Article

Assessment of English as a Lingua Franca and Its Principles: A Research Synthesis

Zari Saeedi*
Allameh Tabataba'i University, Iran

Zia Tajeddin
Tarbiat Modares University, Iran

Fereshteh Tadayon
Allameh Tabataba'i University, Iran

Abstract
Despite the burgeoning interest in English as a Lingua Franca (ELF), there is scant research on the principles that can inform ELF assessment. To address this gap, this mixed-methods research synthesis sought to review and synthesize relevant studies on ELF to outline principles that can serve as a basis for ELF assessment, including test domains, dimensions, and rating rubrics. We drew on three main inclusion criteria to select relevant studies (a) published as journal articles, book chapters, or conference proceedings, (b) appearing during the period from 2000 to 2023, and (c) focused on ELF assessment. The search yielded 58 studies for this research synthesis. Findings revealed nine main principles that can underpin test domains (Using ELF corpora, Prioritizing ELF local context, and Measuring ELF multidimensional competence), test dimensions (Test accommodation, ELF-informed tasks), and rating rubrics (Discipline-informed rating rubric, Nonnative English speakers as raters, Strategy use, and Intelligible pronunciation). Implications are discussed for test developers' awareness of ELF principles that can inform test construct definition, operationalization, and rating criteria.

Keywords
English as a Lingua Franca (ELF), ELF assessment, test domains, test dimensions, rating rubrics, research synthesis

*Corresponding author. Email: saeediz240@yahoo.com

1 Background

Interest in English as a Lingua Franca (ELF) for communication among speakers from different linguacultural backgrounds is becoming increasingly prevalent. The rise of ELF, featuring in the flow of diversity and hybrid use of language among nonnative English speakers (NNESs), has brought to the fore implications for tests that claim to assess English as an indicator of universal proficiency (e.g., Brown, 2014, 2019; Hamid, 2014; Harding & McNamara, 2018; Hu, 2012, 2017; Jenkins, 2007, 2020; Jenkins & Leung, 2019; Leung, 2023; Lowenberg, 2002; Shohamy, 2018, 2019). There are some critics of the one-size-fits-all international language tests that have an intense focus on native English norms, thus failing to meet the contemporary trend of ELF (e.g., Jenkins, 2020; Jenkins & Leung, 2017). The so-called international English language tests demonstrate that native-oriented standards in testing are largely informed by the "communicative turn," which, in turn, is influenced by the standard language variety (i.e., American or British) and mainstream culture (Leung, 2013). Shohamy (2006) stated that by prioritizing native English, international English language tests are abolishing diversity and deemphasizing the "international" definition they claim to render. More recently, Jenkins and Leung (2019) gave an example of an ELF context to argue that assessing language proficiency based on standard native-speaker variations has little correspondence to higher education contexts where the diversity of English use can be seen among NNESs in different disciplines.

This trend of research on ELF counteracts language assessment models inspired by the authority of NESs in designing and making inferences from the tests based on predefined norms. It is argued that high-stakes tests, including Test of English as a Foreign Language (TOEFL), International English Language Testing System (IELTS), and Test of English for International Communication (TOICE), are international in terms of marketing and administration, but not reflecting the diverse use of English among NNESs (Jenkins & Leung, 2019). As there are different varieties of English in any local or international contexts, focusing on one single dominant variety should not be considered as the main principle of assessment design and interpretation (Canagarajah, 2006). Thus, ELF challenges the legitimacy and fairness of these tests on the grounds that they in many respects do not measure the authentic knowledge, ability, or skills of ELF users whose scores rarely mirror their real language knowledge, and are biased or unfair (e.g., Shohamy, 2018). This misalignment is the result of a lack of compatibility between the prescribed models of assessing L2 proficiency premised on native-speaker linguistic norms as the standard on one hand and the dynamic communicative use of English by NNESs in diverse contexts on the other. This incompatibility has a gatekeeping function that prevents many candidates from admission to universities and endangers their professional career as well.

1.1 ELF argument and assessment

In the past decade, the confluence of ELF and language assessment has gained increasing prominence (e.g., Brown, 2019; Harding, 2014; Harding & McNamara, 2018; Hu, 2017; Jenkins, 2020; Jenkins & Leung, 2019; Leung, 2023; Leung et al., 2016; Shohamy, 2018, 2019). A central theme that has emerged is the move beyond the binary demarcation of competence, ability, or knowledge between native and nonnative users. Instead, the focus has shifted toward highlighting the underlying variability that exists within lingua franca communication (Widdowson, 2012), embracing multilingual (Jenkins, 2015) and/or translanguaging competency (Canagarajah, 2006; Shohamy, 2011). Shohamy (2018) directed stakeholders' attention toward the development of an ELF construct and broader contextual characteristics that encompass a comprehensive spectrum of individuals' and groups' languages. This approach contributes to enhancing construct validity, fostering ethical considerations, and creating a more positive influence on pedagogical and assessment practices. According to Larsen-Freeman (2018), Complexity Theory (CT) suggests a non-reductionist approach that seeks patterns in the dynamic performance of language while sustaining stability through reciprocal interactions. In essence, language,

as a complex system, embodies interconnectedness and perpetual dynamism within a given context. In CT, the language repertoires of ELF users never attain ultimate homogeneity, not only due to the heterochronous language-using patterns (Lemke, 2000) required to meet their needs in lingua franca settings but also because they are multifariously engaged in interaction and communication with one another.

Within the academic sphere, where English serves diverse specific and general purposes, the assessment of candidates' English proficiency is predicated upon their alignment with internationally communicative forms of the language (Jenkins, 2020). This approach, while striving to assess language competence on a global scale, inadvertently undermines the intrinsic dynamics of localized linguistic change. The examination of ELFA corpora by Mauranen et al. (2010) elucidated the inherent functioning and autonomy of English within academic lingua franca settings. Contrary to the perception that English operates based on some predetermined "standard" academic English, these analyses reveal that English operates within its distinct paradigm, accommodating the nuanced realities of academic communication. It is crucial to recognize that the concept of English within academia, or academic English, cannot be solely encapsulated by the international reach of language tests. Leung (2023) adds that promoting British or American English as widely accepted norms for assessment has tremendously oversimplified the linguistic diversity within a global context.

Concerning the assessment of English-language proficiency with high-stakes tests, particularly in the context of university admissions, two important concerns have thus far been overlooked. The first concern is the extent to which such tests can authentically assess kinds of language use that align with the actual communicative purposes within academic Target Language Use (TLU) contexts. The second concern is the extent to which these tests can assess the intricate nuances stemming 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 tests, as opposed to being grounded in the tangible realities of language use. Therefore, negligence of these two concerns can yield counterproductive consequences not only for the language itself but also for the future academic pursuits of test candidates (McNamara, 2011).

The inherent mismatch between ELF and assessment criteria, as manifest within the notion of "what is being tested," underscores the incongruity with current and evolving language definitions. This misalignment ultimately strips language tests of their construct validity. Shohamy (2018) highlighted that such tests lack validity in a manner that does not correspond to the actual language knowledge, proficiency, or skills of ELF users, who are assumed to have a lower level of proficiency than those for whom English is a native language. The primary consequence of this underrepresentation of construct validity is that the test scores for those being evaluated fail to accurately reflect their language proficiency. Consequently, the outcome becomes characterized by ethical concerns, unfairness, and a lack of justice.

In English for academic or specific purposes, both IELTS and TOEFL have been designed around the premise of communicative language use that is observed and examined within academic contexts. For instance, the TOEFL iBT test, developed by the ETS, is intended "to gauge the communicative language proficiency of individuals whose first language is not English... in situations and tasks that mirror university life" (Jamieson et al., 2000, p. 10). In essence, test developers draw upon existing models of communicative language ability, which inherently mirror normative references of NESs (e.g., Bachman, 1990; Canale & Swain, 1980; McKay, 2006). Cenoz (2009) and Kirkpatrick (2007) point out that multilingualism and translanguaging are a normal part of routine language use, which should replace monolingual norms prescribed by NESs. However, both IELTS and TOEFL are ingrained in monolingual benchmarks, representing a sharp contrast with bilingual or multilingual speakers whose language uses do not share alignment with norms of assessment inspired by homogenous and native-oriented models of language use.

1.2 The present study

Despite the growth of ELF research, standardized testing models remain resistant to change, continuing to rely on outdated proficiency constructs (Jenkins & Leung, 2019). The problem comes out as international examination boards are framed by established models of communicative language ability, influencing both the "what" of assessment (e.g., construct definition), the "how" of assessment (e.g., diverse facets or characteristics of tests), and standards for rating of speaking and writing assessment. Additionally, the studies have produced inconsistent outcomes as they consider individual aspects one at a time, leaving out various other points that need exploration. Most empirical studies have examined isolated components of ELF assessment—such as intelligibility, discourse strategies, or pronunciation—but have yet to establish a unified model for assessing ELF.

This study sought to address these gaps by systematically identifying principles for ELF-informed assessment models. Specifically, this systematic review provided an initial basis for redefining test constructs, test dimensions, and rating rubrics, which can bring about implications of ELF-oriented testing for language policy and pedagogy (Piller, 2016). By synthesizing findings from recent ELF research, this study called for equitable assessment practices that reflect linguistic diversity rather than prescriptive norms. As Hu (2017) reminds us, such research can "provide further insights needed to inform and extend our understanding of English proficiency and its assessment in a globalized world" (p. 134). Given the centrality and importance of a clear understanding of ELF assessment, this research synthesis aimed to locate, critically appraise, and synthesize relevant studies to arrive at principles that can inform ELF assessment. There are some reasons to apply a mixed-methods research synthesis (MMRS) to extract ELF principles systematically: (a) mixing findings from qualitative and quantitative studies to renders a richer answer to the research question of the present study (Tashakkori & Teddlie, 2003), (b) respecting the contribution of both qualitative and quantitative studies, and (c) amalgamating principles that are sparse in a qualitative or a quantitative study exerted in isolation (Plano Clark et al., 2008). As such, the current study employed MMRS in which a body of studies related to ELF were examined to derive the key principles of how the ELF construct has been defined, operationalized, and rated. The following research question guided this research synthesis:

> RQ: What are the principles of assessing English as a Lingua Franca in terms of relevant studies over the past two decades?

2 Method

2.1 Study identification

As the first phase, we attempted to extract relevant studies from different databases. As stated by In'nami and Koizumi (2010), one of the approaches to locating relevant studies is to conduct a comprehensive search of numerous sources. We conducted an electronic search of the relevant peer-reviewed articles, book chapters, and conference proceedings using different databases: ERIC, Google Scholar, Scopus, and Social Sciences Citation Index. We first started with keyword searching such as (a) EIL, ELF, World Englishes, (b) ELF and assessment, (c) EIL/ELF/World Englishes and language testing proficiency, (d) IELTS/TOEFL and EIL/ELF/World Englishes, (e) Englishes, (f) tests and World Englishes, and (g) English varieties. This extensive search yielded 420 studies, including empirical and non-empirical articles, book chapters, and conference proceedings. Next, we tracked reference lists from previous reviews and empirical articles. Table 1 provides a list of the journals included in the review.

Table 1

Frequency of Journal Articles, Book Chapters, and Conference Proceedings

Studies	Frequency
Articles	**40**
Annual Review of Applied Linguistics	2
Applied Linguistics	4
Applied Linguistics Review	1
Australian Review of Applied Linguistics	1
Canadian Modern Language Review	1
Educational Linguistics	1
ELT Journal	1
Englishes in Practice	2
Frontiers in Psychology	1
Journal of English as a Lingua Franca	1
Journal of language identity and education system	1
Language Assessment Quarterly	7
Language Teaching	2
Language Testing	9
Melbourne Papers in Language Testing	1
Modern Language Journal	1
RELC Journal	1
System	3
TESOL Quarterly	3
Book chapters	**15**
Conference proceedings	**3**

2.2 Inclusion and exclusion criteria

In the present study, journal quality was considered along with two preliminary criteria including indexing (WOS or Scopus) and scope of journal. Besides, the search was conducted through key terms such ELF/EIL/World Englishes in relation to assessment. A total of 19 journals were selected for the further scrutiny of the papers they published. The electronic search initially identified 420 studies, with 250 duplicates removed. To reduce bias, consistent inclusion and exclusion criteria were applied and reviewed multiple times (Norris & Ortega, 2000). Another important point worth mentioning is the inclusion of book chapters and conference proceedings. Both book chapters and conference proceedings represent strict rules of reviewing, whereas PhD theses were excluded due to their lack of formal peer review and editorial process (Page et al., 2021). Non-empirical studies were also included due to theoretical frameworks that helped the purpose of the research (Hammersley, 2001), which was the extraction of ELF assessment principles. In terms of evaluating their "findings," non-empirical studies were assessed based on the clarity, coherence, and rigor of their arguments rather than statistical data. Their "findings" often consisted of synthesized knowledge, conceptual frameworks, or methodological insights rather than quantitative results. To evaluate these contributions, the third researcher and two research assistants examined the proposed theoretical concepts, their logic and the relevance of the studies to this study's main purpose, which was the synthesis of ELF assessment principles. For example, systematic reviews introduced concepts such as symbolic competence, multilingual competence, and

performative competence, which formed the basis for understanding ELF multidimensional competences across studies. Given these considerations, the initial pool of studies associated with EIL/ELF/world Englishes and assessment was based on the journals listed in Table 1. The third author searched each journal individually based on the following criteria:

(a) The study reports on data from 2000 to 2023.

(b) The study is written in English.

(c) The study is published as an article in a journal, a book chapter, or a conference proceeding (since there were not many relevant studies, we included conference proceedings as well).

(d) The study is either empirical or non-empirical in nature.

(e) The study has clear relevance to ELF/World Englishes/EIL and language assessment.

The exclusion criteria for the studies identified were as follows:

(a) Studies are dissertations or an unpublished study, a newsletter, an editorial, an interview, or a book introduction.

(b) Studies are not relevant to WE/EIL/ELF and assessment.

(c) Studies are published in journals that are not indexed in WOS or Scopus.

From identification, as illustrated in Figure 1, the retrieval process rendered 420 studies. A total of 170 not clearly associated with ELF/WE assessment were excluded. The 250 remaining studies went through a meticulous screening of their titles and abstracts. Dissertations, unpublished studies, newsletters, editorials, interviews, and book introductions were also excluded. During more than 10 rounds of searches, the title, abstract, keywords, research questions, methodology, and findings of the studies identified were screened using the abovementioned key terms (e.g., ELF, EIL, World Englishes, assessment, tests, language proficiency) against the inclusion criteria to select possible relevant studies. In this process, 177 studies that did not meet the criteria, including incomplete research reports and unpublished dissertations, were excluded. Moreover, studies reporting issues not associated with our research question on ELF and assessment, including studies on the pedagogical aspects of ELF and not on its assessment, were excluded. Based on the inclusion and exclusion criteria, the data used in this study included 58 studies. Among these, 3 conference proceedings, 15 book chapters, and 40 articles that fulfilled all the inclusion criteria were considered for in-depth analysis.

Figure 1

Procedures for Research Synthesis
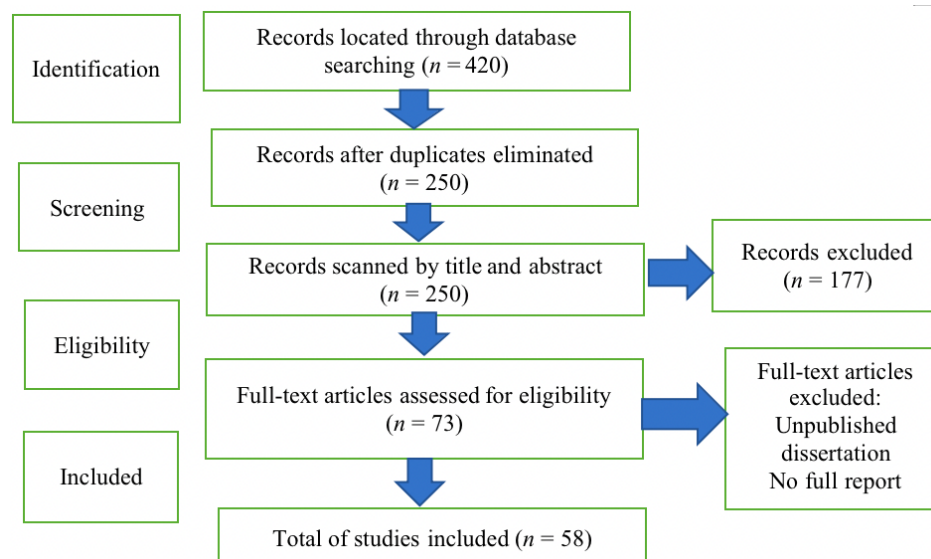
2.3 Data analysis

In this study, the Qualitative Content Analysis (QCA) approach was employed to compare and integrate findings from both empirical and non-empirical studies. As a systematic approach, QCA was employed to comprehensively elucidate the meaning of qualitative data (Schreier, 2012). The method entails the systematic classification of different portions of the data as instances belonging to predefined categories outlined in a coding frame. Notably, QCA requires researchers to focus their analysis on specific facets of the data that align with their research question, thereby distinguishing it from other qualitative data analysis methods.

In this phase of the study, summaries of each of the 58 selected studies were extracted, and key details from each article, book chapter, and conference proceeding were coded. Following the approach outlined by Macaro (2019), three coders, including the third researcher and two research assistants, worked independently to extract the relevant studies and develop a coding scheme informed by ELF principles. NVivo 12 software was used for coding and analysis. The coding process involved rating each article, book chapter, and proceeding based on the following criteria: (a) type of study (e.g., empirical, quantitative/qualitative, or non-empirical, conceptual or review article), (b) research questions, (c) methodological details, and (d) findings. Additionally, concepts suggested by specialists in conceptual studies, such as symbolic competence, intercultural competence, and ELF corpora (e.g., LFC, VOICE, or ELFA), were considered for synthesis. Although these concepts were not directly related to assessment, they were applied to enrich and extract ELF principles effectively. To ensure consistency and agreement, the three coders randomly coded one study from each journal, book chapter, and conference proceeding. Through discussions and clarifications, the coding scheme was refined until consensus was reached, and the agreed-upon coding scheme was used for further analysis of all studies. This systematic approach to coding and synthesis facilitated the identification of key findings and principles related to ELF assessment, drawing from a diverse range of studies and incorporating both empirical and non-empirical sources.

Given that some of the extracted findings were originally reported in a quantitative format, an approach known as qualitizing was used to transform these quantitative findings into qualitative themes. This process, as described by Nzabonimpa (2018) and Onwuegbuzie and Leech (2019), involves analyzing quantitative data to identify codes, categories, and themes. First, there was a familiarization with the results, involving a thorough review of the quantitative findings to understand their significant outcomes, concepts, and trends. Following this, the second step involved coding significant findings, where key quantitative results were assigned codes representing meaningful recurring patterns. The third step was grouping codes into broader thematic categories, organizing the individual codes into larger conceptual groups that captured underlying concepts (Sandelowski & Barroso, 2006). The established coding scheme was applied to analyze diverse aspects of the selected studies, including the type of studies (e.g., articles, book chapters, conference proceedings), research questions, methodological details, and findings. In every quantitative study, the findings were interpreted as themes or explanations, which were then incorporated into the themes that emerged from the qualitative studies. This process continued to have themes grouped into some predefined categories suggested by Elder and Davies (2006). These categories included the "what" of assessment (defining the ELF construct), the "how" of assessment (operationalizing the ELF construct), and rating rubrics. The three coders carefully examined each study entirely, seeking ideas, perspectives, and recommendations that could effectively inform ELF principles. The synthesis process resulted in the emergence of key themes based on ELF domains, dimensions, and rating rubrics, with the aim of achieving coherence across the data. The emerging key themes were thoroughly discussed among the coders, and agreement was reached through the collaborative revision of different drafts of coding. This collaborative process ensured the accuracy and validity of the identified themes.

3 Findings

The review of the past studies has yielded a set of ELF-informed principles for three important aspects of the test design: test domains, test dimensions, and rating rubrics. The findings relevant to each aspect are reported below. Figure 2 shows a summary of ELF assessment principles underpinning three main categories:

Figure 2

ELF Assessment Principles
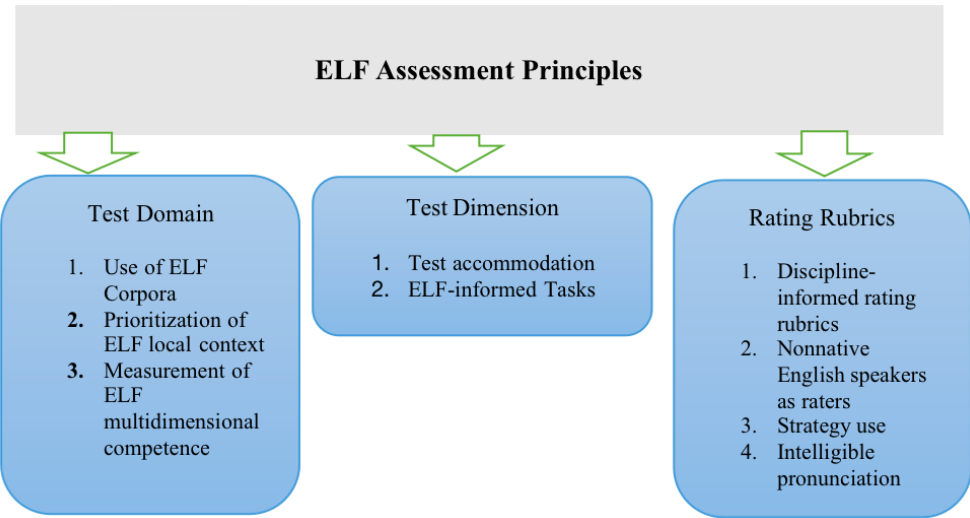
3.1 ELF principles of assessment domains

Table 2 shows three main principles underlying target domains grounded in the identified studies, indicating what should be assessed as part of the ELF construct. These principles include the use of ELF corpora, the prioritization of the ELF local context, and the measurement of ELF multidimensional competence.

3.1.1 Use of ELF corpora

The principle of redefining ELF emphasizes the importance of communicative contexts, advocating for assessments derived from ELF communication domains that incorporate grammatical, lexical, prosodic, phonological, and discoursal features (Elder & Davies, 2006). ELF corpora, such as the Lingua Franca Core (Jenkins, 2000), the Vienna-Oxford International Corpus of English (Seidlhofer, 2004), the English as a Lingua Franca in Academic Settings (Mauranen, 2003), and the Asian Corpus of English (Kirkpatrick, 2010), provide insights into the diverse and dynamic nature of real-world English use. These corpora highlight essential features like phonological intelligibility, pragmatic strategies, and region-specific linguistic traits, offering authentic materials for language assessment. By analyzing real-world interactions, such as in VOICE or ACE, test developers can design realistic and relevant tasks reflecting the communicative needs of ELF users, integrating strategies like code-switching and paraphrasing to ensure mutual understanding (Seidlhofer, 2018). This approach moves away from standardized norms, emphasizing the variability and complexity of ELF communication, while demonstrating consistent patterns in grammatical and discoursal dimensions to revise test constructs (Seidlhofer, 2004). Consequently, these corpora enable the development of flexible and valid assessments tailored to the needs of NNESs.

Table 2

ELF Assessment Principles of Domains

Use of ELF corpora	This principle refers to the role of ELF-informed corpora that showcase a wide range of ELF interactions	Lingua Franca Core (LFC) Corpus of Academic English (ELFA) Vienna-Oxford International Corpus of English (VOICE) Asian Corpus of English (ACE)	(e.g., Elder & Davies, 2006; Jenkins, 2000; Kirkpatrick, 2010; Mauranen, 2003; Seidlhofer, 2004)
Prioritization of ELF local context	This principle explains the importance of the local context that shapes ELF language use.	The local context of participant-driven language use Locally defined EIL approach Englishing (what each learner can do with English)	(e.g., Brown, 2004, 2012, 2014; Canagarajah, 2006; Hall, 2014; Jenkins & Leung, 2017, 2019; Leung, 2023; Leung et al., 2016; McKay & Brown, 2016)
Measurement of ELF multidimensional Competence	This principle moves away from linguistic competence to highlight a broad range of ELF competence.	Intercultural competence Multilingual competence Symbolic competence	(e.g., Brown, 2019; Canagarajah, 2006, 2011, 2013; Cenoz, 2017; Kramsch, 2011; Leung, 2019, 2023; Lopez et al. 2017; McKay, 2011; Shohamy, 2019)

3.1.2 Prioritization of ELF local context

This principle advocates for localized ELF assessments that prioritize specific contexts and purposes over standardized tests like IELTS and TOEFL iBT (Jenkins & Leung, 2019). Localized assessments should consider factors such as disciplines, institutions, NNESs' languages, and communication types like translanguaging and multilingualism (Leung et al., 2016; McKay & Brown, 2016). Localization through needs analysis ensures curricula and tests are tailored to the diverse requirements of ELF users (Brown, 2012; Canagarajah, 2006). Addressing variations among "Englishes"—including test takers' own English and the English used for decision-making—is crucial for test validity (Brown, 2004). This approach emphasizes users' actual performance in context-specific tasks, fostering genuine communication and linguistic diversity (Gu & So, 2015; Hall, 2014). By shifting to smaller, locally focused assessments, ELF users' communicative competence is highlighted, promoting an inclusive understanding of English proficiency (Graddol, 2006; Leung, 2023).

3.1.3 Measurement of ELF multidimensional competence

Multidimensional competence in ELF encompasses linguistic, intercultural, and strategic communication skills, reflecting the importance of assessing users' intercultural strategic competence alongside linguistic abilities (Brown, 2019; Canagarajah, 2006, 2011, 2013; Kramsch, 2011; McKay, 2011; Shohamy, 2019). ELF users demonstrate multicompetence, which emphasizes unique forms, meanings, and functions of ELF rather than native speaker performance, requiring awareness of transcultural communication dynamics (Bassetti & Cook, 2011). Language assessment models should account for the variability

of interactional contexts across regions, disciplines, and time, reflecting specific ELF contexts and multilingual uses (Leung, 2019). This approach recognizes multiple proficiencies in English and other languages, emphasizing translanguaging as an integrated and adaptive linguistic practice (Canagarajah, 2011; Cenoz, 2017). Symbolic competence further enhances assessments by considering cultural meanings and performative competence in intercultural communication (Canagarajah, 2013). Ultimately, ELF competence integrates linguistic and social practices, optimizing communicative effectiveness for NNESs (Leung, 2023; Shohamy, 2018).

3.2 ELF principles of assessment dimensions

The findings of this research synthesis also resulted in the identification of ELF principles of assessment dimensions, which can be synthesized under the two main categories: Test accommodations and ELF-informed tasks. Table 3 illustrates the general features of test tasks upon which the ELF construct can be operationalized into principles.

Table 3

ELF Assessment Principles of Dimensions

Principles	Descriptions	Studies
Test accommodations	This principle refers to different ways that can make tests fairer and easier to test takers.	(Elder & Davies, 2006)
ELF-informed Tasks	This principle demonstrates some features that can be incorporated into test items or tasks to make them more amenable toward NNESs' needs	(e.g., Abeywickrama, 2013; Brown, 2019; Dai & Roever, 2019; Harding, 2008, 2012; Hynninen, 2019; Jenkins & Leung, 2019; Kang et al., 2019, 2023; Kim & Billington, 2016; Major et al., 2002; Matsuura, et al., 2016; Miao, 2023; Nishizawa, 2023; Ockey & French, 2016; Shin et al., 2021; Shohamy, 2018)

3.2.1 Test accommodations

This principle acknowledges that the test construct can remain largely unchanged for ELF users who are potential candidates for high-stakes tests (e.g., IELTS and TOEFL iBT). Minor accommodations can enhance fairness in test delivery, particularly in speaking and reading tests (Elder & Davies, 2006). For example, selecting reading texts that avoid specific topics or genres relevant to particular contexts of use eliminates the need for providing nonnative speakers with background knowledge. Similarly, removing complex grammatical structures or infrequent lexical items that are uncommon for ELF users can be achieved by providing brief introductions to reading texts or using glossaries to exempt nonnative speakers from acquiring difficult lexical knowledge. In speaking tests, involving ELF examiners or interlocutors as interviewers can scaffold test candidates and ensure fair scoring and standard-setting procedures.

3.2.2 ELF-informed tasks

The principles of ELF in language testing have evolved to emphasize authenticity and inclusivity, with scholars proposing various strategies to reflect global English use. Elder and Davies (2006) advocate

performance-based assessments, while Brown (2019) and Harding (2012) suggest integrative and interactive tasks. Studies by Abeywickrama (2013), Dai and Roever (2019), and Kang et al. (2019, 2023) reveal the nuanced effects of non-native accents, shared-L1 advantages, and accent familiarity on listening performance, underscoring the importance of managing these factors in tests. Additionally, Ockey and French (2016) and Miao (2023) highlight how accent strength and familiarity influence comprehension, with shared-L1 and familiar accents often aiding performance. Leung et al. (2016, 2019) emphasize authentic interaction through non-native speakers and locally relevant tasks, and Jenkins (2015) and Shohamy (2018) propose multilingual tasks like role-plays to enhance fairness and relevance. Together, these findings underscore the necessity of adapting test design to reflect diverse, contextualized, and practical ELF use.

3.3 ELF assessment principles of rating rubrics

Following the ELF principles of test domains and dimensions, this study examined rating rubrics from prior studies. As shown in Table 4, four principles can guide the rating procedure of an ELF-informed test: Discipline-informed rating rubrics, Nonnative English speakers as raters, Strategy use, and Intelligible pronunciation.

Table 4

ELF Assessment Principles of Rating Rubrics

Principles	Descriptions	Studies
Discipline-informed rating rubrics	This principle underlines developing a rating rubric based on each specific discipline.	(e.g., Hu, 2012; Jenkins & Leung, 2019; McKay & Brown, 2016; Sato & McNamara, 2019; Shohamy, 2018)
Nonnative English speakers as raters	This principle refers to raters who can be NNESs and have a shared L1 with test takers	(e.g., Elder & Davies, 2006; Hsu, 2016; Hynninen, 2014; Johnson & Lim, 2009; Kim, 2009; Park, 2020; Wei & Llosa, 2015; Winke et al., 2013; Xu et al., 2023; Zhang & Elder, 2011)
Strategy use	This principle refers to the rating of spoken interactions based on the use of communication strategies.	(e.g., Brown, 2019; Harding, 2012, 2014; Kouvdou & Tsagari, 2018; Newbold, 2018; Ockey & Hirch, 2020)
Intelligible pronunciation	This principle highlights intelligible pronunciation required for spoken interactions.	(e.g., Isaacs, 2008, 2018; Isaacs & Harding, 2017; Isaacs & Trofimovich, 2012; Jenkins, 2000)

3.3.1 Discipline-informed rating rubrics

The redevelopment of current rating rubrics necessitates the consideration of context-aware assessment criteria that can effectively capture relevant norms in specific academic disciplines (Shohamy, 2018). As Jenkins and Leung (2019) argue, successful communication should be redefined to genuinely reflect the flexible use of linguistic and semiotic resources by ELF users in diverse contexts. In other words, the dynamic nature of ELF communication cannot be prespecified or graded in terms of marks, bands, or levels, making it meaningless to anticipate real-life task performance levels. For instance, McKay and Brown (2016) explain that the criteria for efficient writing can vary significantly based on the purpose of the writing and context. Additionally, rhetorical conventions can differ greatly across various local communities, and these differences should be taken into account when assessing students' writing.

While in global assessment contexts it seems appropriate to judge students' writing against a prescribed standard, in local contexts, decisions about acceptability—such as grammar and vocabulary—must be contextualized within the locally used variety of English and its norms (Hu, 2012). More interestingly, Sato and McNamara (2019) challenge traditional rating criteria, advocate the use of factors beyond linguistic accuracy, and call for a reevaluation of proficiency test criteria to capture the diverse criteria considered by real-world interlocutors.

3.3.2 Nonnative English speakers as raters

The reciprocal and dynamic nature of ELF communication challenges the notion that competence is solely an individual attribute. Instead, it suggests that competence should be evaluated within group scores and assessed by ELF examiners from diverse L1 backgrounds (Elder & Davies, 2006; Harding, 2012; Hsu, 2016; Hynninen, 2014; Johnson & Lim, 2009; Kim, 2009; Park, 2020; Wei & Llosa, 2015; Winke et al., 2013; Xu et al., 2023; Zhang & Elder, 2011). Elder and Davies (2006) argue that NNES raters are effective judges of ELF users' competence by prioritizing comprehensibility over linguistic norms. These raters should have expertise and experience to recognize and overlook nonstandard linguistic and discourse features if the overall message is comprehensible. For example, new norms might include pluralizing uncountable nouns (e.g., "equipments"), using nonstandard topic-comment structures (e.g., "TV I don't usually watch"), or employing phrasal verbs (e.g., "cope up with"). Research on nonnative raters' linguistic backgrounds highlights the nuanced relationship between their linguistic diversity and the rating of speaking proficiency in high-stakes language assessments. Zhang and Elder (2011) observed similar outcomes across diverse rater groups, while Winke et al. (2013) identified leniency among nonnative raters despite lacking familiarity with test takers' first languages. Studies by Wei and Llosa (2015) and Johnson and Lim (2009) emphasized the role of linguistic background in identifying language features and reducing bias through training. Harding (2012) and Park (2020) explored the influence of accent familiarity, with findings suggesting potential bias and its impact on interrater reliability. Xu et al. (2023) highlighted the importance of linguistic awareness in scoring processes. Collectively, these studies provide a nuanced understanding of the impact of nonnative raters' linguistic backgrounds on the rating of speaking proficiency.

3.3.3 Strategy use

The dominance of prescriptive language norms is particularly challenged in spoken interaction. Localized varieties of English have been developed and successfully used in local interactions, with their own standards emerging over time (Rose & Syrbe, 2018). When assessing English for local use, it is logical to evaluate language output against these localized standards. In regions where multiple localized varieties are relevant for successful communication, all of these should be represented in tests (Brown, 2019). In ELF-oriented assessments, prioritizing pragmatics over grammatical competence leads to norm-defocused, user-centered rating scales (Newbold, 2018). Harding (2012) emphasizes holistic rating criteria and purpose-built tasks that facilitate interaction and accommodation across diverse linguistic backgrounds. Building on this, Ockey and Hirch (2020) propose an ELF-informed oral communication test incorporating criteria such as rhetorical sensitivity, interactional competence, and fairness. Kouvdou and Tsagari (2018) adapt assessment practices to focus on strategic competence over linguistic accuracy, treating linguistic errors as authentic deviations rather than deficiencies. Overall, these studies highlight the multilingual, dynamic nature of ELF communication and underscore the need for rating rubrics that reflect the use of strategies for being communicatively effective.

3.3.4 Intelligible pronunciation

In ELF contexts, the emphasis has shifted from achieving native-like pronunciation to ensuring clear and effective communication, with intelligibility defined as being easily understood regardless of accents (Isaacs & Harding, 2017). This shift has led to a rejection of the nativeness principle in favor of interactional intelligibility standards (Sewell, 2013). Jenkins (2000) developed the LFC, a set of pronunciation features crucial for intelligibility in ELF contexts, excluding less important features like connected speech. Despite critiques about the limited dataset used in its development, the LFC remains relevant in paired and group oral assessments, where interlocutors' pronunciation patterns converge or diverge, especially in same-L1 versus different-L1 pairings. However, Isaacs (2008) explored whether intelligibility was a comprehensive construct for screening international teaching assistants and concluded that it was insufficient for this purpose. Isaacs (2018) developed an L2 English comprehensibility scale designed for formative assessment to guide educators on targeting specific linguistic features at different levels of L2 comprehensibility. Isaacs and Trofimovich (2012) showed that comprehensibility is related to a wide range of linguistic domains, including segmental, prosodic, temporal, lexicogrammatical, and discourse-level measures. The scale includes descriptors articulating the "effortfulness" of understanding L2 speech, supplemented with examples of error types that impede comprehension at lower levels, such as misplaced word stress, sound substitutions, and insufficient stress on important words in a sentence. A key design decision was to explicitly state that "sounding native-like is not expected" at the highest level of the scale, clarifying that students can achieve top proficiency without eliminating all traces of their L1 accent.

4 Discussion

The present study sought to synthesize a set of principles to inform the design of ELF assessments. A systematic review of 58 studies identified three core principles that underlie effective test design: ELF Principles of Assessment Domains, ELF Principles of Assessment Dimensions, and ELF Assessment Principles of Rating Rubrics.

The principles represent different features of assessment domains, dimensions, and rating rubrics. They foster a growing understanding of ELF assessment that transcends traditional norms defined by standard varieties. Instead, they represent ELF use, restructured by the contingent, dynamic, and local linguistic, sociopragmatic, and multidimensional repertoires of translingual and/or multilingual speakers (e.g., Jenkins & Leung, 2017). Referring to CT, Larsen-Freeman (2018) posits that the linguistic repertoires of ELF users never achieve ultimate homogeneity. This is not only due to the dynamic language-using patterns required in lingua franca settings but also because users are engaged in multifaceted interactions and communications. Each corpus has its distinctive features that can inform the test construct. The use of different ELF corpora for test construct definitions may offer a worthwhile understanding of what ELF is and how educated native speakers can showcase the appropriate and real-life language needs of ELF users by distancing from "authorized pathfinders" (Widdowson, 2012, p. 16). It should be noted that features shaping ELF interactions are not non-normative, as Seidlhofer (2004) stressed, because there is regularity in ELF use. It means that regularity can span from grammatical to discursive features and grow into forms and functions to help represent the ELF domains required for the renovation of the test construct. These corpora can be used as a source of authentic materials representing the use of ELF in different contexts and functions and as a basis for developing tests catering to NNESs' purposes and needs. For example, these corpora can provide a basis for assessing listening based on ELF interactions as embodied in VOICE and ELFA, or assessing pronunciations using LFC. Thus, they can represent a more realistic picture of intercultural competencies, communication strategies, and diverse linguistic skills that are mainstream in EIL contexts (Rose & Syrbe, 2018).

The findings also reveal that incorporating local communicative contexts as one of the assessment domains underscores the inherent variability of ELF use, referred to as *Virtual Capacity* (i.e., how ELF communication includes adaptive variability and offers the idea of virtual language as a flexible, emergent resource shaped by context and interlocutors within the encoded language itself) (see Seidlhofer, 2011). Simply put, this principle suggests that ELF users utilize available language resources based on what is possible, even if it has not been formally codified (Widdowson, 1997). Consequently, this emphasizes the flexible nature of ELF as essential for effective communication across diverse local contexts among speakers from various backgrounds(Jenkins et al., 2011). This sheds light on the definition of the assessment domain, which should encompass participant-driven language use, reflecting what ELF users actually do with the language for locally defined assessment purposes. This perspective seeks to strengthen the connection between increased validity, social justice, and fairness by adopting situated language practices and context-sensitive proficiency (Jenkins & Leung, 2019). As Leung (2019) elaborates, the emphasis should shift toward understanding the unique local context and what is "standardly expected" within that environment. In the context of university English language entry tests, this involves examining how English is practically used in each specific setting. In many English-medium universities, this often entails the use of English as a (multi)lingua franca, where students communicate through their various forms of English and leverage their multilingual abilities, a concept known as translanguaging. Similarly, the concept of a locally defined EIL approach denotes the adaptation of teaching objectives and materials to suit specific local needs of NNESs (Brown, 2012).

The findings show that the concept of ELF multidimensional competence has challenged the notion of idealized competence, which arose from the traditional pedagogy of NES' conventions (Baker, 2018). Similarly, Harding and McNamara (2018) question the communicative competence model, as it does not adequately capture the essence of successful ELF communication or fails to identify who qualifies for it. In other words, ELF communicative competence should be defined based on intercultural awareness, performative/symbolic competence, and multilingual/translanguaging competence, which can better reflect the variable representation of ELF communication (e.g., Jenkins, 2015; Leung, 2023; Shohamy, 2011). Brown (2019) points out that the symbolic competence approach (where numerous meanings and diverse interpretations come into play in intercultural communication), the intercultural communication skills approach, and the performative ability approach constitute bottom-up, person-focused methods of testing. Bassetti and Cook (2011) argued that ELF users possessing multicompetence show their communicative competence tilted towards ELF forms, meanings, and functions differently from those defined in terms of native speakers' performance. Multidimensional competence in ELF also extends to English as a Multilingua Franca (Jenkins, 2015), suggesting that language assessment should focus on how well individuals navigate diverse interactions. May (2014) emphasizes the importance of effectively utilizing multilingual skills in teaching and testing, which includes both using ELF and being multilingual. This is the main transformation toward other languages including NNESs' L1 and multilingual repertoire.

The synthesis of ELF principles underpinning the assessment of dimensions reveals that the operationalization of ELF necessitates a significant transformation in input content and task design. This includes various test features such as the test accommodations initially proposed by Elder and Davies (2006), different task characterizations like performance-based assessments and/or interactive goal-oriented tasks, as well as locally defined test tasks (McKay & Brown, 2016). The reviewed studies consistently emphasize that localized standards of a test should be coupled with candidates' self-assessment of their need to use ELF not only for general communication but also for discipline-specific communication, given that the intended use of a test can determine certain linguistic norms (Jenkins & Leung, 2019). As Bachman (2005) argued, the intended uses and potential consequences of a test are the determinants of specific local concerns in language assessment. If the core of ELF assessment is to make true inferences from tests, self-assessment provides more accurate norms of ELF use, which can be influenced and shaped by diverse language and literacy practices in English-medium universities.

In this respect, the test input should draw on local and discipline-specific situated language practices, and the task performance should reinforce self-regulation skills in the candidates (e.g., Hynninen, 2019; Jenkins & Leung, 2019; Shohamy, 2018). In other words, the specific features of the test input should be defined based on candidates' awareness of their needs and requirements of higher education programs, taking into account their self-assessment skills to decide what types of assessment materials, tasks, or activities can help them better fulfill their academic purposes. As argued by Hynninen (2019), assessing competence in ELF should be premised on whose norms should be channeled into the tests and what should be considered acceptable, functional, or good aspects of task performance. Practically, self-assessment, given local practices and a variety of assessment formats and tasks, should inform the development of assessment procedures that can adequately represent NNESs' language proficiency.

This systematic review indicates that ELF rating rubrics necessitate NNESs as examiners who are knowledgeable about ELF users' performance (Cenoz, 2009; Kirkpatrick, 2007). Monolinguals should not be the sole benchmark for assessing multilingual speakers; rather, multilingualism, particularly translanguaging, should be considered a standard language practice. The review also highlights that ELF test performance should not be separately scored due to the varied domains and dimensions, which may surpass static measures of language ability. This review suggests that the principles underpinning ELF assessment can unravel the lack of situatedness and contingencies in traditional rating rubrics by emphasizing scoring a wide-range use of strategies essential for successful communication, developing ELF rubrics for various academic domains, and focusing on task fulfillment and intelligibility. These measures can provide a clearer understanding of performance in ELF communicative contexts and activities. However, there are some reasons for resistance by testing organizations or hidden ideologies inherent in these widely common standards on measuring language proficiency of NNESs. As Jenkins and Leung (2019) argue, the assessment of test-taker performance has been heavily influenced by conventional rules rooted in Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), leading to an idealized conceptualization of prespecified language abilities within the academic context. Given this, it remains unclear how well these standards reflect ELF communication (McNamara, 2011). One of these challenges to test development is driven forward by language education policy (LEP) whereby ideologies come to practice and stakeholders comply (Shohamy, 2006). There is no doubt that language tests are the de facto part of language policies. These tests, above all IELTS and TOEFL, have grown in importance under the complex processes of ideological enforcements and political approbations. This is exactly why these "big tests" spreading out Anglophone practices have called themselves "international" because they are marketed internationally. In effect, LEP promotes a hidden and covert agenda for English tests to appear on the surface as objective, standardized, reliable, and trustworthy instruments of domination (Davies et al., 2003). Second, designing the ELF construct on the flexible and variable norms of communicative effectiveness (Elder & Harding, 2008) may tend to result in a complex picture, which has made testing agencies irreconcilable to prompt dynamic norms in test design. Particularly, there might occur unpredictable consequences for test takers who would not have any good command of what kinds of language use is acceptable, what sources serve as a model for test preparation, or what standards or criteria benchmark the correctness of English use in test performance (Jenkins, 2007). Cargos (2022) found that barriers to integrating Global Englishes include English commodification, linguistic intolerance, and political resistance when viewed as ideological. Baker et al. (2025) showed that English and TESOL in higher education can both marginalize and privilege learners, affecting their academic opportunity. Rose and McKinley (2025) pointed out that assessment washback remains a major obstacle to innovation, highlighting the need for new, context-sensitive assessment standards that reflect the global diversity of English. Despite these challenges, the present review stresses that language assessment should not only focus on the technicalities of language use but also consider the cultural and societal implications of tests that can reshape assessment models and practices to accommodate dynamic, context-sensitive, and multilingual realities of intercultural communication of NNESs.

5 Conclusion and Implications

The present research critically reviewed a body of 58 studies to integrate the conceptualization of ELF principles, which can inform an initial understanding of domains, dimensions, and rating rubrics for assessment purposes. The study highlights that traditional prescriptive norms are insufficient for capturing the complexities of ELF use and calls for assessment constructs that reflect the variability and situated practices of ELF users. The principles call for fairness, validity, and relevance, providing a framework for more equitable and context-sensitive assessment frameworks and practices. These ELF principles thus underscore the necessity to rethink the design of tests that can capture the nuances of bilingual/multilingual NNESs' language ability, thereby representing authentic and effective communication rather than standardized notions of linguistic correctness.

This study has several implications for stakeholders who can leverage the principles of assessing ELF. From a theoretical perspective, this systematic review provides a basis for reevaluating the existing models of test design and creating space to the principles of assessment given that the use of ELF reflects the dynamic and localized nature of language use in bilingual and/or multilingual contexts. This encourages adopting a more evidence- and context-based approach, given the multilingual nature of ELF, for more updated assessment practices and models. Broadening the discourse, the study also calls for increased attention to fairness and ethics in language assessments, stressing that test development practices should not only focus on the technicalities of language use but also consider their cultural and societal consequences (Shohamy, 2018). Moreover, this review serves as a preliminary step in raising awareness among language assessment professionals about the principles essential for evaluating ELF. Such an understanding highlights the social and cognitive plurilithic nature of English, as argued by Hall (2014), urging testers, ELT professionals, and educators to reconsider their assumptions about English within ELF communities. This raises a necessity for ELT professionals to integrate these principles into updated assessment models that better address the realities of ELF communication.

From a practical standpoint, following the insights of Jenkins and Leung (2019), this review suggests a comprehensive range of assessment formats, task designs, and rating criteria that reflect ELF users' performances more accurately. Implementing these principles can enhance test development processes, including construct definition, task creation, and scoring methods. Teacher education can encourage stronger collaboration among educators, practitioners, and researchers to elevate the role of ELF in teacher professional development and classroom-based assessments (Jung & Crookes, 2025; Pan et al., 2025; Sifakis & Kordia, 2025; Shu et al., 2024). Preservice and in-service teachers need to be aware of recent theories underscoring ELF for instructional and assessment practices (Galloway, 2025; Imperial, 2025). Also, English-medium universities are in demand to reconsider their admissions criteria and assessment practices to mirror ELF principles for further relevance and fairness. Such alignment can generate positive washback effects, improving teaching quality and informing teacher training curricula.

This study has some limitations, urging the need for further research. First, the exclusion of PhD theses and/or unpublished papers might have limited the representativeness of findings. Future research should expand the scope of evidence by incorporating a broader range of grey literature and unpublished studies to enhance comprehensiveness (Adams et al., 2016). Second, subjective bias in the interpretation of the results to synthesize the ELF principles along with heterogeneity of included studies for both qualitizing quantitative findings and integrating qualitative results might have resulted in loss of details or nuances, which can impact the validity of interpretations (Onwuegbuzie & Leech, 2019). Future research can employ mixed-methods designs to reduce interpretative bias and preserve data richness (Nzabonimpa, 2018).

References

∗References marked with an asterisk indicate studies included in the research synthesis.

∗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

Adams, J., Hillier-Brown, F. C., Moore, H. J., Lake, A. A., Araujo-Soares, V., White, M., & Summerbell, C. (2016). Searching and synthesizing 'grey literature 'and 'grey information 'in public health: Critical reflections on three case studies. *Systematic Reviews*, *5*, 1-11. https://doi.org/10.1186/s13643-016-0337-y

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

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

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, *2*(1), 1-34. https://doi.org/10.1207/s15434311laq0201_1

Baker, W. (2018). English as a lingua franca and intercultural communication. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a Lingua Franca* (pp. 25-36). Routledge.

Baker, W., Morán Panero, S., Álvarez Valencia, J. A., Alhasnawi, S., Boonsuk, Y., Ngo, P. L. H., ... & Ronzón-Montiel, G. J. (2025). Decolonizing English in higher education: Global Englishes and TESOL as opportunities or barriers. *TESOL Quarterly. 59*(1), 281-309, https://doi.org/10.1002/tesq.3317

Bassetti, B., & Cook, V. J. (2011). Language and cognition: The second language user. In V. J. Cool & B. Bassetti (Eds.), *Language and bilingual cognition* (pp. 143-190). Psychology Press.

∗Brown, J. D. (2004). What do we mean by bias, Englishes, Englishes in testing, and English language proficiency. *World Englishes*, *23*(2), 317-319. https://doi.org/10.1111/j.0883-2919.2004.00355.x

∗Brown, J. D. (2012). EIL curriculum development. In L. Alsagoff, S. L. Mckay, G. Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching English as an international language* (pp. 147-167). Routledge

∗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. (2011). Translanguaging in the classroom: Emerging issues for research and pedagogy. *Applied Linguistics Review*, *2*, 1-28. https://doi.org/10.1515/9783110239331.1

∗Canagarajah, S. (2013). *Translingual practice: Global Englishes and cosmopolitan relations*. Routledge.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1-47. https://doi.org/10.1093/applin/I.1.1

Cargos, K. (2022). From ELT to GELT: The feasibility of Global Englishes language teaching curricular innovation. *JALT Journal*, *44*(1), 7-32. https://doi.org/10.37546/JALTJJ44.1-1

Cenoz, J. (2009). *Towards multilingual education: Basque educational research from an international perspective*. Multilingual Matters.

*Cenoz, J. (2017). Translanguaging in school contexts: International perspectives. *Journal of Language, Identity & Education*, *16*(4), 193-198. https://doi.org/10.1080/15348458.2017.1327816.

*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

Davies, A., Hamp-Lyons, L., & Kemp, C. (2003). Whose norms? International proficiency tests in English. *World Englishes*, *22*(4), 571-584. https://doi.org/10.1111/j.1467-971X.2003.00324.x

*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. http://doi.org/10.2104/aral0834

Galloway, N. (2025). What's in a name? Global Englishes—An umbrella term to address silo mentality or a misunderstood paradigm reinforcing silos? *TESOL Quarterly, 59*(1), 24-48. https://doi.org/10.1002/tesq.333y7

Graddol, D. (2006). *English next: Why global English may mean the end of 'English as a Foreign Language.'* British Council

Gu, L., & So, Y. (2015). Voices from stakeholders: What makes an academic English test 'international'. *Journal of English for Academic Purposes*, *18*, 9-24. https://doi.org/10.1016/j.jeap.2014.10.002

*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

Hamid, M. O. (2014). World Englishes in international proficiency tests. *World Englishes*, *33*(2), 263-277. https://doi.org/10.1111/weng.12084

Hammersley, M. (2001). On 'systematic' reviews of research literatures: A 'narrative' response to Evans & Benefield. *British Educational Research Journal*, *27*(5), 543-554. https://doi.org/10.1080/01411920120095726

*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. (2014). Communicative language testing: Current issues and future research. *Language Assessment Quarterly*, *11*(2), 186-197. https://doi.org/10.1080/15434303.2014.895829

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, N. (2014). The common European framework of reference from the perspective of English as a lingua franca: What can we learn from a focus on language regulation. *Journal of English as a Lingua Franca*, *3*(2), 293-316. https://doi.org/10.1515/jelf-2014-0018

*Hynninen (2019). 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.

Imperial, R. A. (2025). Rethinking Global Englishes and moving toward reparative redress for language-minoritized and racialized TESOL practitioners. *TESOL Quarterly, 59*(1), 310-332. .https://doi. org/10.1002/tesq.3338

In'nami, Y., & Koizumi, R. (2010). Database selection guidelines for meta-analysis in applied linguistics. *TESOL Quarterly*, *44*(1), 169-184. https://doi.org/10.5054/tq.2010.215253

*Isaacs, T. (2008). Towards 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., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*(3), 475-505. https://doi.org/10.1017/S0272263112000150

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 framework: A working paper*. TOEFL Monograph Series Report, 16. Educational Testing Service.

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

Jenkins, J. (2007). *English as a Lingua Franca: Attitude and identity*. Oxford University Press.

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

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., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a Lingua Franca. *Language Teaching*, *44*(3), 281-315. https://doi.org/10.1017/ S0261444811000115

*Jenkins, J., & C. Leung (2017). Assessing English as a Lingua Franca. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment: Encyclopedia of language and education* (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

Jung, H., & Crookes, G. V. (2025). Enacting a Global Englishes and critical pedagogy integrated curriculum in Korean high school English as a foreign language classroom. *TESOL Quarterly*, *59*(1), 163-196. https://doi.org/10.1002/tesq.3332

*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, Y. H. (2009). A G-theory analysis of rater effect in ESL speaking assessment. *Applied Linguistics*, *30*(3), 435-440. https://doi.org/10.1093/applin/amp035

∗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. (2007). *World Englishes: Implications for international communication and English language teaching*. Cambridge University Press.

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

∗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.

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

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.

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture, and Activity*, *7*(4), 273-290. https://doi.org/10.1207/S15327884MCA0704_03

Leung, C. (2013). The "social" in English Language Teaching: abstracted norms versus situated enactments. *Journal of English as a Lingua Franca*, *2*(2), 283-313. https://doi.org/10.1515/jelf-2013-0016

∗Leung, C. (2019). 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. (2023). 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

∗Lopez, A. A., Turkan, S., & Guzman-Orth, D. (2017). Assessing multilingual competence. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment: Encyclopedia of language and education* (pp. 91-102). Springer.

Lowenberg, P. H. (2002). Assessing English proficiency in the expanding circle. *World Englishes*, *21*(3), 431-435. https://doi.org/10.1111/1467-971X.00261

Macaro, E. (2019). Systematic reviews in applied linguistics. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in Applied Linguistics* (pp. 230-239). Routledge.

∗Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: implications for ESL assessment. *TESOL Quarterly*, *36*, 173-190. https://doi.org/10.2307/3588329

∗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., Perez-Llantada, C., & Swales, J. M. (2010). Academic Englishes: A standardised knowledge? In A. Kirkpatrick (Ed.), *The Routledge handbook of World Englishes* (pp. 634-652). Routledge.

∗Matsuura, H., Chiba, R., Mahoney, S., & Rilling, S. (2016). Accent and speech rate effects in English as a Lingua Franca. *System*, *46*, 143-150. http://doi.org/10.1016/j.system.2014.07.015

May, S. (2014). Contesting public monolingualism and diglossia: Rethinking political theory and language policy for a multilingual world. *Language Policy, 13,* 371-393. https://doi.org/10.1007/s10993-014-9327-x

McKay, S. L. (2006). *Researching second language classrooms*. Routledge.

∗McKay, S. L. (2011). English as an international lingua franca pedagogy. In E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (Vol. II, pp. 122-139). Routledge.

∗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

∗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

∗Newbold, D. (2018). "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

∗Nishizawa, H. (2023). Construct validity and fairness of an operational listening test with World Englishes. *Language Testing, 40*(3), 493-520. https://doi.org/10.1177/02655322221137869

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417-528. https://doi.org/10.1111/0023-8333.00136

Nzabonimpa, J. P. (2018). Quantitizing and qualitizing (im-) possibilities in mixed methods research. *Methodological Innovations*, *11*(2). http://doi.org/2059799118789021

∗Ockey G. J., French R. (2016). From one to multiple accents on a test of L2 listening comprehension. *Applied Linguistics*, *37*(5), 693-715. https://doi.org/10.1093/applin/amu060

∗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.

Onwuegbuzie, A. J., & Leech, N. L. (2019). On Qualitizing. *International Journal of Multiple Research Approaches*, *11*(2), 98-131. https://doi.org/10.29034/ijmra.v11n2editorial2

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Moher, D. (2021). Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. *Journal of Clinical Epidemiology*, *134*, 103-112. https://doi.org/10.1016/j.jclinepi.2021.02.003

Pan, L., Li, H., Seargeant, P., & Block, D. (2025). Exploring preservice teachers' translanguaging practices and perceptions in teacher training: A global Englishes perspective. *TESOL Quarterly, 59*(1), 103-135. https://doi.org/10.1002/tesq.3360

∗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.

Plano Clark, V. L., Huddleston-Casas, C. A., Churchill, S. L., O'Neil Green, D., & Garrett, A. L. (2008). Mixed methods approach in family science research. *Journal of Family Issues*, *29*(11), 1543-1566.

Rose, H., & McKinley, J. (2025). Global Englishes and TESOL: An editorial introduction to innovating research and practice. *TESOL Quarterly*, *59*(1), 5-23. https://doi.org/10.1002/tesq.3373

Rose, H. L. & Syrbe, M. (2018). Assessment practices in teaching English as an International Language. In J. I., Liontas (Ed.), *TESOL encyclopedia of English language teaching: Teaching English as an international language* (pp. 1-12). Wiley-Blackwell.

Sandelowski, M., & Barroso, J. (2006). *Handbook for synthesizing qualitative research*. Springer.

∗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

Seidlhofer, B. (2011). Conceptualizing 'English' for a multilingual Europe. In A. Wilton (Eds.), *English in Europe today: sociocultural and educational perspectives* (pp. 133-146). John Benjamins.

Seidlhofer, B. (2018). Standard English and the dynamics of ELF variation. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook on English as a Lingua Franca* (pp. 85-100). Routledge.

Sewell, A. (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 (2019). 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

Shu, D., Yang, S., & Sato, M. (2024). Marrying collective wisdom: Researcher practitioner collaboration in developing ELT textbooks. *Applied Linguistics*, *45*(1), 41-64. https://doi.org/10.1093/applin/amac074

Sifakis, N., & Kordia, S. (2025). Situating reflection within ELF awareness: A practical and evaluative orientation. *TESOL Quarterly*, *59*(1), 333-343. https://doi.org/10.1002/tesq.3284

Tashakkori, A., & Teddlie, C. (2003). Issues and dilemmas in teaching research methods courses in social and behavioral sciences: US perspective. *International Journal of Social Research Methodology*, *6*(1), 61-77.

∗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. (1997). EIL, ESL, EFL: Global issues and local interests. *World Englishes*, *16*(1), 135-146. https://doi.org/10.1111/1467-971X.00054

Widdowson, H. G. (2012). ELF and the inconvenience of established concepts. *Journal of English as a Lingua Franca*, *1*(1), 5-26. https://doi.org/10.1515/jelf-2012-0002

∗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

Zari Saeedi received her PhD from the British University of Trinity College (Constituent College of Cambridge) and is an Associate Professor of Allameh Tabataba'i University. She has taught various BA, MA, and PhD courses, taking part in different national/international conferences, presenting/publishing papers on a range of topics including educational neuro/psycholinguistics and brain functioning, Cognitive language learning and in particular, Brain Based Language Learning (BBLL), culture, technology-assisted language learning, and Role and Reference Grammar Theory of linguistics. She has done some Brain scanning projects in ATU and her published paper and book in Benjamins and Equinox publications, the dictionary of virtual education with a focus on Computer Assisted Language Learning (CALL), Brain-Based CALL paper in the International Journal of Educational Technology in Higher Education, Artificial Intelligence paper published by Routledge, and another research article on Content Development for ESP via Storyline Software published by Elsevier are among some recent publications. ORCID: https://orcid.org/0000-0003-2165-8891

Zia Tajeddin is Professor of Applied Linguistics at Tarbiat Modares University, Tehran, Iran. He co-edits two international journals: *Applied Pragmatics* (John Benjamins) and *Second Language Teacher Education* (University of Toronto Press). He is the co-editor of the Springer book series *Studies in Language Teacher Education*. His research interests center on teacher education, L2 pragmatics, and EIL/ELF pedagogy. His recently published books include *Lessons from Good Language Teachers* (with Carol Griffiths, Cambridge University Press, 2020), *Pragmatics Pedagogy in English as an International Language* (with Minoo Alemi, Routledge, 2021), *Teacher Reflection: Policies, Practices and Impacts* (with Atsuko Watanabe, Multilingual Matters, 2022), *Language Education Programs: Perspectives on Policies and Practices* (with Carol Griffiths, Springer, 2023), *Language Teacher Identity Tensions: Nexus of Agency, Emotion, and Investment* (with Bedrettin Yazan, Routledge, 2024), and *Handbook of Language Teacher Education: Critical Review and Research Synthesis* (with Thomas S. C. Farrell, Springer, 2025). ORCID: https://orcid.org/0000-0002-0430-6408

Fereshteh Tadayon received her PhD in TEFL from Allameh Tabataba'i University, Tehran, Iran. She lectures at Shahid Bahonar and Vali-e-Asr universities, Kerman, Iran. Her research interests include English as a Lingua Franca (ELF) assessment, cognitive processes in language testing, teacher-mediated peer feedback, and academic writing pedagogy. She has conducted research on ELF in high-stakes and low-stakes exam contexts, cognitive diagnostic approaches to feedback and assessment, teacher education, and comparative studies in language testing. Her research has been published in peer-reviewed journals and as book chapters and presented at international academic conferences. ORCID: https://orcid.org/0000-0002-3725-5766