Tabaran Institute of Higher Education
International Journal of Language Testing
Vol. 14, No. 2; 2024 October

ISSN 2476-5880
https://www.ijlt.ir

Assessment Principles of English as a Lingua Franca: Their Realization in Low-Stakes Local English Tests in Iran

Zari Saeedi[1]*, Zia Tajeddin[2], Fereshteh Tadayon[3]

ARTICLE INFO

Article History:
Received: September 2023
Accepted: November 2023

Keywords:
English as a Lingua Franca;
Language testing;
Local English tests;
Low-stakes tests

ABSTRACT

This research paper delved into the critical issue of applying English as a Lingua Franca (ELF) assessment principles in local English language tests used for non-native English speakers in Iranian language institutes. A qualitative content analysis was made on 60 local tests, dissecting them into domains, dimensions, and rating rubrics to scrutinize their alignment with ELF assessment principles. The study unveiled that despite some alignment with ELF assessment principles, key aspects like local communicative context, intercultural competence, and linguistic diversity are often overlooked. In particular, writing and reading tests failed to fully reflect these principles, and listening and speaking assessments showed biases towards native English varieties. The study provides crucial insights for test developers to foster a more nuanced and accurate assessment of non-native English speakers' abilities. Moreover, it highlights the need to embed ELF principles into test construction, argues for broader assessment scopes and a focus on locally relevant tasks, and contributes to more equitable and contextually relevant English language proficiency tests by emphasizing linguistic diversity in assessment frameworks.

1. Introduction

In the past decade, there has been a growing interest in English as a Lingua Franca (ELF) and its significance for language assessment. (Elder & Davis, 2006; Jenkins & Leung, 2017; Shohamy, 2018). This interest extends beyond the representation of nonnative language varieties and encompasses the diverse sociolinguistic realities of language users from different linguacultural backgrounds (e.g., Harding & McNamara, 2018; Jenkins & Leung, 2019). It has profoundly challenged the traditional notion of nativeness and argued for the recognition of variable language use in contexts where established norms and testing practices do not tend to be homogeneous, monolingual, and native-like (e.g., Brown, 2019; Hu, 2012, 2017; Isaacs & Rose, 2021; Jenkins, 2020; Leung, 2022; McNamara, 2011; Shohamy, 2018). However, the persistence of examination boards' reliance on normative benchmarks, rooted in traditional models of communicative language ability, hinders the full integration of ELF awareness in assessment. Leung (2022) cogently argues that the rationale for perpetuating a specific mode of language use promoted by NESs is untenable. Furthermore, assuming that British or American English unilaterally fosters international understanding oversimplifies the intricate interplay

[1] Associate Professor of Applied Linguistics, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran, saeedi.za@gmail.com
[2] Professor of Applied Linguistics, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran, tajeddinz@modares.ac.ir; zia_tajeddin@yahoo.com
[3] PhD Candidate of TEFL, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran, tadayonfereshteh@yahoo.com

Cite this paper as: Saeedi, Z., Tajeddin, Z., & Tadayon, F. (2024). Assessment principles of English as a Lingua Franca: Their realization in low-stakes local English tests in Iran. *International Journal of Language Testing*, *14*(2), 39–56. https://doi.org/ 10.22034/IJLT.2023.416485.1289

of linguistic dynamics within a global context. Similarly, Jenkins and Leung (2019), who adopt an ELF perspective in their appraisal of so-called international language tests, expound that these tests remain confined to native English norms.

The research conducted so far has predominantly delved into the theoretical realm of ELF and/or World Englishes in language assessment. However, practical applications in the context of local test analysis have been lacking. This study aimed to bridge this gap by scrutinizing Iranian low-stakes tests through the lens of ELF principles. By doing so, it not only highlighted the necessity for a nuanced, locally relevant approach to evaluating English language proficiency but also extracted ELF principles from prior studies for analysis. Moreover, the research brought to attention whether these local tests unintentionally perpetuate negative perceptions of nonnative English users' abilities, thereby impacting testing systems in Iranian language institutes. The theoretical significance of this study lies in its contribution to the extraction of ELF principles from existing literature, informing the analysis of test contents and paving the way for a comprehensive restructuring of tests that align with the plurilithic nature of English (Hall, 2014), especially in locally-defined contexts. The practical significance of the study is rooted in its potential to raise awareness among stakeholders about the incorporation of ELF principles in test design, ensuring that tests authentically reflect language use in local settings.

2. Review of Literature

2.1 ELF-Informed Assessment

The emergence of ELF has profoundly impacted language testing and assessment practices, marked by the pioneering work of scholars like Elder and Davies (2006) and subsequent contributions from other scholars (Brown, 2014, 2019; Canagarajah, 2006, 2019; Harding & McNamara, 2018; Hu, 2017; Jenkins, 2020; Jenkins & Leung, 2014, 2017, 2019; Leung, 2022; Mackay & Brown, 2016; McNamara & Shohamy, 2016; Shohamy, 2018). This paradigm shift has been closely intertwined with Complexity Theory (CT), as articulated by Larsen-Freeman (2018). This theory offers a non-reductionist perspective, emphasizing the dynamic, interconnected nature of language, and considers ELF, with its unique traits—novelty, situatedness, and continuous variation within the realm of complex adaptive systems. Within CT, ELF users' language repertoires remain heterogeneous due to the diverse linguistic patterns demanded by lingua franca contexts and multifaceted interactions among users. This heterogeneity underscores the significance of individual agency in regulating language use, enabling identity development and resistance against external linguistic changes (Labov, 2006). Additionally, the innovative nature of ELF finds parallels in child language learning, where initial creative language forms are viewed as valuable innovations rather than errors (Larsen-Freeman, 2018). Consequently, ELF users draw on their linguistic resources, engaging in morphogenesis to create new meanings and expand the English language's semantic potential.

Mauranen's (2018) perspective redefines ELF as a contact language, integrating macro, meso, and micro-level interactions. This view connects the social and cognitive dimensions of language, mirroring the intricate dynamics observed in multilingual settings. Widdowson (2018) challenges the rigidity of standard English and stresses that ELF should observe constant variation and adaptability to diverse communicative contexts and purposes. This stance aligns with sociolinguistic research, revealing the limitations of the standard language ideology prevalent in "Western" linguistic contexts (Milroy, 2001). Jenkins (2015) further emphasizes the importance of Multilingua-Franca in similect-based contact, portraying ELF users as adaptable, navigating different linguistic repertoires in complex cognitive processes. At the micro-level, ELF functions as a dynamic adaptive system, operating alongside communal languages at the macro-level and autonomous dynamic systems at the meso-level. These interconnected layers demonstrate the multifaceted nature of ELF, illustrating its role as a variable contact language across diverse linguistic and interactional levels. The emergence of translingual orientations challenges traditional language constructs, raising questions about the feasibility of conventional assessment methods in the context of ELF. Within international higher education, ELF becomes intertwined with multilingualism, leading to a shift in the description of individuals from non-native speakers of English to local speakers (Jenkins, 2017). This perspective emphasizes the influence of individuals' first language and local contexts on their English proficiency, defying predefined norms and standardized evaluation methods (Mauranen, 2018). Scholarly discussions explore concepts like Multilingua Franca, plurilingualism, metrolingualism, and Focus on

Multilingualism (FoM), highlighting the intricate interplay between multilingualism and ELF (Cenoz & Gorter, 2011). In summary, the intersection of ELF, CT, and multilingualism illuminates the complexity of language use, underscores the continuous innovation within ELF, and emphasizes its adaptability, interconnectedness, and dynamic nature.

Leung (2019) advocates for an approach anchored in the unique local context, accentuating the practical use of English within diverse environments. This recalibration extends to various aspects, including the assessment of translanguaging, where students leverage multilingual skills and accommodate different English forms, and test developers tend to mitigate biases against ELF users and prioritize assessment of pragmatic and discourse strategies rooted in localized practices (Harding & McNamara, 2018; Hu, 2017). The departure from traditional paradigms is evident in prioritizing communicative efficacy over rigid grammatical accuracy, aligning assessment with the actual needs of ELF users (Canagarajah, 2006; Clyne & Sharifian, 2008). Additionally, the concept of "self-regulation in context," as proposed by Jenkins and Leung (2019, p. 104), suggests localized self-assessment tailored to specific university settings, ensuring greater validity and social equity. Hynninen (2019) further calls for an assessment approach that evaluates competencies based on practical needs and diverse linguistic resources, urging a comprehensive understanding of contextual demands. This academic perspective underscores the necessity of reevaluating assessment frameworks to better align with the practical and varied linguistic realities of ELF users in university settings.

Similarly, Shohamy (2019) directly addresses concerns germane to ELF assessment, asserting that ELF, alongside the phenomena of translanguaging and bi-multi-languaging, disrupts conventional paradigms of language assessment. Shohamy argues that, for the majority of individuals globally, L2 is conceived as ELF, multilingual, and multimodal, leading to the emergence of novel and innovative linguistic amalgamations. Regrettably, these inventive amalgamations are often disregarded in English language assessment, which persists in upholding monolingual practices while penalizing the application of the L1. When addressing the assessment of English-language proficiency with high-stakes implications, particularly in the context of university admissions, two pivotal considerations have thus far been overlooked. These concerns encompass the extent to which such assessment can authentically represent the modes of language utilization that align with the actual communicative undertakings in academic target language use contexts. Additionally, they must grapple with the capacity to grasp the intricate nuances emanating from the multidimensional character of academic communication, including both spoken and written language forms (Jenkins & Leung, 2019). These concerns allude to the overarching ideologies that underpin English language assessment, as opposed to being grounded in the tangible realities of language application. This disconnect can yield counterproductive consequences not only for the language itself but also for the future academic pursuits and professional trajectories of test candidates (McNamara, 2011; Shohamy, 2019). This misalignment leads to a lack of construct validity, rendering language tests unable to accurately assess the language knowledge of ELF users. Shohamy (2018) argues that these tests fail to represent the proficiency of non-native speakers, leading to ethical concerns and injustice. The consequence is the perpetuation of global inequalities, favoring native speakers and non-native speakers with 'nativelike' skills while discriminating against those with less proficiency (Piller, 2016, p. 165).

2.2 Purpose of the Study

Several studies have been conducted on various aspects of the assessment of ELF or World Englishes: the role of native and nonnative raters (e.g., Hsu, 2016; Johnson & Lim, 2009; Park, 2020; Tajeddin & Alemi, 2014; Wei & Llosa, 2015; Winke et al., 2013; Xu et al., 2023; Zhang & Elder, 2011); the impact of nonnative accents and/or different speech varieties in listening assessment (e.g., Abeywickrama, 2013; Dai & Roever, 2019; Harding, 2012; Kang et al., 2019; Miao, 2023; Ockey et al., 2016; Shin et al., 2021); intelligible pronunciation in oral proficiency tests (e.g., Bøhn & Hansen, 2017; Isaacs, 2018; Isaacs & Harding, 2017; Kim & Billington, 2016; Sewell, 2013; Suzukida & Saito, 2022); comprehensibility rather than linguistic accuracy in speaking assessment (e.g., Sato & McNamara, 2019), developing ELF-informed tests (e.g., Kouvdou & Tsagari, 2018; Newbold, 2017, 2018; Nishizawa, 2023; Ockey & Hirch, 2020), assessment alternatives (Ghorbanpour et al., 2021; Motallebzadeh & Baghaee Moghaddam, 2011; Pishghadam & Zabihi, 2012; Saif, 2012), and designing new rating scales (Arefsadr et al., 2022; Doosti & Ahmadi Safa, 2021). These studies challenge the concept of an idealized "native-speaker" competence, which has long been serving as a reference point

for assessing non-native English speakers' proficiency, perpetuating linguistic hegemony, and favoring certain English varieties over others. As argued by Leung (2022), proficiency is a construct shaped by authorized stakeholders, often reinforcing language prescriptivism and linguistic inequality. This critique calls for a reevaluation of language assessment standards in view of the diverse forms of English used globally. However, the entrenched tendency of international examination boards to adhere to normative references persists. These references are grounded in the established models of communicative language ability, shaping both the "what" of assessment (e.g., construct definition) the "how" of assessment (e.g., diverse facets or characteristics of tests), and standard-setting procedures heavily reliant on ratings/scorings by the native raters of written and spoken productions.

Despite the theoretical background, the inconclusive nature of empirical studies on ELF in relation to language assessment has left much to consider for the realization of ELF principles in different tests. The studies reviewed above have examined each feature of the test separately, for instance, rating, accent/speech language varieties, intelligibility, communicative effectiveness, and some developed ELF-aware tests to meet their contextual needs for their specific local or international purposes. Moreover, each study has adopted a different kind of test for its analysis, which has caused little generalization to other tests developed for local purposes in a specific context. Thus, there remains uncertainty regarding how all aspects of locally defined tests in terms of their domains, dimensions, and rating rubrics might have been influenced in response to assessment principles derived from relevant ELF research. To address this issue, the present study aimed to examine the content of the local low-stakes tests through two main steps: first, by drawing on ELF principles in past studies; and second, by analyzing the content of the low-stakes local tests based on the three key components of test design, namely the test construct, test tasks, and scoring rubrics. In the present study, local low-stakes tests, also called classroom-based tests, are "used to assess what students know or can do in the language concerning what is being taught in a specific classroom or program" (M. Kay & Brown, 2016, p. 80). The findings of such a study have the potential to reshape these tests to enhance their fairness, generate a greater impact, and expand the understanding of authentic language use in similar contexts. The following research question was posed to address the purpose of this study:

RQ. How do the domains, dimensions, and rating rubrics of local English language tests in Iran align with the assessment principles of ELF?

3. Method

Drawing on Mayring (2004), the present study used qualitative content analysis (QCA) for the sampling of test selection and data analysis based on systematic and well-defined steps. The design of this study was informed by descriptive and deductive approach to analyze the alignment of local tests in terms of ELF assessment principles (e.g., Elo & Kyngäs, 2008; Schrier, 2012).

3.1 Test Selection

As the aim was to analyze low-stakes local English tests in Iran, local English tests used across proficiency levels were selected. The tests were a major part of classroom-based assessment that measured L2 learners' achievement at the end of each term. Therefore, the study sought to investigate the alignment of these tests with ELF assessment principles in the local context.

We carefully considered several key criteria for the sampling procedure. These criteria included (a) the selection of local institutes, (b) the types of tests administered, (c) the number of tests conducted, (d) the different levels of the tests, and (e) the recent tests utilized in the institutes. Given the substantial number of private English language centers in Iran, we deliberately focused on five well-established institutes that had a presence across multiple cities, with at least 10 branches. Another factor we took into account was the great number of tests, for which we decided to draw on only upper-intermediate and advanced levels. Therefore, the selected institutes were requested to provide general English classroom tests at three upper-intermediate and three advanced levels, resulting in a total of six proficiency levels. Regarding the types of local general English tests, we utilized tests used for final exams. Each institute was given the opportunity, upon agreement, to provide two different tests for each of the six proficiency levels, totaling 60 distinct tests along with relevant rating rubrics. Finally, we also requested that the tests used be recent, specifically those designed and implemented when the study was in progress. By incorporating tests from different institutes, this study sought to examine variations in

test design, content, and evaluation practices across multiple language centers. The process of collection of the tests can be illustrated as follows:

- Tests were sorted and organized according to their levels and respective institutes, which helped researchers keep track of which tests belonged to each institute.
- All the tests were checked to ensure they had a consistent format and structure, regardless of the institute. Any tests with a disorganized structure for informal assessment were excluded, as they did not represent the best attempts of each institute.
- To compare each test, three major categories were considered: domains (what skills and subskills each test measures), dimensions (the types of tasks, questions, or texts used to assess each skill), and rating rubrics.
- The upper-intermediate levels were grouped as level 1, level 2, and level 3. The same applies to the advanced levels, which were divided into advanced 1, advanced 2, and advanced 3.
- The main focus was on tests of general proficiency, so any tests designed for specific purposes were initially excluded during the sorting process.
- We specifically focused on tests designed for final exams because the institute principals reported that these tests were the complete version and well-structured.
- We intentionally requested each institute to provide us with their most recent version of tests, which was in 2023. However, they assured us that their tests had not undergone significant changes even before the required time, and were representative and sufficient samples of local tests.

3.2 Test Content Analysis

This study employed a deductive approach rooted in Schreier's (2012) content-analytical method to scrutinize local English tests. These tests were analyzed based on three main components: domains (representing the assessed constructs), dimensions (indicating assessment tasks and texts), and rating rubrics (encompassing scoring criteria). Guided by existing ELF principles, the analysis focused on key language skills—speaking, writing, reading, and listening. Domains encompassed overarching constructs like reading comprehension, with dimensions specifying tasks within each domain, such as reading passages. Rating rubrics delineated criteria for assessment. The analysis process involved meticulous coding by the first author and two assistants. Regular discussions ensured coding consistency and resolved discrepancies. This rigorous approach, informed by ELF principles, yielded robust insights into the test content and structure. Despite limited test specifications, the study assessed the alignment between these components and ELF principles. The study aimed to discern how each skill was assessed and operationalized, thereby enhancing the nuanced understanding of the findings in the context of the following ELF principles:

- ELF communicative interactions (e.g., Elder & Davis, 2006; Jenkins, 2000; Kirkpatrick, 2010; Mauranen, 2003; Seidlhofer, 2004)
- local communicative contexts (e.g., Brown, 2014; Hall, 2014; Jenkins & Leung, 2017, 2019; Leung, 2019; Leung et al., 2016; McKay & Brown, 2016)
- multidimensional ELF competence (e.g., Canagarajah, 2006; Jenkins, 2015; Kramsch, 2011; Leung, 2022; Leung et al., 2016; Shohamy, 2018)
- test accommodations (Elder & Davis, 2006)
- Task variety **(**e.g., Brown, 2014; 2019; Dai & Roever, 2019; Harding, 2012; Kang et al., 2019; Newbold, 2018; Nishizawa, 2023; Ockey et al., 2016; Shin et al., 2021)
- discipline-specific assessment tasks (e.g., Hynninen, 2019; Jenkins & Leung, 2019; Shohamy, 2018)
- successful ELF communication strategies (e.g., Chan, 2021; Chopin & Bayyurt, 2015; Harding, 2012, Kim & Billington, 2016**;** Kouvdou & Tsagari, 2018; Ockey & Hirch, 2020)
- task fulfillment/intelligibility criteria (e.g., Abeywickrama, 2013; Elder & Harding, 2008; Isaacs, 2008, 2018; Kim & Billington, 2016; Sewell, 2013)**.**

The coding frame underwent trial coding and modifications until an agreement was reached on whether ELF principles could account for the units of analysis in both test samples. The analysis employed a meticulous and systematic approach, incorporating three coders, inter-coder reliability

checks, and references to existing literature to ensure accuracy in interpreting the test content based on the coding scheme and ELF principles. To ensure the trustworthiness of the results, a thick description of each unit of analysis was followed through coding and constant comparative analysis. The coders achieved demonstrable agreement on the interpretation drawn from the analysis of tests, supported by a substantial agreement (89%), and resolved any discrepancies through discussion. The validity of the inferences drawn from the analysis was assessed through structural validity. As Mayring (2004) argued, structural validity assesses the backing of analysts' abductive inferences from categorization in the coding process. In this study, multiple rounds of coding and recoding were conducted, providing support for the final interpretations. The categorization of test domains, dimensions, and rating rubrics in relation to ELF principles was explained and clarified by referring to ELF assessment in past studies.

4. Results

The content analysis of low-stakes local tests shows how far the ELF assessment principles have been reflected in the local tests' domains, dimensions, and rating rubrics. There are four main subsections: writing tests, speaking tests, reading tests, and listening tests. First, ELF principles that have made up the test domains are as follows: (a) *ELF communicative interactions*: it refers to the role of ELF-informed corpora that show a wide range of ELF interactions; (b) *local communicative contexts*: it explains the importance of the local context that shapes ELF language use; and (c) *multidimensional ELF competence*: it moves away from linguistic competence to highlight a broad range of ELF competence. Second, tasks included in each test were analyzed using three ELF principles: (a) *test accommodations:* referring to different ways that can make tests fairer and easier for test takers; (b) *task variety:* demonstrating a range of tasks that can be used instead of discrete point items in a test; and (c) *locally discipline-specific assessment tasks:* showing some general instances of how to define locally-discipline-specific tasks. Third, this study examined rating rubrics based on three principles that can guide the rating procedure of an ELF-informed test: (a) *strategies for successful ELF communication* (accommodation, negotiation, and maintaining smooth interaction); (b) *intelligible pronunciation*; and (c) *task fulfillment*. Table 1 shows a summary of the content analysis of tests inspired by ELF principles:

Table 1
The Analysis of Local Low-stakes Tests Based on ELF Assessment Principles

	ELF principles	**Writing tests**	**Speaking tests**	**Reading tests**	**Listening tests**
Test domain	ELF communicative interactions (only in speaking and listening)		The ability to communicate using different nonnative varieties of English		Lack of ELF interactions within relevant corpora
	Local communicative contexts	To a limited extent	A certain degree of alignment with local communicative contexts	Comprehension of monocultural and social topics	Conversations between interlocutors from native varieties of English
	Multidimensional ELF competence (except reading)	A fraction of NNESs' writing competence	To some extent		Lack of diverse ELF competence among interlocutors
Test dimension	Test accommodation			Misalignment	Misalignment

	(only reading and listening)				
	Task variety	Limited types of tasks	Limited range of tasks	Limited range of tasks	Limited range of tasks
	Locally discipline-specific assessment tasks	No	Locally-defined tasks	No	No
Rating rubric	Strategies for successful ELF communication (only listening and speaking)		Yes		No
	Intelligible pronunciation (only speaking)		Its importance in speaking rating scales		
	Task fulfillment	Its presence in addition to other criteria	Its presence in addition to other criteria	Linguistic accuracy	Linguistic accuracy

4.1 Writing Tests

An analysis of classroom-based writing tests revealed that a wide range of task types were utilized to assess students' writing abilities. These tasks included writing essays, emails, short stories, film reviews, leaflets, advertisements, reports, articles, integrated tasks (such as reading to write), and personal profiles. These diverse task types effectively evaluated the writing proficiency of NNES students across various prompts. However, it should be noted that there was a lack of precise test specifications, making it challenging to determine the exact domain for analysis. Nonetheless, the eye-catching features of these test types provide an understanding of the domain of writing tests used in language institutes. The examples are as follows: (a) Write an email to a friend. Describe a problem you have and ask for advice, (b) Write a short story ending with the saying '*Every cloud has a silver lining'*, (c) Write an informal article on the topic: *How to reduce the stress in your life*, (d) Write a film review. Include the information below, (e) Write a leaflet advising about using public transport in your town or country. Use the prompts below to help you, (f) Read the advert and write a letter to the organization. Give details about yourself and ask for more information about the things circled, (g) Read the extract from a newspaper. Then write a 200–300-word letter to the editor, using the paragraph plan as a guide. Use the appropriate layout and register in your letter, (h) Imagine your ideal job. Write your profile for it (the details don't have to be true) in 250–300 words. Think about these points and use the paragraph plan as a guide, (i) Write a report on the financial situation of RTV Music Channel, based on the extract from the Profit and Loss Account below, and (j) Describe an event that changed your life.

The local communicative context and multidimensional competence, which are fundamental principles of ELF, were systematically overlooked in all existing writing tests. The presence of diverse writing tests itself is indicative of the recognition of the specific requirements and needs of EFL test takers, thereby necessitating a broader spectrum of the writing domain. However, in real-life scenarios, NNESs who actively employ ELF as their means of communication draw upon a rich repertoire of resources, such as their multilingual or bilingual capabilities, context-dependent knowledge, and specialized competence in specific academic or professional domains. It is these multifaceted resources that empower NNESs to meet the intricate demands and standards associated with writing at the university level, including writing academic articles.

Regrettably, the prevailing domain defined for writing tests tended to narrowly delimit the scope of writing, largely overlooking the broader dimensions of NNESs' writing abilities. This limited approach failed to acknowledge NNESs' ability to interpret and adapt their writing skills to various contexts, thus undermining their writing proficiency. Consequently, the writing tests merely gauged a

fraction of NNESs' writing competence, disregarding their capacity for Englishing, which encompasses the use of English in locally defined tasks that align with their needs. For instance, NNESs might need to write a resume to cater to their intercultural communication requirements, yet this aspect was rarely addressed in the existing writing tests. By embracing this approach, writing tests can authentically reflect the realities and complexities of NNESs' writing experiences, contributing to a deeper intercultural understanding and enabling them to successfully engage in diverse writing practices for education, business, and migration purposes.

The task types mentioned earlier are indicative of the writing tasks commonly found in classroom-based tests across language institutes in Iran. These tests primarily aimed to assess general learning objectives rather than discipline-specific writing tasks. Consequently, the scope of writing assessments in this context did not extensively cover the principle of *local discipline-specific assessment tasks*. In considering the features of multiple tasks, although these task types or *the task variety principle* encompassed various aspects of writing, there was a limited exploration of a broader range of performance assessment tasks. For instance, at the upper-intermediate or advanced levels, students had the opportunity to create writing portfolios to demonstrate their skills and achievements. Likewise, integrated writing and reading tasks often involved a basic integration of the two components, having authenticity. An example of such a task is reading a newspaper extract and writing a 200-300-word letter to the editor, utilizing a provided paragraph plan as a guide.

This study also undertook to examine rating rubrics designed to guide writing within the ELF-informed principle of task fulfillment. The central tenet of these rubrics revolved around the prioritization of task fulfillment over the stringent scrutiny of linguistic accuracy. Inherent to the principle is the recognition that ELF proficiency, particularly in the written form, is primarily concerned with the effective transmission of ideas and communication, rather than the meticulous adherence to linguistic norms. These criteria can be explained as follows: (a) appropriateness refers to writing clear, detailed information on a wide variety of general and familiar topics, using functional language to express opinions; (b) completeness addresses all aspects of the topic; (c) accuracy means using a good range of vocabulary, collocations, and functions; (d) clarity alludes to writing with clarity on most topics, presenting clearly and logically using transition words; and (e) complexity covers a good variety of vocabulary, sentence structures, and functions on some abstract and cultural topics.

The first criterion of the writing rubrics was the appropriateness and completeness of written responses. It underscored the ability to convey clear and comprehensive information on a diverse array of familiar topics, to write a purpose-driven description, narrative, or argument, replete with nuanced details both in favor of and against a particular viewpoint, utilizing complex linguistic forms. Second, while the rubrics endorsed linguistic accuracy to facilitate clear communication, they acknowledged that minor errors in unfamiliar or highly abstract contexts do not impair overall comprehension. Precision in vocabulary selection, adept utilization of collocations and grammatical structures, and the ability to articulate abstract and cultural ideas represent focal points of the rating rubrics. Notably, the rubrics demonstrate the understanding that undue emphasis on linguistic rigidity may be incongruous with the ELF principle. Third, clarity, as per the rubrics, encompasses the logical organization of ideas, judicious application of transition words, and the incorporation of requisite details and supporting materials to enhance clarity. The rubrics further recognize the utility of cohesive devices in interconnecting arguments across paragraphs, while acknowledging that such cohesion may occasionally manifest awkwardly, particularly within the complex milieu of ELF writing. Finally, the rubrics afford due consideration to the construct of complexity within written responses. While the complexity, in this context, hinges upon the adept deployment of an array of vocabulary and sentence structures, the ELF principle reiterates that linguistic nuance should be a secondary consideration to effective task fulfillment within the ambit of ELF.

4.2 Speaking Tests

Based on the ELF principle of communicative interactions, it was observed in speaking tests that the ability to communicate effectively occurred primarily among individuals using different nonnative varieties of English. This means that despite the presence of various nonnative English varieties, the speakers demonstrated their oral proficiency by employing either American or British English, for instance. The tasks administered in these tests consisted of comparable questions taken

from a wide range of examples, highlighting specific domains and serving the purpose of evaluating their competence in a set of general topics. In essence, the emphasis was on assessing their ability to engage in effective communication rather than adhering strictly to native English norms. The speaking tests also primarily gauged the learners' ability to engage in discussions related to specific conventional topics, rather than exploring their proficiency across various domains of oral communication for different purposes. For example, some interview questions revolved around important holidays in their country and how they are celebrated. This approach highlights a certain degree of alignment with local communicative contexts, emphasizing a more local and familiar domain of discourse. However, it did not necessarily capture the broader concept of "Englishing", which refers to evaluating what L2 learners can accomplish with English as a tool, rather than just how they can use it.

In speaking tests, the inclusion of interactive tasks provided students with opportunities to engage in meaningful communication and employ pragmatic strategies like accommodation, repairing, and seeking clarification. These tasks were designed to assess their proficiency in discussing various topics. However, it is important to acknowledge that the current format of speaking tests does not comprehensively encompass the wide array of communicative tasks encountered in different real-life contexts. The task types employed in speaking tests primarily resembled interview-style interactions, wherein students responded to prompts and engaged in dialog with their nonnative examiner whose L1 was Persian. While these interactions offer valuable insights into students' language abilities and their capacity to express themselves on common subjects, they might not capture the nuances of specialized tasks that are specific to particular local contexts. The limitations of the interview format in speaking tests raise concerns about the tests' validity and their ability to fully evaluate the learners' communicative competence in real-world scenarios. By focusing predominantly on a limited range of tasks, speaking tests failed to include important aspects of language use, such as the ability to navigate domain-specific tasks or adapt communication strategies to accommodate different interlocutors and contexts.

Analysis of rating criteria for all local speaking tests brought to light the following features: (a) communication: the skill to understand, engage, and express oneself effectively; (b) intelligible pronunciation; (c) vocabulary: accurate and efficient word usage; (d) conversation strategy: skillful and natural utilization of conversational techniques; (e) fluency: the speed, flow, naturalness, and comfort in using words; (f) appropriateness: confidently employing a diverse range of vocabulary, word combinations, and some advanced functions, and having the ability to express pertinent ideas and opinions without evident word-searching; (g) completeness: addressing all aspects of the conversation thoroughly and employing a wide vocabulary repertoire, including idiomatic expressions; (h) accuracy: confidently and precisely communicating on a broad range of topics, with occasional errors when tackling complex abstract subjects that don't hinder comprehension; (i) quality of interaction: confidently expressing ideas and opinions with precision, actively engaging in complex discussions, consistently following up on statements and inferences, selecting language suitable for the context and audience, and rarely requiring clarification, occasionally employing paraphrasing when necessary; and (j) complexity: delivering viewpoints through extended stretches of language, discussing a variety of subjects, effectively linking ideas and arguments with advanced language, experiencing occasional pauses while searching for expressions, and using stress and intonation effectively to convey meaning. Given intelligible pronunciation, this ELF principle prioritizes the ability to accommodate diverse accents and dialects, emphasizing the need for intelligible pronunciation to facilitate smooth interaction among speakers from different language backgrounds. The analysis of rating rubrics of speaking tests indicated the importance of intelligible pronunciation but may focus more on evaluating a speaker's ability to comprehend and express themselves clearly within the context of a specific local accent. Likewise, the ELF principle in terms of communication strategies (such as accommodation and negotiation) for successful communication value fluency which enables NNESs to have the natural flow of conversation and adaptability to different communication styles. The same is true for local tests to assess conversation strategy, including how effectively a speaker can use communication strategies in a conversation, not only in terms of speed and flow but also in terms of adaptability to the local context of interactions.

The main difference comes from assessing grammatical accuracy in speaking tests. ELF principles shown in the rating rubric do not explicitly include grammar and vocabulary but place a great

emphasis on effective communication and accommodation. This reflects the nature of ELF, where communication might involve speakers with varying levels of grammar and vocabulary proficiency. However, local speaking tests include separate criteria for assessing grammar and vocabulary, aiming for the correct and effective use of these linguistic elements regardless of variation within the local context. The ELF principle of task fulfillment, meaning the ability to meet communication goals in a given context, can vary widely in international interactions. Similarly, local speaking tests assess completeness, accuracy, and the ability to self-correct within the framework of the specific language or dialect being tested. Also, both place a strong emphasis on the quality of interaction, focusing on precision in communication and the ability to handle complex lines of argument in an international context where participants may have diverse linguistic backgrounds. In fact, local speaking tests also consider the quality of interaction but may do so within the context of local topics and discussions.

4.3 Reading Tests

By examining reading passages in detail, it became evident that these reading tests assessed the proficiency of EFL learners in comprehending a range of common subjects derived from books, magazines, or websites. These were common passages in classroom-based tests: *What books mean to them, Personality tests, Survival stories, Brain change, Live long, Keep healthy, Earth day, Shops/supermarket, Celebrity is the answer to anything, Romance language, Stone-age mini skirt discovered, A brief history of TV, Lost in translation, A world without clean water,* and *The Salem witch trials.*

The examples of reading topics indicate that these tests assess the comprehension of learners specifically on monocultural and social topics. However, this approach fell short of adhering to the principle of ELF, which emphasizes the importance of *local communicative contexts*. In other words, these tests evaluated skills that may not be realistic for L2 learners, particularly those learning English in expanding circle contexts and using English primarily for understanding texts rather than solely focusing on inner-circle contexts. For instance, one of the reading tests at the upper intermediate level assessed comprehension on topics such as "Is celebrity the answer to anything?" and "The British petroleum oil disaster," which highlights the dominance of native English varieties within the classroom reading tests. While assessing L2 learners' understanding of various topics might be seen as promoting intercultural comprehension, it does little to showcase the abilities of learners in locally oriented domains. In this context, learners are not required to possess knowledge of a single domain of comprehension that primarily focuses on inner-circle countries. Such a domain is neither necessary nor relevant for their local purposes of comprehension. Additionally, the analysis of the reading tests revealed a focus on monocultural competence rather than intercultural competence. The tests did not incorporate multiple texts from different English varieties, thus limiting the assessment of learners' abilities in engaging with diverse linguistic contexts. Notably, there were no reading tests that incorporated multiple texts from different varieties of English. This limitation indicates a missed opportunity to assess learners' abilities in engaging with diverse linguistic contexts and to promote intercultural competence.

As to reading dimensions, the reading tasks encompassed several primary formats, including multiple-choice questions, gap-filling, short answers, matching, and locating headings. There are two noteworthy aspects to consider. First, the texts used for these tasks exclusively focused on native varieties of English, such as American or British English, which aligns with the language required for providing answers. Moreover, such an exclusive focus on native-oriented varieties overlooked the non-native varieties of English that L2 learners may use or encounter. These non-native varieties may be integral to locally-defined texts employed by stakeholders within that specific context. Second, the selection of texts is largely based on specific subjects, providing cultural and social information about situations in inner-circle countries. However, these features did not fully align with the three main principles of ELF. For example, *test accommodation* should be made if understanding any given text relies on cultural background knowledge or unfamiliar lexical items. Additionally, the texts reported here tended to overlook the existence of multiple Englishes, particularly those demanded for locally-defined reading purposes, which can be complemented by the use of the learners' L1. Bilingual or multilingual learners, using Persian as their L1 who possess competencies beyond English were not given a space to demonstrate their comprehension using their L1 within the testing context.

4.4 Listening Tests

The findings indicate that the interactions occurring in ELF within relevant corpora, such as Vienna-Oxford International Corpus of English (VOICE), did not receive sufficient attention in the listening sections of all tests. The conversations, monologs, or discussions between interlocutors predominantly featured native varieties of English (e.g., Australian, American, and British) without adequately acknowledging the local communicative contexts.

While these tasks provided examples of listening comprehension, they overlooked the crucial consideration of how local contexts should be incorporated into the assessment. As a result, the importance of local communicative contexts was largely neglected in the listening sections of the tests. There is a pressing question regarding the extent to which local contexts should be taken into account and how they can be appropriately integrated into the listening sections. The current focus on native English varieties failed to capture the diverse range of linguistic encounters that L2 learners are likely to encounter in real-world situations. By neglecting local communicative contexts, the tests failed to assess learners' ability to comprehend and navigate diverse accents, speech patterns, and cultural nuances that are intrinsic to ELF interactions. Thus, the current listening sections of the tests did not adequately recognize the interactions in ELF that exist within relevant corpora. The tasks primarily featured native English varieties, disregarding the importance of local communicative contexts.

In the context of local listening tests, an analysis is typically conducted to assess the accommodations provided in the test, particularly regarding the representation of different accents. However, it is worth noting that the accents predominantly featured in these tests tended to be native accents, potentially biased against the diverse range of nonnative accents that learners might encounter in real-life situations. The discrepancy between the accents represented in local listening tests and the diverse array of nonnative accents found in real-life communication settings raises concerns about the validity and authenticity of such tests. By primarily featuring native accents, these tests helped create a biased assessment environment that does not fully reflect the challenges and realities faced by NNESs when listening to and comprehending different accents in their local communities. When it comes to task variety in local listening tests, the commonly utilized formats included multiple choice, short answer, matching, and other similar question types. While these formats served the purpose of evaluating listening comprehension skills to some extent, they did not adequately mirror the types of tasks that nonnative students would typically encounter in local contexts.

During the examination of the reading and listening test rating scales, we observed that the rating criteria were stringent by requiring accurate spelling for gap-filling and short-answer items. Furthermore, candidates were restricted to using English only, without the option to utilize their first language or substitute synonyms for the correct words. There were no criteria such as task fulfillment; rather, precise spelling or selecting the accurate options was the sole benchmark.

5. Discussion

This study analyzed local low-stakes tests, focusing on the principles of ELF assessment in writing, reading, listening, and speaking. The findings unveiled significant shortcomings in the ability of these tests to gauge the proficiency of NNESs in Iranian language institutes. Specifically, writing tests lacked precision and contextual relevance, and reading tests predominantly centered around monocultural themes, neglecting intercultural competence. Speaking tests primarily assessed communication effectiveness, overlooking broader aspects of oral communication, and listening tests failed to incorporate local accents and contexts. This disparity highlights the pressing need for a broader range of context-specific assessment tasks to ensure effective evaluation. The study also revealed a mismatch between ELF principles, which emphasize intelligible pronunciation and effective communication across diverse language backgrounds, and local tests, which primarily assess grammar and vocabulary within a specific local context. Writing tests, according to ELF principles, should prioritize effective communication, whereas current local tests tend to emphasize linguistic accuracy. This disparity underscores the difference in the tests' objectives, with ELF principles accommodating linguistic diversity and focusing on international communication, while local tests prioritize correctness based on native speaker norms. The observed deviation from ELF principles in local test formats might be attributed to the underlying theoretical frameworks upon which these tests were built. However, a multi-component view of language, decomposing language into discrete components, may struggle to

adequately account for the intricate and dynamic interrelationships inherent in the communicative use of language (Bachman, 1990). The study suggests that perceiving language proficiency as a fixed construct does not capture its variability across diverse contexts and users, as argued by Mahboob and Dutcher (2014).

Given the comparison of writing tests against ELF principles, interestingly, the diversity of writing task types observed in classroom-based tests showcases a commendable effort to assess the writing proficiency of NNESs across various prompts. However, the recognition of NNESs' writing abilities in locally defined contexts, as underscored by ELF, is unfortunately lacking in existing tests. By narrowing the scope of writing assessments to general domains, these tests fail to acknowledge the multifaceted resources that NNESs employ in real-life situations. Moreover, the limited approach disregards the capacity for "Englishing," using English for locally defined tasks aligned with learners' actual writing needs. As explained by (Hu, 2012), local contexts require assessments within the parameters of locally used English variants and their norms. This lack of harmony contrasts with the notion that the practical use of English should be informed by the unique local context (Leung, 2019). In local contexts, local norms should inform classroom-based writing assessment, considering varying conventions across locally defined communities (McKay & Brown, 2016). The analysis revealed a clear underrepresentation of ELF principles in evaluating NNESs' ability to produce written work for academic purposes in English-medium settings. It clearly showed the absence of ELF principles in evaluating NNESs' academic written work in English-medium settings. Integrated writing tasks, in line with Llosa and Malone (2019), offered a more comprehensive representation of NNESs' writing proficiency across different tasks. Some conformity was found between ELF's principle of task fulfillment and certain rating criteria like completeness and appropriateness. However, the emphasis on linguistic accuracy as the primary criterion contradicts the heterogeneous linguistic patterns and individual agency of NNESs in regulating their language use variably, as suggested by Larsen-Freeman (2018).

The speaking tests, while aligning with the ELF principle of effective communication across non-native English varieties (Brown, 2019; Jenkins & Leung, 2019), primarily focused on interview-style interactions, potentially missing the intricacies of specialized tasks in local contexts. This limitation underscores the necessity for speaking tests to encompass a broader array of tasks, ensuring a comprehensive evaluation of learners' communicative skills. Despite these limitations, some aspects of the speaking tasks reflect real-life scenarios, incorporating communication strategies for negotiation and aligning with the principles of successful ELF communicative effectiveness (Matsumoto, 2011; Vettorel, 2018). The incorporation of strategies for successful ELF communication and accommodation signifies a positive step toward capturing learners' proficiency in facilitating mutual comprehension. This is in line with the study by Ockey and Hirch (2020), who developed an oral ELF-informed placement test and explained that rating rubrics should reflect the interactional competence of NNESs by focusing on their comprehensibility, adaptability to different speech varieties, and resolving communication breakdowns. However, it is crucial to acknowledge that these tasks might not fully assess one's ability to handle the linguistic demands in professional or academic disciplines. The speaking topics appeared less academically oriented, raising questions about their suitability for assessing NNESs' English proficiency in an academic context. While evaluation crit ria consider communication strategies, intelligibility (Isaacs & Harding, 2017; Suzukida & Saito, 2022), and task fulfillment (Sato & Macnamara, 2019), the excessive emphasis on grammar, vocabulary, and pronunciation overshadows the importance of local communicative contexts and domain-specific tasks. Despite the awareness of ELF principles, the current rating rubrics prioritize preconstructed grammatical accuracy over emergent grammatical appropriacy (Canagarajah, 2014). This suggests the need for a more nuanced approach in aligning assessment criteria with the multifaceted nature of language use in diverse contexts.

The analysis of reading and listening tests revealed a clear focus on native English varieties, which diverges from the ELF principle of embracing local communicative contexts. This misalignment contrasts sharply with the idea of tailoring assessment to localized uses and benchmarking language knowledge against context-specific norms, as advocated by Galloway and Rose (2018). The tests were designed with a limited comprehension scope defined by general norms rather than local ones. The findings underscore the prevalence of monocultural topics, minimal test accommodations like

glossaries, limited locally-suited themes, and the requirement for English-only responses in short-answer questions. Moreover, linguistic accuracy takes precedence over task fulfillment in scoring gap-filling questions, deviating from suggestions to incorporate test accommodations (Elder & Davis, 2006) and the choice of local texts for assessment (Mackay & Brown, 2016). The discrepancy is further highlighted by Shohamy's (2019) emphasis on using NNESs' L1 in tests to avoid reinforcing monolingual practices. Despite this, the exclusive use of English for responses in current listening and reading tests disregards Multilingua Franca (Jenkins, 2017) and the concept of multilingualism (Cenoz & Gorter, 2011), while NNESs possess diverse linguistic repertoires that enable them to draw on their varied languages to fulfill test requirements.

The analysis of pronunciations in listening tests also shows the presence of native English accents. This result contradicts the idea put forth by Galloway and Rose (2018) that audio materials should include speakers of relevant language variants and that EIL learners should understand various L2 accents. Despite evidence from various studies (e.g., Harding, 2012; Isaacs, 2018; Kang et al., 2023), demonstrating the need for diverse accents and pronunciations, the tests persisted in their concentration on native English varieties. This lack of harmony overlooks the linguistic complexities NNESs encounter in real-life interactions and disregards ELF's emphasis on local communicative contexts and diverse linguistic encounters (e.g., Blommaert, 2010; Shohamy, 2018). The disparity between ELF principles and the test content arises from the fact that the tests, shaped by the ideologies and policies of testing systems, can marginalize ELF speakers by penalizing them, thereby limiting their opportunities (e.g., McNamara, 2011). Scholars such as Canagarajah (2019) and Shohamy (2011) advocate for a shift from solely emphasizing native English and English-only competence to recognizing broader ELF competencies like intercultural or multilingual competence across various domains. Acknowledging these wider competencies is crucial for a fair and inclusive evaluation of language proficiency, aligning with the intricate and multifaceted nature of language use in diverse global contexts.

6. Conclusion and Implications

This study highlighted the intricate relationship between traditional assessment models and the evolving nature of language proficiency, emphasizing the importance of aligning assessment practices with the principles of ELF. It argued for the need to integrate local communicative contexts (Jenkins & Leung, 2019), domain-specific tasks (Shohamy, 2018), and nonnative English varieties (Brown, 2019) into assessment frameworks to create a more meaningful evaluation of language abilities. Additionally, the study emphasized the influence of testing institutes' ideologies and policies, calling for a reevaluation of language assessment practices to accommodate linguistic diversity and promote a more equitable approach to assessing language proficiency in each testing context (Shohamy, 2006).

The findings provide implications for stakeholders, particularly in relation to incorporating ELF assessment principles into local test construction. One implication is the role of stakeholders in understanding and implementing the nuances of ELF in tests. Elder and Davies (2006) emphasize that stakeholders, including test developers, educators, and policymakers, should actively engage with the different shades of ELF, even if they may not always align perfectly with each other. This understanding and awareness of ELF principles among stakeholders are crucial for effectively integrating them into test design and ensuring their applications over time. For example, teachers should be aware of the limitations of current tests and help students understand the discrepancies between test performance and actual academic or professional performance. They can support students in self-assessment and reflection, encouraging them to consider the language demands of specific situations and adapt their language use accordingly. Local institutes offering language courses can collaborate with local test developers to bridge the gap between classroom instruction and test expectations, aligning teaching practices with the requirements of language tests. Moreover, test developers should consider the contingent variability and flexibility of ELF principles when designing language tests. This may involve reconfiguring test formats and tasks to better reflect the dynamic nature of language use in real-life situations. They can include speaking and writing tasks that require negotiation strategies, intercultural competence, and multilingual skills, going beyond a narrow focus on native-like English proficiency. Test developers should regularly review and update their assessment criteria to reflect the fluid and non-linear progression of language development.

This study has a few limitations. It primarily focused on low-stakes tests, leaving room for research on how teachers interpret and implement the findings. The study excluded other test types, such as placement exams, and its specific context might not apply universally to all ELF settings. Cross-cultural comparative studies are vital to understanding variations in ELF assessment across different contexts. Moreover, the study did not explore the perspectives of test-takers and assessors. Exploring stakeholders' views can uncover intricate details of ELF testing. Finally, exploring the pedagogical implications of ELF assessment could inform instructional strategies aligned with each setting.

Acknowledgments
We are grateful to the reviewers for their comments, which helped us improve the quality of the manuscript.

Declaration of Conflicting Interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article.

References

Abeywickrama, P. (2013). Why not non-native varieties of English as listening comprehension test input. *RELC Journal*, *44*(1), 59-74. https://doi.org/10.1177%2F0033688212473270

Arefsadr, S., Babaii, E., & Hashemi, M. R. (2022). Why IELTS candidates score low in writing: Investigating the effects of test design and scoring criteria on test-takers' grades in IELTS and World Englishes essay writing tests. *International Journal of Language Testing*, *12*(2), 145-159. https://doi.org/10.22034/ijlt.2022.157131

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Blommaert, J. (2010). *The sociolinguistics of globalization*. Cambridge University Press.

Bøhn, H., & Hansen, T. (2017). Assessing pronunciation in an EFL context: Teachers' orientations towards nativeness and intelligibility. *Language Assessment Quarterly*, *14*(1), 54-68. https://doi.org/10.1080/15434303.2016.1256407

Brown, J. D. (2014). The future of world Englishes in language testing. *Language Assessment Quarterly*, *11*(1), 5-26. https://doi.org/10.1080/15434303.2013.869817

Brown, J. D. (2019). World Englishes and international standardized English proficiency tests. In C. L. Nelson, Z. G. Proshina, & D. R. Davis (Eds.), *The handbook of World Englishes* (pp. 703-724). Wiley Blackwell.

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly*, *3*(3), 229-242. https://doi.org/10.1207/s15434311laq0303_1

Canagarajah, S. (2014). In search of a new paradigm for teaching English as an international language. *TESOL Journal*, *5*(4), 767-785. https://doi.org/10.1002/tesj.166

Canagarajah, S. (2019, March 9-12). Assessing lingua franca interactions as performative. In L. Harding (Chair), *Assessing lingua franca competence* [Symposium]. AAAL Conference, Atlanta, Georgia, United States. https://www.aaal.org/aaal-2019-con

Chan, J. Y. H. (2021). Bridging the gap between ELF and L2 learners' use of communication strategies: Rethinking current L2 assessment and teaching practices. *System*, *101*, 102609. https://doi.org/10.1016/j.system.2021.102609

Cenoz, J., & Gorter, D. (2011). Focus on multilingualism: A study of trilingual writing. *The Modern Language Journal*, *95*(3), 356-369. https://doi.org/10.1111/j.1540-4781.2011.01206.x

Chopin, K., & Bayyurt, Y. (2015). Reconceptualizing norms for language testing: Assessing English language proficiency from within an ELF framework. In S. Akcan & Y. Bayyurt (Eds.), *Current perspectives on pedagogy for English as a Lingua Franca* (pp. 193-204). De Gruyter Mouton.

Clyne, M., & Sharifian, F. (2008). English as an international language: Challenges and possibilities. *Australian Review of Applied Linguistics*, *31*(3), 28.1-28.16. https://doi.org/10.2104/aral0828

Dai D. W., & Roever C. (2019). Including L2-English varieties in listening tests for adolescent ESL learners: L1 effects and learner perceptions. *Language Assessment Quarterly*, *16*(1), 64-86. https://doi.org/10.1080/15434303.2019.1601198

Doosti, M., & Ahmadi Safa, M. (2021). Fairness in Oral Language Assessment: Training Raters and Considering Examinees' Expectations. *International Journal of Language Testing*, *11*(2), 64-90.

Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics*, *26*, 282-304. https://doi.org/10.1017/S0267190506000146

Elder, C., & Harding, L. (2008). Language testing and English as an international language constraints and contributions. *Australian Review of Applied Linguistics*, *31*(3), 34.1-34.11. https://doi.org/10.2104/aral0834

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107-115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Galloway, N., & Rose, H. (2018). Incorporating global Englishes into the ELT classroom. *ELT Journal*, *72*(1), 3-14. https://doi.org/10.1093/elt/ccx010

Ghorbanpour, E., Abbasian, G. R., & Mohseny, A. (2021). Assessment alternatives in developing L2 listening ability: Assessment FOR, OF, AS learning or integration? (assessment x̄ approach). *International Journal of Language Testing*, *11*(1), 36-57.

Hall, C. J. (2014). Moving beyond accuracy: From tests of English to tests of 'Englishing'. *ELT Journal*, *68*(4), 376-385. https://doi.org/10.1093/elt/ccu016

Harding, L. (2008). Accent and academic listening assessment: A study of test-taker perceptions. *Melbourne Papers in Language Testing*, *13*(1), 1-33. https://doi.org/10.3316/aeipt.174066

Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing*, *29*(2), 163-180. https://doi.org/10.1177%2F0265532211421161

Harding, L., & McNamara, T. (2018). Language assessment: The challenge of ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a Lingua Franca* (pp. 570-583). Routledge.

Hu, G. (2012). Assessing English as an International Language. In L. Alsagoff, S. L. McKay, G. Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching English as an International Language* (pp. 131-151). Routledge.

Hu, G. (2017). The challenges of World Englishes for assessing English proficiency. In E. Low & A. Pakir (Eds.), *World Englishes: Rethinking paradigms* (pp. 118-135). Routledge.

Hsu, T. H. L. (2016). Removing bias towards World Englishes: The development of a Rater Attitude Instrument using Indian English as a stimulus. *Language Testing*, *33*(3), 367-389. https://doi.org/10.1177/0265532215590694

Hynninen (2019, March 9-12). Language regulation and assessment of English for Lingua Franca use. In L. Harding (Chair), *Transition, mobility, validity: English as a (Multi)Lingua Franca perspectives on language assessment* [Symposium]. 40th Language Testing Research Colloquium, Auckland, New Zealand. https://www.iltaonline.com/page/ltrc2018 landing.

Isaacs, T. (2008). Toward defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Canadian Modern Language Review*, *64*(4), 555-580. https://doi.org/10.3138/cmlr.64.4.555

Isaacs, T. (2018). Shifting sands in second language pronunciation teaching and assessment research and practice. *Language Assessment Quarterly*, *15*(3), 273-293. https://doi.org/10.1080/15434303.2018.1472264

Isaacs, T., & Harding, L. (2017). Pronunciation assessment. *Language Teaching*, *50*(3), 347-366. https://doi.org/10.1017/S0261444817000118

Isaacs, T., & Rose, H. (2021). Redressing the balance in the native speaker debate: Assessment standards, standard language, and exposing double standards. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3041

Jenkins, J. (2000). *The phonology of English as an international language. New models, new norms, new goals*. Oxford University Press.

Jenkins, J. (2006). Current perspectives on teaching World Englishes and English as a Lingua Franca. *TESOL Quarterly, 40*(1), 157-181. https://doi.org/10.2307/40264515

Jenkins, J. (2015). Repositioning English and multilingualism in English as a Lingua Franca. *Englishes in Practice, 2*(3), 49-85. http://dx.doi.org/10.1515/eip-2015-0003

Jenkins, J. (2017). Not English but English-within-multilingualism. In S. Coffey & U. Wingate (Eds.), *New directions for research in foreign language education* (pp. 63-78). Routledge.

Jenkins, J. (2020). Where are we with ELF and language testing? An opinion piece. *ELT Journal, 74*(4), 473-479. https://doi.org/10.1093/elt/ccaa045

Jenkins, J., & Leung, C. (2014). English as a Lingua Franca. In A. J. Kunnan (Eds.), *The companion to language assessment* (pp. 1605-1616). Wiley-Blackwell.

Jenkins, J., & Leung, C. (2017). Assessing English as a Lingua Franca. In E. Shohamy, I. Or, & S. May (Eds), *Language testing and assessment* (pp. 103-117). Springer.

Jenkins, J., & Leung, C. (2019). From mythical 'standard' to standard reality: The need for alternatives to standardized English language tests. *Language Teaching, 52*(1), 86-110. https://doi.org/10.1017/S0261444818000307

Johnson, J. S., & Lim, G. S. (2009). The influence of rater language background on writing performance assessment. *Language Testing, 26*(4), 485-505. https://doi.org/10.1177%2F0265532209340186

Kang O., Thomson R., & Moran M. (2019). The effects of international accents and shared first language on listening comprehension tests. *TESOL Quarterly, 53*(1), 56-81. https://doi.org/10.1002/tesq.463

Kang, O., Yan, X., Kostromitina, M., Thomson, R., & Isaacs, T. (2023). Fairness of using different English accents: The effect of shared L1s in listening tasks of the Duolingo English test. *Language Testing, 40*(4), 1-27. https://doi.org/10.1177/02655322231179134

Kim, H., & Billington, R. (2016). Pronunciation and comprehension in English as a lingua franca communication: Effect of L1 influence in international aviation communication. *Applied Linguistics, 39*(2), 135-158. https://doi.org/10.1093/applin/amv075

Kirkpatrick, A. (2010). *English as a Lingua Franca in ASEAN: A multilingual model.* Hong Kong University Press.

Kramsch, C. (2011). The symbolic dimensions of the intercultural. *Language Teaching, 44*(3), 354-367. https://doi.org/10.1017/S0261444810000431

Kouvdou, A., & Tsagari, D. (2018). Towards an ELF-aware alternative assessment paradigm in EFL contexts. In N. Sifakis & N. Tsantila (Eds.), *English as a Lingua Franca for EFL* contexts (pp. 227-246), Multilingual Matters.

Labov, W. (2006). *The social stratification of English in New York City*. Cambridge University Press.

Larsen-Freeman, D. (2018). Complexity and ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a Lingua Franca* (pp. 51-60). Routledge.

Leung, C. (2019, March 9-12). Farewell to the Phantom of Standardization. In L. Harding (Chair), *Transition, mobility, validity: English as a (Multi)Lingua Franca perspectives on language assessment* [Symposium]. 40th Language Testing Research Colloquium, Auckland, New Zealand. https://www.iltaonline.com/page/LTRC2018Landing

Leung, C. (2022). Language proficiency: From description to prescription and back. *Educational Linguistics, 1*(1), 56-81. https://doi.org/10.1515/eduling-2021-0006

Leung, C., Lewkowicz, J., & Jenkins, J. (2016). English for academic purposes: A need for remodeling. *Englishes in Practice, 3*(3), 55-73. https://doi.org/10.1515/eip-2016-0003

Llosa, L., & Malone, M. E. (2019). Comparability of students' writing performance on TOEFL iBT and in required university writing courses. *Language Testing, 36*(2), 235-263. https://doi.org/10.1177/0265532218763456

Mahboob, A., & Dutcher, L. (2014). Dynamic approach to language proficiency-A model. In A. Mahboob (Eds.), *Englishes in multilingual contexts* (pp. 117-136). Springer.

Matsumoto, Y. (2011). Successful ELF communications and implications for ELT: Sequential analysis of ELF pronunciation negotiation strategies. *The Modern Language Journal*, *95*(1), 97-114. https://doi.org/10.1111/j.1540-4781.2011.01172.x

Mauranen, A. (2003). The corpus of English as lingua franca in academic settings. *TESOL Quarterly*, *37*(3), 513-527. https://doi.org/10.2307/3588402

Mauranen, A. (2018). Conceptualizing ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a Lingua Franca* (pp. 7-24). Routledge.

Mayring, P. (2004). Qualitative content analysis. In U. Flick, E. von Kardoff, & I. Steinke (Eds.), *A companion to qualitative research* (pp. 159-176). SAGE.

McKay, S. L., & Brown, J. D. (2016). *Teaching and assessing EIL in local contexts around the world*. Routledge.

McNamara, T. (2011). Managing learning: Authority and language assessment. *Language Teaching*, *44*(4), 500-515. https://doi.org/10.1017/S0261444811000073

McNamara, T., & Shohamy, E. (2016). Language Testing and ELF: Making the Connection. In M.-L. Pitzl & R. Osimk-Teasdale (Eds.), *English as a Lingua Franca: Perspectives and prospects: Contributions in honor of Barbara Seidlhofer* (pp. 227-234). De Gruyter.

Miao, Y. (2023). The relationship among accent familiarity, shared L1, and comprehensibility: A path analysis perspective. *Language Testing*, *40*(3), 723-747. https://doi.org/10.1177/02655322231156105

Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics*, *5*(4), 530-555. https://doi.org/10.1111/1467-9481.00163

Motallebzadeh, K., & Baghaee Moghaddam, P. (2011). Models of language proficiency: A reflection on the construct of language ability. *International Journal of Language Testing*, *1*(1), 42-48.

Newbold, D. (2017). Co-certification: A close encounter with ELF for an international examining board. *Journal of English as a Lingua Franca*, *6*(2), 367-388. https://doi.org/10.1515/jelf-2017-0017

Newbold, D. (2018). 12. ELF in Language Tests. In N. Sifakis & N. Tsantila (Eds.), *English as a lingua franca for EFL contexts* (pp. 211-226). Multilingual Matters.

Nishizawa, H. (2023). Construct validity and fairness of an operational listening test with World Englishes. *Language Testing*. https://doi.org/10.1177/02655322221137869

Ockey, G. J., & Hirch, R. R. (2020). A step toward the assessment of English as a Lingua Franca. In G. J. Ockey & B. A. Green (Eds.), *Another generation of fundamental considerations in language assessment* (pp. 9-29). Springer.

Ockey, G. J., Papageorgiou, S., & French, R. (2016). Effects of the strength of accent on an L2 interactive lecture listening comprehension test. *International Journal of Listening*, *30*(1-2), 84-98. https://doi.org/10.1080/10904018.2015.1056877

Park, M. S. (2020). Rater effects on L2 oral assessment: Focusing on accent familiarity of L2 teachers. *Language Assessment Quarterly*, *17*(3), 231-243. https://doi.org/10.1080/15434303.2020.1731752

Piller, I. (2016). *Linguistic diversity and social justice: An introduction to applied sociolinguistics*. Oxford University Press.

Pishghadam, R., & Zabihi, R. (2012). Establishing a life-language model of proficiency: A new challenge for language testers. *International Journal of Language Testing*, *2*(2), 93-109.

Saif, S. (2012). On the Need for a New Model of Language Proficiency and its Implications for Language Testing: A Review of Pishghadam and Zabihi (2012). *Iranian Journal of Language Testing*, *2*(2).

Sato, T., & McNamara, T. (2019). What counts in second language oral communication ability? The perspective of linguistic laypersons. *Applied Linguistics*, *40*(6), 894-916. https://doi.org/10.1093/applin/amy032

Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE.

Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, *24*, 20-39. https://doi.org/10.1017/S0267190504000145

Sewell, A. J. (2013). Language testing and international intelligibility: A Hong Kong case study. *Language Assessment Quarterly*, *10*(4), 423-443. https://doi.org/10.1080/15434303.2013.824974

Shin, S.-Y., Lee, S., & Lidster, R. (2021). Examining the effects of different English speech varieties on an L2 academic listening comprehension test at the item level. *Language Testing*, 38(4), 580-601. https://doi.org/10.1177/0265532220985432

Shohamy, E. (2006). *Language policy: Hidden agendas and new approaches.* Routledge.

Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, *95*(3), 418-429. https://doi.org/10.1111/j.1540-4781.2011.01210.x

Shohamy, E. (2018). ELF and critical language testing. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a Lingua Franca* (pp. 583-593). Routledge.

Shohamy, E. (2019, March 9-12). ELF and multilingualism in search of valid assessment. In L. Harding (Chair), *Transition, mobility, validity: English as a (Multi)Lingua Franca perspectives on language assessment* [Symposium]. *40th Language Testing Research Colloquium*, Auckland, New Zealand. https://www.iltaonline.com/page/LTRC2018Landing

Suzukida, Y., & Saito, K. (2022). What is second language pronunciation proficiency? An empirical study. *System*, *106*(1). https://doi.org/10.1016/j.system.2022.102754

Tajeddin, Z., & Alemi, M. (2014). Pragmatic rater training: Does it affect non-native L2 teachers' rating accuracy and bias? *Iranian Journal of Language Testing*, *4*(1), 66-83.

Vettorel, P. (2018). ELF and communication strategies: Are they taken into account in ELT materials? *RELC Journal*, *49*(1), 58-73. https://doi.org/10.1177/0033688217746204

Wei, J., & Llosa, L. (2015). Investigating differences between American and Indian raters in assessing TOEFL iBT speaking tasks. *Language Assessment Quarterly*, *12*(3), 283-304. https://doi.org/10.1080/15434303.2015.1037446

Widdowson, H. G. (2018). Historical perspectives on ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a Lingua Franca* (pp. 101-112). Routledge.

Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background is a potential source of bias in rating oral performance. *Language Testing*, *30*(2), 231-252. https://doi.org/10.1177/0265532212456968

Xu, Y., Huang, M., Chen, J., & Zhang, Y. (2023). Investigating a shared-dialect effect between raters and candidates in English speaking tests. *Frontiers in Psychology*, *14*, 1143031. https://doi.org/10.3389/fpsyg.2023.1143031

Zhang, Y., & Elder, C. (2011). Judgments of oral proficiency by non-native and native English-speaking teacher raters: Competing or complementary constructs. *Language Testing*, *28*(1), 31-50. https://doi.org/10.1177%2F0265532209360671