lt288oa.

Chow, A., & Leung, P. (2011). Assessment for learning in language classrooms. In R. Berry and B. Adamson (Eds.), *Assessment reform in education, education in the Asia-Pacific region: Issues, concerns and prospects 14* (pp. 135-154). Netherlands: Springer. http://dx.doi.org/ 10.1007/978-94-007-0729-0_10.

Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds). *The Cambridge Guide to Second Language Assessment,* pp. 20-29. Cambridge University Press.

Crooks, T. (2001, September 13-15). *The validity of formative assessments*. Paper presented to the British Educational Research Association Annual Conference, University of Leeds. http:// www. leeds. ac. uk/educol/documents/00001862. htm.

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305–334. https://doi.org/10.1191/0265532204lt286oa.

Dehqan, D., & Asadian Sorkhi, S.R. (2020). Pre-service and in-service teachers' knowledge and practice of assessment literacy: A dweller in an ivory tower. *Issues in Language Teaching, 9* (2), 347- 375. https://doi.org/10.22054/ilt.2021.57203.556.

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into practice*, *55* (2), 153–159. http://dx.doi.org/10.1080/00405841.2016.1148989.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Estaji, M., & Fassihi, S. (2016). On the relationship between the implementation of formative assessment strategies and Iranian EFL teachers' self-efficacy: Do gender and experience make a difference? *Journal of English Language Teaching and Learning*, 8 (18), 65-86. https://elt.tabrizu.ac.ir/article_5494.html.

Farhady, H., & Hedayati, H. (2009). Language assessment policy in Iran. *Annual Review of Applied Linguistics, 29,* 132- 141. http://dx.doi.org/10.1017/S0267190509090114.

Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9 (1), 1-14. http://dx.doi.org/10.1186/s40468-019-0078-7.

Foroozandeh, E., & Forouzani, M. (2015). Developing school English materials for the new Iranian educational system. In C. Kennedy (Ed.), *English language teaching in the Islamic Republic of Iran: Innovations, trends and challenges* (pp. 59-72). London: British Council. https://www.teachingenglish.org.uk/article/english-language-teachingislamic-republic-iran-innovations-trends-challenges

Fulcher, G., & Davidson, F. (2007) *Language testing and assessment: An advanced resource Book*. Oxford: Routledge.

Ganji, M., Ketabi, S., & Shahnazari, M. (2018). English teacher training courses in Iranian private language institutes: Issues and options. *Issues in Educational Research*, 28(2), 367- 384. http://www.iier.org.au/iier28/ganji.pdf.

Gholamnejad, R., & Raeisi-Vanani, A. (2021). English language teaching in Iranian mainstream schools: Pedagogical, societal and government policy environment. *Issues in Educational Research*, 31 (1),111-129. 129. http://www.iier.org.au/iier31/gholaminejad.pdf.

Ghorbani, M. R., & Neissaari, M. (2015). Workshop effect of the Iranian Konkour on senior high school students' EFL learning activities. *Iranian Journal of Language Testing*, 5 (1), 1-28.

Glaser, R. (1994). Instructional technology and the measurement of learning outcomes: Some questions. *Educational Measurement Issues and Practice*, 13 (3), 6-8. https://doi.org/10.1111/j.1745-3992.1994.tb00561.x.

Gronlund, N. E. (2006). *Assessment of student achievement* (8th ed.). Boston: Pearson.

Hazim Jawad, A. (2020). Examination of Iraq EFL teachers' attitudes, intentions, and practices regarding formative assessment. *International Journal of Language Testing*, 10 (2), 145-166. Retrieved from https:// www.ijlt.ir/article_119160.html.

Holliday, A. (1996). Developing a sociological imagination: Expanding ethnography in international English language education. *Applied Linguistics*, 17(2), 234-255. http://dx.doi.org/10.1093/applin/17.2.234.

Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: Impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia*, 12 (1), 40. http://dx.doi.org/10.1186/s40468-022-00191-4.

Jalilzadeh, K., Alavi, S. M., & Siyyari, M. (2022). Comparing language assessment literacy and challenges of Iranian EFL teachers: TEFL vs non-TEFL background. *Journal of Language and Translation*, 12 (4), 177-196. https//doi.org/ 10.30495/TTLT.2022.696885.

Karimi, M. N., & Shafiee, Z. (2014). Iranian EFL teachers' perceptions of dynamic assessment: Exploring the role of education and length of service. *Australian Journal of Teacher Education*, 39 (8), 143-162. http://dx.doi.org/10.14221/ajte.2014v39n8.10.

Mardani, M., & Tavakoli, M. (2011). Beyond reading comprehension: The effect of adding a dynamic assessment component on EFL reading comprehension. *Journal of Language Teaching and Research*, 2(3), 688-696. http://dx.doi.org/10.4304/jltr.2.3.688-696.

McMillan, J. (Ed.). (2013). *SAGE handbook of research on classroom assessment*. Los Angeles, CA: SAGE.

McMillan, J. (2015). Sınıf içi değerlendirme, etkili ölçütlere dayalı etkili bir öğretim için ilke ve uygulamalar (A. Arı, Trans. Ed.). Eğitim Yayınevi

Miles, M. b., Huberman, A.M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourceboo*k (3rd ed). Los Angeles, CA: Sage.

Momeni, A., & Nushi, M. (2022). Iranian EFL teachers' beliefs about dynamic assessment: Does context make a difference? *Language Related Research*, 13(3), 403-423. http://dx.doi.org/10.52547/LRR.13.3.16.

Morrow, C. K. (2018). Communicative language testing. In J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 342–350). NJ: Wiley.

Naeini, J., & Duvall, E. (2012). Dynamic assessment and the impact on English language learners' reading comprehension performance. *Language Testing in Asia*, 2(2), 22–41. https://doi.org/10.1186/2229-0443-2-2-22.

Naka, L. (2023). The impression of formative assessment in the immediate EFL learning improvement: Investigation of pre-service teachers' percipience in English for teachers' course. *International Journal of Language Testing*, 13 (1), 234-259. https://dx.doi.org/10.22034/IJLT.2023.384356.1230

National Curriculum Document. (2013). *Ministry of Education of the Islamic Republic of Iran* (in Persian). Tehran, Iran: Ministry of Education. http://mebtedaei.yazdedu.ir/documents/ Google Scholar.

Nitko, A. (2001). *Education assessment of students* (3rd ed.). NJ: Prentice-Hall, Inc .

Öz, H. (2014). Turkish teachers' strategies of assessment for learning in the English as a foreign language classroom. *Journal of Language Teaching and Research, 5*(4), 775-785. https://dx.doi.org/10.4304/jltr.5.4.775-785.

Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal, 6(1),* 25-44.

Ozan, C., & Kıncal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory and Practice*, 18, 85–118. https://dx.doi.org/10.12738/estp.2018.1.0216.

Paivandi, S. (2012). *Education in the Islamic Republic of Iran and perspectives on democratic reforms*. London: Legatum Institute.

Pishghadam, R., Barabadi, E., & Kamrood, A. M. (2011). The differing effect of computerized dynamic assessment of L2 reading comprehension on high and low achievers. *Journal of Language Teaching and Research*, 2(6), 1353-1358. http://dx.doi.org/10.4304/jltr.2.6.1353-1358.

Pishghadam, R., & Shayesteh, S. (2012). Conceptions of assessment among Iranian EFL teachers. *The Iranian EFL Journal,* 8(3), 9-23.

Razavipour, K., & Rezagah, K. (2018). Language assessment in the new English curriculum in Iran: managerial, institutional, and professional barriers. *Language Testing in Asia, 8* (1), 1-18. http://dx.doi.org/10.1186/s40468-018-0061-8.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing,* 21(3), 249-258. http://dx.doi.org/10.1191/0265532204lt283ed.

Rea-Dickins, P., & Gardner, S. (2000). Snares and silver bullets: disentangling the construct of formative assessment. *Language Testing,* 17(2), 215-243. http://dx.doi.org/10.1191/026553200673636547.

Riazi, A.M., & Razavipour, K. (2011). (In) Agency of EFL teachers under the negative backwash effect of centralized tests. *International Journal of Language Studies*, 5(2), 123-142. Retrieved from http://drive.google.com/file/d/1yVX-yW2yrhm1N2b8ccljCzf7jgDTinaO/view.

Richards, J. C. (2001). Program factors in effective foreign and second language teaching. *Journal of Southeast Asian Education*, 2(2), 373–412.

Saad, M. R. B. M., Sardareh, S. A., & Ambarwati, E. K. (2013). Iranian secondary school EFL teachers' assessment beliefs and roles. *Life Science Journal*, 10(3), 1638-1647.

Sardabi, N., & Koosha, M. (2015). New perspectives in PROSPECT: an assessment of strengths and weaknesses of Iranian second year junior high school textbooks. *Research in English Language Pedagogy*, 3 (2), 57–70. Retrieved from https://relp.isfahan.iau.ir/article_533631.html.

Sathik, M., & Jonathan, S. G. (2013). Effect of facial expressions on student's comprehension recognition in virtual educational environments. *SpringerPlus*, 2, 1-9. http://dx.doi.org/10.1186/2193-1801-2-455.

Scriven, M. (1991). Chapter II: Beyond formative and summative evaluation. *Teacher College Record*, 92 (6), 19-64. https://doi.org/10.1177/016146819109200603.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M.K., & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement,* 21(4), 295-314. http://dx.doi.org/10.1080/08957340802347647.

Spiller, D. (2012). *Assessment matters: Self-assessment and peer assessment*. Hamilton, New Zealand.

Stiggins, R. J. (1991). Assessment literacy. The *Phi Delta Kappan*, 72(7), 534–539. http://www.jstor.org/stable/20404455.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2006). *Classroom assessment for learning: Doing it right—using it well*. Princeton: Educational Testing Service

Wang, X. (2017). A Chinese EFL teacher's classroom assessment practices. *Language Assessment Quarterly*, 14 (4), 312-327. http://dx.doi.org/10.1080/15434303.2017.1393819.

Watmani, R., Asadollahfam, H., & Behin, B. (2020). 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*, 10 (2), 129-144. Retrieved from https://www.ijlt.ir/article_118025.html.

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

Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia-Social and Behavioral Sciences, 116,* 2998-3004. http://dx.doi.org/10.1016/j.sbspro.2014.01.696.

Zolfaghari, F. & Ahmadi, A. (2016). Assessment literacy components across subject matters. *Cogent Education*, 3 (1), 1252561. http://dx.doi.org/10.1080/2331186X.2016.1252561.