Revista EDUCATECIENCIA Journal
E-ISSN: 2683-2836
CD-ISSN: 2007-6347
Tepic, Nayarit; México

Vol. 34 Núm. 2 (2026)
Julio-diciembre 2026
https://doi.org/10.58299/edutec.v34i2.464
Pp. 66-96

Rethinking standardized English assessment in public higher education: a contextualized and discipline-specific framework

Repensando la evaluación estandarizada del inglés en educación superior pública: un marco contextualizado y disciplinar

Benjamín Topete López
Universidad Autónoma de Nayarit
Unidad Académica de Educación y Humanidades
Tepic, Nayarit, México
Benjamin.topete@uan.edu.mx.
https://orcid.org/0009-0008-2289-7724

Marcela García Ramos
Universidad Autónoma de Nayarit
Unidad Académica de Educación y Humanidades
Tepic, Nayarit, México
Marcela.garcia@uan.edu.mx
https://orcid.org/0000-0003-3618-4799

Martha Xolyanetzin Rodríguez Villarreal
Universidad Autónoma de Nayarit
Unidad Académica de Educación y Humanidades
Tepic, Nayarit, México
xolyanetzinrodriguez@uan.edu.mx
https://orcid.org/0009-0004-3876-4310

Nadia Sarahi Uribe Olivares
Universidad Autónoma de Nayarit
Secretaria Académica
Tepic, Nayarit, México
Nadia.uribe@uan.edu.mx.
https://orcid.org/0000-0003-2521-9634

Abstract

Research problem: the limited representation of linguistic practices in English assessments. **General objective**: to examine the relationship between standardized assessment and discipline-specific linguistic needs at a Mexican public university, in order to identify the implications of a contextualized model. **Research classification**: a qualitative, interpretive, descriptive, and cross-sectional case study was conducted. **Population and sample**: purposive sampling involving 25 students and 4 faculty members from Nursing, Applied Linguistics, and Early Childhood Education. **Research instruments**: a 22-item Likert questionnaire, written reflections, and semi-structured interviews were used. **Instrument validation**: the questionnaire was designed specifically for this study and did not undergo formal validation; its design was guided by its alignment with the theoretical framework and research questions. **Data analysis**: a combination of descriptive trends and thematic analysis. **Results**: findings highlighted the academic value of English, diverse linguistic trajectories, and a perceived gap between certification and disciplinary practices. **Conclusions**: the study concludes by proposing the MCEAF as a multi-level assessment architecture.

Keywords: educational policy; educational testing; higher education; language instruction; language policy.

Resumen

Problema de investigación: representación limitada de prácticas lingüísticas por parte de las evaluaciones de inglés. **Objetivo general**: examinar, en una universidad pública mexicana, la relación entre evaluación estandarizada y necesidades lingüísticas disciplinares, para identificar implicaciones de un modelo contextualizado. **Clasificación de la investigación**: se desarrolló un estudio de caso cualitativo, interpretativo, descriptivo y transversal. **Población y muestra**: con muestreo intencional de 25 estudiantes y 4 docentes de Enfermería, Lingüística Aplicada y Educación Infantil. **Instrumentos de investigación**: se utilizaron un cuestionario Likert de 22 ítems, reflexiones escritas y entrevistas semiestructuradas. **Validación de instrumentos de investigación**: el cuestionario se diseñó para esta investigación y no contó con validación; su alineación con el marco teórico y las preguntas de investigación orientó su diseño. **Análisis de datos**: combinó tendencias descriptivas y análisis temático. **Resultados**: evidenciaron el valor académico del inglés, trayectorias lingüísticas heterogéneas y una brecha percibida entre certificación y prácticas disciplinares. **Conclusiones**: se concluye proponiendo el MCEAF como arquitectura de evaluación multinivel.

Palabras clave: control escolar; enseñanza de idiomas; enseñanza superior; política educacional; política lingüística.

Publicación de acceso abierto bajo la licencia Creative Commons 4.0 Atribución-No Comercial (CC BY-NC 4.0 Internacional).

Recibido: agosto 14, 2026 * Aprobado: septiembre 30, 2026 * Publicado: octubre 06, 2026

Introduction

Research problem

The internationalization of higher education has repositioned English from a conventional foreign-language subject toward a strategic resource for academic mobility, academic collaboration, postgraduate study, and participation in global knowledge production. Public universities may use standardized examinations to obtain comparable evidence of English proficiency, particularly where certification and internationalization are institutional priorities. Recent bibliometric evidence indicates sustained growth in English language assessment research and increasing attention to more holistic and inclusive approaches (Yang & Wang, 2025).

The institutional usefulness of standardized assessment does not, however, resolve whether evidence from a general proficiency examination represents the language practices students must perform in their academic programs. Assessment research distinguishes the quality and validity of measurement from questions concerning the usefulness of assessment evidence, including authenticity, contextual relevance, and intended consequences (Bachman & Palmer, 2010; Douglas, 2000; Yang & Wang, 2025). In higher education, this distinction becomes particularly relevant when students must interpret scholarly literature, write disciplinary texts, present specialized information, or communicate in professional settings. The central problem addressed in this study is therefore the limited representation of discipline-specific academic language practices within a general assessment architecture.

This problem is especially salient in public higher education, where students may enter university with unequal access to English instruction and different multilingual or transnational learning trajectories (Piller & Bodis, 2024). Such differences can result in profiles in which general communicative experience does not necessarily correspond to academic literacy, specialized discourse knowledge, or formal academic writing ability. A single proficiency score may consequently conceal important variation in how students can use English for university study. From an equity perspective, recent research has also examined how language proficiency requirements can interact with unequal linguistic histories and language ideologies in higher education.

Disciplinary diversity adds another dimension to the problem because academic communities organize communication through different genres, terminology, rhetorical practices, and professional expectations. Research on English for Specific Academic Purposes

emphasizes aligning language provision with disciplinary learning and future professional practices, including collaboration between language specialists and discipline teachers (Kim & Suh, 2024; Yung & Fong, 2024). These principles suggest that assessment evidence should also attend to the communicative practices students are expected to perform in their fields. An ELF perspective further draws attention to intelligibility, adaptability, and negotiation of meaning within multilingual academic environments (Jenkins, 2015; Saeedi et al., 2025).

The institutional issue is therefore not simply the selection of a particular standardized test, but the way assessment evidence is interpreted and connected with academic needs. Recent scholarship on language assessment literacy emphasizes that university personnel require knowledge and principles for designing, interpreting, and making informed decisions from language assessment data (Vogt et al., 2024; Lam & Gayton, 2025; Cui et al., 2022). Research conducted in the Mexican higher education context likewise identifies assessment literacy and program level assessment as important institutional concerns (Banitz, 2022). The present study responds to this problem by examining how standardized assessment can coexist with diagnostic and discipline sensitive evidence within a public university context. Previous research at the participating university has likewise examined English as a lingua franca in academic communication and identified challenges related to institutional support and coordination (Topete López et al., 2025).

Objective

To examine, at a Mexican public university, how standardized English language assessment relates to discipline specific linguistic needs by analyzing students' and faculty members' perceptions of the academic uses of English and of existing assessment practices, in order to identify pedagogical and institutional implications for a contextualized, multilevel English language assessment framework.

Specific objectives

- To examine students' and faculty members' perceptions regarding the role of English in accessing disciplinary knowledge and supporting higher education internationalization.
- To identify the tensions that students and faculty members perceive between standardized English language assessment and the authentic linguistic demands of their academic disciplines.

- To analyze how students' diverse linguistic and educational trajectories shape their academic participation and perceptions of their English language competence.
- To explore the perceived relevance of diagnostic and contextualized English language assessment informed by ESP and ELF principles and its potential implications for institutional English language policy.

Research questions

RQ1. How do students and faculty members perceive the role of English in accessing disciplinary knowledge and supporting higher education internationalization?

RQ2. What tensions do students and faculty members identify between standardized English language assessment and the authentic linguistic demands of their academic disciplines?

RQ3. How do students' diverse linguistic and educational trajectories shape their academic participation and perceptions of their English language competence?

RQ4. How do students and faculty members perceive the relevance of diagnostic and contextualized English language assessment informed by ESP and ELF principles, and what institutional policy implications do they identify?

Hypothesis

It is hypothesized that the perceived relevance of diagnostic and discipline-specific contextualized English language assessment among students and faculty can provide a basis for developing institutional English language policies that retain the comparability of standardized certification while incorporating assessment practices aligned with the linguistic demands of academic and professional fields. Given the qualitative and interpretive design, this hypothesis is treated as a guiding proposition examined through participants' perceptions and triangulated evidence rather than as a statistically tested causal hypothesis.

Theoretical framework

The theoretical framework integrates language assessment theory, English as a Lingua Franca (ELF), English for Specific Purposes (ESP), and institutional language policy. These perspectives provide complementary lenses for interpreting standardized assessment, disciplinary language use, and institutional responses to linguistic diversity.

Language assessment in higher education

In higher education, language assessment can serve multiple institutional functions, including certification, academic mobility, progression decisions, student support, and quality processes. Standardized examinations are especially useful when institutions require evidence that can be compared across students and settings. Their usefulness, however, depends on the relationship between the intended use of scores and the construct represented by assessment tasks. Recent scholarship continues to emphasize the importance of contextual relevance, authenticity, diagnostic value, and educational consequences alongside measurement quality (Aryadoust, 2023; Yang & Wang, 2025). Formative assessment perspectives also reinforce the value of using assessment information to support learning and instructional improvement (Black & Wiliam, 1998).

Bachman and Palmer (2010) conceptualize useful assessment through validity, reliability, authenticity, interactiveness, impact, and practicality, while Douglas (2000) emphasizes Target Language Use (TLU) situations. Contemporary work on language assessment literacy further highlights the importance of stakeholders' capacity to interpret assessment information and make informed decisions from it (Vogt et al., 2024; Cui et al., 2022). Recent work on formative and ELF-informed assessment also emphasizes the educational use of assessment evidence and communicative effectiveness (Dewi & Slamet, 2025; Caprario, 2025). In the present study, these principles are extended to disciplinary contexts: if students are expected to interpret academic literature, write disciplinary texts, or communicate specialized information, those practices need a visible place in the evidence used to understand academic language competence.

The framework consequently differentiates three institutional questions: what level of English proficiency students can demonstrate through internationally comparable evidence; what linguistic support they may require locally; and how effectively they can perform the academic language practices of their disciplines. This functional differentiation provides the conceptual bridge to the MCEAF and avoids treating one score as a complete representation of academic language competence.

English as a Lingua Franca (ELF)

ELF provides a lens for understanding English as a shared resource among multilingual users. Academic communication in multilingual settings is not defined exclusively by proximity to a native-speaker norm; it also involves intelligibility, flexibility, strategic adaptation, and

negotiation of meaning (Adriansen et al., 2023; Jenkins, 2015; Seidlhofer, 2011; Saeedi et al., 2025). ELFA-aware EAP research likewise emphasizes aligning academic English provision with multilingual university communication (Liu, 2024).

For assessment purposes, this perspective shifts attention from whether students reproduce a particular linguistic norm toward whether they can accomplish meaningful academic communicative tasks. Recent empirical work has explored ELF-oriented assessment of spoken communication and communication-strategy development in university contexts (Caprario, 2025; Dewi & Slamet, 2025). Research on disciplinary writing in EMI contexts further highlights the importance of making language practices and assessment tasks sensitive to disciplinary context (Kim & Suh, 2024). In the present study, ELF therefore informs the interpretation of participants' emphasis on communication, access to disciplinary knowledge, and academic participation across linguistic backgrounds.

Structural accuracy is not treated as irrelevant. Rather, grammatical and lexical knowledge is considered alongside learners' capacity to use English intelligibly and effectively in the communicative situations that matter in higher education. This broader view helps explain why a general proficiency score may provide useful evidence of overall language ability while still offering only partial evidence of discipline-specific communicative competence.

English for Specific Purposes (ESP) and discipline-specific language assessment

ESP begins from the communicative requirements of particular academic and professional communities. Recent ESAP research shows the value of connecting language instruction with disciplinary learning and future professional practices, including collaboration between language specialists and subject teachers (Kim & Suh, 2024; Yung & Fong, 2024). Academic writing and disciplinary discourse also vary across contexts, reinforcing the need for discipline-sensitive assessment (Kim & Suh, 2024). In university settings, students therefore encounter disciplinary genres, terminology, rhetorical practices, and interactional expectations that vary across fields.

The concept of authentic TLU is particularly useful because it connects assessment tasks with the situations in which students are expected to use English (Douglas, 2000). Recent ESAP research also shows that disciplinary learning and future professional practices can inform course design and encourage collaboration between language specialists and discipline teachers (Kim & Suh, 2024; Yung & Fong, 2024). Within the proposed framework, discipline-sensitive tasks therefore complement rather than replace international certification.

The contribution of ESP to the MCEAF is consequently operational as well as theoretical: discipline-specific needs inform the selection of assessment tasks, the interpretation of performance, and the connection between assessment results, curriculum, and academic support. This approach also avoids reducing contextualization to technical vocabulary alone by emphasizing genres, purposes, interactional demands, and disciplinary participation.

Language policy and educational equity

Language assessment also has an institutional policy dimension because decisions about what is assessed can influence access to academic opportunities, recognition of linguistic resources, and the distribution of language support. Recent research conceptualizes language policy and assessment as institutional practices that interact with linguistic diversity, language ideologies, and access in higher education (Fan et al., 2022; Piller & Bodis, 2024; Shirahata & Lahti, 2023). Research in Mexico likewise documents the absence of uniform language assessment criteria and identifies concerns among university language teachers regarding assessment practices (Banitz, 2022).

The policy dimension is particularly relevant in public universities serving heterogeneous student populations. When assessment assumes comparable prior access to English, a uniform requirement may obscure differences in students' previous educational histories. A contextualized diagnostic component can instead make these differences visible and provide information for differentiated institutional support. Recent EAP research similarly argues that equity-oriented language support requires coordination across individual, departmental, and institutional levels (Bakogiannis & Papavasiliou, 2025, 2026).

From an equity perspective, contextualized assessment is therefore not proposed simply as a more authentic testing format. Its institutional value lies in combining evidence of current abilities with information that can guide differentiated support, curriculum planning, faculty development, and internationalization. The MCEAF incorporates this policy dimension by connecting measurement with institutional action.

Conceptual framework: The Multilevel Contextualized English Assessment Framework (MCEAF)

The MCEAF integrates four conceptual foundations. Language assessment theory contributes validity, authenticity, and purposeful use of evidence; ELF contributes a multilingual view of communicative effectiveness; ESP connects assessment with disciplinary practices; and

language policy places assessment within institutional decisions concerning equity and internationalization. Recent work on diagnostic assessment, ELF assessment, and inclusive EAP provides converging support for these principles (Devos et al., 2024; Saeedi et al., 2025; Bakogiannis & Papavasiliou, 2025, 2026).

The distinctive feature of the MCEAF is the functional differentiation of assessment levels. International examinations provide comparable evidence of general proficiency; institutional diagnostic assessment identifies linguistic profiles and support needs; and discipline-specific assessment examines performance on authentic academic tasks. These components address different purposes and therefore coexist within a coordinated architecture. Their relationship is consistent with recent developments toward more holistic, multilingual, and context-sensitive approaches to language assessment (Yang & Wang, 2025).

The architecture allows universities to retain the external value of recognized certification while generating locally meaningful evidence for teaching, curriculum planning, academic support, and disciplinary development. The contribution of the framework lies in institutionalizing these functions within one assessment system rather than treating standardized and contextualized assessment as mutually exclusive alternatives.

Table 1 summarizes the theoretical perspectives, their contributions to the study, and their implications for language assessment.

Table 1. Theoretical Framework Underpinning the Study.

Theoretical perspective	Contribution to the study	Implications for language assessment
English as a Lingua Franca (ELF)	Functional use of English in multilingual contexts	Prioritize intelligibility and communicative effectiveness
English for Specific Purposes (ESP)	Discipline-specific language use	Develop contextualized assessment models
Language Policy	Educational equity and institutional inclusion	Avoid homogeneous assessment assumptions
Formative Assessment	Diagnostic assessment and learning support (Black & Wiliam, 1998)	Promote the pedagogical use of assessment results (Black & Wiliam, 1998)

Source: developed by the authors.

Method

Classification and design of the research

This study adopted a qualitative, interpretive, descriptive, cross-sectional case-study design at a public university in Mexico. The qualitative case-study approach was selected because it permits an in-depth examination of a bounded phenomenon in its institutional context and is well suited to integrating participants' accounts with contextual evidence (Creswell & Poth, 2024). Questionnaire responses were incorporated as supplementary descriptive evidence rather than as inferential quantitative data. The design therefore prioritizes contextual understanding, convergence across sources, and interpretation of how English assessment is experienced within academic programs.

Study population and sample

The study population comprised undergraduate students and faculty members connected with Nursing, Applied Linguistics, and Early Childhood Education at the participating university. The purposive sample included 25 undergraduate students and four faculty members. Students were selected according to their enrollment in the participating programs, while faculty members were selected because of their involvement in teaching, academic support, or assessment. Participation was voluntary and based on direct experience with English learning, teaching, academic support, or language assessment.

Table 2 presents the distribution of student and faculty participants across the three academic programs.

Table 2. Participant distribution.

Academic program	Students	Faculty
Nursing	15	1
Applied Linguistics	5	2
Early Childhood Education	5	1
Total	25	4

Source: developed by the authors.

Participation was entirely voluntary, and confidentiality and participant privacy were protected throughout the study. All participants provided informed consent before data collection.

Research instruments

Three complementary instruments/procedures were used. A researcher-developed 22-item Likert-scale questionnaire documented students' perceptions; open-ended written reflections elicited accounts of English learning experiences and academic language demands; and a

semi-structured interview guide elicited faculty perspectives on assessment practices and discipline-specific language needs. The three sources were designed for complementary purposes so that descriptive questionnaire tendencies could be interpreted together with qualitative accounts rather than treated as independent quantitative evidence.

Likert-scale questionnaire

The questionnaire used a five-point response scale from 1 (Strongly Disagree) to 5 (Strongly Agree). It addressed six conceptual domains derived from the theoretical framework and the four research questions. The instrument was used exclusively as supplementary descriptive evidence and was not intended to measure English proficiency psychometrically or to generate inferential estimates.

Table 3 maps the questionnaire domains and item ranges to their intended analytic purposes.

Table 3. Questionnaire domains and item distribution.

Domain	Items	Purpose
Previous English-learning experiences	1	Document prior exposure and learning trajectories
Motivation and academic/professional relevance	2–4	Examine reasons for learning English and perceived relevance
Academic English use and needs	5–9	Identify academic practices and perceived language needs
Standardized and institutional assessment	10–14	Explore perceptions of current assessment practices
ELF and communicative orientation	15–18	Examine views on intelligibility, communication, and English varieties
Institutional support, discipline-specific assessment, and internationalization	19–22	Examine perceived institutional responses and internationalization needs

Source: developed by the authors.

Students' written reflections

Students were invited to produce brief written reflections describing their previous English-learning experiences, academic uses of English, linguistic challenges encountered during their studies, and institutional support they considered necessary. The reflections were treated as qualitative accounts and analyzed together with interview data and questionnaire tendencies.

Semi-structured interviews

Semi-structured interviews were conducted with faculty members to explore current English assessment practices, discipline-specific linguistic competencies, perceptions of standardized English examinations, pedagogical adaptations to students' proficiency levels, and the

relationship between English assessment and institutional internationalization. The semi-structured format provided consistency across interviews while allowing participants to elaborate on contextual experiences.

Research variables (analytic categories)

Because the study is qualitative and interpretive, statistical variables were not modeled or tested. The study therefore reports four analytic categories that organize the evidence: academic role of English, perceived relationship between standardized assessment and disciplinary needs, linguistic trajectories, and relevance of contextualized assessment and institutional support. These categories function as analytical categories rather than psychometric variables.

Operationalization of variables (analytic categories)

Table 4 operationalizes the analytic categories through observable indicators and the corresponding data sources.

Table 4. Operationalization of analytic categories.

Analytic category	Operational focus	Indicators	Data source
Academic role of English	How English is used in study and internationalization	Academic reading; access to literature; presentations; international communication	Questionnaire; reflections; interviews
Standardized assessment and disciplinary needs	Perceived fit between general tests and authentic academic demands	Reading; academic writing; specialized presentations; professional communication	Questionnaire; reflections; interviews
Linguistic trajectories	Variation in prior access and English-learning experiences	Formal/informal learning; bilingual or transnational exposure; confidence and participation	Questionnaire; reflections
Contextualized assessment and institutional support	Perceived relevance of diagnostic, discipline-specific, and policy responses	Diagnostic assessment; discipline-specific English; language support; institutional policy	Questionnaire; reflections; interviews

Source: developed by the authors.

Instrument validation

The 22-item questionnaire was researcher-developed from the theoretical framework and the four research questions, and its item domains were explicitly mapped to the study's analytic purposes (table 3). No formal psychometric validation procedure, expert-judgment panel, or pilot-study validation is documented in the available study record. Accordingly, the questionnaire is not presented as a validated measure of English proficiency; it is reported only as supplementary descriptive evidence. The written reflections and interview guide were treated as

qualitative elicitation procedures, with credibility supported through triangulation across data sources. This limitation is carried forward to the implications for future research, where expert review and pilot testing are recommended before any future psychometric use. Argument-based validation scholarship also emphasizes linking validation evidence to intended score use and construct interpretation (Dai et al., 2024).

Data collection and analysis

Data collection took place at the participating public university in Mexico between February and April 2026. The student questionnaire was administered electronically through a digital platform, while faculty interviews were conducted individually in face-to-face or online formats according to participant availability. Interviews lasted approximately 30 minutes and were conducted in English and Spanish. Participation was voluntary, confidentiality and anonymity were maintained, and the collected information was used exclusively for academic research purposes.

Table 5 summarizes the techniques, participants, purposes, and types of data used in the study.

Table 5. Data collection techniques.

Technique	Participants	Purpose	Type of data
Likert-scale questionnaire	Undergraduate students	Explore perceptions regarding English language assessment	Descriptive questionnaire evidence
Open-ended written reflections	Undergraduate students	Examine English-learning experiences and academic language demands	Qualitative data
Semi-structured interviews	University faculty	Explore assessment practices and discipline-specific linguistic needs	Qualitative data

Source: developed by the authors.

Questionnaire responses were examined descriptively to identify response tendencies across the five-point scale. Because individual numerical responses were not retained in the available analytical record, the study does not report percentages, means, or inferential statistics. Qualitative data from written reflections and faculty interviews were analyzed thematically following the six-phase approach associated with Braun and Clarke (2021), with attention to contemporary guidance on transparent reporting of reflexive thematic analysis (Braun & Clarke, 2024): familiarization with the data; initial coding; identification of emerging themes; review and refinement of thematic categories; definition and naming of themes; and interpretation in relation to the research questions and theoretical framework.

Methodological triangulation was used to compare convergent and divergent patterns across the questionnaire, reflections, and interviews. The analysis was iterative, with interpretations repeatedly checked against the original data and discussed among the research team. Such explicit attention to transparency and methodological consistency is also emphasized in recent qualitative-methodological literature (Malmqvist et al., 2024).

Trustworthiness

Trustworthiness was strengthened through methodological triangulation, a detailed account of the institutional context and participants, transparent reporting of the research process, and systematic comparison of evidence across data sources. The coding process involved repeated engagement with the dataset, grouping related codes into broader themes, and reviewing themes against the original evidence and research questions. The research team discussed emerging interpretations and compared evidence from different sources before finalizing themes and implications. These procedures were used to strengthen transparency and methodological coherence, consistent with recent guidance on qualitative quality and reflexive reporting (Braun & Clarke, 2024; Malmqvist et al., 2024).

Researcher reflexivity

Because the study was conducted within the researchers' institutional context, the research team recognized the potential influence of insider positioning on data collection and interpretation. The researchers maintained reflexive attention to assumptions regarding English language assessment and disciplinary language needs. Throughout the analytical process, interpretive decisions were discussed among team members, and evidence from the different data sources was compared before themes and implications were finalized. This reflexive process was intended to strengthen transparency and reduce the influence of individual assumptions on interpretation.

Results and discussions

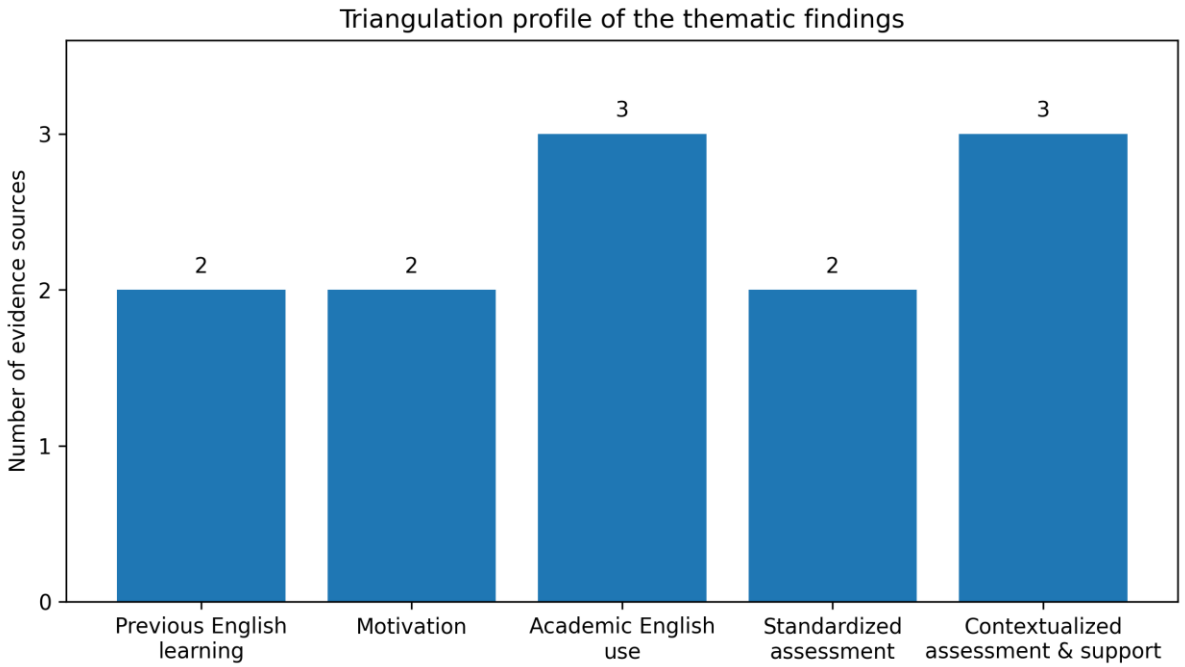
Results

The findings are organized in the order of the research questions. Four overarching dimensions emerged from the thematic analysis of students' written reflections and faculty interviews and were interpreted alongside descriptive tendencies in the questionnaire. Because item-level numerical responses were not preserved, the results report response tendencies and triangulated qualitative evidence rather than percentages or inferential statistics.

The questionnaire showed predominantly positive response tendencies regarding the academic and professional relevance of English. Students recognized English as important for academic development, professional opportunities, and access to academic texts. Responses also showed positive orientations toward diagnostic assessment, discipline-related English courses, and the inclusion of English in institutional internationalization efforts, whereas perceptions of existing standardized assessment were more mixed. The evidence-source profile in Figure 1 summarizes the triangulation base for the principal thematic findings.

Figure 1 shows the number of independent evidence sources contributing to each thematic finding.

Figure 1. Triangulation profile of the thematic findings.



Source: developed by the authors.

Note. Bars indicate the number of independent evidence sources contributing to each thematic finding (questionnaire, written reflections, and/or faculty interviews); they do not represent participant frequencies or percentages.

RQ1. Academic role of English

Students' questionnaire responses, written reflections, and faculty interview accounts indicate that English functions primarily as a means of accessing disciplinary knowledge and participating in international academic environments. Academic reading is a recurrent demand, particularly when students are required to interpret scholarly concepts, prepare presentations, or locate specialized information. One student explained that understanding the concepts in English was necessary to complete a compare-and-contrast presentation, while another stated: “I can understand English in everyday situations, but academic texts are different.” These accounts provide direct evidence of a distinction between everyday comprehension and academic language processing.

The questionnaire also showed a communicative and functional orientation toward English. Items 15–18 addressed intelligibility, international communication, and exposure to different English varieties, and the response pattern was predominantly supportive of a functional orientation. The written reflections reinforce this broader view through students' descriptions of English as useful for academic presentations, scholarly reading, information searches, and participation in international environments. These converging sources provide empirical grounding for interpreting English through an ELF-informed lens rather than solely through general proficiency categories.

RQ2. Tensions between standardized assessment and disciplinary demands

A recurrent finding is the perceived mismatch between standardized English examinations and the linguistic practices students encounter in their degree programs. Across the written reflections and faculty interviews, participants identified academic and disciplinary reading, production of academic texts, presentations of specialized information, and professional or academic communication as important practices that are only partially represented in general proficiency examinations. One student reported difficulty understanding academic texts even when individual words were familiar, providing concrete evidence of the distinction between lexical recognition and disciplinary meaning-making.

Faculty interview data further indicate that existing assessment is experienced largely through institutional compliance and certification purposes, while attention to students' ability to use English in academic and professional contexts is more limited. One faculty participant emphasized the need for a more consistent institutional framework connecting English teaching, assessment, and the purposes of individual academic programs. Because this observation

reflects one participant's perspective, it is treated as illustrative rather than representative of all faculty members. The convergence with student reflections nevertheless strengthens the interpretation of a perceived construct gap.

RQ3. Linguistic and educational trajectories

The evidence reveals substantial variation in students' previous exposure to English. The questionnaire and written reflections describe trajectories that include informal learning, sustained formal instruction, bilingual or transnational educational experiences, and more limited prior access to English. One student described learning through music, later combining self-directed learning with formal classes, while also reporting that previous instruction was not always effective for real-life language use. This account illustrates how prior learning experiences can shape the relationship between formal language knowledge, confidence, and functional use.

The reflections also show that students use English for scholarly reading, presentations, writing assignments, information searching, and translation or interpretation activities, but experience difficulties with academic vocabulary, reading comprehension, contextual interpretation, and sustained oral fluency. The convergence of questionnaire and reflection data supports the conclusion that a single proficiency score may conceal meaningful differences in students' academic language needs. The result is therefore interpreted as a rationale for institutional diagnostic assessment rather than as evidence of a deficit associated with any particular student group.

RQ4. Relevance of contextualized assessment and institutional support

Evidence from the questionnaire, written reflections, and faculty interviews indicates a recurring perceived need for a stronger institutional approach to English language education. Participants' responses include support for discipline specific English assessment, academic language support programs, diagnostic procedures, and a more explicit institutional language policy. These findings are supported by the triangulation pattern summarized in Table 7, where the need for contextualized assessment and institutional support is represented across all three data sources.

Faculty accounts also suggest that English instruction could be more effectively connected to the academic practices of individual degree programs through ESP- and EAP- oriented approaches. The finding is therefore not interpreted as a rejection of standardized certification.

Rather, participants' accounts support a complementary model in which international certification provides general proficiency evidence, institutional diagnostic assessment identifies support needs, and discipline specific assessment examines performance on authentic academic tasks.

Taken together, these findings provide qualitative support for the guiding proposition that the perceived relevance of contextualized and discipline-sensitive assessment may provide a basis for institutional English language policy development. This evidence does not demonstrate a causal effect of the MCEAF, but it identifies a plausible institutional relationship that warrants empirical examination through future implementation studies.

Table 6 synthesizes the major thematic dimensions, principal findings, and corresponding educational implications.

Table 6. Major themes emerging from the thematic analysis.

Thematic dimension	Main finding	Educational implication
Academic use of English	Academic and disciplinary reading is a predominant use of English.	Greater integration of ESP- and EAP-oriented instruction.
Standardized assessment	Existing assessments are perceived as only partially representing academic language needs.	Development of contextualized and complementary assessment evidence.
Linguistic trajectories	Students display considerable differences in previous English-learning experiences.	Implementation of institutional diagnostic assessment.
Institutional language policy	Evidence indicates perceived need for greater institutional coordination.	Development of a comprehensive university language policy.

Source: developed by the authors.

Discussion

RQ1 and the academic function of English

The RQ1 findings align with ELF and ELFA-aware EAP perspectives that treat English as a functional resource for participation in multilingual academic environments (Liu, 2024; Saeedi et al., 2025; Dewi & Slamet, 2025). Students' descriptions of academic reading, presentations, and information access are consistent with research emphasizing the role of English in disciplinary knowledge construction and communication. Recent ELF research also highlights communicative adaptability and intelligibility as relevant dimensions of multilingual academic communication (Saeedi et al., 2025; Dewi & Slamet, 2025). The present findings therefore

extend the functional interpretation of English to institutional assessment, where evidence needs to represent the uses for which students require the language.

RQ2 and the construct of standardized assessment

The perceived mismatch identified in RQ2 is consistent with assessment theory that distinguishes general proficiency evidence from evidence of performance in specific Target Language Use situations (Bachman & Palmer, 2010; Douglas, 2000). Recent scholarship likewise identifies contextualization, authenticity, and diagnostic purpose as continuing priorities in language assessment research (Yang & Wang, 2025). The findings are also compatible with recent ESAP assessment research showing that assessment tasks and expectations vary across disciplinary and professional contexts (Kim & Suh, 2024; Yung & Fong, 2024). The MCEAF translates this theoretical distinction into a coordinated assessment architecture rather than proposing the elimination of standardized certification.

RQ3 and educational equity

The RQ3 findings highlight the importance of interpreting language assessment evidence against unequal educational and linguistic trajectories. Piller and Bodis (2024) show how higher-education proficiency requirements can become connected with language ideologies and processes of inclusion and exclusion, while recent EAP scholarship argues for equitable language support that responds to student diversity across institutional levels (Bakogiannis & Papavasiliou, 2025, 2026). In the present case, diagnostic assessment is therefore conceptualized as an information-to-action mechanism: evidence about students' starting points should inform academic support, advising, and curriculum planning without turning linguistic diversity into a deficit label.

RQ4 and institutional language policy

The RQ4 findings support an institutional interpretation of language assessment in which test evidence is connected to curriculum, support, faculty development, and internationalization. Language assessment literacy research emphasizes that university personnel need not only technical knowledge of assessment but also the capacity to interpret results and make informed educational decisions (Vogt et al., 2024; Lam & Gayton, 2025). Inclusive EAP research further indicates that equitable support requires coordination beyond the individual classroom and into departmental and institutional structures (Bakogiannis & Papavasiliou, 2025, 2026). The MCEAF incorporates these implications by assigning distinct purposes to certification, diagnosis, and discipline-specific assessment.

In relation to the guiding proposition, the findings suggest that the perceived relevance of contextualized assessment may constitute an initial basis for institutional policy development. However, because the present study did not implement the MCEAF or evaluate policy change over time, this relationship should be understood as a qualitative proposition rather than evidence of causal influence.

Toward a multilevel model of English language assessment

Taken together, the findings support a three-level assessment architecture in which each component addresses a distinct institutional purpose: international proficiency certification provides internationally comparable evidence; institutional diagnostic assessment identifies linguistic profiles, strengths, and support needs; and discipline-specific academic assessment examines performance through authentic communicative tasks aligned with individual programs. Recent diagnostic assessment research supports the value of connecting post-admission screening with language support, while ELF assessment research highlights the importance of local context and discipline-informed tasks (Devos et al., 2024; Saeedi et al., 2025). The three levels are therefore conceptually complementary rather than competing forms of assessment.

Figure 2 presents the three-level MCEAF architecture proposed to connect international certification, institutional diagnosis, and discipline-specific assessment.

Figure 2. Multilevel Model for English Language Assessment in Public Higher Education.



Source: developed by the authors.

Note. The figure represents the conceptual architecture proposed by the authors from the complementary functions identified in the study.

Empirical basis for the MCEAF

The empirical evidence maps onto the three proposed levels of the MCEAF. Participants' recognition of the value of English certification supports the role of international proficiency evidence; heterogeneous learning trajectories and reported academic language difficulties

support an institutional diagnostic function; and descriptions of academic reading, writing, presentations, specialized terminology, and professional communication support discipline-specific contextualized assessment. The framework therefore emerges from the complementary functions identified across the different evidence sources rather than from a rejection of standardized certification.

Faculty professional development

A further institutional requirement is faculty development in language assessment literacy and capacity for contextualized assessment. The usefulness of contextualized assessment depends on how tasks are designed, implemented, interpreted, and connected to curriculum. Recent scholarship defines language assessment literacy as encompassing the knowledge, skills, principles, and practices required to design assessment tasks and make informed decisions from assessment information (Vogt et al., 2024); research with university teachers further identifies professional development and test-design experience as pathways for strengthening assessment literacy (Cui et al., 2022; Zhang et al., 2026).

Table 7 links the main empirical findings to the evidence sources that support each dimension of the MCEAF.

Table 7. Empirical findings supporting the MCEAF.

Dimension	Main finding	Evidence
Previous English-learning experiences	Students displayed heterogeneous trajectories, including informal learning, formal instruction, and bilingual or transnational experiences.	Questionnaire; written reflections
Motivation for learning English	Students associated English with academic development, professional opportunities, and internationalization.	Questionnaire; written reflections
Academic use of English	English was used for academic literature, specialized information, presentations, and assignments; difficulties remained in reading and academic vocabulary.	Questionnaire; reflections; faculty interviews
Perceptions of standardized assessment	Participants described standardized examinations as providing useful general evidence but only partial representation of academic language needs.	Written reflections; faculty interviews
Need for contextualized assessment and institutional support	Participants identified discipline-specific assessment, diagnostic evaluation, and institutional language policy as relevant needs.	Questionnaire; reflections; faculty interviews

Source*:* developed by the authors.

Theoretical implications

The principal theoretical contribution is the integration of language assessment, ELF, ESP, and language policy within a single institutional architecture. The findings suggest that these perspectives can be connected through a differentiated evidence system in which different forms of assessment serve different purposes. This reframes the relationship between standardized and contextualized assessment from a choice between alternatives to a problem of functional coordination. Recent assessment scholarship supports this shift by emphasizing contextual relevance, authenticity, diagnostic value, and educational consequences (Yang & Wang, 2025).

The study also extends discussions of validity by emphasizing the relevance of assessment evidence to the academic practices for which students need English. Douglas's (2000) TLU perspective provides the conceptual basis for connecting task demands with intended use, while ELF assessment research adds local context, strategic communication, and discipline-informed task considerations (Saeedi et al., 2025). Recent EAP assessment research further demonstrates that academic genres and task demands need to be considered when developing discipline-sensitive evidence (Kim & Suh, 2024; Yung & Fong, 2024). The MCEAF therefore provides a conceptual bridge between general proficiency evidence and discipline-specific communicative performance.

The study applies ESP principles beyond discipline-specific teaching by proposing their use in discipline sensitive institutional assessment. Needs analysis can inform not only what students learn but also which performances should count as evidence of academic language competence. This is consistent with recent ESAP research emphasizing coordination between language specialists and discipline teachers (Yung & Fong, 2024) and with recent inclusive EAP research calling for alignment across curricular and institutional levels (Bakogiannis & Papavasiliou, 2025, 2026).

Practical implications

First, universities can complement international certification with institutional diagnostic assessment at entry. A diagnostic component can identify linguistic profiles, previous learning experiences, and broad academic language needs, generating information for advising, curriculum planning, and targeted support. Evidence from post-admission diagnostic assessment studies suggests that screening can be linked to language support when institutional procedures are designed for this purpose (Devos et al., 2024).

Second, academic programs can progressively incorporate assessment tasks that reflect recurrent communicative practices, including academic and disciplinary reading, research summaries, disciplinary presentations, technical documentation, and professional communication. The aim is not to create a separate high-stakes English examination for every program, but to collect meaningful evidence of the language practices that matter within each disciplinary context. Recent ESAP and EAP research supports this connection between language provision, disciplinary learning, and assessment tasks (Kim & Suh, 2024; Yung & Fong, 2024).

Third, institutional language policy should connect assessment with curriculum design, faculty development, academic mobility, student support, and internationalization. Language assessment literacy development can strengthen staff capacity to interpret scores appropriately and use evidence educationally (Vogt et al., 2024; Lam & Gayton, 2025; Cui et al., 2022). Inclusive EAP research further supports coordinated institutional action across micro, meso, and macro levels (Bakogiannis & Papavasiliou, 2025, 2026).

Table 8 summarizes the response patterns and their corresponding educational implications.

Table 8. Educational implications of the empirical findings.

Dimension	Questionnaire response pattern	Educational implication
Previous English-learning experiences	Mixed response pattern	Institutional diagnostic assessment should identify diverse linguistic backgrounds and guide differentiated language support.
Motivation for learning English	Predominantly positive response pattern	English functions as academic and professional capital within higher education.
Academic use of English	Predominantly positive response pattern	Discipline-specific language demands support ESP- and EAP-oriented curricula.
Perceptions of standardized assessment	Mixed response pattern	General proficiency tests can be complemented by contextualized assessment evidence.
Need for contextualized assessment and institutional support	Predominantly positive response pattern	Universities can develop inclusive language policies linked to contextualized assessment and support.

Source: developed by the authors.

Note. Because individual numerical responses were not retained, the response-pattern labels describe the overall tendency observed in the available questionnaire evidence and do not represent predefined statistical thresholds.

Pedagogical and institutional implications

The pedagogical and institutional implications point toward an integrated assessment ecology rather than a certification-centered practice. At the pedagogical level, English can be connected with authentic academic activity through EAP and ESP approaches that develop students' capacity to read, interpret, write, present, and interact within disciplinary communities. At the institutional level, assessment, curriculum, student support, and internationalization need to be coordinated so that assessment evidence results in instructional responses. This orientation is compatible with recent inclusive EAP research emphasizing systemic rather than isolated responses to student diversity (Bakogiannis & Papavasiliou, 2025, 2026).

At the institutional level, the findings support an explicit language policy connecting internationalization with educational equity. Such a policy could coordinate English teaching, assessment practices, faculty development, academic mobility, and international research collaboration. For the UAN, the MCEAF can be considered as a proposed architecture for such coordination, while its actual implementation and effects remain to be established empirically.

- English language teaching;
- Language assessment practices;
- Faculty professional development;
- Academic mobility initiatives;
- International research collaboration.

A multilevel institutional assessment model

Based on the findings, the participating university could progressively implement the MCEAF as a three-level institutional model in which each level has a clearly defined purpose, user group, and decision function. Level 1, International Certification, provides comparable evidence for mobility, postgraduate study, and institutional requirements. Level 2, Institutional Diagnostic Assessment, identifies linguistic strengths, previous learning experiences, and broad academic language needs at entry and uses the information to guide advising, support programs, and curriculum planning. Level 3, Discipline-Specific Academic Assessment, uses authentic tasks to examine performance in academic and professional practices within each discipline.

Examples of Level 3 tasks include reading and synthesizing academic articles, writing research abstracts, interpreting technical documentation, participating in professional

presentations, analyzing discipline-specific terminology, and engaging in collaborative academic communication. These tasks establish the direct connection between assessment and disciplinary practice. The proposed three-level architecture therefore enables international comparability to coexist with locally meaningful and pedagogically actionable evidence.

Conclusions

This study examines the relationship between standardized English language assessment and discipline-specific linguistic needs in a Mexican public university. The findings indicate that standardized examinations remain useful for internationally comparable proficiency evidence, but participants perceive that they do not by themselves represent the full range of academic and professional language practices required across disciplines. English functions primarily as a resource for accessing disciplinary knowledge, while students' learning trajectories reveal substantial variation in prior exposure and academic language needs.

The study therefore supports a complementary assessment architecture that combines international certification, institutional diagnostic assessment, and discipline-specific academic assessment. The MCEAF represents the principal conceptual contribution because it organizes these forms of evidence around different institutional purposes rather than treating standardized and contextualized assessment as competing alternatives. This interpretation is consistent with recent work on diagnostic assessment, language assessment literacy, ELF assessment, and inclusive EAP (Devos et al., 2024; Vogt et al., 2024; Saeedi et al., 2025; Bakogiannis & Papavasiliou, 2025, 2026; Cui et al., 2022).

The findings provide qualitative support for this guiding proposition, insofar as participants' perceptions of the relevance of contextualized assessment were accompanied by calls for greater institutional coordination and language policy. However, the study does not establish a causal effect of MCEAF implementation, since the framework was not implemented or evaluated longitudinally.

A methodological limitation is that the researcher-developed questionnaire is not formally validated through expert judgment, pilot testing, or psychometric procedures and is therefore used only as supplementary descriptive evidence. Future studies should strengthen the evidence base by conducting formal content validation and pilot testing of any institutional diagnostic or discipline-specific instruments before using them for measurement. Further research should also examine implementation across additional disciplines and universities,

using larger samples, mixed methods, and longitudinal designs to determine the framework's applicability and educational consequences.

More broadly, the findings position English assessment as part of institutional language policy. When assessment evidence informs curriculum, student support, faculty development, and internationalization, it can function as an educational and strategic mechanism rather than as an isolated administrative requirement. The proposed framework is therefore best understood as a testable institutional design whose transferability and effectiveness require further empirical examination.

References

Adriansen, H. K., Juul-Wiese, T., Madsen, L. M., Saarinen, T., Spangler, V., & Waters, J. L. (2023). Emplacing English as lingua franca in international higher education: A spatial perspective on linguistic diversity. *Population, Space and Place, 29*(2), Article e2619. https://doi.org/10.1002/psp.2619

Aryadoust, V. (2023). The vexing problem of validity and the future of second language assessment. *Language Testing, 40*(1), 8–14. https://doi.org/10.1177/02655322221125204

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world.* Oxford University Press.

Bakogiannis, A., & Papavasiliou, E. (2025). Exploring inclusive teaching practices of English for Academic Purposes (EAP) in higher education (HE): Recommendations for practice. *Journal of English for Academic Purposes, 76*, 101538. https://doi.org/10.1016/j.jeap.2025.101538

Bakogiannis, A., & Papavasiliou, E. (2026). Exploring inclusive teaching practices of English for Academic Purposes (EAP) in higher education (HE): A systematic review and narrative synthesis. *Higher Education, 91*, 1525–1553. https://doi.org/10.1007/s10734-025-01483-3

Banitz, B. (2022). Language assessment in Mexico: Exploring university language teachers' backgrounds, practices, and opinions. *System, 110*, 102898. https://doi.org/10.1016/j.system.2022.102898

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide.* SAGE.

Braun, V., & Clarke, V. (2024). Supporting best practice in reflexive thematic analysis reporting in Palliative Medicine: A review of published research and introduction to the Reflexive Thematic Analysis Reporting Guidelines (RTARG). *Palliative Medicine, 38*(6), 608–616. https://doi.org/10.1177/02692163241234800

Caprario, M. (2025). ELF communication strategies: A pedagogical intervention pilot study. *TESOL Quarterly, 59*(1), 359–377. https://doi.org/10.1002/tesq.3309

Creswell, J. W., & Poth, C. N. (2024). *Qualitative inquiry and research design: Choosing among five approaches (5th ed.).* SAGE.

Cui, Y., Liu, Y., Yu, H., & Gao, Y. (2022). Developing English teachers' language assessment literacy in an EAP reform context through test design: A case study. *System, 109*, 102866. https://doi.org/10.1016/j.system.2022.102866

Dai, D. W., Vu, T., Knoch, U., Lim, A. S., Malone, D. T., & Mak, V. (2024). Expanding Kane's argument-based validity framework: What can validation practices in language assessment offer health professions education? *Medical Education, 58*(12), 1462–1468. https://doi.org/10.1111/medu.15452

Devos, N. J., Nizonkiza, D., & Lynch, S. (2024). Developing and validating a post-admission screening-diagnostic assessment procedure to offer language support in diploma programs. *Canadian Journal of Applied Linguistics, 27*(1), 51–77. https://doi.org/10.37213/cjal.2024.33259

Dewi, O. T. S., & Slamet, J. (2025). Formative ELF-based assessment of spoken communication via learning management system in a culturally diverse classroom. *Discover Education, 4*, 516. https://doi.org/10.1007/s44217-025-00970-0

Douglas, D. (2000). *Assessing languages for specific purposes.* Cambridge University Press.

Fan, J., Frost, K., & Jin, Y. (2022). Local English testing in China's tertiary education: Contexts, policies, and practices. *Language Testing, 39*(3), 453–473. https://doi.org/10.1177/02655322211070839

Jenkins, J. (2015). *Global Englishes: A resource book for students (3rd ed.).* Routledge.

Kim, S.-Y., & Suh, Y.-M. (2024). Disciplinary writing in EMI courses: Faculty beliefs and practices in the Korean higher education context. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688241270752

Lam, D. M. K., & Gayton, A. M. (2025). '6.5 sounds a bit better than 6.0?': A case for embedding language assessment literacy in university teaching staff professional development. *Higher Education* Research & Development, 44(4), 976–991. https://doi.org/10.1080/07294360.2024.2439854

Liu, L. (2024). Exploring the feasibility of incorporating ELFA awareness into the development of EAP pedagogy in a UK pre-sessional course. *Journal of English as a Lingua Franca, 13*(2), 233–260. https://doi.org/10.1515/jelf-2025-2001

Malmqvist, J., Atterström, A., & Swärd, A.-K. (2024). But was it trustworthy? Methodological experiences from a study of a hard-to-reach group of students in need of a flexible research approach. *International Journal of Qualitative Methods, 23*, 1–13. https://doi.org/10.1177/16094069241234805

Piller, I., & Bodis, A. (2024). Marking and unmarking the (non)native speaker through English language proficiency requirements for university admission. *Language in Society, 53*(1), 1–23. https://doi.org/10.1017/S0047404522000689

Saeedi, Z., Tajeddin, Z., & Tadayon, F. (2025). Assessment of English as a Lingua Franca and its principles: A research synthesis. *International Journal of TESOL Studies, 7*(3), 130–152. https://doi.org/10.58304/ijts.250711

Seidlhofer, B. (2011). *Understanding English as a lingua franca.* Oxford University Press.

Shirahata, M., & Lahti, M. (2023). Language ideological landscapes for students in university language policies: Inclusion, exclusion, or hierarchy. *Current Issues in Language Planning, 24*(3), 272–292. https://doi.org/10.1080/14664208.2022.2088165

Topete, B., García, M., & Rodríguez, M. X. (2025). English as a foreign language in higher education: Case study of a public university in México. *Revista Veritas de Difusão Científica, 6*(3), 4018–4040. https://doi.org/10.61616/rvdc.v6i3.1139

Vogt, K., Bøhn, H., & Tsagari, D. (2024). Language assessment literacy. *Language Teaching, 57*(3), 325–340. https://doi.org/10.1017/S0261444824000090

Yang, Z., & Wang, P. (2025). Current status and research trend of English language assessment: A bibliometric analysis. *Language Testing* in Asia, 15, Article 11. https://doi.org/10.1186/s40468-024-00317-w

Yung, K. W.-H., & Fong, N. (2024). Revamping an English for specific academic purposes course for problem-based learning: Reflections from course developers. *Journal of English for Academic Purposes, 69*, 101386. https://doi.org/10.1016/j.jeap.2024.101386

Zhang, L., Zheng, Y., & Yang, D. (2026). Exploring Chinese university English teachers' language assessment literacy: A mixed-method study. *Asia Pacific Journal of Education, 46*(1), 322–340. https://doi.org/10.1080/02188791.2024.2354684

Appendix A. Declarations

Ethics approval and consent to participate

Formal ethical review was not required under the institutional procedures applicable to the registered research project “Inglés como lengua extranjera en nivel superior” (SIP25-005), administered through the Secretaría de Investigación y Posgrado of the participating university. The study nevertheless followed institutional and ethical requirements for research involving human participants. All participants provided written informed consent before data collection. Participation was voluntary, participants were informed of their right to withdraw without academic or professional consequences, and all data were treated confidentially and reported anonymously.

Informed consent

All participants were informed about the study and participated voluntarily. Written consent was obtained before data collection.

Conflict of interest

The authors declare no conflict of interest.

Funding

This research received no specific funding from any public, commercial, or not-for-profit funding agency.

Appendix B. Statement on the use of AI tools

Introduction: ChatGPT was used as a supporting tool to organize and format bibliographic references in accordance with American Psychological Association (APA, 7th edition) standards. The selection, relevance, and verification of bibliographic sources were carried out manually by the authors using the original sources.

Method: generative artificial intelligence was not used for participant selection, data collection, instrument administration, data processing, or the methodological analysis of the research.

Results: generative artificial intelligence was not used for the processing, interpretation, or presentation of the obtained results.

Discussion: generative artificial intelligence was not used to interpret the findings or to compare the results with the scientific literature.

Conclusions: generative artificial intelligence was not used to draft the conclusions.

Figures: ChatGPT was used as a supporting tool to format and structure Figure 2, “A multi-level model for English language assessment in public higher education,” based on the conceptual

proposal developed by the authors. The conceptual content of the figure was defined by the authors and subsequently reviewed and validated manually.

Authorship responsibility: the use of generative artificial intelligence was limited to the functions described above. All references were manually verified against their original sources, and the authors reviewed and validated the final content of the figure. The authors are responsible for the content, interpretation, results, and conclusions presented in the manuscript.