The Impact of Collaborative In-service Training on EFL Teachers' Language Assessment Literacy, Perceptions and Practices

Amir Mashhadi *   **Hoora Dehghani** **

Abstract

This mixed-methods study sought to examine any significant impact of Iranian English as a Foreign Language (EFL) teachers' collaboration in rating speaking tests and brainstorming theoretical concepts of language assessment (LA) on their literacy, perceptions, and practices of speaking assessment. Forty-one Iranian EFL teachers selected through purposeful sampling were asked to fill out a questionnaire, including 20 components of LA, so as to measure their knowledge and perceptions of LA components. Then, 21 of the teachers volunteered to score the video-projected speaking performance of 12 international students in the Key English Test (KET) using KET speaking rating scales. Having received six in-service speaking assessment training sessions based on their collaboration, the 21 participating teachers completed the same questionnaire and scored the same videos again. An ANCOVA was used to compare the quantitative data from the questionnaire and teachers' scoring of KET speaking tests before and after the treatment and also to investigate any significant differences between the novice and experienced Iranian EFL teachers in terms of their knowledge and perception of LA, and the assessment of KET speaking tests. Moreover, the 21 volunteered teachers were interviewed to further discover their perceptions about the in-service speaking assessment training course they attended. The findings revealed that the collaboration of the Iranian EFL teachers in rating speaking tests and brainstorming theoretical concepts of LA significantly impacted their knowledge, perceptions, and practices of LA. However, the effects were unequal and different for novice and experienced EFL teachers.

---

* Assistant Professor, Shahid Chamran University of Ahvaz, mashhadi.scu@gmail.com, Corresponding Author
** M.A., Islamic Azad University, North Branch, Tehran, houra.dehghani@gmail.com

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

90

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

Language assessment (LA) training for teachers is deemed necessary for the development of rigid rubrics for measuring learners' progress and thereby providing accurate interpretations of assessment results (Homayounzadeh & Razmjoo, 2021; Popham, 2013). Language Assessment Literacy (LAL) is defined as the ability of English as a Foreign Language (EFL) teachers to analyze, comprehend, and process assessment data, such as scores, in order to support their classroom teaching and students' learning (Inbar-Lourie, 2008). According to Lam (2015), LAL as an instructional ability includes knowledge of assessment principles and mastery of assessment skills such as test construction and use. By the same token, Malone (2008) described LAL as the competence that EFL teachers must possess to conduct assessment-related tasks.

The productive abilities of speaking and writing are more challenging for teachers to assess consistently among the four language skills because they are primarily affected by various aspects, compromising the reliability, fairness, and applicability of scores (Luoma, 2004). Teachers must therefore be capable of accurately grading learners' speaking output, particularly in classroom-based speaking assessments, utilizing efficient rating scales.

Several studies have already addressed teachers' insufficient LAL (Fulcher, 2012; Lam, 2015; López Mendoza & Bernal Arandia, 2009; Tsagari & Vogt, 2017). According to López Mendoza and Bernal Arandia (2009), LA training courses for EFL teachers can positively modify their concepts and views concerning language assessment. The operational factors that influence teachers' knowledge of assessment theories and practices should be explored for that purpose. Berry, Sheehan, and Munro (2019) believe that further research studies should be conducted to investigate the possible relationship between teachers' conceptions of assessment theories and their actual practices so as to ensure the effectiveness of the training they receive.

Although there have been several attempts to research LAL among EFL teachers worldwide, the most recent ones conducted in Iran, such as Coombe, Vafadar and Mohebbi (2020), Firoozi, Razavipour, and Ahmadi (2019), and Watmani, Assadollahfar, and Behin (2020), mainly dealt with the status of LAL among Iranian EFL teachers in terms of their perceptions and knowledge. Accordingly, there is a gap in terms of researching the impact of LAL training on Iranian EFL teachers' knowledge, perception, and actual performance, which this study was going to fill. In other words, this study adopted a mixed-methods approach rather than a correlational or survey approach to investigate any significant effect of Iranian EFL teachers' collaboration in rating speaking

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

91

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

tests and brainstorming theoretical concepts of LA on their literacy, perceptions, and practices of speaking assessment.

The following research questions were thus formulated to address the research objectives:

1. Does the collaborative in-service assessment training significantly impact Iranian EFL teachers' knowledge and perceptions of LA?
2. Does the collaborative in-service assessment training significantly impact Iranian EFL teachers' performance in assessing speaking?
3. Do novice and experienced Iranian EFL teachers differ significantly in terms of their knowledge, perception and performance of LA after the treatment?
4. How do the participating teachers perceive the collaborative in-service assessment training course?

Literature Review

LAL

LAL roots in assessment literacy (AL) as a general concept in education, defined as "an individual's understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions" (Stiggins, 1991, p. 17). LAL is also considered as teachers' and other professionals' technical information to accomplish language assessment-related tasks (Fulcher, 2012; Taylor, 2013). Inbar-Lourie (2012) argues that it is crucial to specify the required levels and areas of AL for different individuals involved in assessment-related processes in order to improve their AL. Identifying the comprising components of LAL as a highly demanded issue has also been highlighted in previous research. For instance, Davies (2008) maintained that AL has three key components of skills (mastery of developing and analyzing tests), knowledge (linguistic and statistical theories), and principles ("the proper use of language tests, their fairness, and impact, including questions of ethics and professionalism") (p. 335).

Taylor (2013) proposed a more comprehensive list of AL elements, including theoretical competence, practical know-how, knowledge of principles and notions, language training, sociocultural qualities, practices within different localities, personal outlooks/conceptions, grading, and policymaking. On the other hand, it is clear that the extent and distribution of LAL components are unequal for different teachers depending on their roles in language teaching (Kim, Chapman, Kondo, & Wilmes, 2020). In a similar strand, Deluca and Bellara (2013) believe that teachers participating in assessing students'

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

92

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

learning need to be aware of the current assessment theoretical as well as practical issues for directing, supporting, and improving learning.

However, a number of studies conducted in different educational settings, including the type of programs and teacher characteristics, have shown teachers' lack of proper LAL in assessing student language proficiency (Lam, 2015; Vogt & Tsagari, 2014; Tsagari & Vogt, 2017). Accordingly, several researchers explored possible reasons behind teachers' inadequate LAL. For instance, Lam (2015) reported a mismatch between the knowledge base provided by assessment training organizations and the needs of trainees (i.e., pre-service teachers). Others claim that teachers have sufficient LAL in theory but fail to apply it in reality (Herrera Mosquera, Macas, & Fernando, 2015; Lan & Fan, 2019; Looney, Cumming, van Der Kleij, & Harris, 2018; Stiggins, 2007). Scarino (2013), in turn, referred to varying perceptions of AL as another factor affecting the quality of teachers' assessments of students' achievement.

Other studies have underlined the importance of assessment training for in-service teachers (Bangert & Kelting-Gibson, 2007; Volante & Fazio, 2007; Mertler, 2009; Inbar-Lourie, 2012). Similarly, Tsagari and Vogt (2017) proposed that the actual settings and practical limits that affect the assessment enactment of training teachers be taken into account in the curriculum of such training courses. In effect, teachers should develop rigid rubrics to integrate theories with practice in assessing students' success in the classroom (DeLuca, Klinger, Pyper & Woods, 2015; Jonsson, Lundahl & Holmgren, 2015). Moreover, self and peer reflection of EFL teachers' performance in assessment has recently gained a lot of prominence and attention within interactive and collaborative frameworks (William & Thompson, 2007; Vogt & Tsagari, 2014; Berry et al., 2019) in response to inconsistencies in teachers' LAL in theory and practice.

Intended to improve teachers' LAL, LA training needs to be included as a basic topic within a school-based cooperative work agenda rather than a one-off program (Malone, 2008, 2013; Herrera Mosquera et al., 2015). This can be, in turn, accomplished by assisting teachers in developing their assessor personality within solo or group assessment training courses (Xu & Brown, 2016). As far as speaking assessment is concerned, one could claim that teachers should have the opportunity to reflect on their own performance within a framework that is delineated based on their needs, using an ongoing training method that incorporates elements of AL in theory and practice rather than relying on teachers' subjective knowledge base.

Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 41(3), Summer 2022, pp. 89-115	**93** **Amir Mashhadi**

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

LAL and Rating Speaking

Rating is one of the important aspects of the assessment process since it significantly impacts the state of evoked induction from the tests (Bachman & Palmer, 1996; Nowroozi & Amerian, 2020). Several studies in recent years have emphasized the importance of rating speaking as a skill that plays an important role in language learning (Rashvand, Semiyari, & Ahangari, 2019). Consistent assessment of speaking usually happens through an interaction between an examiner and a student. In effect, the examiner needs to consider an array of facets in students' oral performance, including discourse components (e.g., accuracy, fluency, and intonation) as well as demographic characteristics of the interlocutor and examinee. Although the evaluation of speaking is difficult given its complexities, it is nonetheless plausible (Bijani & Khabiri, 2017; Fulcher, 2003).

The subjectivity of the evaluations of raters has long been regarded as a serious problem in speaking assessment (Ahmadi, 2019). The reliability of scores in speaking assessment can ensure the authenticity of scores for future decision-making (e.g., learners' accurate placement). The validity of speaking scores is another important factor that exhibits the interrelationship of scales with the goals and descriptions of the concepts being assessed. As Han (2016) puts it, several elements jeopardize the validity of scores during the rating process. Thus, constructing or applying authentic classroom assessment seems to be an integral part of the teachers' mission (Allal, 2013), given that it can directly contribute to students' learning and success. Accordingly, it might be argued that classroom assessment requires a sound level of assessment literacy, which a teacher can obtain through teacher training courses, among others (Luma, 2004; Popham, 2009; Bijani & Lu, 2019).

Rating speaking is commonly facilitated through rating scales as measures, which assess students' language proficiency according to a set of predetermined levels. Regarding inconsistencies in assessment scores, some researchers claim that discrepancies in teachers' perceptions of assessment (how and what to score) lead to diverse scores (Cheng & Wang, 2007; Rashvand et al., 2019). However, discrepancies in interpretations of scales can be minimized if teachers are properly trained in interpreting rating scales and not relying on their own self-perceptions (Orr, 2002). Therefore, altering teachers' professional personalities as assessors appears to be essential in assessment research (Xu & Brown, 2016).

Alderson and Clapham (1995) proposed that scorers' preparation involves some sort of interaction and socialization so that their communal cognition of standard descriptors of rating scales develops gradually. Research has also reported on the beneficial outcomes

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

94

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

of teachers' contribution during professional development courses to discussing and examining each other's notions and practices, as well as to coming to a complementary phase of standard perception and eliminating inconsistencies in assessment-related performances (Sato, Wei, & Darling-Hammond, 2008; Popham, 2009; Koh, 2011; Wiliam & Thompson, 2007; Allal, 2013; Berry et al., 2019).

Examining different perspectives of EFL teachers and speaking examiners, Baker and Riches (2018) concluded that teacher collaboration in assessment training courses, which allows teachers to share their perspectives and skills, leads to an improvement in their AL. In another study, Adie (2013) created an online medium to assist teachers in modifying and developing their professional perceptions of assessment standards by exposing them to the perspectives and understandings of other teachers.

Overall, these studies concentrated on changing teachers' perceptions of assessment through teacher associations, though they failed to examine the impact of the trainers' enactments in a real-life context. Given the gap between assessment theory and practice, both assessment knowledge and practice should be integrated into the educational curriculum (Gu, 2015; Inbar-Lourie, 2012; Lam, 2015; Lan & Fan, 2019). In effect, teachers may be completely familiar with some concepts like reliability, validity, or rubrics, though they are incapable of applying them in a real context (Lan & Fan, 2019).

Method

This study was intended to investigate any significant impact of Iranian EFL teachers' collaboration in rating speaking tests and brainstorming theoretical concepts of LA on their literacy, perceptions, and practices of speaking assessment by adopting a mixed-methods approach. The data from the LAL questionnaire and the teachers' scores on KET speaking tests before and after the treatment constituted the quantitative part of the research, whereas the data elicited from the teachers' representative perceptions about the collaborative in-service speaking assessment training formed the qualitative part.

Participants

A total of 41 Iranian EFL teachers (20 experienced and 21 novices; 18 males and 23 females) were selected through purposeful sampling from 5 branches of a private language learning institute in Tehran, Iran, to examine their knowledge and perceptions of LA. They were aged 19 to 45, with varied years of teaching experience ranging from 1 to 17 years. All teachers had Teaching English as a Foreign Language (TEFL) degrees or certifications from Cambridge CELTA or teacher training courses (TTCs) from

	95
Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 41(3), Summer 2022, pp. 89-115	**Amir Mashhadi**

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

accredited institutes in Iran or other countries. Besides, all teachers had passed a two-credit course or a module on language testing, but none had taken courses on LA, specifically speaking assessment. The teachers obtained informed consent for participation in this study prior to the treatment. Table 1 shows the participants' demographic information, including their gender, years of teaching experience, educational level, and type of participation in this study.

Table 1

Participants' demographic information

Age range	Male	Female	Number of novice vs. experienced teachers	Educational certificate(s)	Number of volunteered participants rating the speaking tests and attending the interview
Before 20	1	2	Novice (3)	*TTC (Teacher Training Course) Certificate (3)	2
20s	11	15	Novice (18)	*TTC (Teacher Training Course) Certificate (23)	14
			Experienced (8)	*B.A. English Teaching (6) *B.A. Linguistics (1) *M.A. English Literature (2) *Cambridge CELTA (1)	
30s	4	5	Experienced (9)	*TTC (Teacher Training Course) Certificate (7) *B.A. English Translation Studies (1) *Cambridge CELTA (3) * M.A. English Teaching (1) *Ph.D. Linguistics (1)	4
40s	2	1	Experienced (3)	*TTC (Teacher Training Course) Certificate (3) * M.A. English Teaching (1) *Ph.D. English Teaching (2) * Cambridge CELTA (2)	1

Instruments

Three instruments used are as follows: 1) a LAL questionnaire that examined Iranian EFL teachers' knowledge and perceptions of LAL components; 2) a number of videos of students taking the Cambridge KET level A2 speaking test; and 3) an interview protocol.

LAL Questionnaire

To assess the Iranian EFL teachers' knowledge and perceptions of LAL components, a questionnaire including all LAL components (i.e., 18 components) as listed in Lan and Fan (2019) and the other two LAL components as cited in Kremmel and Harding (2020)

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

96

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

was administered. In effect, the questionnaire (Appendix A) elicited teachers' demographic information, familiarity with LAL components, and perceptions of the same components for LAL. The 20 components of LAL were rated twice (i.e., once to measure the participating teachers' literacy and once more to measure their perception of these components) before and after the treatment. To evaluate the accuracy of the selected items of the questionnaire for the measurement of the participants' knowledge and perceptions of LAL, the absolute value of the factor load of each of the items and their significance was computed using a t-test to be greater than 0.7. As a result, the homogeneity of the components of the questionnaire was approved, and no components needed to be removed.

Moreover, Cronbach's alpha value for both knowledge and perception of the LAL components of the questionnaire was calculated to be 0.75. The composite reliability for both parts of the questionnaire was also examined, which turned out to be 0.91 for knowledge, and 0.82 for the perception of the LAL questionnaire items. Given that the items of the questionnaire used in this study included LAL items adapted from two other questionnaires (i.e., Lan & Fan, 2019; Kremmel & Harding, 2020), the validity of the questionnaire items was evaluated using the convergent and discriminant validity through the PLS (Partial Least Squares) model. In terms of the convergent validity, the average variance extracted (AVE) index was used, and the estimated values were found to be 0.562 for knowledge and 0.650 for the perception of the LAL components.

According to the discriminant validity with the Fornell-Larcker criterion, the square root of the average variance extracted ($\sqrt{AVE}$) from the knowledge and perception parts of the LAL questionnaire as latent variables were 0.75 and 0.80, respectively. The discriminant validity was confirmed as both measures were greater than the maximum correlation of each latent variable with other latent variables (i.e., 0.59). Furthermore, the cross-validated communality (CV Com) for knowledge and perception of the LAL components of the questionnaire was calculated to be 0.53 and 0.40, respectively, indicating the approved quality of the measurement model.

Students' KET speaking videos

A number of videos of real Cambridge KET speaking tests at the A2 level were downloaded from YouTube (https://www.youtube.com/hashtag/mrmanhcambridge) for two purposes: (1) educational purposes; i.e., having the 21 volunteered teachers (9 experienced and 12 novice teachers; 10 males, 11 females) score students' speaking tests collaboratively during the training sessions; and (2) as a means for assessment, i.e.,

| | Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 41(3), Summer 2022, pp. 89-115 | **97** |
| | | **Amir Mashhadi** |

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

measuring the teachers' ability to score students' speaking (i.e., LAL in practice). Each video included an examiner, an observer who did not have any interaction with the students, and two students of the same proficiency level taking the speaking test and following the examiner's instructions in alternating order. The Cambridge KET speaking test is divided into four sections. In the first section, students were supposed to answer some general demographic questions. In the second section, the examiner established a common scenario in everyday life (e.g., buying a gift for a friend, deciding which language to study, etc.) on which the students were expected to converse. In the third section, each student had to describe a picture (e.g., people working in a laboratory, on a farm, in an airport, etc.) without interacting with their partner. Based on what they had seen in the photographs, the students were encouraged to interact with their partner and, for instance, describe their preferred jobs with convincing arguments, then ask the partner for their choice and reasons.

Interview

In order to triangulate the quantitative data from the questionnaires and speaking test scores with the qualitative data, the mentor of the course conducted a semi-structured interview asking the 21 volunteered teachers to reflect on the speaking assessment training course focusing on their collaboration and report any additional perceptional and psychological effects it had on them. Two of the participants wrote down their comments, while the rest sent their voice recordings and exhibited additional dimensions of the collaborative in-service speaking assessment training before and after taking the training course.

Data Collection and Analysis

In phase 1 of this study, the researchers used a two-part questionnaire that included 20 components of LAL adapted from Fan and Lan (2019) and Kremmel and Harding (2020). Having ensured the reliability and validity of the two parts of the questionnaire the researchers asked teachers to rate each LAL component on a scale of 0 to 5 in terms of their knowledge and perception of LAL components (i.e., their perception of the importance of LAL components). Then, 21 of the teachers volunteered to watch the recorded videos of 12 international students taking KET speaking tests at the A2 level, using the KET speaking rating scale. Grammar, vocabulary, intonation, and interactive skills were considered on this scale. The teachers were, indeed, asked to score the students' performance in the KET speaking test according to the descriptions in the rating

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

98

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

scale to ensure they used the main criteria of the comprising elements of speaking assessment based on a scale of 0 to 5.

Seventy days later, in phase 2, the 21 teachers who had volunteered to score the performances of 12 international students in the KET speaking section using KET speaking rating scales participated in the six-session in-service speaking assessment training course during which one of the researchers of this study explained the LA concepts to the participants. She also played some videos of Cambridge KET speaking tests that were similar but not identical to those in phase 1. Then, she asked the participant teachers to collaboratively score the video-projected KET speaking sample performances in groups of 5 or 6. The mentor also asked them to support their ideas by interpreting the instructions of the KET A2 level rating scale with the principles of LAL. Finally, the representatives of the two groups shared the results with the course mentor and the rest of the participants. Moreover, in rare instances of dispute or disagreement among the participants within groups or within the scope of the class, the mentor made the ultimate decision, offering logical educational support.

In phase 3, and after the passage of five months from phase 1, the same questionnaire was rated by the 41 participants. In addition, the 21 teachers scored the same students' performances in the KET speaking test as in phase 1. Furthermore, in response to an open-ended interview regarding their subsequent perceived psychological and educational implications of the treatment, the 21 volunteered instructors wrote or spoke about their views on the speaking assessment training via WhatsApp or in person.

Having garnered the data from the teachers' taking the questionnaires and scorings the speaking tests before and after the treatment, the researchers ran ANCOVA with the aim of examining: (1) any significant change in the Iranian EFL teachers' literacy and perceptions of the LAL components before and after the treatment; and (2) any significant change in the teachers' scoring of the students' KET speaking tests using the KET speaking rating scale. Furthermore, using ANCOVA and comparing descriptive statistics related to the questionnaires and the speaking scores, the researchers compared the experienced teachers' knowledge, perceptions, and practices of speaking assessment with those of novice ones before and after the in-service collaborative training course. Moreover, the comments of the 21 volunteered teachers in the interview were recorded, transcribed, and analyzed descriptively and organized into three main thematic categories: course efficacy, self-efficacy, and life-long learning.

Results

Table 1 includes the LAL questionnaire's observed mean scores and standard deviations, representing the novice and experienced teachers' LAL and perceptions of LAL comprising components before and after the treatment.

Table 2

Descriptive statistics for experienced vs. novice teachers' LAL and perceptions of LAL

LAL				Perceptions of LAL			
Novice		*Experienced*		*Novice*		*Experienced*	
before treatment	after treatment	Before treatment	after treatment	Before treatment	after treatment	Before treatment	after treatment
1.23 (0.87)	2.28 (1.63)	3.42 (0.66)	3.82 (0.21)	2.78 (1.12)	3.25 (0.73)	3.32 (0.78)	3.87 (0.12)

As indicated in Table 2, the observed mean for LAL and teachers' perceptions of LAL components increased after the training course. Further inferential analysis was conducted to examine if the difference in the mean scores was significant for the novice and experienced teachers. Having analyzed the data, the researchers reported the findings based on the four research questions. To address the first research question, the researchers compared the scores of the two parts of the questionnaire after the treatment (i.e., the teachers' LAL and perception of LAL components) by running an ANCOVA in which the teachers' initial level of LAL and their perception of LAL were the covariates.

Table 3

ANCOVA results for the teachers' LAL

	df	*F*	*p*	**Partial Eta Squared**
LAL	2	273.80	0.00	0.63
Teaching experience and LAL	2	7.56	0.00	0.17

As shown in Table 3, there was a significant difference between the teachers' LAL before and after the treatment ($F = 273.80$, $p = 0.00 < 0.05$, $\eta 2 = 0.63$). Hence, it can be concluded that teachers' LAL was significantly and positively altered after taking part in the collaborative in-service assessment training course. Considering the observed eta square, it was also found that the difference was large. In addition, the results ($F = 7.56$, $p = 0.00 < 0.05$, $\eta 2 = .17$) revealed a significant difference between the experienced and

		100
Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 41(3), Summer 2022, pp. 89-115		**Amir Mashhadi**

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

novice teachers' LAL after the treatment, while based on the observed eta square, the difference was minimal.

Table 4

ANCOVA results for the teachers' perceptions of LAL

Source	df	F	p	*Partial Eta Squared*
Perceptions of LAL	2	219.03	0.00	0.51
Teaching experience and Perceptions of LAL	2	6.64	0.00	0.14

Table 4 reveals that teachers' perceptions of LAL components were significantly different before and after the treatment ($F = 219.03$, $p = 0.00 < 0.05$, $\eta2 = 0.14$). This, in turn, indicates that teachers' participation in the in-service assessment training course significantly and positively altered their perceptions of LAL components. Based on the observed eta square, the difference was considered to be large. After the treatment, a significant difference was also identified in the novice and experienced teachers' perceptions of LAL components ($F = 6.64$, $p = 0.00 < 0.05$, $\eta2 = 0.14$), while the difference was minimal based on the eta square observed.

To answer the second research question, an ANCOVA was run to compare the 21 Iranian EFL teachers' scoring of the three aspects of speaking performance (i.e., grammar and vocabulary, pronunciation, and interactive skills) of the 12 students before and after the treatment. The covariates in this analysis were the teachers' scores to the speaking tests before the treatment.

Table 5

Descriptive statistics for the teachers' scoring of KET speaking tests

Assessment in practice			
Novice		*Experienced*	
before treatment	after treatment	before treatment	after treatment
3.89 (0.91)	3.69 (0.75)	4.06 (0.61)	3.89 (0.70)

As it is shown in Table 5, the observed means for the novice and experienced teachers' scores to the students' speaking performances decreased after the treatment. It is worth noting that although the standard deviation of the novice teachers' ratings of the students' speaking performance decreased after the treatment, that of the experienced

teachers remained almost unchanged. This demonstrates that both groups of teachers could better calibrate their ratings after receiving the treatment.

Table 6

ANCOVA for scoring KET speaking tests

Source	df	F	p	*Partial Eta Squared*
Scoring speaking test	2	283.32	0.00	0.48
Teaching experience and scoring speaking test	2	6.90	0.01	0.12

The data in Table 6 demonstrate a significant difference between the teachers' scores of students' speaking tests before and after the treatment (F = 283.32, p = 0.00 < 0.05, η2 = 0.48). As a result, the assessment methods used by the teachers to score students' speaking performances changed significantly and positively after they participated in the training course. Considering the observed eta square, it was revealed that the difference was very large. The results (F = 6.90, p = 0.01 < 0.05, η2 = 0.12) also showed a significant difference between the experienced and novice teachers' scoring of speaking tests after the treatment, while based on the observed eta square, the difference was minimal.

To address the third research question, the difference between the novice and experienced teachers' LAL and perceptions of LAL components were considered. As represented in Table 3 (F = 219.03, p = 0.00 < 0.05, η2 = 0.14), and Table 4 (F = 6.64, p = 0.00 < 0.05, η2 = 0.14), this change was significant for all the teachers, and both their LAL and perceptions of LAL components improved, as indicated by the improvements in the observed means in Table 2. However, the effect size of experience was minor regarding the observed eta squares.

The results in this section show that, while the experienced teachers' overall LAL and perceptions of LAL components were higher than those of the novice teachers before and after the treatment, the in-service assessment training course based on teacher collaboration in rating speaking tests resulted in more changes in novice teachers' LAL. As in Table 2, the observed mean score of LAL for the novice and experienced teachers rose from 1.23 to 2.28 and from 3.42 to 3.82, respectively. On the other hand, for the novice and experienced teachers, the observed mean score of the perception of LAL components increased from 2.87 to 3.25 and from 3.32 to 3.87, respectively.

Following the quantitative analysis, a qualitative analysis of the interview data was performed to answer the fourth research question. The findings are presented in the

following sections according to the emerging thematic categories identified in the interview data.

Course Efficacy

The novice and experienced teachers taking part in the semi-structured follow-up interview commented that they had learned a lot about LA, and that their perceptions of LAL had broadened significantly after receiving the treatment. They predicted that the in-service training focusing on teacher collaboration would positively impact their classroom teaching and their scoring of student speaking performances. For instance, they believed that after taking the training course, they could use more suitable criteria and measures for scoring student speaking performances and provide them with proper remedial feedback and instruction.

> What I learned in this course was very valuable since it allowed me to restructure my knowledge of L2 speaking assessment. The fact is that now I understand how to assess my students' speaking objectively. I know 'how' to apply what I learned in this course to my classroom instruction and assessment (Novice Teacher 3).

> There were some broad definitions of assessment that became more understandable for me during the course. Furthermore, I currently see assessment as an ongoing process which is in a close relationship with teaching (Experienced Teacher 1).

The diverse interpretations and, at times, conflicting ideas and notions about speaking assessment raised by the Iranian EFL teachers led to their brainstorming about LA concepts and principles, which broadened their knowledge of speaking assessment. As one teacher put it,

> The discrepancy between the teachers' notions of speaking assessment and their differing interpretations of the rating scales expanded my knowledge of speaking assessment and led to brainstorming ideas and notions related to assessment among the teachers (Novice Teacher 11).

Self-efficacy

The teachers also reported that they felt more confident about scoring student speaking performances after receiving the treatment. Three teachers opined that they used speaking rating scales more efficiently than before as a result of the training they received on speaking assessments. Another teacher maintained that, using the rating scales,

I can now assess students' speaking within a specific language teaching context according to their proficiency level, and I may no longer be perplexed. I assume that when I score speaking test results or even the classroom performance of my students, I will no longer rely on my intuition and that my grading will be more objective and fairer (Experienced Teacher 2).

It was also reported that

Being reminded of the scoring rubrics by my colleagues while discussing their practicality and calibration was more useful than reading about them in books. Brainstorming speaking assessment concepts and elements expanded our confidence in how to apply assessment principles (Experienced Teacher 9).

Lifelong Learning

Besides, teachers believed that although their knowledge of LA was broadened after they took part in the training course, they still needed to expand it, especially to keep up with other rubrics and advancements in the field.

I learned a lot about speaking assessment from the collaborative discussions and practices that happened in the course, and I found them useful and informative. However, there is a long way ahead to the mastery of speaking assessment, and we need to keep ourselves updated from now on (Novice Teacher 6).

Discussion

This study investigated whether an in-service assessment training course based on the collaboration of Iranian EFL teachers in rating speaking tests and brainstorming theoretical concepts of LA had any impact on their literacy, perceptions, and practices of speaking assessment. This study also considered if the training course could lead to significant differences between novice and experienced Iranian EFL teachers in terms of their LAL and perceptions of LAL, and their practices in rating speaking. As its third objective, a semi-structured follow-up interview was conducted with the volunteer teachers in order to discover how they perceived the speaking assessment training course they attended.

The results showed that the collaboration of Iranian EFL teachers in sharing the concepts of LA and rating students' speaking generally significantly impacted their literacy, perceptions, and practices of LA, specifically speaking assessment. The findings are consistent with prior studies by Baker (2018) and Wiliam and Thompson (2007),

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

104
Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

which found that collaboration among language teachers can improve their LAL. It also confirms the findings of Popham (2009), who found that (1) cooperation among teachers with similar goals and (2) appropriate professional training can have considerable effects on teachers' insights and perceptions of LA. The results also demonstrated that combining theoretical knowledge and principles with practices in teacher training courses increases their LAL, which is in line with Giraldo's (2019) claims.

According to the findings of the interviews, teachers experienced a positive shift in their LAL and perception of LAL components, which, as Berry et al. (2019) pointed out, may be enhanced and restored through the theory-practice connection. In addition, teachers also concluded that LAL is an inseparable part of classroom teaching and practice. In other words, they reached a different and more constructive view of formative assessment, based on which they could detect students' weaknesses and provide suitable training tailored to their individual needs. This solidly confirms Gotwals and Birmingham's (2016) standpoint that teacher preparation courses that include three elements of planning, teaching/assessing, and reflecting/revising can improve teachers' perceptions and practices of formative assessment in the classroom setting.

The findings also revealed that the teaching experience had a significant role in improving teachers' perception and literacy of LA components and their assessment practices, though its impact was minimal. That is, while novice teachers took more advantage of this course, experienced teachers were more proficient in terms of LAL, perception of LAL, and speaking assessment before and after the treatment. This may be because the more experienced the teachers were, the more opportunities they had to make a theory-practice connection with the issues raised in the course. This result is in line with what Farhady and Tavassoli (2018) claimed; teachers' experience and their LA knowledge are associated.

Based on the findings, although the training course contributed to improving LAL, perception of LAL components, and assessment practices in both experienced and novice teachers, these effects were unequal in such a way that they benefited the experienced teachers more in terms of their perceptions of LAL. That is, the level of improvement in the experienced teachers' perceptions was greater than those of their literacy and practice of LAL. On the other hand, the general level of changes in the LAL and assessment practices of the novice teachers was more significant than those in the experienced ones, which shows that the novice teachers benefited more from the training course in terms of LAL and practices.

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 41(3), Summer 2022, pp. 89-115	**105** Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

A further interesting result in this study was that the teachers' given scores to the students' speaking tests declined significantly as their literacy, perceptions, and practices of LAL components significantly rose. This could indicate a major shift in their LAL and perception of LAL and their awareness of the criteria for assessing students' speaking performance after the treatment. On the other hand, based on the mean of the speaking assessment scores (as shown in Table 5), the experienced teachers were more generous than novice ones, as they gave higher scores to student speaking tests before and after the treatment. In other words, novice teachers were prone to underestimating the students' speaking performance. Moreover, the findings from the interview confirm that the collaboration of teachers for scoring student speaking performances was highly helpful and that the relevant concepts of LAL were more tangible when applied in practice collaboratively. This outcome is consistent with Baker's (2018) conclusion, which highlights teacher collaboration and teamwork as a rewarding and helpful process for teachers. However, the findings of this study contradict those of Bijani and Lu (2019), who claimed that after speaking assessment training courses, there was no significant difference in the speaking scores of novice and experienced teachers, and that novice teachers represented more reliable speaking scores than the professional ones.

Based on the qualitative analysis from the interviews, the teachers were keen to gain more in-depth knowledge of assessments even after the training course. They believed that the process of teacher assessment training should be ongoing to help EFL teachers keep up with the most recent developments in LAL. It is also concluded that the improvement in teachers' LAL and perceptions of LAL would seemingly raise their confidence and self-efficacy in assessing speaking tests. This conclusion is based on the teachers' statements that after receiving the treatment, they felt more confident in assessing students' speaking by relying on their knowledge rather than their intuition in rating students' speaking. The teachers' increased confidence in rating speaking was also attributable to their collaboration in rating speaking tests and brainstorming theoretical knowledge of LA, which led to a more calibrated speaking assessment, particularly among novice teachers.

Conclusion

This study aimed to respond to the need to deal with the assessment-related practices of Iranian EFL teachers in the classroom. It also addressed the need for action research directed at the effect of a professional development course incorporating theory and practice in speaking assessment (Baker, 2018; Berry et al., 2019; Herrera Mosquera et al.,

Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 41(3), Summer 2022, pp. 89-115	**106** Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

2015; Tsagari & Vogt, 2017). More precisely, this study was intended to examine the demonstrated LAL (i.e., assessment of speaking) as practiced in the classroom by Iranian EFL teachers. The outcome of the research revealed that the collaboration of novice and experienced Iranian EFL teachers and the establishment of a theory-practice connection in a collaborative context led to positive outcomes for both parties and encouraged lifelong learning, adding up to the teachers' confidence and self-efficacy.

The results of this study also have a number of implications for EFL teachers, teacher trainers, and administrators of teacher training courses as follows:

Firstly, in order to develop teachers' perception, literacy, and practices of LA, particularly speaking assessment, training courses focusing on the collaboration of teachers would be highly beneficial for both novice and experienced teachers. Such courses would be more fruitful if a needs analysis of novice and experienced teachers was performed, and a goal-oriented in-service training course was developed accordingly. As the findings showed, for experienced teachers, considering assessment-related perceptions and for novice ones, bridging theory and practice activities would be more beneficial.

Secondly, the number and length of assessment training courses should be extended for novice teachers compared to experienced ones as, based on this study, experienced teachers' knowledge of assessment in theory and practice was already higher than that of the novices before and after the treatment. Thus, by speeding up or lengthening the training sessions for novice teachers, this mismatch is likely to be reduced to some extent. Thirdly, more experienced teachers must conduct sensitive and critical assessment-related activities. Based on this research, experience benefited Iranian EFL teachers' perception, literacy, and practices of LA, so hiring more experienced teachers for assessment-related activities would lead to more valid and reliable results. Finally, a fundamental shift in trainers' perspectives on LAL is also required, with a shift toward theory-practice connections through the development of practice-oriented vocational and on-the-job training programs that foster collaboration among would-be teachers, novice and experienced ones.

There were some limitations to this research. For one thing, the number of novice and experienced Iranian EFL teachers who took part in this study was limited, therefore, the findings may not be applicable to the entire EFL teaching community. Although the teachers came from 5 branches of a private language learning institute in Tehran and had a wide range of academic degrees, professional positions, and teaching experience, future studies may benefit from a larger sample of EFL teachers. A further limitation of this

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

107

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

study was the possible effect of the teachers' bias in estimating their own LAL via a self-rating questionnaire. This form of assessment could be supplemented with a literacy exam, allowing for a more accurate portrayal of teachers' LAL. A log or journal could also be used to trace the theory-practice relationship in teachers' assessment practices in classroom evaluation strategies.

Acknowledgments

Declaration

References

Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 91-106.

Ahmadi, A. (2019). A Study of Raters' Behavior in Scoring L2 Speaking Performance: Using Rater Discussion as a Training Tool. *Issues in Language Teaching*, *8*(1), 195-224.

Alderson, J. C., & Clapham, C. (1995). Assessing student performance in the ESL classroom. *Tesol Quarterly*, *29*(1), 184-187.

Allal, L. (2013). Teachers' professional judgment in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 20-34.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, *35*(4), 557-581.

Bangert, A., & Kelting-Gibson, L. (2007). Teaching principles of assessment literacy through teacher work sample methodology. *Tep Vol 19-N3*, 351.

Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, *73*(2), 113-123.

Bijani, H. (2019). Evaluating the effectiveness of the training program on direct and semi-direct oral proficiency assessment: A case of multifaceted Rasch analysis. *Cogent Education*, *6*(1), 1670592.

Bijani, H., & Khabiri, M. (2017). The impact of raters' and test takers' gender on oral proficiency assessment: A case of multifaceted Rasch analysis. *Teaching English as a Second Language*, *36*(3), 1-32.

Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, *4*(1), 85-107.

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: what do we need to learn, unlearn, and relearn? *Language Testing in Asia*, *10*, 1-16.

Davies, A. (2008). Textbook trends in teaching language testing. *Language testing*, *25*(3), 327-347.

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

108

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, *64*(4), 356-372.

DeLuca, C., Klinger, D., Pyper, J., & Woods, J. (2015). Instructional rounds as a professional learning model for systemic implementation of assessment for learning. *Assessment in education: principles, policy & practice*, *22*(1), 122-139.

Farhady, H., & Tavassoli, K. (2018). Developing a language assessment knowledge test for EFL teachers: A data-driven approach. *Iranian Journal of Language Teaching Research*, *6*(3), 79-94.

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.

Fulcher, G. (2003). Testing second language speaking. Pearson Education.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113-132.

Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *How*, *26*(1), 35-61.

Gotwals, A. W., & Birmingham, D. (2016). Eliciting, identifying, interpreting, and responding to students' ideas: Teacher candidates' growth in formative assessment practices. *Research in Science Education*, *46*(3), 365-388.

Gu, L. (2015). Language ability of young English language learners: Definition, configuration, and implications. *Language Testing*, *32*(1), 21-38.

Han, Q. (2016). Rater cognition in L2 speaking assessment: A review of the literature. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 16(1), 1-24*.

Herrera Mosquera, L., & Macías V, D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, *17*(2), 302-312.

Homayounzadeh, Z., & Razmjoo, S. A. (2021). Examining assessment literacy in practice in an Iranian context: Does it differ for instructors and learners? *Teaching English as a Second Language*, *40*(2), 1-45.

Inbar-Lourie, O. (2008). Language assessment culture. *Encyclopedia of language and education*, *7*, 285-300.

Inbar-Lourie, O. (2012). Language assessment literacy. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-9). John Wiley & Sons.

Jonsson, A., Lundahl, C., & Holmgren, A. (2015). Evaluating a large-scale implementation of Assessment for Learning in Sweden. *Assessment in education: Principles, policy & practice*, *22*(1), 104-121.

Kim, A. A., Chapman, M., Kondo, A., & Wilmes, C. (2020). Examining the assessment literacy required for interpreting score reports: A focus on educators of K–12 English learners. *Language Testing*, *37*(1), 54-75.

Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, *22*(3), 255-276.

Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, *17*(1), 100-120.

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 41(3), Summer 2022, pp. 89-115	**109** Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, *32*(2), 169-197.

Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, *61*, 112-122.

Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualizing the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 442-467.

López Mendoza, A. A., & Bernal Arandia, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile: Issues in Teachers Professional Development*, *11*(2), 55-70.

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

Malone, M. E. (2008). Training in language assessment. *Encyclopedia of language and education*, *7*, 225-239.

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, *30*(3), 329-344.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving schools*, *12*(2), 101-113.

Nowruzi, M., & Amerian, M. (2020). Exploring the factors Iranian EFL institute teachers consider in grading using personal construct theory. *Journal of Teaching Language Skills*, *38*(4), 123-164.

Orr, M. (2002). The FCE speaking test: Using rater reports to help interpret test scores. *System*, *30*(2), 143-154.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into practice*, *48*(1), 4-11.

Popham, W. J. (2013). Classroom assessment: What teachers need to know (7th ed.). Boston, MA: Pearson.

Rashvand Semiyari, S., & Ahangari, S. (2019). The Impact of Task Types and Rating Methods on Iranian EFL Learners' Speaking Scores. *Research in English Language Pedagogy*, *7*(2), 187-208.

Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, *45*(3), 669-700.

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.

Stiggins, R. (2007). Assessment through the student's eyes. *Educational leadership*, *64*(8), 22.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, *72*(7), 534-539.

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language testing*, *30*(3), 403-412.

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, *6*(1), 41-63.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374-402.

Volante, L., & Fazio, X. (2007). Exploring Teacher Candidates' Assessment Literacy: Implications for Teacher Education Reform and Professional Development. *Canadian Journal of Education*, *30*(3), 749-770.

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.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Lawrence Erlbaum Associates.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162.

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 41(3), Summer 2022, pp. 89-115	**111** Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

Appendix A

Language Assessment Literacy (LAL) Survey:

The purpose of this survey is to examine teachers' knowledge and perception of Language Assessment Literacy. Please provide honest answers to the items and <u>do not</u> leave any items unanswered. A definition has been provided for those words with a number in front of them.

Part I: General Information:

1. Name: -------------------

2. Gender: Male ☐   Female ☐

3. Age: ----------------

4. Please select your current educational level:

High School Graduate ☐   BA degree ☐   MA degree ☐   Doctorate ☐

5. Major: --------------------

6. Years of teaching experience: ------------------

Part II. Questions about classroom-based language assessment literacy

 Please specify <u>your level</u> in the following aspects

 Note:

Illiteracy: the ignorance of language assessment concepts and methods.

Nominal literacy: understanding that a specific term relates to assessment, but may indicate a misconception.

Functional literacy: comprehensive understanding of basic terms and concepts.

Procedural and conceptual literacy: understanding central concepts of the field, and using knowledge in practice.

Multidimensional literacy: knowledge extending beyond ordinary concepts including philosophical, historical and social dimensions of assessment.

	Aspect	Illiteracy	Nominal Literacy	Functional Literacy	Procedural and Conceptual Literacy	Multi-dimensional Literacy
1	Preparing classroom tests					
2	Using ready-made tests from textbook packages or from other sources					
3	Giving feedback to students based on information from tests/assessment					
4	Using self- or peer-assessment					
5	Using informal, continuous, non-test type of assessment					

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

112

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

6	Using the language Portfolio[1], an adaptation of it or some other portfolio					
7	Giving grades					
8	Finding out what needs to be taught/learned					
9	Placing students onto courses, programs, etc.					
10	Awarding final certificates (from school/program; local, regional or national level)					
11	Testing/assessing receptive skills (reading/listening)					
12	Testing/assessing productive skills (speaking/writing)					
13	Testing/assessing microlinguistic aspects (grammar/vocabulary)					
14	Testing/Assessing integrated language skills					
15	Testing/assessing aspects of culture					
16	Establishing reliability of tests/ assessment					
17	Establishing validity of tests/assessment					
18	Using statistics to study the quality of tests/assessment					
19	Selecting appropriate rating scales[2]					
20	Using rating scales to score speaking performances					

[1] A portfolio is a collection of papers, documents, and records that represent a student's work.

[2] A rating scale is a written list of performance criteria that permits the teacher more than two choices (e.g., poor, fair, good, excellent) to judge student performance of each criterion.

Part III: Questions on teachers' perceptions of classroom-based language assessment literacy:

Please specify your perceived <u>level of importance</u> of the following aspects of assessment:

	Aspect	Not At All Important	Low Importance	Neutral	Important	Very Important
1	Preparing classroom tests					
2	Using ready-made tests from textbook packages or from other sources					
3	Giving feedback to students based on information from tests/assessment					
4	Using self- or peer-assessment					
5	Using informal, continuous, non-test type of assessment					
6	Using the language Portfolio[1], an adaptation of it or some other portfolio					
7	Giving grades					
8	Finding out what needs to be taught/learned					
9	Placing students onto courses, programs, etc.					
10	Awarding final certificates (from school/program; local, regional or national level)					

Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(3), Summer 2022, pp. 89-115

113

Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

	Aspect	Not At All Important	Low Importance	Neutral	Important	Very Important
11	Testing/assessing receptive skills (reading/listening)					
12	Testing/assessing productive skills (speaking/writing)					
13	Testing/assessing microlinguistic[2] aspects					
14	Testing/assessing integrated language skills[3]					
15	Testing/assessing aspects of culture					
16	Establishing reliability[4] of tests/assessment					
17	Establishing validity[5] of tests/assessment					
18	Using statistics to study the quality of tests/assessment					
19	Selecting appropriate rating scales[6]					
20	Using rating scales to score speaking or writing performances					

[1] A Portfolio is a collection of papers, documents, and records that represent a student's work.

[2] Microlinguistic aspects like grammar and vocabulary concern themselves with the study of language systems in the abstract, without regard to the meaning or notional content of linguistic expressions.

[3] An integrated language assessment is based on testing two or more language skills simultaneously as opposed to discrete-point testing.

[4] Test reliability refers to the degree to which a test is consistent and stable in measuring what it is intended to measure.

[5] Test validity is the extent to which a test accurately measures what it is supposed to measure.

[6] A rating scale is a written list of performance criteria that permits the teacher more than two choices (e.g., poor, fair, good, excellent) to judge student performance of each criterion.

Thank you for your participation

Appendix B

Speaking rating sheet for A2 level of KET exam

Cambridge English: Key (LEVEL A2) SPEAKING GRAMMAR & VOCABULARY	Name of student: Name of teacher:
1-Does the speaker use simple grammatical forms with sufficient control?	Good Not good
2-Does the speaker use simple grammatical forms with a good degree of control?	Good Not good
3-Does the speaker use appropriate vocabulary to talk about everyday situations?	Good Not good
Comment 1:	Comment 2:

Cambridge English: Key (LEVEL A2) SPEAKING PRONUNCIATION	Name of student: Name of teacher:
1-Are the utterances mostly clear? Can the speaker be mostly understood?	Good Not good
2-Does the speaker show limited control of intonation?	Good Not good
3-Does the speaker show limited control of word and sentence stress?	Good Not good
4-Are individual sounds mostly clear?	Good Not good
Comment 1:	Comment 2:

Cambridge English: Key (LEVEL A2) INTERACTIVE COMMUNICATION	Name of student: Name of teacher:
1-Can the speaker maintain simple exchanges with the interlocutor (Part 1)?	Good Not good
2-Does the speaker react appropriately to what the interlocutor or other candidate says	Good Not good
3-Does the speaker need any prompting or support?	Good Not good
Comment 1:	Comment 2:

	Teaching English as a Second Language Quarterly (TESLQ) (Formerly Journal of Teaching Language Skills) 41(3), Summer 2022, pp. 89-115	**115**
		Amir Mashhadi

THE IMPACT OF COLLABORATIVE IN-SERVICE TRAINING ON EFL

Appendix C

KET speaking rating scale for level A2:

A2	Grammar and Vocabulary	Pronunciation	Interactive Communication
5	• Shows a good degree of control of simple grammatical forms. • Uses a range of appropriate vocabulary when talking about everyday situations.	• Is mostly intelligible, and has some control of phonological features at both utterance and word levels.	• Maintains simple exchanges. • Requires very little prompting and support.
4	Performance shares features of Bands 3 and 5.		
3	• Shows sufficient control of simple grammatical forms. • Uses appropriate vocabulary to talk about everyday situations.	• Is mostly intelligible, despite limited control of phonological features.	• Maintains simple exchanges, despite some difficulty. • Requires prompting and support.
2	Performance shares features of Bands 1 and 3.		
1	• Shows only limited control of a few grammatical forms. • Uses a vocabulary of isolated words and phrases.	• Has very limited control of phonological features and is often unintelligible.	• Has considerable difficulty maintaining simple exchanges. • Requires additional prompting and support.
0	Performance shares features of Bands 1.		