

Moving toward a Democratic Assessment Framework: Iranian EFL Teachers' Critical Language Assessment Literacy

Ali Momeni^{1*}, Esmaeel Ali Salimi²

Received: 8 October 2022

Accepted: 28 December 2022

Abstract

In spite of the resurgence of interest in teachers' language assessment literacy (LAL), teachers' critical language assessment literacy (CLAL) has not garnered satisfactory prominence. This constitutes a substantial void in the extant corpus of literature that this study will address. To this illumination, a cross-sectional survey research approach utilizing the Critical Language Assessment Literacy (CLAL) scale (Tajeddin et al., 2022) was employed to accomplish the stated purpose of the study. The current study conveniently selected 120 Iranian EFL teachers in different language institutes and universities in Tehran and Kermanshah province. To analyze data, descriptive (e.g., mean and standard deviation) and inferential statistics (e.g., one sample t-test) were employed to illustrate a comprehensive answer to the study's primary research question. The results of the one-sample t-test and the mean value for all CLAL items showed that teachers have moderate or low knowledge of CLAL factors. In addition, the findings signposted that effective teachers must retain high linguistic competence, teaching/learning cognition, and knowledge of critical language assessment literacy factors; this could be accomplished by including these factors in teacher training programs. These findings displaying shortcomings and gaps in the critical language assessment literacy (CLAL) level of Iranian EFL teachers hold many theoretical and pedagogical implications for teacher development in the realm of assessment and testing at the practice, policy, and professional development levels.

Keywords: Assessment literacy, critical language assessment, critical language assessment literacy, Iranian EFL teachers' assessment literacy

1. Introduction

Within the past couple of decades, assessment has been acknowledged to be one of the most radical considerations of educational environments because it exerts a substantial effect on the efficacy of education and, consequently, learning (Azizi, 2022; Momeni, 2022; Rezai et al., 2021; Sharma, 2020). Language assessment has gained noticeable momentum in a multitude of educational systems nationwide owing to the fact that the efficiency of every educational system is contingent upon the caliber of its assessment cornerstones (Stiggins, 1999). In concert with Lee and Butler's (2020) investigation, language evaluation appears to serve a distinctive function not just in teaching but also within the community, exhibiting the prominence of language as a major mode of interaction and information exchange. With the

¹ PhD Candidate of Applied Linguistics, Allameh Tabataba'i University, Tehran, Iran.
ali.momenniii@gmail.com; momeni_ali@atu.ac.ir (corresponding author)

² Assistant Professor of Applied Linguistics, Department of English Language and Literature, Allameh Tabataba'i University, Tehran, Iran, casalimi@atu.ac.ir

same token, a large array of literature bears witness to the concern that assessment must be contemplated as an attribute that influences deep learning (Coombe et al., 2020; Kirkwood & Prive, 2008), enhances the second language (L2) learners' motivation (Brookhart & Bronowicz, 2003; Birenbaum, 1996), reinforces L2 learners' self-concept (Black & Wiliam, 2010), develops teacher professionalism (Abell & Siegel, 2011; Engelsen & Smith, 2014), increases sociocultural awareness within multiple discourses (Willis, Adie, & Klenowski, 2013), and raises L2 learners' knowledge of quality assessment (Smith et al., 2013). As Malone (2013) rightly signposted, robust, well-administered assessment feeds students, teachers, and other testing partners with essential information regarding student achievement and the amount to which instructional learning outcomes have been reached. In this light, assessment knowledge appears to be a fundamental component of instructors' assessment competency, practice, and process (Lan & Fan, 2019; Kyttälä et al., 2022). As part of their ongoing professional development, teachers are required to be knowledgeable about crucial assessment concerns since their assessment literacy (AL) has a significant influence on their course of action in classrooms (DeLuca & Klinger, 2010; Leung, 2014; Mertler, 2009). It is erroneous to assume that if somebody can teach exceptionally effectively, they also exhibit a high magnitude of literacy in evaluating language learners (Davies, 2008; Spolsky, 1978) because as public-school educators in the Popham (2004) study maintained, assessment is "a complex, quantitative arena well beyond the comprehension of mere mortals" (p. 82). In addition, assessment literacy encompasses comprehension and effective application of assessment procedures, as well as awareness of the theoretical and philosophical grounds of measuring students' performance (DeLuca & Klinger, 2010; Stiggins 2002; Volante & Fazio, 2007). To put it more tellingly, the term "assessment literacy" (Stiggins, 1997) has become endorsed to designate the spectrum of competencies and information that interested parties require to navigate the emerging landscape of assessment into which we have been thrown (Fulcher, 2012). Empirical studies have associated assessment literacy with considerable yields in student learning and pedagogical enhancements (Campbell & Collins, 2007; Wilson et al., 2001). Contemporary concepts of assessment emphasize the incorporation of assessment procedures and outcomes as educational protocols that encourage self-regulated learning and the implementation of metacognitive competencies (Black & Wiliam, 2010; Earl, 2003). Good classroom assessment empowers teachers to make appropriate conclusions about the accomplishment of each student, disseminate this information to students and parents, and target subsequent instruction (Brookhart, 1999). Ineffective classroom assessment, on the contrary side, reduces reliability and validity, culminating in misguided and unsuitable instructional decisions. In consonance with the socially co-constructed characterization, LAL necessitates teachers to comprehend and scrutinize the function and importance of assessment practices as a function of their effect and the position of teacher learning prospects in a particular sociocultural, political, educational, and philosophical milieu (Coombe et al., 2020; Fulcher, 2012; Levi & Inbar-Lourie, 2020; O'Loughlin, 2013; Scarino, 2017; Yan & Fan, 2020).

In the pursuit of words of wisdom offered by the post-method timeframe and the conducive contributions of critical pedagogy for the interface of language assessment, the conventional approach of testing has lately been disputed, and a novel framework of assessment, known as critical or democratic language assessment (CLA), has emerged (Akbari, 2008; Kumaravadivelu, 1994; Sun, 2021). CLA seeks to link language testing to the geopolitical, educational, and cultural contexts in which it is frequently conducted or/and controlled for specific objectives (Shohamy, 2001a, 2001b, 2017; Spolsky, 1995). CLA, as characterized by Shohamy (2001a), is especially relevant to the inspection of assessment purposes and outcomes in social and educational environments. Messick (1981), who

recommended an expansion of construct validity as a substantial aspect of test quality, seems to have argued that the connections between assessment, education, and society comprise his central argument. This expansion includes characteristics of test usage, such as values, influence, and intentional or unintentional outcomes. CLA offers a paradigm change in the field of language assessment and testing by including new benchmarks, notably systemic, consequential, interpretative, and ethical indications of validity, which help put into account the validity of the test applications (Shohamy, 2007). In this line, as a forerunner in the discipline, Shohamy (2001b) provided an intricate framework encompassing CLA concepts and urged assessment practitioners to entail these crucial ideas into language assessment. Unfortunately, she did not adequately illustrate how to implement these concepts into practical sophistication.

In addition, further investigation is necessary on the notion of CLA in terms of language teachers' awareness of CLA, the competencies needed for the conduct of this assessment, and the underlying concepts (Tajeddin et al., 2022). Furthermore, as research on English as a foreign language (EFL) teachers' LAL reveals that CLA is one of the gaps in their knowledge foundation (Vogt & Tzagari, 2014), further inquiry on CLA as a vital component of LAL is still warranted. Consequently, this study has theoretical significance since it can give insight into the level of CLA among Iranian EFL teachers. Meanwhile, in the extant body of literature, there is a dearth of study that explores the amount of CLAL, constituting a major gap that this study seeks to fulfill. To this end, the primary purpose of the current study is to examine the CLAL proficiency of Iranian EFL teachers.

2. Review of the Related Literature

2.1. Theoretical Framework

The conceptualization of 'language assessment literacy' (LAL) is frequently utilized as a subcategory of assessment literacy (AL), which has been the emphasis of assessment and testing-oriented general educational research during the last two decades (Rezagah, 2022). It is also particularly notable that the concern for LAL has been currently an ongoing trend of inquiry, and countless studies have been executed on various aspects of LAL, including LAL in line with the teachers' assessment skills and knowledge (e.g., Al-Bahlani, 2019; Jawhar & Subahi, 2020; Latif, 2021; Rauf & McCallum, 2020); LAL as the complexity of teachers' perceived assessment training requirements (e.g., Brown et al., 2019; Giraldo, 2019), and the sophistication of demand for LAL development among different stakeholders (e.g., Kremmel & Harding, 2020; O'Loughlin, 2013).

Since Stiggins' (1991) influential manuscript on 'assessment literacy,' which he defines as a person's capacity to assess the weaknesses and strengths of an assessment and implement such awareness in decision-making regarding student performance, the trend of language assessment literacy has emerged as a point of contention. Drawing upon the literature, it can be claimed that LAL is a distinguishable field of inquiry due to the complexities implicated in assessing linguistic and communicative competence, knowledge, and skills (Levi & Inbar-Lourie, 2020; Plake & Impara, 1996; Vogt et al., 2020). The aforementioned characterization of AL is consistent with Davies' (2008) meticulous investigation of main movements in language assessment textbooks, in which he observed that their materials are shifting from skills and knowledge approach to a principles-based approach. This is also applicable to language assessment programs (Jeong, 2013; Jin, 2010), which merit partial but insufficient priority to assessment fundamentals and ramifications (Giraldo, 2018). As Davies (2008) highlighted, skills involve training in vital and acceptable techniques, including test analysis, item writing, statistics, and software programs accountable for test analysis, administration, and reporting. Knowledge refers to a foundation

in assessment, linguistic competency, and creating the scenario. The principles address crucial concerns in language assessment, such as assessment outcomes, washback, impact, test fairness, and ethical concerns. Establishing a more critical standpoint on the assessment knowledge substrate of stakeholders implicated in language assessment, Inbar-Lourie (2008a, 2008b) outlined LAL as the capability to raise and respond to key considerations concerning the rationale for assessment, the suitability of the instrument being deployed, testing settings, and what would occur based on test findings.

2.2. *Critical Language Assessment (CLA)*

In tandem with the historical sophistication on the ground of language testing and measurement, it is safe to postulate that measurement techniques are the instruments for the instillation and encroachment of doctrines and powers of specific communities, a manifestation designated as CLA, which is strongly entrenched in critical pedagogy and critical applied linguistics as antithetical to traditional language assessment/testing which adopts a "hierarchical classroom structure of teacher over students" (Crawford, 1978, p. 91; Norton & Toohey, 2004; Shohamy, 1998). In addition, the sources of power at CLT appear at three interfaces, namely the state (e.g., the bureaucracy of testing), discourse (e.g., the imposition of tests by unequal individuals), and ideology (e.g., what is deemed right or wrong) (Tollefson, 1995). Critical pedagogy, conceptualized as the characterization of the ramifications and utilization of tests in society and education, is a paradigm that seeks to challenge the legitimacy of repressive authorities in societal structure (Freire, 1970; Giroux & Bosio, 2021; Spolsky, 1995). More prominently, critical pedagogy holds a practice-leveled stance where critical pedagogy is not a collection of theories but a methodology of conducting teaching and learning and behavior inspired by a particular orientation toward society and classrooms (Canagarajah, 2005). Pioneered by leading figures in the field (e.g., Spolsky, Pennycook, Kramersch, and Shohamy), CLT strives to empower individuals by equipping them with the key tools and skills necessary to view the world through dynamic and transformational glasses, not as a static structure but as a reality in transition and development (Freire, 1970, 1973; Rasmussen, 1996). In this frame of reference, education is, therefore, the exercise of liberation from top-down decision-making processes, with all representatives cooperatively attending and coordinating to combat social prejudices and unfairnesses, enabling the empowerment and social acceptance of people from diversified socioeconomic and cultural contexts to have their voices and thoughts heeded (McLaren, 2020). As such, under this assessment foundation that acknowledges test-takers to be people, the right of people is valued, and it is asserted that the framework should embrace the doctrine of the role of the people, by the people, and for the people (Estaji & Ghiasvand, 2022; Taylor, 1998). In lockstep with the before-alluded standpoint, Javidanmehr and Rashidi (2011) introduced the complexity of the "right of test takers" and Shohamy (2001a) put it as the constitution of test-takers as "black boxes". In parallel with the metaphor of test-takers as "black boxes", Shohamy (2001) acknowledged test-takers as the "true victims" of tests since they are compelled to pursue the authority of tests without interrogating its contents and their connections to their existing level of understanding (Bourdieu, 1991). Additionally, critical pedagogues are driven to assume that educational institutions are not divorced from the social, political, cultural, and economic realities of learners (Giroux, 2020), but rather that they reflect the social dynamics in which they operate. Owing to the extant marginalization and discrimination in light of race, social class, and gender in social institutions, similar viewpoints and behaviors are naturally repeated in educational systems (Tajeddin et al., 2022).

Inspired by Giroux, and other critical pedagogues, Shohamy (1997) was among the first scholars to meticulously apply the rules and concepts of critical theory and critical pedagogy into the domain of language testing in order to "minimize, limit and control the powerful uses of tests" (p. 131). According to her, CLA is a field of study located within the framework of critical pedagogy (Shohamy, 2017). In his American Association for Applied Linguistics plenary presentation entitled *Critical Language Testing and Beyond*, Shohamy (1997) argued for the integration of critical ideology within the area of language assessment. In fact, Shohamy (2001a, 2001b) endeavored to promote the awareness of both practitioners and researchers that language assessment is a type of social praxis that incorporates cultural, educational, social, and ideological norms. Afterward, several assessment professionals had the same viewpoint (e.g., Fulcher, 2012; Inbar-Lourie, 2008b; Kunnan, 2018; Lynch, 2001; Schissel, 2019). In compliance with this claim, language assessment has transformational potential and may be used to result in further educational and societal alterations that benefit those in charge (Brown, 2019; Rea-Dickins, 2008; Shohamy, 2017). As Shohamy (2001a) maintained, there is abundant confirmation that they are effective instruments that are frequently implemented in an unethical and undemocratic manner for disciplinary objectives and the implementation of diverse policy goals. Likewise, as Shohamy (2001a) rightly remarked, test findings are bestowed with economic significance and can therefore serve as a tool for managing knowledge.

In this regard, the utilization of exams as instruments of power violates the ideals and standards of democratic values. The assessment approach that the CLA demands is the one that can reform curriculum and teaching in a genuinely democratic manner, not one that supports the purposes of those in power and demeans, trivializes, and marginalizes those of others (Fetterman et al., 1996). Additionally, the CLA assessment mechanism incorporates another characteristic of CLT, namely dialogic interaction in which teacher and student roles are shared, and all voices are heard (Keesing-Style, 2003; Nevo, 1996). Consequently, CLA is also known as democratic assessment. Based on democratic assumptions, CLA seeks to restrict and prevent the misuse of tests and other types of measurement apparatuses and processes (Tajeddin et al., 2022). Therefore, assessment concepts drawn from democratic practices can reframe and restructure assessment procedures that may prejudice specific individuals and groups. Shohamy (2001b) established a definitive model of CLA standards in a determined attempt to establish the ground for the adoption of democratic assessment techniques in the language assessment area.

2.3. *The Purpose and Research Gap*

Although attempts have been performed to introduce sociocultural–political–ethical implications to the frontline of crucial LAL boundary lines (e.g., Brindley, 2001; Fulcher, 2012; Inbar-Lourie, 2008a; Pill & Harding, 2013; Taylor, 2013), more inquiry is necessary to broaden LAL qualifications (Inbar-Lourie, 2013a; Taylor, 2013) and, more saliently, to "better operationalize LAL conceptualizations" (Stabler-Havener, 2018, p. 18) by focusing on CLA benchmarks to promote more accurate measures to evaluate teachers' LAL and CLAL as part of their assessment knowledge foundation. In the spirit of the Stabler-Havener (2018) research, developing more effective measures of LAL will be contributive for teacher educators to (1) uncover strong points and weaknesses in teachers' LAL and (2) promote professional development alternatives to remedy the observed deficiencies. Due to the fact that teachers' LAL necessities differ considerably from setting to setting (Stabler-Havener, 2018; Vogt et al., 2020), more contextually sensitive devices to quantify EFL teachers' LAL level can offer a clearer understanding of their assessment knowledge foundation and, more particularly, the weak aspects of their LAL. This assessment can guide more relevant

programs for teacher education. In fact, the extant body of knowledge illustrates that all instructors lack LAL proficiency, and they unquestionably require incentives to increase their assessment expertise while striving to satisfy specific assessment demands (Vogt et al., 2020; Vogt & Tsagari, 2014). In addition, research on EFL instructors' LAL substantiate that teachers' LAL levels are insufficient for engaging in effective assessment procedures (e.g., Lam, 2015; Sultana, 2019; Vogt & Tsagari, 2014), let alone their CLAL, which is one of the weakest facets of their assessment knowledge basis. Studies also demonstrate that non-native English-speaking instructors (those whose L1 is not English) require training in many features of LAL (Firoozi et al., 2019; Gottheiner & Siegel, 2012). In addition, in the context of Iran, there is a severe lack of studies with regard to the EFL teachers' CLAL proficiency which renders an important gap. In this way, the demand for more investigation on teachers' LAL, and more especially on EFL instructors' CLAL as part of their LAL, was the impetus for the current study. The current study is sophisticated with the following research question.

1. What is the critical language assessment literacy level of Iranian EFL teachers, as measured by Critical Language Assessment Literacy (CLAL) scale (Tajeddin et al., 2022)?

3. Research Methodology

3.1. The Research Design

The current study is concerned with the quantitative measurement of data accumulation, and a questionnaire was employed to gather the data. More specifically, a cross-sectional survey research strategy was adopted to address the primary research question of the study. The rationale for this survey design rests in the interface that neither treatment nor the learning process that the participants may have undergone as a relevant factor is engaged in the study (Hatch & Lazaraton, 1991). To this end, the study aimed to collect data from its participants by employing a validated questionnaire (Tajeddin et al., 2022).

3.2. Participants

The participants of the current study were 120 Iranian EFL teachers in different language institutes and universities in Tehran and Kermanshah province. In the paper-based version, these teachers were targeted through their language institutes/universities, and in the online format, through their LinkedIn, Research Gate, Email, or Telegram groups. Due to the fact that participants voluntarily answered the questionnaire, this study adopted convenience sampling, the most common sampling technique in L2 research (Dornyei, 2007). Note should be taken that the teachers' respective field was Teaching English as a Foreign Language (TEFL), English Literature, Linguistics, or English Language Translation Studies, or they were involved in programs leading to these degrees. In addition, 38% of the participants were female, while 61% were male, and their ages ranged from 20 to 51 years old. Table 1 summarizes the participant information that is essential for the purpose of this research.

Table 1

Participants' Demographic Information

Variable	Category	Frequency	Percentage
Gender	Female	46	38.3%
	Male	74	61.7%
Age	20-30	44	36.7%
	31-40	40	33.3%
	41-50	20	16.7%
	x>51	16	13.3%

Teaching Experience	X<5 years	24	20%
	6-10 years	37	30.8%
	11-15 years	29	24.2%
	16-20 years	12	10%
	21-25 years	11	9.2%
	X>26 years	7	5.8%

3.3. Instrumentation

In this study, the Critical Language Assessment Literacy (CLAL) scale (Tajeddin et al., 2022) was employed to collect the data. The questionnaire was composed of six sections entitled (a) demographic information of the participants (7 items); (b) teachers' knowledge of assessment objectives, scopes, and types (15 items); (c) assessment use consequences (14 items); (d) fairness (4 items); (e) assessment policies (3 items); and (f) national policy and ideology (2 items). The items in the demographic information section asked for documentation about the gender, age, educational background, English teaching experience, and the context and level of teaching of the participants. The questionnaire included 38 Likert-type items (ranging from strongly disagree to strongly agree): 1= *strongly disagree*, 2= *disagree*, 3= *neither agree nor disagree*, 4= *agree*, and 5= *strongly agree*. The participants were asked to show their critical language assessment literacy competence on the aforementioned scale. The scale was validated through exploratory factor analysis for construct validity and Cronbach's alpha for estimating internal consistency in the original paper (Tajeddin et al., 2022), and the results met widely accepted indicators of construct validity and reliability. The researcher calculated Cronbach's alpha reliability for the entire questionnaire and reported it to be 0.75, as shown in Table 2, which details the reliability of all questionnaire components. According to the outcomes of this factor-wise reliability test in Table 2, all of the factors/components have the necessary internal consistency, and the research questionnaire is reliable.

Table 2
Reliability of the Five Factors of CLAL

Factors	Items Number	Cronbach Alpha ($\alpha > 0.7$)	
Factor 1	Assessment objectives, scopes, and types	15	.840
Factor 2	Assessment use consequences	14	.866
Factor 3	Fairness	4	.915
Factor 4	Assessment policies	3	.883
Factor 5	National policy and ideology	2	.893
Total		38	.753

In addition, the content validity of the questionnaire was reinforced by soliciting the viewpoints of experts in English language testing and assessment (as assessed by their publications). The feedback from experts was encouraging, and based on their view, the questionnaire was appropriate for the study's objectives. To more appropriately contextualize the assessment procedures in English language teaching, and since all of the teachers had an advanced level of English proficiency, the questionnaire was administered in English, and no translation was necessary.

3.4. Procedure and Data Analysis

In order to enhance the response rate of the questionnaire, both an online platform (e.g., Google Forms) and a paper-based format were deployed. After the selection and

validation of the questionnaire, it was sent to EFL teachers in various contexts. In paper-based format, the researcher attended various universities and language institutions to administrate and gather responses from teachers. In addition, in the online format, participants were targeted via their profiles on LinkedIn, Research Gate, Instagram, Telegram, and Whats App, and a link to the questionnaire was sent to them so that they could submit their responses online. As research has shown that a little financial motivation can increase response rates (Ary et al., 2006; Saunders & Mitchell, 2004), the author gave participants the opportunity of selecting one of the accompanying rewards as a mark of appreciation for spending the time to answer the questionnaire. They may select between a 10,000 Toman code for phone charging, a pdf download of two practicum course booklets that were not available on the internet, or a two-hour online counseling session on how to utilize the SPSS program for applied linguistics inquiry. Despite the researcher's endeavors in disseminating the questionnaire to 400 teachers in various contexts through both platforms, only 120 participants responded, and this number was reduced to 100 since some responses were incomplete. The collected data from the questionnaire were analyzed using the SPSS package program (Version 21.0) by applying the contribution of both variable-centered (t-test) and person-centered approaches (Howard & Hoffman, 2018). To put it another way, both descriptive (e.g., mean and standard deviation) and inferential statistics (e.g., t-test) were employed to illustrate a comprehensive answer to the study's primary research question.

4. Results and Discussion

To address the study's research question, the researcher analyzed each section of the questionnaire to determine the level of literacy for each component of critical language assessment.

4.1. Assessment Objectives, Scopes, and Types

The second section of the questionnaire examined teachers' knowledge of assessment objectives, scopes, and types. The constituents of this factor are listed in Table 3. In the spirit of the data illustrated in Table 3, the means of separate items varied in a range of 4.69 for the high mean to 3.37 for the low mean. Items 15, 12, 11, 5, and 10 received the highest means for teachers' assessment objectives, scopes, and types, with mean scores of 4.69, 4.43, 3.93, 3.93, and 3.92, correspondingly. The data also revealed that Items 4, 3, and 2 had the lowest means, with 3.37, 3.52, and 3.54 mean scores, correspondingly. Some CLAL items in Table 6 (e.g., 6, 7, and 10) may intersect with overall LAL constructions outlined in the literature, primarily since it can be challenging to draw a straightforward line between general LAL constructs and CLA-related capabilities. In fact, LAL constructions interact with one another as opposed to being autonomous of one another. As a case in point, awareness of crucial concerns impacts the activities and methods of test designers. Interestingly, the highest frequency in the "strongly disagree" column in Table 3 is attributable to item four, which could testify to the fact that multiple-choice items cannot be completely disregarded in the context of Iran, as they are appropriate for tests whose results would be utilized to render very critical decisions about the future of the testees. Despite the merits of multiple-choice items, the use of multiple-choice items in this context may not be in line with the view that language tests should promote interpretive methodologies for assessment (in accordance with CLA principles), permitting for "different meanings and interpretations rather than a single absolute truth" (Shohamy, 2001a, p. 377). However, the justification for the multiple choice and objective items in the Iranian context can be bounded to the concern that "in a country where connections are imperative in doing anything, it is difficult to find other fairer ways than the college entrance exams" (Kirkpatrick & Zang, 2011, p.41) which are multiple choice and objective in nature. On the

other hand, item 15 was exposed to the highest frequency in the "strongly agree" column of Table 3, indicating that the assessment system in Iran is mostly dependent on the regulations imposed on test-developers that might not be in lockstep with the educational principles and benchmarks of standard testing. The sophistication of this regulation-oriented assessment system can be reflected in Babaii et al., (2020) statement that "Iranian L2 teachers often experience barriers in executing their preferred assessment practices. This dilemma seems to be mostly a function of the way assessment is understood in this context" (p.13).

Table 3
Items Corresponding to the Assessment Objectives, Scopes, and Types Factor

Item	Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	M	SD
1	Language teachers should critically analyze which and whose policies tests serve	2 (%1.7)	28 (%23.3)	8 (%6.7)	54 (%45)	28 (%23.3)	3.65	1.128
2	The nature of knowledge that tests measure should be critically analyzed.	4 (%3.3)	30 (%25)	7 (%5.8)	55 (%45.8)	24 (%20)	3.54	1.166
3	Test developers need to take the view of different stakeholders such as teachers, students, and parents into consideration.	7 (%5.8)	29 (%24.2)	6 (%5)	51 (%42.5)	27 (%22.5)	3.52	1.243
4	Language teachers should challenge the widely used traditions in language assessment, like multiple-choice tests.	11 (%9.2)	32 (%26.7)	4 (%3.3)	47 (%39.2)	26 (%21.7)	3.37	1.329
5	Traditional tests, such as multiple-choice items, do not pay much attention to test-takers' attitudes toward some test methods.	4 (%3.3)	17 (%14.2)	7 (%5.8)	48 (%40)	44 (%36.7)	3.93	1.139
6	Tests should lead to language teachers' knowledge about assessment and their professional development.	0 (%0)	8 (%6.7)	4 (%3.3)	46 (%38.3)	62 (%51.7)	4.35	.837
7	Tests should improve language learners' proficiency in all language skills and sub-skills.	5 (%4.2)	15 (%12.5)	8 (%6.7)	55 (%45.8)	37 (%30.8)	3.87	1.115
8	Alternative assessments (such as portfolios, self-assessment, and peer assessment) should be used to provide a clearer picture of language learners' performance and ability.	6 (%5)	19 (%15.8)	7 (%5.8)	61 (%50.8)	27 (%22.5)	3.70	1.135
9	Alternative assessments (such as portfolios, self-assessment, and peer assessment) can reduce test-takers' stress and anxiety.	5 (%4.2)	16 (%13.3)	9 (%7.5)	59 (%49.2)	31 (%25.8)	3.79	1.099
10	Test-takers' preferences for test format and item types should be considered in test construction.	4 (%3.3)	14 (%11.7)	7 (%5.8)	58 (%48.3)	37 (%30.8)	3.92	1.066
11	A handful of test methods can provide better evidence for important decisions about test-takers than a single test method.	8 (%6.7)	10 (%8.3)	8 (%6.7)	51 (%42.5)	43 (%35.8)	3.93	1.168

12	In language testing, the role of students and teachers as two important test parties should be taken into consideration.	0 (%0)	5 (%4.2)	7 (%5.8)	39 (%32.5)	69 (%57.5)	4.43	.786
13	In many cases, test content rather than course syllabus determines what to teach and how to teach	2 (%1.7)	27 (%22.5)	10 (%8.3)	38 (%31.7)	43 (%35.8)	3.78	1.205
14	A test-taker should be assessed using many classroom test tasks rather than only the final exam.	9 (%7.5)	14 (%11.7)	11 (%9.2)	64 (%53.3)	22 (%18.3)	3.63	1.137
15	Language assessment not only belongs to test developers but rather to other stakeholders such as teachers, students, and parents	0 (%0)	1 (%0.8)	7 (%5.8)	20 (%16.7)	92 (%76.7)	4.69	.619
	Mean	4.47 (%3.72)	17.67 (%14.73)	7.33 (%6.1)	49.73 (%41.45)	40.8 (%34)	3.87	.574

4.2. Assessment Use Consequences Factor

In compliance with the findings represented in Table 4, participants' awareness of assessment consequences or consequential validity was below mean, with item means ranging between 3.33 and 2.39, clearly stating that the teachers' knowledge of assessment use consequences is average and below average. This result appears to support the results of the study, which demonstrated that the majority of EFL teachers have low LAL competencies (e.g., Lam, 2015; Razavipour & Rezagah, 2018; Sultana, 2019). According to the data, items 12 (3.33), 11 (3.14), and 13 (2.98) exhibited the greatest mean scores, whereas items 8, 7, and 6 received the lowest means, accordingly. In addition, item six in Table 4, had the highest frequency of 32 in the "strongly disagree" column, implying a consideration that teachers adopt a negative attitude toward examining the consequences of tests for students due to a lack of time in the classroom to undertake such a task. The literature offers robust confirmation of teachers' lack of time (e.g., Mohammadi et al., 2020; Takrouni, 2022; Vakili & Ebadi, 2022), which is exacerbated in the university context since "classes in universities are highly crowded, whereas classes in language institutes are much smaller" (Momeni & Nushi, 2022, p.413). In the meanwhile, items 11 and 12 received the highest frequency of 24 and 35, respectively, in the "strongly agree" column of Table 4, which refers to the washback effect of tests (e.g., Alderson & Wall, 1993; Binnahedh, 2022), and it is aligned with the viewpoint that "implementations of new testing systems should take place only if they are going to impose positive changes in the educational system" (Hatipoğlu, 2016, p. 143). Moreover, "the continuous examination of the quality of tests to reduce the negative effects of their uses and interpretations" necessitates increasing the assessment literacy levels of teachers through long-term assessment training or teacher development courses (Ahangari & Alizadeh, 2015; Shafaghi & Estaji, 2020; Sun & Zhang, 2022) which has been underscored by a multitude of researchers, including LaVelle and Donaldson (2010), who assert that "an extended period of training is necessary to master the evaluation-specific skills and knowledge necessary to provide quality service to clients, and be socialized into the professional frameworks, standards, and ethical guidelines" (p. 10).

Table 4
Items Corresponding to the Assessment Use Consequences Factor

Item	Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	M	SD
1	Test developers need to understand the negative and positive consequences of tests	25 (%20.8)	44 (%36.7)	8 (%6.7)	24 (%20)	19 (%15.8)	2.73	1.407
2	Test developers are responsible for the consequences of tests.	16 (%13.3)	58 (%48.3)	8 (%6.7)	22 (%18.3)	16 (%13.3)	2.70	1.287
3	Inappropriate language tests lead to psychological consequences such as test-takers' stress and anxiety.	22 (%18.3)	50 (%41.7)	8 (%6.7)	20 (%16.7)	20 (%16.7)	2.72	1.385
4	Inappropriate language tests lead to social consequences for test-takers, such as deprivation of qualified students from their own rights.	24 (%20)	48 (%40)	8 (%6.7)	24 (%20)	16 (%13.3)	2.67	1.356
5	Inappropriate language assessment leads to injustice and a lack of fairness.	20 (%16.7)	44 (%36.7)	8 (%6.7)	26 (%21.7)	22 (%18.3)	2.88	1.409
6	The positive and negative consequences of tests for test-takers should be examined carefully.	32 (%26.7)	44 (%36.7)	9 (%7.5)	18 (%15)	17 (%14.2)	2.53	1.396
7	Measures should be taken to protect test-takers from the misuse of tests by authorities.	30 (%25)	50 (%41.7)	8 (%6.7)	23 (%19.2)	9 (%7.5)	2.43	1.261
8	Tests have an important function because their results have significant implications for test-takers and language education	31 (%25.8)	49 (%40.8)	9 (%7.5)	24 (%20)	7 (%5.8)	2.39	1.232
9	Tests should have a positive impact on the learning outcomes of language learners	21 (%17.5)	52 (%43.3)	9 (%7.5)	30 (%25)	8 (%6.7)	2.60	1.226
10	Teaching practices are largely determined by language test demands and test results	20 (%16.7)	55 (%45.8)	9 (%7.5)	22 (%18.3)	14 (%11.7)	2.63	1.284
11	Teaching and testing are closely connected to each other.	18 (%15)	32 (%26.7)	9 (%7.5)	37 (%30.8)	24 (%20)	3.14	1.404
12	Continuous examination of the quality of tests helps reduce the negative effects of their uses and interpretations	18 (%15)	26 (%21.7)	9 (%7.5)	32 (%26.7)	35 (%29.2)	3.33	1.469
13	Gaining a better understanding of how tests are used is essential to controlling and reducing the negative effects of test uses and interpretations.	17 (%14.2)	40 (%33.3)	11 (%9.2)	33 (%27.5)	19 (%15.8)	2.98	1.350
14	All participants affected by the test results should be informed about the harmful effects and consequences of tests.	13 (%10.8)	54 (%45)	10 (%8.3)	32 (%26.7)	11 (%9.2)	2.78	1.217
	Mean	21.93 (%18.28)	46.14 (%38.45)	8.79 (%7.32)	26.21 (%21.84)	16.93 (%14.11)	2.75	.739

4.3. Fairness Factor

Factor 3 (4 items) comprised factors linked to fairness, as illustrated in Table 5. Fairness in language testing is conceptualized as the availability of equitable instructional opportunities for all individuals in society, regardless of their race, religion, gender, ideology, culture, or political beliefs. Fairness and validity are highly correlated in language testing (e.g., Kane, 2010; Kane & Bridgeman, 2017; Kunnan, 2010; Xi, 2010), as "social consequences, such as impartiality and justice of actions and comparability of test consequences, are at the core of fairness" (Xi, 2010, p. 147).

In concert with the data in Table 5, the items with the highest and lowest average scores were 3 and 1, with values of 2.77 and 2.52, correspondingly. The average mean for items in the domain of fairness is less than 3, demonstrating teachers' average or low-oriented understanding of concerns linked to assessment fairness. This finding is in accordance with the findings of Sarani et al. (2014), who reported that Iranian EFL instructors have a limited understanding of critical components of language pedagogy and occasionally breach the tenets of critical pedagogy. This limited awareness and knowledge can be attributed to teacher preparation programs and curricula that have not effectively addressed critical pedagogy and post-method language teaching doctrines (Alibakhshi & Rezaei-Mezajin, 2013; Sarani et al., 2014) and/or critical concerns in language assessment as a component of instructors' LAL. Additionally, this low-oriented knowledge or understanding is adequately displayed by the overall mean of the "disagree" column, which is 47 and higher than the overall means of the other columns (e.g., agree; strongly agree). Similarly, in the "strongly agree" column of Table 5, item 4 earned the highest frequency of 27, which might be attributed in part to the teachers' lack of awareness of crucial issues in language assessment. This concern is also partially related to the fear that teachers in Iran are constrained by the regulations of their respective institutes or universities and that these restrictions may not be welcomed by teachers because they diminish their voice and render a top-down approach to the educational system (see Gan & Lam, 2020; Sheehan & Munro, 2019), which signifies that school administrators 'on behalf of the people define how students' knowledge must be assessed and evaluated (Sadeghi & Jabbarnejad, 2012). More importantly, item 2 in column of table 4 received the highest frequency of 17 in this column, reflecting that the assessment system of teachers in the Iranian context is influenced by the sophistication of racism and gender bias. This explains that the assessment outcomes of EFL teachers may be polluted by the gender of their students, and this issue is substantially showcased in the extent literature on EFL teaching, learning, and assessment (see Abobaker et al., 2021; Ahmed & Ganapathy, 2021; Wallace & Qin, 2021, for more examples). An epitomized testimony to this point can be traced in the statement of one of the participants of the Rezai et al., (2022) that: "I admit that my teaching and testing practices may be affected by my students' gender. For example, sometimes, I give implicitly a higher score to the female students who have an attractive appearance. I think this is a part of our nature driving our attention to absorbing things and individuals" (p. 9).

Table 5
Items Corresponding to the Fairness Factor

Item	Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Mean	St.D
1	Some rating scales might unfairly favor specific test-takers with particular language proficiency.	24 (%20)	55 (%45.8)	8 (%6.7)	21 (%17.5)	12 (%10)	2.52	1.270
2	Some test tasks might unfairly	26	41	7	29	17	2.75	1.404

	favor male or female test-takers.	(%21.7)	(%34.2)	(%5.8)	(%24.2)	(%14.2)		
3	Content of international tests, such as TOEFL and IELTS, should be analyzed from a critical perspective to understand the likely cultural and ideological biases in them.	19 (%15.8)	45 (%37.5)	8 (%6.7)	41 (%34.2)	7 (%5.8)	2.77	1.242
4	Language test content should be based on the representation of the multiple religious, ethnic, and gender groups of society.	27 (%22.5)	47 (%39.2)	9 (%7.5)	30 (%25)	7 (%5.8)	2.52	1.250
	Mean	24 (%20)	47 (%39.17)	8 (%6.7)	30.25 (%25.21)	10.75 (%8.96)	2.64	1.032

Notes: TOEFL=Test of English as a Foreign Language; IELTS = International English Language Testing System

4.4. Assessment Policies Factor

The three-item factor 4 reflected assessment policies. Table 6 comprises items linked with assessment policies. The interface of CLA assessment policies requires teachers and assessment producers to establish a critical perception of whose interests and policies are addressed by assessment procedures and their outcomes. Table 6 discloses that all items pertaining to assessment policies had averages below 3, suggesting that teachers' expertise in this domain was average or low. This upshot is incongruous with the concerns of language testing academics, who are certain that test designers and test-takers must be acutely mindful of the political dimensions of the employment of tests in settings like immigration, citizenship, asylum, and scholarship (Brown, 2019; Shohamy, 2001a, 2017). The findings may also imply that they did not receive enough simplification about the policy element of language tests during their earlier in-service training classes; hence, they are exposed to a lack of relevant understanding of test policies and the ideological, cultural, and political interests they fulfill. In addition, while the third item showed the greatest mean score (2.88), the second item got the lowest mean score (2.57). A closer examination of Table 6 shows that item 2 earned the highest frequency of 52 in the "disagree" column, further confirming the poor level of assessment policies awareness on the part of Iranian EFL teachers. In addition, items 1 and 3 served for the highest frequency of 18 in the "strongly agree" column of Table 6. This finding could be interpreted as underscoring the complexity of assessment and testing, which is impacted by diversified institutional, political, cultural, and educational policy dynamics of the social environment, which influence teachers' assessment policy, attitudes, practices, and knowledge base (Latif & Wasim, 2022; Looney et al., 2017; McNamara, 2001; Scarino, 2013).

Table 6
Items Corresponding to the Assessment Policies Factor

Item	Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	M	SD
1	Because of the use of tests in contexts such as immigration, asylum, citizenship, or scholarship, test-takers sometimes serve the political interests of governments.	19 (%15.8)	45 (%37.5)	9 (%7.5)	29 (%24.2)	18 (%15)	2.85	1.358
2	Tests are tools used within a	23	52	8	27	10	2.57	1.261

	context of social and ideological contexts.	(%19.2)	(%43.3)	(%6.7)	(%22.5)	(%8.3)		
3	Language teachers should have a critical awareness of which and whose interests and policies tests serve.	23 (%19.2)	38 (%31.7)	7 (%5.8)	34 (%28.3)	18 (%15)	2.88	1.403
	Mean	21.67 (%18.06)	45 (%37.5)	8 (%6.7)	30 (%25)	15.3 (%12.75)	2.77	1.079

4.4. National Policy and Ideology

The two-item factor 5 corresponds to national policy and ideology. The associated items are depicted in Table 7. Table 7 reveals that the second item, with the national policy and ideology factor, possessed a mean value greater than 3, whereas the first item had a mean value less than 3, denoting that Iranian EFL teachers' assessment procedures are strongly dominated by the rules and regulations that their institutes or universities enforce. To this end, "under the influence of micro- and macro-level contextual variables, teachers find themselves working in a culture of certainty and compliance" marked by certain pre-identified criteria and boundaries in the form of norms, conventions, policies, and rules that guide them in how and what they can and cannot practice in terms of assessments" (Latif & Wasim, 2022, p. 15). This finding is in lockstep with those of Brown et al.'s (2019) study, which manifests that teachers' multifaceted assessment beliefs, practices, and perspectives reflect the diversification of their societal and cultural backgrounds and that this diversification influences their classroom educational and assessment procedures (also see Rogers et al., 2007; Troudi et al., 2009).

Table 7

Items Corresponding to the National Policy and Ideology Factor

Item	Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	M	SD
1	The government's political and ideological policies affect the administration and interpretation of language tests used for selection, placement, and achievement purposes.	19 (%15.8)	43 (%35.8)	9 (%7.5)	34 (%28.3)	15 (%12.5)	2.86	1.330
2	The government's educational policies affect language test administration and interpretation.	18 (%15)	32 (%26.7)	10 (%8.3)	40 (%33.3)	20 (%16.7)	3.10	1.368
	Mean	18.5 (%15.42)	37.5 (%31.25)	9.5 (%7.92)	37 (%30.83)	17.5 (%14.58)	2.98	1.191

4.5. Answering the Research Question

The primary research question was: what is the critical language assessment literacy level of Iranian EFL teachers, as measured by the Critical Language Assessment Literacy (CLAL) scale (Tajeddin et al., 2022)? It is of paramount significance to highlight that the level is determined by the mean value of all CLAL items in the adopted questionnaire among the Iranian EFL teacher participants. To this end, to provide the answer to the research question, a one-sample t-test is used. The rationale for applying a one-sample t-test is rooted in the interface that there were five factors in the questionnaire, and these five factors construct the variable of the current study, which is CLAL of Iranian EFL teachers. If the level of this variable is to be determined, the values, which are all in the form of decimal numbers, must be investigated; therefore, the results cannot be analyzed using a five-point

Likert scale. To this aim, in order to measure the degree of CLAL proficiency among Iranian EFL teachers, the one sample t-test is performed. This test differs from the paired or independent samples t-test in that there is only one variable or group, and no comparison will be made. The results of one sample t-test are shown in Tables 8 and 9.

Table 8
The Results of One-sample T-test for CLAL

	t	df	Sig. (2-tailed)	Test Value = 3		
				Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Total	5.808	119	.000	.19539	.1288	.2620

According to Table 8, the significance level is less than 0.05 standard error, and the t-statistic is more than 1.96 ($t(119)=5.808$); therefore, the null hypothesis centered on the equality of the mean and (test value=3) is rejected. The fact that both (upper) and (lower) have positive values demonstrates that the mean is greater than the value (test value). To assure the accuracy of this result, Table 9's findings are displayed.

Table 9
The Results of One-sample Statistics for CLAL

	N	Mean	Std. Deviation	Std. Error Mean
Total	120	3.1954	.36851	.03364

Table 9 clearly clarifies that the average level of literacy of teachers was measured (3.19), together with the standard deviation (0.368). Due to the small value of the standard deviation, it is plausible to draw the conclusion that the data are close to the mean and have limited dispersion. In conclusion, Table 9 demonstrates that the mean value for all CLAL items is 3.19, implying that teachers have moderate or low knowledge of CLAL.

5. Conclusion and Implication

The contributing purpose of the current research sought to investigate the CLAL competency of Iranian EFL teachers. To this illumination, a cross-sectional survey research strategy was utilized to achieve the stated purpose of the study. Needless to say, the absence of needed AL can "cripple the quality of education" (Popham, 2009, p. 43) and is considered "professional suicide" (Popham, 2004, p. 82). As a necessary component of their ongoing professional development, strengthening teachers' literacy in language assessment key constituents enables them to participate successfully in assessment and enhance teaching and learning effectiveness. In this respect, the findings of this study are consistent with previous research on teachers' pedagogical knowledge (e.g., DeLuca et al., 2016), which confirmed that effective teachers must retain high linguistic competence and teaching/learning cognition, as well as knowledge of assessment, its objectives, scopes, and types in particular. In this investigation, it was discovered that language instructors' LAL must take into consideration the relationship between the scope of language assessment and its influence on stakeholders, including instructors, learners, test creators, and curriculum and materials programmers. Accordingly, both language instructors and other stakeholders must be cognizant that language evaluation should assist in strengthening and safeguarding the

liberties of individuals involved in the assessment procedures of powerful organizations. In reality, language evaluation is currently recognized as an integrated process in sight of LAL inquiry findings (Butler et al., 2021). It must be consistent with evaluations of empowerment and democratic assessment, wherein dialogical and collaborative techniques, as well as groups of stakeholders, collaborate in order to strengthen the program (Fetterman, 2019; Shohamy, 2001a).

In light of the findings of the study, teachers' low awareness of fairness factors in language assessment necessitates the concern that fairness requires the incorporation of all influenced individuals in the evaluation and the undertaking of assessment in a way as to mitigate bias, thus optimizing ethical concerns to ensure power associations (Fulcher, 2015; Kane, 2010) as characterized by Foucault (1982), "mobile," "reversible," and "reciprocal" (as cited in Lynch, 2001, p. 365). Hence, every probable cause of bias that leaves a systematic uneven, and unjust influence on particular populations at any level of the evaluation process should be compensated for, rectified, eliminated, or substantially reduced. This result corroborates Tierney's (2016) claim that "democratic values now call for inclusive educational systems that recognize and support student diversity" (p. 5).

In addition, the study theoretically attached documentation to the body of literature on CLAL in the context of Iran, which has not been researched sufficiently in this realm. The findings can also lay the groundwork for the fact that as "teachers have the capacity to be agents of change where school policies related to testing and grading are concerned" (Clark-Gareca, 2019, p. 56), they should have autonomy over classroom assessment procedures. Teachers, teacher trainers, and policymakers may benefit from the results of the current study. Teachers may be viewed as the true protagonists in language assessment and testing since a substantial percentage of testing transpires in the classroom, highlighting their importance as a stakeholder group (Tsagari, 2021). Teachers can profit from the results by enhancing their knowledge and assessment literacy, particularly in regard to critical issues of language assessment to improve their students' academic performance. Teacher trainers can play a crucial function in directing student teachers in teacher preparation programs and assisting with the implementation of CLAL in their professions. In contrast to teachers, who have limited control over their materials, university teacher trainers are allowed to compile and pick their resources. Consequently, they would make a significant contribution to facilitating this transition to critical-oriented assessment practice. The position of policymakers is also unquestionably fundamental because they are concerned with promoting language assessment literacy (Kremmel & Hrding, 2020). They can review what has been accomplished so far and the results gained to give much scholarly attention to CLAL in the realm of testing and teaching. As it is the case with other research studies, the generalizability of the current study's outcomes is compromised by certain restrictions. As the sample size of this study was limited, a greater sample size is required to increase the validity of the results. Furthermore, analyzing the actual assessments conducted in classrooms through observation can provide a broader portrayal of classroom assessments in practice, which can help researchers obtain additional data resources for acquiring a deeper understanding of the nature and impact of critical language assessment practices on language learning. Accordingly, triangulation and mixed methods methodologies are required to garner further evidence on the principle of CLA.

List of Abbreviations

Assessment Literacy (AL), Critical Language Assessment (CLA), Critical Language Assessment Literacy (CLAL), English as a foreign language (EFL), English Language Teaching (ELT), English as a Second Language (ESL), Teaching English to Speakers of

Other Languages (TESOL), Teaching English as a Foreign Language (TEFL), Statistical Package for Social Sciences (SPSS)

Acknowledgments

The author(s) wish to thank all the participants of the study for their cooperation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone, (Eds.), *The professional knowledge base of science teaching* (pp. 205-221). Springer: Dordrecht. https://doi.org/10.1007/978-90-481-3927-9_12
- Abobaker, R. M., Elsayed Khalil, S., Mubarak Merghani, M., Mahadeen, A., Abdelraheem, E. G., & Hamdan-Mansour, A. M. (2021). E-learning success factors from the perspective of academic staff at nursing and education colleges during COVID-19 pandemic: A comparative study. *Educational Sciences: Theory and Practice*, 21(3), 1–10. <https://doi.org/10.12738/jestp.2021.3.001>
- Ahmed, A. A. A., & Ganapathy, A. (2021). Creation of automated content with embedded artificial intelligence: a study on learning management system for educational entrepreneurship. *Academy of Entrepreneurship Journal*, 27(3), 1–10
- Akbari, R. (2008). Postmethod discourse and practice. *TESOL Quarterly*, 42, 641–652. <https://doi.org/10.1002/j.1545-7249.2008.tb00152.x>
- Al-Bahlani, S. M. (2019). *Assessment literacy: A study of EFL teachers' assessment knowledge, perspectives, and classroom behaviors* (Unpublished doctoral dissertation). The University of Arizona, UA Campus Repository, University Libraries. <https://hdl.handle.net/10150/633240>
- Alderson, J.C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–29. <https://doi.org/10.1093/applin/14.2.115>
- Alibakhshi, G., & Rezaei Mezajin, S. (2013). On the consequences of the violation of critical pedagogy principles from Iranian EFL teacher trainers' perspectives. *Iranian Journal of Applied Language Studies*, 5(2), 1–28. <https://doi.org/10.22111/IJALS.2015.1875>
- Ary, D., Jacobs, L. C., Razavieh, A., & Sorenson, C. (2006). *Introduction to research in education*. Belmont, CA: Thomson Higher Education.
- Azizi, Z. (2022). Fairness in assessment practices in online education: Iranian university English teachers' perceptions. *Language Testing in Asia*, 12(1), 1-17. <https://doi.org/10.1186/s40468-022-00164-7>
- Babaii, E., Molana, K., & Nazari, M. (2020). Contributions of assessment-related critical incidents to language teacher identity development. *Innovation in Language Learning and Teaching*, 1-16. <https://doi.org/10.1080/17501229.2020.1824234>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Binnahedh, I. A. (2022). E-assessment: Wash-back effects and challenges (examining students' and teachers' attitudes towards E-tests). *Theory and Practice in Language Studies*, 12(1), 203-211. <https://doi.org/10.17507/tpls.1201.25>
- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes, and prior knowledge* (pp. 3-29). Dordrecht, Netherlands: Springer.
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90. <https://doi.org/10.1177/003172171009200119>
- Bourdieu, P (1991). *Language and symbolic power*. Cambridge, MA: Harvard University Press.
- Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, 18(1), 5-13. <https://doi.org/10.1111/j.1745-3992.1999.tb00002.x>
- Brookhart, S. M., & Bronowicz, D. L. (2003). I don't like writing. It makes my fingers hurt: Students talk about their classroom assessments. *Assessment in Education: Principles, Policy & Practice*, 10(2), 221–242. <https://doi.org/10.1080/0969594032000121298>
- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126–136). Cambridge: Cambridge University Press.
- Brown, G. T. L. (2019). Is assessment for learning really assessment? *Frontiers in Education*, 4(64), 1-7. <https://doi.org/10.3389/feduc.2019.00016>
- Brown, G. T. L., Gebriel, A., & Michaelides, M. P. (2019). Teachers' conceptions of assessment: A global phenomenon or a global localism. *Frontier. Education*, 4(16). <https://doi.org/10.3389/feduc.2019.00016>
- Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 38(3), 429–455. <https://doi.org/10.1177/0265532221992274>
- Campbell, C., & Collins, V. L. (2007). Identifying essential topics in general and special education introductory assessment textbooks. *Educational Measurement: Issues and Practice*, 26(1), 9-18. <https://doi.org/10.1111/j.1745-3992.2007.00084.x>
- Canagarajah, S. (2005). Critical pedagogy in L2 learning and teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 931–949). Lawrence: Lawrence Erlbaum Associates.
- Clark-Gareca, B. (2019). Assessing pre-service teachers' assessment literacy: Building a foundation. In E. White, & T. Delaney (Eds), *Handbook of research on assessment literacy and teacher-made testing in the language classroom* (1st ed., pp. 42-57). IGI Global
- 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>
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P.

- Davidson, B. O'Sullivan, & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 20–29). Cambridge: Cambridge University Press.
- Crawford, L. M. (1978). *Paulo Freire's philosophy: Derivation of curricular principles and their application to second language curriculum design* (Unpublished doctoral dissertation). University of Minnesota. Retrieved from ProQuest Digital Dissertations. (AAT 7911991)
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- DeLuca, C., & Klinger, D.A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438. <https://doi.org/10.1080/0969594X.2010.516643>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/S11092-015-9233-6>
- Dornyei, Z. (2007). *Research methods in applied linguistics: Qualitative, quantitative, and mixed methodologies*. Oxford: Oxford University Press.
- Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Engelsen, K., Smith, K. (2014). Assessment literacy. In C. Wyatt-Smith, V. Klenowski, & P. Colbert, (Eds.), *Designing assessment for quality learning. The enabling power of assessment* (pp. 91-107). New York, NY: Springer Dordrecht. https://doi.org/10.1007/978-94-007-5902-2_6
- Estaji, M. & Ghiasvand, F. (2022). *Language testing and assessment: A comprehensive guide*. Allameh Tabataba'i University: Allameh Tabataba'i University Press.
- Fetterman, D., Kaftarian, S. & Wandersman, A. (1996). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage.
- Fetterman, D. (2019). Empowerment evaluation: A stakeholder involvement approach. *Association of Health Promotion Professionals*, 30(2), 137-142. <https://doi.org/10.1002/hpja.243>
- Foucault, M. (1982). The subject and power. In H. Dreyfus & P. Rabinow (Eds.), *Michel Foucault: Beyond structuralism and hermeneutics* (pp. 777-795). Chicago: University of Chicago Press.
- Freire, P. (1970). *Pedagogy of the oppressed* (Bergman, R., Trans). New York: Herder & Herder.
- Freire, P. (1973). *Education: The practice of freedom*. Writers and Readers.
- Fulcher, G. (2010). *Practical language testing*. London: Hodder Education
- 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). *Re-examining language testing: A philosophical and social inquiry*. London and New York: Routledge.
- Gan, L., & Lam, R. (2020). Understanding university English instructors' assessment training needs in the Chinese context. *Language Testing in Asia*, 10(1), 1-18. <https://doi.org/10.1186/s40468-020-00109-y>
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>

- Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *SciELO Analytics*, 26(1), 35–61.
<https://doi.org/10.19183/how.26.1.481>
- Giroux, H. A., & Bosio, E. (2021). Critical pedagogy and global citizenship education. In E. Bosio (Ed.), *Conversations on global citizenship education: Perspectives on research, teaching, and learning in higher education* (pp. 1-10). London: Routledge.
- Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education*, 23(5), 531-557.
<https://doi.org/10.1007/s10972-012-9278-z>
- Hatch, E. M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House Publishers.
- Hatipoğlu, Ç. (2016). The impact of the university entrance exam on EFL education in Turkey: Pre-service English language teachers' perspective. *Procedia-Social and Behavioral Sciences*, 232, 136-144. <https://doi.org/10.1016/j.sbspro.2016.10.038>
- Howard, M. C., & Hoffman, M. E. (2018). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods*, 21(4), 846-876. <https://doi.org/10.1177/1094428117744021>
- Inbar-Lourie, O. (2008a). 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. (2008b). Language assessment culture. In E. Shohamy (Ed.), *Language testing and assessment* (Vol. 7, 2nd ed., pp. 285–300). Springer.
- Inbar-Lourie, O. (2013a). Guest editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301–307. <https://doi.org/10.1177/0265532213480126>
- Inbar-Lourie, O. (2013b, November). *Language assessment literacy: What are the ingredients?* [Paper presentation] 4th CBLA SIG Symposium Programme, University of Cyprus, Nicosia, Cyprus.
- Javidanmehr, Z., & Rashidi, N. (2011). Critical language assessment: Students' voices at the heart of educational system. *Elixir Psychology*, 37, 3740-3746.
- Jawhar, S. S., & Subahi, A. M. (2020). The impact of specialty, sex, qualification, and experience on teachers' assessment literacy at Saudi higher education. *International Journal of Learning, Teaching and Educational Research*, 19(5), 200–216.
<https://doi.org/10.26803/ijlter.19.5.12>
- 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>
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555–584.
<https://doi.org/10.1177/0265532209351431>
- Kane, M. (2010). Validity and fairness. *Language Testing*, 27(2), 177-182.
<https://doi.org/10.1177/0265532209349467>
- Kane, M., & Bridgeman, B. (2017). Research on validity theory and practice at ETS. In R. E. Bennett & M. von Davier (Eds.), *Advancing human assessment: The*

- methodological, psychological and policy contributions of ETS* (pp. 489-552). Springer: Springer Nature.
- Kirkpatrick, R., & Zang, Y. (2011). The negative influences of exam-oriented education on Chinese high school students: Backwash from classroom to child. *Language Testing in Asia*, 1(3), 1-10. <https://doi.org/10.1186/2229-0443-1-3-36>
- Kirkwood, A., & Price, L. (2008). Assessment and student learning: A fundamental relationship and the role of information and communication technologies. *Open Learning: The Journal of Open, Distance and E-learning*, 23(1), 5-16. <https://doi.org/10.1080/02680510701815160>
- Keesing-Style, L. (2003). The relationship between critical pedagogy and assessment in teacher education. *Radical Pedagogy*, 5(1). Retrieved Oct 22, 2010, from http://radicalpedagogy.icaap.org/content/issue5_1/03_keesingstyles.html
- 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. <https://doi.org/10.1080/15434303.2019.1674855>
- Kumaravadivelu, B. (1994). The postmethod condition: (E)merging strategies for second/foreign language teaching. *TESOL Quarterly*, 28, 27–48. <https://doi.org/10.2307/3587197>
- Kunnan, A. J. (2010). Test fairness and Toulmin's argument structure. *Language Testing*, 27(2), 183-189. <https://doi.org/10.1177/0265532209349468>
- Kyttälä, M., Björn, P. M., Rantamäki, M., Lehesvuori, S., Närhi, V., Aro, M., & Lerkkanen, M. K. (2022). Assessment conceptions of Finnish pre-service teachers. *European Journal of Teacher Education*, 1-19. <https://doi.org/10.1080/02619768.2022.2058927>
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. <https://doi.org/10.1177/0265532214554321>
- 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. <https://doi.org/10.1016/j.stueduc.2019.03.003>
- Latif, M. W., & Wasim, A. (2022). Teacher beliefs, personal theories and conceptions of assessment literacy—a tertiary EFL perspective. *Language Testing in Asia*, 12(1), 1-22. <https://doi.org/10.1186/s40468-022-00158-5>
- Latif, M. W. (2021). Exploring tertiary EFL practitioners' knowledge base component of assessment literacy: Implications for teacher professional development. *Language Testing Asia*, 11(19). <https://doi.org/10.1186/s40468-021-00130-9>
- LaVelle, J. M., & Donaldson, S. I. (2010). University-based evaluation training programs in the United States 1980-2008: An empirical examination. *American Journal of Evaluation*, 31(1), 9-23. <https://doi.org/10.1177/1098214009356022>
- Lee, J., & Butler, Y. G. (2020). Reconceptualizing language assessment literacy: Where are language learners? *TESOL Quarterly*, 54(4), 1098–1111. <https://doi.org/10.1002/tesq.576>
- Leung, C. (2014). Classroom-based assessment issues for language teacher education. In A. J. Kunnan, (Ed.), *The companion to language assessment* (pp. 1510-1519). Chichester, UK: Willey Blackwell.
- Levi, T., & Inbar-Lourie, O. (2020). 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>

- Looney, A., Cumming, J., Kleij, F. V. D., & Harris, K. (2017). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442–467.
<https://doi.org/10.1080/0969594X.2016.1268090>
- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18(4), 351–372.
<https://doi.org/10.1177/026553220101800403>
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344. <https://doi.org/10.1177/0265532213480129>
- McNamara, T. K. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18(4), 333–349. <https://doi.org/10.1177/026553220101800402>
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101–113.
<https://doi.org/10.1177/1365480209105575>
- Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher*, 10(9), 9–20. <https://doi.org/10.3102/0013189X010009009>
- Mohammadi, S., Babaii, E., Hashemi, M. (2020). Examining Iranian EFL teachers' perceptions regarding the importance and application of dynamic assessment. *Foreign Language Research Journal*, 9(4), 1305-1338.
<https://doi:10.22059/jflr.2019.264775.544>
- Momeni, A. (2022). Online assessment in times of COVID-19 lockdown: Iranian EFL teachers' perceptions. *International Journal of Language Testing*, 12(2), 1-24.
<https://doi:10.22034/ijlt.2022.157122>
- Momeni, A., & Nushi, M. (2022). Iranian EFL teachers' beliefs about dynamic assessment: Does context make a difference? *Language Related Research*, 13(3), 403-428.
<https://doi.org/10.29252/LRR.13.3.15>
- Nevo, D. (1996). *School based evaluation*. Oxford: Pergamon University Press.
- Norton, B., & Toohey, K. (2004). Critical pedagogies and language learning: An introduction. In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 1-17). Cambridge: Cambridge University Press.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380. <https://doi.org/10.1177/0265532213480336>
- 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>
- Popham, W. J. (2004). All about accountability/ why assessment literacy is professional suicide. *Educational Leadership*, 62(1), 82-83.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4-11. <https://doi.org/10.1080/00405840802577536>
- Plake, B., & Impara, J. C. (1996). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement and adjustment* (pp. 53–68). New York, NY: Academic Press.
- Rauf, M., & McCallum, L. (2020). Language assessment literacy: Task analysis in Saudi universities. In L. L. McCallum, & C. Coombe (Eds.), *The assessment of L2 written English across the MENA region: A synthesis of practice* (pp. 13–41). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-53254-3_2
- Rasmussen, D.M. (1996). Critical theory and philosophy. In D.M. Rasmussen (Ed.), *The handbook of critical theory* (pp. 11-38). Oxford: Blackwell.

- Rea-Dickins, P. (2008). Classroom-based language assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education, Language testing and assessment* (pp. 257–271). New York, NY: Springer.
- Razavipour, K., & Rezagah, K. (2018). Language assessment in the new English curriculum in Iran: Managerial, institutional, and professional barriers. *Language Testing in Asia*, 8(9).
<https://doi.org/10.1186/s40468-018-0061-8>
- Rezai, A., Namaziandost, E., Miri, M., & Kumar, T. (2022). Demographic biases and assessment fairness in classroom: Insights from Iranian university teachers. *Language Testing in Asia*, 12(1), 1-20. <https://doi.org/10.1186/s40468-022-00157-6>
- Rezai, A., Alibakhshi, G., Farokhipour, S., & Miri, M. (2021). A phenomenographic study on language assessment literacy: Hearing from Iranian university teachers. *Language Testing in Asia*, 11(1), 1-25. <https://doi.org/10.1186/s40468-021-00142-5>
- Rezagah, K. (2022). Iranian EFL teachers' assessment literacy knowledge: The impact of the educational system on teachers' classroom assessment practices. *International Journal of Language Testing*, 12(2), 76-94.
<https://doi.org/10.22034/ijlt.2022.157126>
- Rogers, W. T., Cheng, L., & Hu, H. (2007). ESL/EFL instructors' beliefs about assessment and evaluation. *Canadian and International Education*, 36(1), 39–61.
<https://doi.org/10.5206/cie-eci.v36i1.9088>
- Sadeghi, K., & Jabbarnejad, L. (2012). An evaluation of EFL program objectives at tertiary level in Iran. *Asean Journal of Teaching and Learning in Higher Education (AJTLHE)*, 4(2), 1-16. <http://www.ukm.my/jtlhe/>
- Sarani, A., Alibakhshi, G., & Molazehi, H. (2014). On application of critical pedagogy principles by ELT instructors and subject teachers: A case of Iranian universities. *Research in Applied Linguistics*, 5(2), 41–56. https://rals.scu.ac.ir/article_11011.html
- Saunders, J. D., & Mitchell, V. (2004). Prepaid monetary incentive effects on mail survey research. *Journal of Business Research*, 57(4), 347–350.
[https://doi.org/10.1016/S0148-2963\(02\)00280-1](https://doi.org/10.1016/S0148-2963(02)00280-1)
- 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>
- Scarino, A. (2017). Developing assessment literacy of teachers of languages: A conceptual and interpretive challenge. *Papers in Language Testing and Assessment*, 6(1), 18–40.
- Sharma, M. R. (2020). Perceptions on language testing and assessment: A case study of B. Ed students' in Nepal. *Journal of Advances in Humanities and Social Sciences*, 6(1), 27-33. <https://doi.org/10.20474/jahss-6.1.4>
- Sheehan, S., & Munro, S. (2019). *Classroom assessment: the development of teacher's cognition*. London: British Council.
- Shohamy, E. (2001a). Democratic assessment as an alternative. *Language Testing*, 18(4), 373–391. <https://doi.org/10.1191/026553201682430094>
- Shohamy, E. (2001b). *The power of tests: A critical perspective on the uses of language tests*. Longman.
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education*, 14(1), 117–130. <https://doi.org/10.1080/09695940701272948>
- Shohamy, E. (2017). ELF and critical language testing. In J. Jenkins, M. Dewey, & W. Baker (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 583–593). Routledge.

- Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford: Oxford University Press.
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9, Article 1. <https://doi.org/10.1186/s40468-019-0077-8>
- Sun, X. (2021). Revisiting postmethod pedagogy: Adopting and adapting Socratic circle to secondary EFL teaching. *TESOL Journal*, 12(3), e601. <https://doi.org/10.1002/tesj.601>
- Sun, H., & Zhang, J. (2022). Assessment literacy of college EFL teachers in China: Status quo and mediating factors. *Studies in Educational Evaluation*, 74, 101-117. <https://doi.org/10.1016/j.stueduc.2022.101157>
- Smith, C. D., Warsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students' assessment literacy. *Assessment and Evaluation in Higher Education*, 38(1), 44–60. <https://doi.org/10.1080/02602938.2011.598636>
- Spolsky, B. (1978). Introduction: Linguists and language testers. In B. Spolsky (Ed.), *Approaches to language testing* (pp. v–x). Center for Applied Linguistics.
- Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford: Oxford University Press.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.
- Stiggins, R. (1997). *Student-centered classroom assessment*. Upper Saddle River, NJ: Prentice Hall.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23–27. <https://doi.org/10.1111/j.1745-3992.1999.tb00004.x>
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765. <https://doi.org/10.1177/003172170208301010>
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 1-14. <https://doi.org/10.1186/s40468-019-0077-8>
- Taylor, C. (1998). The dynamics of democratic exclusion. *Journal of Democracy*, 9(4), 143-156. <https://doi.org/10.1353/jod.1998.0068>
- Tajeddin, Z., Khatib, M., & Mahdavi, M. (2022). Critical language assessment literacy of EFL teachers: Scale construction and validation. *Language Testing*, 1-30. <https://doi.org/10.1177/02655322211057040>
- 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>
- Takrouni, A., & Assalahi, H. (2022). An inquiry into EFL teachers' perceptions of integrating Student self-assessment into teaching academic writing at a Saudi university. *Theory and Practice in Language Studies*, 12(3), 471-481. <https://doi.org/10.17507/tpls.1203.06>
- Tierney, R. D. (2016). Fairness in educational assessment. In M. A. Peters (Ed.), *Encyclopedia of educational philosophy and theory* (pp. 793-798). Springer. https://doi.org/10.1007/978-981-287-588-4_400
- Tollefson, J. W. (1995). Introduction: Language policy, power, and inequality. In J.W. Tollefson. (Ed.), *Power and inequality in language education* (pp. 1-8). Cambridge: Cambridge University Press.

- Tsagari, D. (2021). Gauging the assessment literacy levels of English language teachers in Norway. *The European Journal of Applied Linguistics and TEFL*, 10(1), 161-191. <https://link.gale.com/apps/doc/A674712462/AONE?u=anon~25fd71e2&sid=googleScholar&xid=8af72d4d>
- Troudi, S., Coombe, C., & Al-Hamily, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546–555. <https://doi.org/10.1002/j.1545-7249.2009.tb00252.x>
- Vakili, S., & Ebadi, S. (2022). Exploring EFL learners' developmental errors in academic writing through face-to-Face and Computer-Mediated dynamic assessment. *Computer Assisted Language Learning*, 35(3), 345-380. <https://doi.org/10.1080/09588221.2019.1698616>
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- Vogt, K., Tsagari, D., & Spanoudis, G. (2020). What do teachers think they want? A comparative study of inservice language teachers' beliefs on LAL training needs, *Language Assessment Quarterly*, 17(4), 386–409. <https://doi.org/10.1080/15434303.2020.1781128>
- 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.
- Wallace, M. P., & Qin, C. Y. (2021). Language classroom assessment fairness: perceptions from students. *LEARN Journal: Language Education and Acquisition Research Network*, 14(1), 492–521. <https://so04.tci-thaijo.org/index.php/LEARN/index>
- Wilson, S.M., Floden, R.E., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations*. Washington, DC: University of Washington, Center for the Study of Teaching and Policy.
- 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>
- Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27(2), 147–170. <https://doi.org/10.1177/0265532209349465>
- Yan, X., & Fan, J. (2020). Am I qualified to be a language tester? Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219-246. <https://doi.org/10.1177/0265532220929924>