

# Teacher Assessment Literacy in L2 Speaking: Exploring Novice and Experienced Iranian EFL Teachers' Perceptions and Practices

**Masoomeh Estaji**

([mestaji74@gmail.com](mailto:mestaji74@gmail.com))

Allameh Tabataba'i University, Iran

**Pegah Ghashghaie Nezhad**

([pegah.ghashghaie1993@yahoo.com](mailto:pegah.ghashghaie1993@yahoo.com))

University of Victoria, Canada

---

## Abstract

The way teachers conceptualize speaking and assess it in the classroom has always been a matter of controversy among researchers and teacher educators. This debate is further complicated when the variable of teacher experience comes into play. The present phenomenological study aimed to explore the perceptions of speaking assessment practices among experienced and novice Iranian English as Foreign Language (EFL) teachers. To this end, eight Iranian EFL teachers with different teaching experiences were interviewed in two groups of four, including two novice and two experienced teachers in each group. The results of the inductive thematic analysis revealed that the experienced and novice teachers differed in how they defined speaking components, paid specific attention to them, assessed these components, and evaluated their results. The findings are discussed in detail and practical implications are outlined for EFL practitioners and test developers to create well-designed rubrics that include speaking components, which will ultimately contribute to the quality and fairness of scoring criteria.

---

## Share and cite

Estaji, M., & Ghashghaie Nezhad, P. (2025). Teacher assessment literacy in L2 speaking: Exploring novice and experienced Iranian EFL teachers' perceptions and practices. *Electronic Journal of Foreign Language Teaching [e-FLT]*, 22(1), 4–24. <https://doi.org/10.56040/esgh2211>

---

## 1 Introduction

The necessity of classroom assessment is widely acknowledged by stakeholders involved in the learning or teaching context (Berry et al., 2019; Estaji & Ghiasvand, 2023; Wang et al., 2023). To highlight the importance of assessment, Brinke et al. (2007) assert that the assessment process lies at the heart of education due to its direct impact on the students' learning processes. In response to such emphasis on assessment, Popham (2011) suggests that it is vital for teachers to possess assessment literacy, which consists of “an individual's understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions” (p. 267). He further explains that assessment literacy should be regarded as instructional proficiency when it comes to preparing teachers. Assessment literacy in this study refers to teachers' knowledge, skills, and practices in

designing, implementing, and scoring various types of assessment methods in the class.

With the prominence of assessment literacy, language assessment literacy (LAL) has come to the fore in second language (L2) educational milieus due to the pivotal role of language teachers in the internal and external assessment of learners (Zimina, 2018). In recent years, LAL has garnered noticeable attention among researchers who have sought to unravel the language assessment competence of teachers through different measures (e.g., Coombe et al., 2020; Levi & Inbar-Lourie, 2019). It is predominantly considered to be an awareness of drawing on different assessment techniques in a given context and at an appropriate time, which enables language teachers to develop language tests, interpret the test results, and make informed decisions based on the results (Inbar-Lourie, 2013; Pill & Harding, 2013).

L2 speaking assessment, as an integral part of second language education, contributes to the evaluation of the learning process and subsequently the enhancement of learning and teaching practices (Poehner & Inbar-Lourie, 2020; Yan et al., 2021). Recognizing the complexity of assessing speaking skills, several studies have highlighted the difficulty of aligning assessment targets with appropriate tasks and instruments (Luoma, 2004; O'Sullivan, 2006). To respond to the challenges of speaking assessment, diverse approaches and methods have been introduced to facilitate the assessment process (De Jong, 2023). The appropriate incorporation of diverse tasks and methods into the speaking assessment procedure requires teachers' L2 speaking assessment literacy, which has recently gained momentum in L2 education (e.g., Duque-Aguilar, 2021; Koizumi, 2022; Lim, 2018).

Notwithstanding the significance of gaining insights into teachers' LAL, there seems to be a paucity of research on teachers' speaking assessment literacy in the EFL context. More specifically, few researchers have delved into the assessment literacy of Iranian EFL teachers, their oral assessment perceptions and practices to unravel different factors affecting speaking assessment literacy. Furthermore, the question of whether experienced and novice teachers differ in terms of their perceptions or practices of oral assessment has not been explored by previous research. The variation in experience and the difficulties that Iranian EFL teachers have in assessing oral skills also warrant empirical research. Hence, to bridge the aforementioned gaps in knowledge, this study sought to investigate the speaking assessment literacy of language teachers in an EFL context. Additionally, we took a step forward to probe the possible differences between novice and experienced teachers' perceptions of speaking assessment. The findings may yield valuable insights for teacher educators that could help to enhance the LAL of teachers in EFL settings. Furthermore, the results can assist EFL test developers in designing diverse rubrics and measures for assessing speaking.

## 2 Literature review

### 2.1 Language assessment literacy

Teachers' assessment literacy has gained momentum in both general education and L2 education (Davies, 2008). It is a discipline-based literacy, which integrates teachers' pedagogical content knowledge with their assessment knowledge specific to a certain discipline (Zolfaghari & Ahmadi, 2016). Within this framework, LAL is conceptualized as an amalgamation of skills including knowledge of different assessment methods, appropriate use of those methods at a given time and setting, construction of language tests, and analysis of the test results (Coombe et al., 2020; Pill & Harding, 2013). In this regard, Davies (2008) proposes a "skill + knowledge" approach with "skill" referring to the practical side of assessment and "knowledge" addressing the "relevant background in measurement and language description" (p. 328).

Furthermore, assessment literacy is "a dynamic context dependent social practice" (Willis et al., 2013, p. 242). It is intertwined with the input teachers receive from the stakeholders, namely students, colleagues, policymakers, and administrators (Coombe et al., 2020; Fleer, 2015). Additionally, Fulcher (2012) categorizes LAL into (a) practical knowledge, (b) theoretical and procedural knowledge, and (c) socio-historical understanding, with the practical knowledge being the core component of LAL. In another categorization, Taylor (2013) classifies LAL into eight aspects: (1) knowledge of

theory, (2) technical skills, (3) principles and concepts, (4) language pedagogy, (5) sociocultural values, (6) local practices, (7) personal beliefs, and (8) scores and decision making (p. 410).

Despite the fact that assessment literacy should be a major component of teacher qualification, language teacher educators are not adequately trained in the principles and implementation of assessment (Coombe et al., 2020; Taylor, 2009). As Webb (2009) emphasizes, “facilitating change in teachers’ assessment practice is not so much a resource problem” (p. 3). Rather, it is about “helping teachers develop a ‘designers’ eye’ for selecting, accepting, and designing tasks” (p. 3). In this regard, Popham (2009) proposes a framework consisting of two types of assessment for promoting assessment literacy: classroom assessment and accountability assessment. His framework shows the role of internal and external mandates and factors in determining teachers’ practices in the classroom. Technically, classroom assessment refers to “those formal and informal procedures that teachers employ in an effort to make accurate inferences about what their students know and can do” (Popham, 2009, p. 6).

However, accountability assessments “are those measurement devices, almost always standardized, used by governmental entities such as states, provinces or school districts to ascertain the effectiveness of educational endeavors” (Popham, 2009, p. 6). Popham further elaborates that classroom assessment, even if used perfunctorily, might have an impact on student learning. As Kleinsasser (2005) puts it, one of the challenging aspects of language assessment courses is the ability to link theoretical concepts to practical applications. In the same way, Scarino (2013) argues that for teachers to effectively apply their knowledge to a practical setting, it is important to comprehend the underlying principles of various approaches and connect such understanding to their professional practice within their specific context. Considering these points, the investigation of teachers’ awareness of language assessment practices and their understanding of assessment literacies merits attention. However, how EFL language teachers perceive their assessment literacy in different language skills has received scant attention.

## ***2.2 L2 speaking assessment: conceptualizations, approaches, and practices***

Speaking assessment has been deemed a daunting task requiring a lot of effort and resources (Fan & Yan, 2020) due to its transient and context-based nature as well as its different forms during assessment (De Jong, 2023). In this sense, when assessing speaking skills, diverse factors such as tasks, scoring criteria, raters, and interlocutors should be considered (Fulcher, 2015). Additionally, the assessment of speaking abilities in a second language presents certain difficulties, including determining language proficiency, precluding cultural biases, and ensuring the accuracy of training procedures (Sanchez, 2006). Given the significance of speaking assessment, diverse assessment practices, namely formative and summative approaches, have been introduced into the field of language education (Koizumi, 2022). Both approaches serve cardinal functions in reporting students’ progress and their later admission to related professions.

As pointed out by Koizumi (2022), to actualize L2 speaking formative and summative assessment, diverse techniques, including oral presentations, peer work, utilizing rubrics for evaluating oral performance, providing feedback, and using scores and comments, are employed. Summative speaking assessment can be exploited for making higher-level decisions and examining speaking ability comprehensively (Negishi, 2020), while formative speaking assessment provides regular diagnostic information and aims to foster learning and teaching processes by recognizing students’ speaking strengths and weaknesses and making changes to teachers’ lesson plans and teaching practices (Koizumi, 2022). Moreover, summative assessment may lead to a positive washback effect (Lee et al., 2020). In this sense, studies have ascertained that drawing on both formative and summative assessment in speaking assessment may yield fruitful results (Munoz & Alvarez, 2010). As Koizumi (2022) proclaims, the simultaneous exploitation of formative and summative speaking assessment can offer ample opportunities for students to enhance their speaking ability by altering their learning techniques. It can also aid teachers in summatively evaluating students’ efforts and their overall achievements in speaking skills.

### ***2.3 Factors influencing L2 speaking assessment practices: Issues and challenges***

The assessment practices that teachers adopt while evaluating learners' L2 speaking are affected by myriad factors. As identified by Zhang and Burry-Stock (2003), in-service teachers who have undergone assessment and measurement training exhibit superior assessment skills when compared to their counterparts who have not received such training. Their study demonstrated the complexity and flexibility of teachers' assessment knowledge and literacy over time and instruction. Quilter and Gallini (2000) state that a teacher may opt for assessments that do not align well with a specific assessment objective due to a lack of training in standardized testing associated with positive attitudes toward alternative assessment. Additionally, teachers' attitudes are likely to be situation-specific (Fang, 1996). In this regard, Quilter and Gallini (2000) note that although attitudes may be specific, they are not always consistent with behavior. For example, teachers may believe that standardized tests are pointless, but they still have to include them in their classroom.

Similarly, Jamil et al. (2012) indicate that the classroom practice of pre-service teachers can be influenced by their personality and self-efficacy. Teachers' assessment practices are also affected by the context in which they are embedded (Frey & Fisher, 2009; Gu, 2014). More importantly, teaching experience is thought to play a pivotal role in the speaking assessment practices of teachers (Mertler & Campbell, 2004). Furthermore, Xu and Liu (2009) accentuate the role of prior assessment practices in applying current assessment techniques and making future assessment plans. In this sense, Crusan et al. (2016) contend that teachers' experience exerts an influence on their writing assessment practices. However, the role of experience in L2 speaking assessment has not been explored to date, which necessitates further investigation of how novice and experienced teachers' perceptions differ in terms of L2 speaking assessment.

### ***2.4 Previous studies on teacher assessment literacy and L2 speaking assessment***

With the growing significance of speaking assessment in L2 educational milieus, the important role of assessment practices has received a surge of interest (e.g., Koizumi, 2022; Negishi, 2020; Thuy & Nga, 2018). Nevertheless, a thorough examination of previous studies reveals that LAL in general has gained more attention while speaking assessment literacy has been barely explored. For instance, a study conducted by Ahmadi and Mirshojaiee (2016) involved a semi-structured interview to assess the level of assessment literacy among 20 Iranian English teachers. They reported that the participants' assessment literacy was unsatisfactory regarding issues such as assessment bias, validity evidence, assessment reliability, and formative assessment. This might be due to the educational context of Iran, where assessment training is rare for teachers and facilities supporting the assessment of oral skills are limited.

Taking a quantitative rather than a purely qualitative approach, Razavipour et al. (2011) conducted a study where 53 EFL secondary school teachers were given a test to measure their assessment literacy, along with a questionnaire regarding their English language teaching practices. The results revealed that there was a low level of assessment knowledge among the teachers. Additionally, regardless of their assessment literacy level, they tended to prioritize teaching and testing based on external tests. Teachers' lack of assessment literacy might be the cause of such tendencies in Iran. Top-down approaches to education in Iran may also be the cause of such low assessment literacy. In a study conducted by Jeong (2013), the impact of instructors on the content and structure of language assessment courses was analyzed. The results indicated that there were notable variations in course content depending on the instructors' backgrounds. Furthermore, individuals who were not language testers exhibited lower levels of confidence when it came to instructing technical assessment abilities in comparison with language testers. It was shown that they tended to concentrate on matters related to classroom assessment.

As regards speaking assessment literacy, Thuy and Nga (2018) investigated how EFL teachers perceived English speaking assessment conducted in the classroom. They focused on three main components, namely the teachers' overall comprehension of speaking assessment, the types of tasks

utilized in speaking assessment, and assessment execution. Forty-two EFL teachers filled out a questionnaire, and five of them attended the interview sessions. The results demonstrated that the teachers generally exhibited an appropriate perception considering the three aspects under investigation. However, their understanding of oral portfolios as a type of speaking assessment was found to be limited.

Likewise, Özdemir-Yılmaz and Özkan (2017) examined how English language teachers at the tertiary level perceived and implemented speaking assessment in the Turkish context. The case study was conducted based on the principles of Personal Construct Theory, and information was gathered through the use of repertory grids, semi-structured interviews, and classroom observation. The participants were found to hold distinct viewpoints regarding speaking assessment, which underscores the need for teachers to receive training on classroom assessment to enhance their understanding of standard practices related to in-class speaking assessments. In a recent study, Duque-Aguilar (2021) probed speaking assessment approaches during a teacher education program. To this end, four in-service teachers were selected, and their speaking assessment approaches were examined. The data were collected through interviews, observations, and document analysis. The results divulged teachers' preferences for summative assessment during speaking assessment.

Although a body of research into assessment literacy exists in general education, assessment literacy is still in immediate need of investigation in the EFL context. More specifically, to the best of the researchers' knowledge, there has been to date no examination of the speaking assessment literacy skills of Iranian EFL teachers, and their perspectives and implementation of oral assessments have been underexplored. Furthermore, novice and experienced teachers' perceptions or practices of speaking skill merit further investigation to determine whether the teachers differ in terms of their speaking assessment literacy. Therefore, the present study sought to provide answers to the subsequent research questions using interview data:

1. How do novice and experienced Iranian EFL teachers perceive speaking assessment?
2. How do Iranian EFL teachers assess their students' oral proficiency?

### **3 Methods**

#### ***3.1 Participants and setting***

The participants of this qualitative study were eight Iranian EFL teachers, working in different private language institutes in the Isfahan province. In Iran, English language education is mostly concerned with teaching communicative skills and teachers' pedagogical expertise rather than their assessment literacy and practice. Hence, the broader issue of assessment literacy aligns with specific challenges and gaps in the Iranian educational system. Assessment literacy gaps among teachers with varying experiences at institutes justify the study of novice and experienced EFL teachers' perspectives. The target respondents in this study were chosen through purposive sampling based on their teaching and assessment experience in L2 speaking courses. This type of sampling was used to glean data from information-rich respondents based on pre-specified criteria in line with the purpose of the investigation.

In the sample, there were four males and four females divided into two groups. The age of the participants varied from 20 to 35 years old. The participants had BA, MA, and PhD degrees in English language teaching (see Table 1). This background contributed to the study by providing data from teachers whose major was L2 education and who were familiar with common speaking assessment theories and practices. It is noteworthy that the participants had different teaching experience levels ranging from 0-2 years (novice) to above 3 years (experienced). This classification fits with Gatbonton's (2008) typology that distinguishes between novice and experienced teachers. Prior to commencing the data collection procedure, the participants provided their informed consent. Their privacy and ethical concerns were carefully considered. Confidentiality of identity and responses was observed by using codes and labels (e.g., N1, E2) to mask teachers' identities.

**Table 1. Summary of the participants' demographic information**

<b>Demographics</b>	<b>N</b>	<b>%</b>
<b>Gender</b>		
Female	4	50
Male	4	50
<b>Age</b>		
20-25	3	37.5
26-30	3	37.5
31-35	2	25
<b>Years of teaching</b>		
0-2	4	50
3-8	2	25
+10	2	25
<b>Academic qualifications</b>		
BA	3	37.5
MA	4	50
PhD	1	12.5

### **3.2 Instruments**

The data for this study were gleaned through two semi-structured panel discussions held in line with what was proposed by Schafer (1991) and Xu and Brown (2016) about fundamental knowledge of assessment literacy. Each focus group included four teachers with the same teaching experience level answering 20 open-ended English items. In the panel discussion sessions, the teachers shared and exchanged ideas about assessment literacy in L2 speaking and practices among themselves. The content validity of the items was checked and approved by a team of three experts in applied linguistics, who were university professors teaching assessment courses in Iran. The language of the items was commented upon, and redundant questions were recommended for removal from the question pack.

### **3.3 Data collection procedure**

This qualitative phenomenological study aimed to gather deep and rich data from a group of individuals about a specific phenomenon or event (Mackey & Gass, 2005). It was designed to provide a thick description of Iranian EFL teachers' experiences with L2 speaking assessment practices, focusing on different aspects of assessment literacy and L2 speaking. Phenomenology was suited to the goal of the study because it allowed the researchers to delve deeply into the participants' points of view and perspectives in great depth. The perceptions and lived experiences of participants could be captured through this research design (Dörnyei, 2007).

Prior to the panel discussions, the participants were divided into two focus groups and briefed on the nature and aim of the guided discussions. They were divided into gender-specific panels for their convenience and because it was believed that teachers of the same gender might have similar understandings and concerns about L2 assessment practices. At the start of each session, the researchers provided an overview of the topics and subtopics that would be covered. The focus groups attended one panel discussion per week, with a total of eight sessions held to evaluate their oral literacy skills—four for each group.

The researchers organized the meetings and led the discussion sessions by initiating the conversation and then stepping back. In the words of Wray and Bloomer (2012), they acted as “facilitators” and “coordinators” of the panel discussion sessions and clarified the topics and issues when necessary. The researchers maintained an outsider positionality during the discussions and bracketed their own perceptions and experiences of L2 speaking assessment, both as current teachers and previous

L2 learners. However, throughout the process, the researchers were reflexive about the data collection and research goals. To preclude bias, the researchers used a protocol and only posed prompts and encouraged the participants to elaborate on their responses without playing a direct role in the data generation and teachers’ responses. The time spent on the focus groups varied for each session. However, the mean time spent per panel discussion session was around 20 to 40 minutes. The panel discussion sessions were documented through audio recordings, which were subsequently organized and transcribed.

3.4 Data analysis

After the data were collected, the process of arranging and preparing them for analysis commenced. Braun and Clarke’s (2006) thematic analysis model was used to analyze the panel discussion data (see Figure 1). To become more familiar with the data, the researchers first listened to the audio recordings of the panel discussions and gained a general understanding of the data. Then initial and preliminary codes were created after listening to the audio recordings twice. The codes were quick explanations and ideas about the L2 speaking assessment perceptions and practices of the teachers. Next, the extracted codes were combined in order to form larger themes in line with the research questions of the study. After identifying a set of themes, the researchers reviewed each theme in relation to the objectives of the study. In the next step, labels and phrases were selected to report each theme. In the last step, the researchers created a detailed report of the obtained findings with sufficient supporting evidence from the panels and explanations for the readers. The themes were given to a second coder to examine inter-coder reliability. The results of Cohen’s Kappa Coefficient revealed a value of .79, which is greater than .70, hence representing a substantial agreement (Altman, 1999). To add credibility to the study, the whole data analysis process was audited by an expert qualitative researcher from a state university in Tehran with a postdoctoral research degree in applied linguistics. Additionally, trustworthiness and rigor were ensured by providing a detailed account of the participants’ background, research context, design, and research process for confirmability and transferability maxims, which are pivotal in qualitative research.

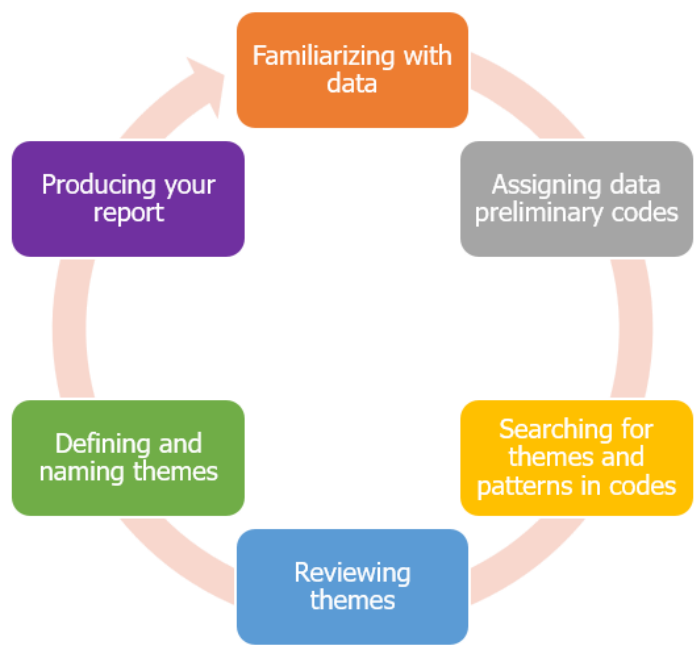


Fig. 1. Representation of Braun and Clarke’s (2006) thematic analysis model

The researchers were reflexive and maintained an outsider perspective throughout the data collection phases. The data analysis took 28 days from data transcription to theme generation and reporting the final outcomes.

## 4 Findings

The first research question aimed to probe teachers' perceptions concerning speaking assessment. Through inductive analysis, the teachers' comments were gathered and classified, resulting in significant themes related to their perspectives on speaking assessment. Subsequently, a frequency analysis was conducted to identify the most common responses given by the EFL teachers for each interview question. The following sections are arranged in accordance with the major themes obtained from the analysis. A sample of the coding process is presented in Table 2.

**Table 2. Thematic analysis sample**

Global theme	Organizing theme	Basic theme	Frequency
Pronunciation	"For lower levels, we pay attention to intonation."	Intonation	5
	"What distinguishes English from Persian is stress in English words."	Stress	7

### 4.1 Components of speaking and teaching them

The first four interview questions dealt with speaking components and the necessity of teaching them.

#### *Question 1: What are the components of speaking?*

The results showed that pronunciation (correct production of sounds, intonation, rhythm, and stress) was the EFL teachers' first concern. In addition, they prioritized vocabulary knowledge and fluency. Novice teachers attached more importance to vocabulary knowledge and fluency, while experienced teachers emphasized pronunciation and interactional efficiency. Novice teachers also reported that confidence and the level of extemporaneous speaking were important speaking components.

#### *Question 2: Which components are more significant and why?*

Novice teachers valued vocabulary knowledge as the most important component. They contended that knowledge of vocabulary, collocations, and idioms was indispensable to speaking proficiency. The other components that novice teachers considered highly important were fluency, pronunciation, and accuracy. However, experienced teachers ranked pronunciation and interactional efficiency as first and second, respectively.

**E4:** *We can divide speaking components into two parts. The first parts are for lower levels; we pay attention to stress and intonation, pronunciation, accent, and these matters. At higher levels, we pay attention to how to negotiate, how to argue, and how to do critical thinking.*

**N1:** *But as far as you say communication is important, you know that fluency is a matter in communication. If you are not fluent, you interrupt the interlocutor from comprehension, and communication is hindered.*

#### *Question 3: Do you give more weight to components that play a more important role in determining speaking?*

Teachers mainly believed that learners' level of language proficiency and their age should be considered among the criteria for assessing learners' speaking competence. Experienced teachers placed more emphasis on the assessment of components they valued more and also considered learners' level. Furthermore, they highlighted the importance of learners' motivation.

**N3:** *It depends on the level; age is sometimes important too.*

**E2:** *Naturally, we have to. We are supposed to give weight to those which are more important.*

*Question 4: Do you think speaking components should be taught and why?*

Teachers also indicated that some elements, including rules governing social interaction (turn-taking and politeness), topic management, organization, awareness of non-verbal behavior (volume, intonation, pronunciation, eye contact), and interactional efficiency, needed to be taught explicitly. However, fluency and vocabulary knowledge should be taught implicitly.

**E3:** *I think the components should be taught. Pronunciation can be taught, for example, by how to pronounce /ð/ or /θ/. Some students are unaware of the fact that there is a difference between these two sounds.*

**N2:** *Some components, such as fluency, can be acquired.*

## 4.2 Speaking tasks

Interview questions 5 and 6 revolved around the identification of effective speaking tasks that would result in higher student achievement.

*Question 5: What speaking tasks should be used and why?*

Results showed that teachers favored open-ended discussions, followed by picture description and storytelling. They contended that such activities were similar to real-life situations and were more authentic.

**E1:** *Yeah, open-ended discussions.*

**N4:** *I use storytelling. For example, they read a story, and after that, they come to the front of the class and tell us the summary of the story, and also I use role plays in different situations. I also use interviews.*

*Question 6: What speaking tasks lead to higher achievement of students and why?*

Again, open-ended discussions were the most preferred task. Both novice and experienced teachers put open-ended discussions in first place. They thought that group collaboration provided maximum opportunities for learners to speak a foreign language. Next, they favored storytelling and picture description as authentic activities for improving communicative competence.

**N3:** *Yeah, interaction and communication with the students. I mean students with students, students with the teacher.*

## 4.3 Rating scale

The next four interview questions focused on the teachers' rating scale for speaking.

*Question 7: Which rating scale should a teacher use in the speaking assessment?*

Experienced teachers reported that they considered diverse criteria in their assessment of speaking. They argued that analytic assessment provided diagnostic information about learners' speaking abilities. In contrast, novice teachers believed in both analytic and holistic assessment. Two of them reported that they only paid attention to the overall impression, whereas others were inclined toward analytic assessment.

*N2: I think both analytic and holistic rating scales should be used together. Sometimes you want to categorize your students as good or bad speakers, and sometimes you want to have a more detailed analysis of their speaking.*

*Question 8: Do you think teachers' knowledge and experience play a role in better rating students' speaking ability?*

Teachers believed that their knowledge and experience were of paramount importance in assessing learners' speaking ability effectively.

*Question 9: What strategies should a teacher apply to avoid a subjective rating scale?*

Novice teachers tended to apply collaborative assessment and used a checklist to avoid subjective assessment, while experienced teachers mostly preferred a checklist.

**E2:** The institute can provide the teachers with a checklist with many factors included, and all the teachers should use it. In this way, there is uniformity between teachers.

**N1:** I think they can get some help from other colleagues, the expert ones, especially experienced ones, and they can search and consult some standard rating scales.

*Question 10: Do you think the speaking assessment of your students and the rating scale you apply are perfect indicators of your students' language ability and language use?*

Teachers contended that classroom assessment could not efficiently assess learners' proficiency levels. Moreover, they said that test tasks were not truly representative of real-life language.

**E4:** *I think assessment, in general, is not a clear manifestation of their ability because sometimes students may be introverted, but still, they have perfect knowledge of English, so our assessment doesn't say much.*

#### **4.4 Assessment approaches**

Assessment approaches utilized by the teachers represented the theme extracted from responses to the following questions.

*Question 11: Do you think alternative assessment of students, like self- and peer assessment, is important in determining their speaking?*

Both groups of teachers believed that self-assessment was conducive to learning. Experienced teachers felt that self-assessment encouraged learners to develop more responsibility for their learning, alleviated teachers' assessment burden, and promoted learning. According to them, peer assessment led to embarrassment and hindered improvement.

**E3:** *Self-assessment is really beneficial while peer assessment leads to embarrassment and fear of speaking in front of classmates.*

*Question 12: Should a teacher ever take advantage of self-assessment or peer assessment while assessing students' speaking?*

Teachers argued that peer and self-assessment could not be relied on solely. They believed that teachers themselves should assess learners' proficiency levels.

**N3:** *No, it's not enough to only pay attention to peer assessment while you are scoring them since it is not valid, it is not authentic.*

**E1:** *I think that the peers don't have sufficient knowledge to assess.*

*Question 13: Which assessment approach leads to a more accurate rating? Teacher, peer, or self-assessment?*

**E3:** *These assessment approaches are all important, but the most reliable I suppose will be teacher assessment as teachers are fully aware of the scoring rubrics and the students' strengths and weaknesses.*

**N3:** *I guess, teacher assessment is more effective and accurate in L2 contexts because teachers are more knowledgeable than students.*

As indicated in their remarks, both novice and experienced teachers ranked teacher assessment as the most accurate form of assessment.

#### **4.5 Planning speaking assessment**

The next theme, extracted from the following interview questions, focused on teachers' plans and strategies to assess students' speaking.

*Question 14: Do you ever plan to assess your students' speaking? And if yes, what type of plan do you specifically use?*

Experienced teachers explained that they usually planned for appropriate age-related activities, including opportunities for purposeful speaking across various contexts. However, novice teachers indicated that they did not specifically plan to assess learners' speaking ability. They would usually listen to their learners and keep informal notes of their proficiency levels.

**N4:** *Planning is mostly in my mind. I know what I'm going to do, but the questions are mostly raised in the class because the class situation triggers some questions. I plan to know which part I'm going to assess and which questions I want to ask, but I don't write these questions, so I set the purpose for myself, but I don't write it on paper.*

**E1:** *I do plan it, actually. I go to the classroom intending to assess my students today, and sometimes actually, they are aware of it.*

*Question 15: Do you think such plans affect the selection of tasks you employ?*

Teachers reported that planning for the assessment provided a framework for the assessment process, outlined the content, provided variety for the assessment and general guidelines, and helped to define the assessment structure.

**E3:** *Yes, it provides a framework for the selection of future tasks.*

**N2:** *It would add to creativity and variation in the tasks.*

*Question 16: How do you assess your students' speaking?*

Experienced teachers favored integrated assessment that focused on the examinees' comprehension of the input and their capacity to produce language. They emphasized the integration of language skills with speaking tasks and the design of tasks that combined reading, listening, writing, and speaking. Novice teachers reported that they concentrated on learners' interaction and spoken production.

**E4:** *Yes, and reading provides us with a lot of speaking tasks because I usually integrate the skills or I give them a topic, they write about it, and next session, they talk about it. So, different activities can turn into speaking.*

**N3:** *Assessing students' speaking is a good idea, but I let them know the criteria beforehand.*

*Question 17: What challenges or problems do you encounter in assessing your students' speaking?*

Teachers indicated that speaking ability included subjective features, such as pronunciation and fluency, which could be evaluated differently. Furthermore, they argued that learners' anxiety directly influenced their performance, which then hindered the assessment process. Additionally, a lack of specific standards was a problem in speaking assessment.

**E2:** *Subjectivity and also their feelings.*

**N1:** *For example, they may get stressed when assessed.*

#### **4.5 Assessment feedback, score interpretation, and ethics**

There were also themes pertaining to teachers' feedback, score interpretation, and ethical issues in speaking assessment, which were extracted from the following interview questions.

*Question 18: Do you believe that you offer feedback to your students in your assessment, and if yes, what type of feedback do you use?*

Teachers reported that they preferred evaluative feedback in which they highlighted learners' weaknesses and strengths. Furthermore, they indicated that they employed descriptive and corrective feedback.

**E1:** *When it comes to feedback, we should be honest and realistic, but I also take into consideration their self-confidence, their motivation, and emotions, so I prefer both supportive and evaluative feedback.*

**N3:** *I make clear to them what their areas of weakness are and which points need to be worked on more. I explain it to them and give them some ways to overcome their weakness.*

*Question 19: How do you interpret the assessment results?*

Novice teachers were inclined to interpret the results against their own standards. However, experienced teachers tended to follow the institute's policy in their interpretation of results.

**E3:** *It depends on the policy of the system. If it is an institute or a university.*

**N2:** *At the beginning of the term, I clarify the components and subcomponents to myself and my students. First, I tell them that stress, intonation, fluency, and accuracy are the most important parts and tell them that I focus on these components so everything is clear to my students and me. So at the end of the term, it's crystal clear to my students which components are more emphasized.*

*Question 20: Do you consider ethical responsibilities concerning using and storing and archiving the assessment results?*

Experienced teachers emphasized that teachers had to keep learners' educational information confidential. They noted that ethics played an important role in education, while novice teachers did not seem to prioritize ethical issues.

**E1:** *We should do this with a lot of caution. For example, females do not like others to know their marks, but males do not care about their marks being released. I think gender can play an important role in ethical issues because girls have serious competition, but males do not have much.*

**N4:** *No, I don't care. It's not important.*

The ensuing discussion is structured around the four primary components of creativity (Torrance, 1998): fluency, flexibility, originality, and elaboration.

## 5 Discussion

The primary impetus for conducting this study was to examine Iranian EFL teachers' perceptions and practices concerning speaking assessment. In addition, this study intended to delineate the similarities and differences that existed between novice and experienced teachers in this respect and aimed to clarify how teaching experience would affect their perceptions and practices related to assessment. The analysis of the comments indicated that teachers considered pronunciation, fluency, lexical knowledge, interpersonal competence, accuracy, prior knowledge, structure, topical knowledge, self-esteem, and extemporaneous speaking as the components of speaking proficiency. These dimensions showcased the complexity of assessing oral skills and the multi-layered nature of assessment literacy. The findings demonstrated the theoretical and practical components of Fulcher (2012) and Taylor (2013) models of assessment literacy.

Additionally, the study aligns with Davies' (2008) "skill + knowledge" approach to assessment literacy. The findings confirmed that the participants had a sufficient awareness of the theory and practice of speaking assessment. Empirically, the findings are compatible with those reported in the extant literature (Babaii et al., 2015; Butler et al., 2000; Plough et al., 2010; Zhang & Elder, 2011), which argues that different factors are at play when assessing oral skills in L2 education. In addition, Kuo (2011) found that learners believed classroom interaction was enhanced by factors such as effective communication, grammatical precision, and corrective feedback. A justification for these findings could be the participants' prior teaching, testing, and research experiences with L2 speaking. Their high assessment literacy and strong assessment identity may also explain the findings (Estaji & Ghiasvand, 2023).

Both novice and experienced teachers reported a variety of evaluation methods for these components. For experienced teachers, pronunciation was the first concern, while novice teachers prioritized vocabulary knowledge. Novice teachers considered confidence and extemporaneous speaking as components of speaking proficiency. This finding can be compared with Joe et al. (2011), who categorized confidence and extemporaneousness as construct-irrelevant components. They found that both experienced and novice teachers considered factors that were not related to the intended purpose of assessment and individual preferences when making their decisions.

Similarly, according to Yin's (2010) research, teachers possess individualized beliefs and thoughts during the assessment process. The emphasis on pronunciation by experienced EFL teachers is also in tune with Bai and Yuan (2019), who found that intonation and pronunciation were considered critical in teaching and testing L2. This finding diverges from the literature, which mostly reports that novice teachers focus more on superficial aspects of speaking (pronunciation and accent) compared to experienced teachers who highlight features beyond surface-level (e.g., Borg, 2006;

Richards, 2015). The EFL context of Iran and teachers' different pedagogical expertise levels compared to prior studies can explain this divergent finding. The idea of native speakerism might also be a reason for such a different conceptualization in this study.

The study is also in agreement with previous studies that reported variations in assessing speaking skills in EFL contexts (e.g., De Jong, 2023; Fan & Yan, 2020). Teachers' personality traits and personal concerns may be the reasons behind this finding. Experienced teachers highlighted pronunciation probably because they were aware of the importance of correct pronunciation and accent in the globalized world and international encounters. However, novice teachers emphasized vocabulary knowledge, possibly due to their limited knowledge and experience in the field, as they were at the beginning of their career journey. Teacher preferences and experiences as current teachers and previous learners may also explain the emphasis on these aspects of oral assessment.

In their speaking assessment, the teachers valued tasks that were authentic and similar to real-life speaking tasks. In their opinion, group collaboration provided maximum opportunities for learners to speak a foreign language. Both novice and experienced teachers placed open-ended discussions in the first and second places, and they valued storytelling and picture description as authentic activities for improving communicative competence. This is in line with Thuy and Nga (2018), who reported that interviews, role plays (open-ended discussions), and picture descriptions are preferred for evaluating speaking abilities. The participants' high level of knowledge and familiarity with communicative approaches to L2 education and the needs of the current era may justify the findings. Authenticity was highlighted by both groups due to the significance of this element in L2 assessment theories and practices. The findings can be attributed to the teachers' university education or their practical experiences and concerns for real-life practices in the classroom. This interpretation shows that the participants had moved away from traditional assessment to real-life and performance-based practices.

Experienced teachers reported that they considered diverse criteria in their speaking assessment. They argued that analytic assessment provided diagnostic information about learners' speaking ability. However, novice teachers were inclined to use both analytic and holistic assessment. Some of them put sole emphasis on the overall impression, whereas others were inclined toward analytic assessment. These findings are compatible with the rubric-related skills of assessment literacy and LAL conceptualizations mentioned in the background of the study (Stiggins, 2001; Xu & Brown, 2016). However, the results contrast with those of Joe et al. (2011), who found that experienced teachers heavily relied on holistic and intuitive evaluations. This result is in line with that of Kim (2015), who found that experienced teachers considered each aspect of the performance as reflected in the rating scale, while novice and developing teachers paid selective attention to some features.

Gui (2012), Shi (2001), and Kim (2009) also found that native EFL teachers' judgments were more elaborate and detailed than those of non-native teachers. The results can be ascribed to teachers' assessment experience and LAL level, especially regarding different types of assessment rubrics. Rubric awareness and use are signs of assessment literacy in speaking skills, a finding reported by Koizumi (2022). It seems that both groups had comparable views of using rubrics due to their personal tendencies and previous teacher training courses at their institutes. Experienced teachers had moved away from holistic scoring, probably because of their understanding that more critical elements should be considered in assessment. Trial and error practices may also explain the findings.

In addition, teachers indicated that speaking proficiency consists of subjective features, such as pronunciation and fluency, which can elicit diverse evaluation procedures from different teachers. This finding is consistent with that of Bond (1995), who reported that accentedness may be perceived and weighed differently and can lead to diverse evaluations. The participants' knowledge about and care for the native-speakerism paradigm that highlights accent may explain this finding. Furthermore, teachers argued that learners' anxiety directly influenced their performance, which hindered the assessment process. A lack of specific standards was indicated as a problem in speaking assessment as well.

These results are comparable with Mottet and Beebe's (2006) assertion that the perception of immediacy, command of the listening audience, and sincerity are other subjective features that are

sources of assessment bias. The findings imply that L2 assessment is a situated practice in which various internal and external factors are at play. Subjective features, attributes, beliefs, and attitudes determine how a language skill such as speaking is assessed (Gu, 2014; Jamil et al., 2012). The participants' knowledge of the social dimensions of L2 assessment and factors that lead to differential scoring in speaking skills can also justify the findings of the present study. Their participation in conferences and workshops related to L2 assessment may explain this nuanced understanding.

Another finding was that both groups of teachers believed that self-assessment was conducive to learning. Experienced teachers thought that self-assessment encouraged learners to develop more responsibility for their learning, reduced the load of evaluations for teachers, and promoted learning. The existing literature, for example, De Grez et al. (2010), Leger (2009), Orsmond and Merry (2011), and Orsmond et al. (2013), supports the idea that self-assessment is a useful tool for enhancing learners' understanding and promoting their learning. However, in a secondary school setting, self-assessment and peer assessment were deemed unsuitable methods for evaluating oral proficiency (Grada, 2014). In this study, the teachers indicated that peer and self-assessment are unreliable; instead, teachers should assess learners' proficiency levels. This notion contrasts with Orsmond et al.'s (2000) study, which concluded that self-assessment is more conducive to long-lasting learning than teachers' assessment. The participants' previous experiences and tendencies for self-assessment practices may explain the findings. Another reason could be their formative assessment knowledge and practice, because self- and peer assessment are two common practices in the alternative assessment trend. The quest for cultivating agency and autonomy in assessment may possibly explain why the teachers highlighted self-assessment. Furthermore, it seems that they had previously received training on alternative assessment practices in L2 education and oral language skills. That is why they displayed identical conceptualizations regarding assessment.

This study found that experienced and novice teachers were different in their perceptions of speaking components, allocation of attention toward them during evaluations, assessment methods employed, and interpretation of performance ratings. Novice teachers expressed a tendency to adhere to personal preferences and incorporate construct-irrelevant factors into their decision-making processes, as indicated by their verbal reports. These findings suggest that EFL teachers lack a deep understanding of the language use model on which the assessment scales are based. This shallow competency may stem from the failure of teachers' professional training programs to prepare assessment-literate teachers.

As Stiggins (2001) argues, investments for improvement in teachers' professional development and commitment to higher assessment competency remain scarce, or investors fail to effectively prepare teachers for the successful fulfillment of their assessment roles. Teachers' lack of proper and sufficient LAL professional development programs may be the reason behind the findings. This inference is in line with Jeong (2013) and Ahmadi and Mirshojaiee (2016), who suggest that the absence of professional development courses is the common cause of EFL teachers' low assessment literacy. Probably because of this lack of support, novice teachers talked about construct-irrelevant factors when assessing speaking skills.

Increased communication between experienced and novice teachers is required to compensate for this situation. Exchanging information about the learners' performance and opinions about evaluative criteria will benefit both groups. This means that collaborative assessment programs and collective efforts to develop L2 speaking assessment practices are needed. In addition, teacher education courses and programs are required to provide prospective teachers with the knowledge of what assessment criteria should be included in assessing speaking. Teachers should be exposed to examples of how experts make judgments and justify their decisions.

One of the factors that probably affects the judgment of the teachers in charge of the oral assessment of EFL learners is rubrics. Besides their benefits, numerous studies are denouncing the usage of rubrics, spanning from the overall notion of rubrics (e.g., Koh et al., 2017) to the argument that they are inadequately structured and/or ineffectively executed, resulting in rubrics being more detrimental than beneficial (e.g., Wilson, 2007). The differences in the views of novice and experienced teachers concerning oral assessment might be related to their competency or preferences in using

rubrics or their greater emphasis on one of the components. For instance, a teacher might regard the content as significant in real interactions, while another might focus on accent or intonation. Barkaoui (2011) compared novice and experienced teachers regarding the assessment and use of rubrics and found that the experienced ones were more stringent in this regard. It appears that the participants had collectively perceived assessing assessment rubrics as context- and teacher-specific and as a cause of unreliability when assessing L2 speaking. They possibly preferred self-made rubrics for their own classes rather than one-size-fit-all rubrics taken from others' works. This is an indication of strong assessment agency and developed assessment identity among Iranian EFL teachers (Estaji, 2024b; Estaji & Ghiasvand, 2023; Ghiasvand et al., 2023).

## 6 Conclusion and implications

The aim of this study was to examine the views of and methods that Iranian EFL teachers employed in assessing speaking skills and to determine whether their level of experience had an impact on these views and methods. The findings showed differences between the experienced and novice teachers in terms of their perceptions of how to assess students' speaking abilities and the ways in which they carried out such assessments, which influenced their choice of grading scale and evaluation of the results. Based on the outcomes, it is concluded that like other areas, teaching experience also plays a crucial role in teachers' assessment perceptions and practices. As assessment literacy and assessment identity are dynamic and shifting constructs (Estaji, 2024a; Estaji & Ghiasvand, 2023), growth in teachers' teaching experience meaningfully affects their state and status among EFL teachers. It can also be claimed that skill-specific assessment literacy requires further practical and scientific attention in L2 education. Teachers need literacies beyond general assessment literacy, which has dominated the literature for decades.

The findings can contribute to theoretical and practical aspects of L2 teaching and assessment. The study informs theories and models of assessment literacy and LAL by connecting them to specific language assessment practices, especially speaking skills. Measurement tools can be developed based on the findings of the study. Practically, EFL teachers can use the findings as starting points to revise and strengthen their oral assessment practices in EFL contexts. Additionally, teachers can understand the criticality of their assessment literacy in speaking courses and modify their approaches to assessment and mindsets about assessment. Both Iranian and international L2 teachers are encouraged to pay equal attention to assessment literacy in the same way they prioritize pedagogical aspects of the L2 teaching profession. The findings may lead to the creation or refinement of more effective oral assessment methods that are culturally appropriate and aligned with EFL teachers' needs, literacy level, and contextual specificity.

Teacher educators may also benefit from this phenomenological research by proposing assessment literacy-related training courses, where various aspects of oral assessment are explained and taught to novice EFL teachers. The difficult aspects of assessing oral skills can be discussed in training programs. Likewise, the courses can explicitly address operational and effective methods to assess L2 speaking skills in EFL settings. Challenges and coping techniques to handle speaking assessment practices can also be integrated into such programs. Moreover, based on the findings, professional learning communities or collaborative workshops can be developed, where EFL teachers can continue to discuss and refine their understanding of assessment literacy.

Offering assessment-oriented courses can shape teachers' assessment identity as professional L2 assessors as well (Estaji & Ghiasvand, 2022, 2023). In effective teacher education programs, teachers will develop high levels of assessment literacy that are aligned with their current pedagogical practices. As traditional approaches to foreign language teaching have been replaced with new models that emphasize the mastery of interactional skills when learning a foreign language, the assessment of foreign language knowledge needs to be consistent with the teaching. Inbar-Lourie (2008) argues that the key abilities that individuals can develop through assessment training will "reflect current views about the social role of assessment in general and language assessment in particular and contemporary views about the nature of language knowledge" (pp. 396–397).

Teachers' professional development courses need to support teachers' assessment knowledge and skills, and they should include new notions of effective and up-to-date assessment. These courses and programs should equip teachers with the necessary proficiency to supervise and assess students' learning and achievement. As Williams (2006) indicates, tailoring professional development courses and programs to the needs of in-service teachers guarantees successful assessment. Furthermore, the success of the teaching and learning process relies on the collaboration of key players in English Language Teaching, such as syllabus designers, materials developers, and teacher educators. To ensure this success, it is essential for these stakeholders to develop and disseminate thorough sets of scoring criteria. Foreign language policymakers can use the findings to support EFL teachers' assessment literacy development, especially in relation to oral skills, by changing the current teaching-driven curriculum and materials. Special testing and assessment materials can be given to teachers to augment the construction of their assessment literacy and assessment identity. A balance should be struck between pedagogy and assessment in L2 education policies because teaching and testing are two sides of the same coin (Estaji & Ghiasvand, 2019; Heaton, 1988).

Despite these implications, the study was limited in terms of scope and sample size. Future studies can involve a larger sample to offer a more comprehensive image of the oral assessment skills and literacy level of EFL teachers. The design was purely qualitative; surveys, observations, and narratives could be used to triangulate the interpretations. Another limitation was that the data in this study were collected at a single point in time through focus group interviews. Future studies can collect data in different phases to see how oral skills' assessment literacy changes over time in longitudinal explorations. The role of demographics in perceiving and practicing oral assessment was also overlooked in this study. It is advisable to examine this limitation in the future too. The Iranian educational context was the focus of this study, but future cross-cultural studies are recommended to examine the influence of socio-cultural factors in assessment literacy and practices in oral skills. Finally, the interplay of assessment agency, identity, and literacy could be an engaging line of research for enthusiastic L2 researchers.

## References

- Ahmadi, A., & Mirshojaei, S. B. (2016). Iranian English language teachers' assessment literacy: The case of public school and language institute teachers. *The Iranian EFL Journal*, 12(2), 6–32.
- Altman, D. G. (1999). *Practical statistics for medical research*. Hall/CRC Press.
- Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2015). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing*, 33(3), 411–437. <https://doi.org/10.1177%2F0265532215590847>
- Bai, B., & Yuan, R. (2019). EFL teachers' beliefs and practices about pronunciation teaching. *ELT Journal*, 73(2), 134–143. <https://doi.org/10.1093/elt/ccy040>
- Barkaoui, K. (2011). Do ESL essay raters' evaluation criteria change with experience? A mixed-methods, cross-sectional study. *TESOL Quarterly*, 44(1), 31–57. <https://doi.org/10.5054/tq.2010.214047>
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, 73(2), 113–123. <https://doi.org/10.1093/elt/ccy055>
- Bond, L. (1995). Unintended consequences of performance assessment: Issues of bias and fairness. *Educational Measurement Issues and Practice*, 14(4), 21–24. <https://psycnet.apa.org/doi/10.1111/j.1745-3992.1995.tb00885.x>
- Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language Teaching Research*, 10(1), 3–31. <https://doi.org/10.1191/1362168806lr182oa>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brinke, D. J., Bruggen, J. V., Harmans, H., Burgers, J., Giesbers, B., Koper, R., & Latour, I. (2007). Modeling assessment for re-use of traditional and new types of assessment. *Computers in Human Behavior*, 23(6), 2721–2741. <https://doi.org/10.1016/j.chb.2006.08.009>
- Butler F. A., Eignor D., Jones S., McNamara T., & Suomi B. K. (2000). *TOEFL 2000 speaking framework: A working paper* (TOEFL Monograph No. 20). Educational Testing Service.
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(3), 1–16. <https://doi.org/10.1186/s40468-020-00101-6>

- Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43–56. <https://doi.org/10.1016/j.asw.2016.03.001>
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- De Grez, L., Valcke, M., & Berings, D. (2010). Student response system and learning oral presentation skills. *Procedia - Social and Behavioral Sciences*, 2(2), 1786–1789. <https://doi.org/10.1016/j.sbspro.2010.03.985>
- De Jong, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, 9, 541–560. <https://doi.org/10.1146/annurev-linguistics-030521-052114>
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
- Duque-Aguilar, J. F. (2021). Teachers' assessment approaches regarding EFL students' speaking skill. *Profile: Issues in Teachers' Professional Development*, 23(1), 161–177. <https://doi.org/10.15446/profile.v23n1.85964>
- Estaji, M. (2024a). Perceived need for a teacher education course on assessment literacy development: Insights from EAP instructors. *Asian-Pacific Journal of Second and Foreign Language Education*, 9(50), 1–27. <https://doi.org/10.1186/s40862-024-00272-2>
- Estaji, M. (2024b). Unveiling the impact of language assessment teacher education on EFL teachers' assessment agency and professional development trajectory: perceptions, opportunities, and challenges. *Language Testing in Asia*, 14(20), 1–28. <https://doi.org/10.1186/s40468-024-00292-2>
- Estaji, M., & Ghiasvand, F. (2019). The washback effect of IELTS examination on EFL teachers' perceived sense of professional identity: Does IELTS-related experience make a difference? *Journal of Modern Research in English Language Studies*, 6(3), 83–103. <https://doi.org/10.30479/jmrels.2019.11123.1391>
- Estaji, M., & Ghiasvand, F. (2022). Teacher assessment identity in motion: The representations in e-portfolios of novice and experienced EFL teachers. *Issues in Language Teaching*, 11(2), 33–66. <https://doi.org/10.22054/ilt.2022.70302.741>
- Estaji, M., & Ghiasvand, F. (2023). Tracing the developmental trajectories of EFL Teacher Assessment Identity (TAI) through web-based audio diaries. *The Qualitative Report*, 28(9), 2681–2715. <https://doi.org/10.46743/2160-3715/2023.5903>
- Fan, J., & Yan, X. (2020). Assessing speaking proficiency: A narrative review of speaking assessment research within the argument-based validation framework. *Frontiers in Psychology*, 11, 330. <https://doi.org/10.3389/fpsyg.2020.00330>
- Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, 38(1), 47–65. <https://doi.org/10.1080/0013188960380104>
- Fleer, M. (2015). Developing an assessment pedagogy: The tensions and struggles in re-theorizing assessment from a cultural-historical perspective. *Assessment in Education: Principles, Policy and Practice*, 22(2), 224–246. <https://doi.org/10.1080/0969594X.2015.1015403>
- Frey, N., & Fisher, D. (2009). Using common formative assessments as a source of professional development in an urban American elementary school. *Teaching and Teacher Education*, 25(5), 674–680. <http://dx.doi.org/10.1016/j.tate.2008.11.006>
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Fulcher G. (2015). Assessing second language speaking. *Language Teaching*, 48(2), 198–216. <https://doi.org/10.1017/S0261444814000391>
- Gatbonton, E. (2008). Looking beyond teachers' classroom behaviors: Novice and experienced ESL teachers' pedagogical knowledge. *Language Teaching Research*, 12(2), 161–182. <https://doi.org/10.1177/1362168807086286>
- Ghiasvand, F., Jahanbakhsh, A., & Sharifpour, P. (2023). Designing and validating an assessment agency questionnaire for EFL teachers: An ecological perspective. *Language Testing in Asia*, 13(41), 1–22. <http://dx.doi.org/10.1186/s40468-023-00255-z>
- Grada, T. K. A. (2014). *An investigation into Libyan EFL novice secondary school teachers' current knowledge and practice of speaking assessment: A sociocultural perspective* [Doctoral dissertation, University of Exeter].
- Gu, P. Y. (2014). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a Foreign Language in a Chinese secondary school. *Assessment in Education: Principles, Policy and Practice*, 21(3), 286–305. <http://dx.doi.org/10.1080/0969594X.2013.836076>
- Gui, M. (2012). Exploring differences between Chinese and American EFL teachers' evaluations of speech performance. *Language Assessment Quarterly*, 9(2), 186–203.

- <http://dx.doi.org/10.1080/15434303.2011.614030>
- Heaton, J. B. (1988). *Writing English language tests: A practical guide for teachers of English as a second or foreign language*. Longman.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402. <https://doi.org/10.1177/0265532208090158>
- Inbar-Lourie, O. (2013). Language assessment literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2923–2931). Blackwell.
- Jamil, F. M., Downer, J. T., & Pianta, R. C. (2012). Association of pre-service teachers' performance, personality, and beliefs with teacher self-efficacy at program completion. *Teacher Education Quarterly*, 39(4), 119–138.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30(3), 345–362. <https://doi.org/10.1177/0265532213480334>
- Joe, J. N., Harmes, J. Ch., & Hickerson, C. A. (2011). Using verbal reports to explore rater perceptual processes in scoring: A mixed methods application to oral communication assessment. *Assessment in Education: Principles, Policy & Practice*, 18(3), 239–258. <https://doi.org/10.1080/0969594X.2011.577408>
- Kim, H. J. (2015). A qualitative analysis of rater behavior on an L2 speaking assessment. *Language Assessment Quarterly*, 12(3), 239–261. <http://dx.doi.org/10.1080/15434303.2015.1049353>
- Kim, Y. H. (2009). An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing*, 26(2), 187–217. <https://doi.org/10.1177/0265532208101010>
- Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect*, 20(3), 77–102.
- Koh, K., Carol-Ann Burke, L. E., Luke, A., Gong, W., & Tan, C. (2017). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264–288. <http://dx.doi.org/10.1177/1362168816684366>
- Koizumi, R. (2022). L2 speaking assessment in secondary school classrooms in Japan. *Language Assessment Quarterly*, 19(2), 142–161. <https://doi.org/10.1080/15434303.2021.2023542>
- Kuo, I. (2011). Student perceptions of student interaction in a British EFL setting. *ELT Journal*, 65(3), 281–290. <https://doi.org/10.1093/elt/ccq063>
- Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, 33(2), 124–140. <https://doi.org/10.1080/08957347.2020.1732383>
- Leger, D. S. (2009). Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals*, 42(1), 158–178. <http://dx.doi.org/10.1111/j.1944-9720.2009.01013.x>
- Levi, T., & Inbar-Lourie, O. (2019). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168–182. <https://doi.org/10.1080/15434303.2019.1692347>
- Lim, G. S. (2018). Conceptualizing and operationalizing second language speaking assessment: Updating the construct for a new century. *Language Assessment Quarterly*, 15(3), 215–218. <https://doi.org/10.1080/15434303.2018.1482493>
- Luoma, S. (Ed.) (2004). *Assessing speaking*. Cambridge University Press.
- Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum.
- Mertler, C. A., & Campbell, C. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(1), 49–64.
- Mottet, T.P., & Beebe, S.A. (2006). The relationships between student responsive behaviors, student socio-communicative style, and instructors' subjective and objective assessments of student work. *Communication Education*, 55(3), 295–312. <http://dx.doi.org/10.1080/03634520600748581>
- Munoz, A. P., & Alvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33–49. <https://doi.org/10.1177/0265532209347148>
- Negishi, M. (2020). Teiki shiken ga mondaida [Term tests are important and problematic]. *ELPA Vision*, 6(1). <https://elpa.or.jp/elpavision/>
- Orsmond, P., Maw, S. J., Park, J. R., Gomez, S., & Crook, A. C. (2013). Moving feedback forward: Theory to practice. *Assessment & Evaluation in Higher Education*, 38(2), 240–252. <https://doi.org/10.1080/02602938.2011.625472>
- Orsmond, P., & Merry, S. (2011). Feedback alignment: Effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education*, 36(2), 125–136. <https://psycnet.apa.org/doi/10.1080/02602930903201651>

- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23–38.  
<https://doi.org/10.1080/02602930050025006>
- O'Sullivan, B. (2006). *Modelling performance in oral language tests: Language testing and evaluation*. Peter Lang.
- Özdemir-Yilmazer, M., & Özkan, Y. (2017). Speaking assessment perceptions and practices of English teachers at tertiary level in the Turkish context. *Language Learning in Higher Education*, 7(2), 371–391.  
<https://doi.org/10.1515/cercles-2017-0015>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381–402. <https://doi.org/10.1177/0265532213480337>
- Plough, I. C., Briggs, S. L., & Van Bonn, S. (2010). A multi-method analysis of evaluation criteria used to assess the speaking proficiency of graduate student instructors. *Language Testing*, 27(2), 235–260.  
<https://doi.org/10.1177%2F0265532209349469>
- Poehner, M. E., & Inbar-Lourie, O. (Eds.). (2020). *Toward a reconceptualization of second language classroom assessment: Praxis and researcher-teacher partnership*. Springer.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265–273. <https://doi.org/10.1080/08878730.2011.605048>
- Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator*, 36(2), 115–131. <https://doi.org/10.1080/08878730009555257>
- Razavipour, K., Riazi, A., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *Journal of English Language Teaching*, 4(1), 156–161. <https://doi.org/10.5539/elt.v4n1p156>
- Richards, J. C. (2015). The changing face of language learning: Learning beyond the classroom. *RELJ Journal*, 46(1), 5–22. <https://doi.org/10.1177/0033688214561621>
- Sanchez, L.E. (2006). Bilingualism/second-language research and the assessment of oral proficiency in minority bilingual children. *Language Assessment Quarterly an International Journal*, 3(2), 117–149.  
[https://doi.org/10.1207/s15434311laq0302\\_3](https://doi.org/10.1207/s15434311laq0302_3)
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327.  
<https://doi.org/10.1177/0265532213480128>
- Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. *Educational Measurement: Issues and Practice*, 10(1), 3–6. <https://doi.org/10.1111/j.1745-3992.1991.tb00170.x>
- Shi, L. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing*, 18(3), 303–325. <https://doi.org/10.1177/026553220101800303>
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5–15. <https://doi.org/10.1111/j.1745-3992.2001.tb00065.x>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36.  
<http://dx.doi.org/10.1017/S0267190509090035>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412. <https://doi.org/10.1177/0265532213480338>
- Thuy, N. H. H., & Nga, T. T. T. (2018). An investigation into EFL teachers' perceptions of in-class English speaking assessment. *VNU Journal of Foreign Studies*, 34(2), 125–139.  
<https://doi.org/10.25073/2525-2445/vnufs.4251>
- Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023). Chinese EFL teachers' writing assessment feedback literacy: A scale development and validation study. *Assessing Writing*, 56, 1–16. 100726.  
<http://dx.doi.org/10.1016/j.asw.2023.100726>
- Webb, D. C. (2009). Designing professional development for assessment. *Educational Designer*, 1(2), 1–26.
- Williams, C. (2006). Teacher judgements of the language skills of children in the early years of schooling. *Child Language Teaching and Therapy*, 22(2), 135–154. <http://dx.doi.org/10.1191/0265659006ct304oa>
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241–256.  
<https://doi.org/10.1007/s13384-013-0089-9>
- Wilson, M. (2007). Why I won't be using rubrics to respond to students' writing. *The English Journal*, 96(4), 62–66. <https://doi.org/10.2307/30047167>

- Wray, A., & Bloomer, A. (2012). *Projects in linguistics and language studies: A practical guide to researching language*. Hodder Education.
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493–513. <https://doi.org/10.1002/j.1545-7249.2009.tb00246.x>
- Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), 228–260. <https://doi.org/10.1080/0969594X.2021.1884042>
- Yin, M. (2010). Understanding classroom language assessment through teacher thinking research. *Language Assessment Quarterly*, 7(2), 175–194. <http://dx.doi.org/10.1080/15434300903447736>
- Zhang, Y., & Elder, C. (2011). Judgments of oral proficiency by non-native and native English-speaking teacher raters: Competing or complementary constructs? *Language Testing*, 28(1), 31–50. <https://doi.org/10.1177/0265532209360671>
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323–342. [https://doi.org/10.1207/S15324818AME1604\\_4](https://doi.org/10.1207/S15324818AME1604_4)
- Zimina, E. V. (2018). Teacher involvement in the Russian national exam in foreign languages. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language tests* (pp. 245–262). Springer.
- Zolfaghari, F., & Ahmadi, A. (2016). Assessment literacy components across subject matters. *Cogent Education*, 3, 1252561. <http://dx.doi.org/10.1080/2331186X.2016.1252561>

---

## About the Authors

**Masoomeh Estaji** (<https://orcid.org/0000-0002-8014-9491>) is a Professor of Applied Linguistics at Allameh Tabataba'i University (ATU), Tehran, Iran. She earned the Top Researcher Award at ATU in 2018, 2020, and 2022. She has presented and published numerous papers on methodology, language testing, and second language acquisition in various national and international journals, such as *Teaching and Teacher Education*, *Educational Assessment*, *Language Learning in Higher Education*, *Reading Psychology*, *Instructional Science*, *Educational Media International*, *The Asian ESP*, *Asia TEFL*, and *Classroom Interaction*. Her research interests include language testing and assessment, English for Specific Purposes, and teacher education.

**Pegah Ghashghaie Nezhad** (<https://orcid.org/0000-0003-2658-1985>) is currently pursuing her second master's degree in Applied Linguistics with a fellowship at the University of Victoria, where she also serves as a teaching assistant. She holds a Master's degree in Teaching English as a Foreign Language from Allameh Tabataba'i University and has 13 years of experience as an English teacher and IELTS instructor. She was a Lecturer at Sobh-e-Sadegh Institute of Higher Education. Her research focuses on classroom discourse, teacher education, bilingualism, and translanguaging.