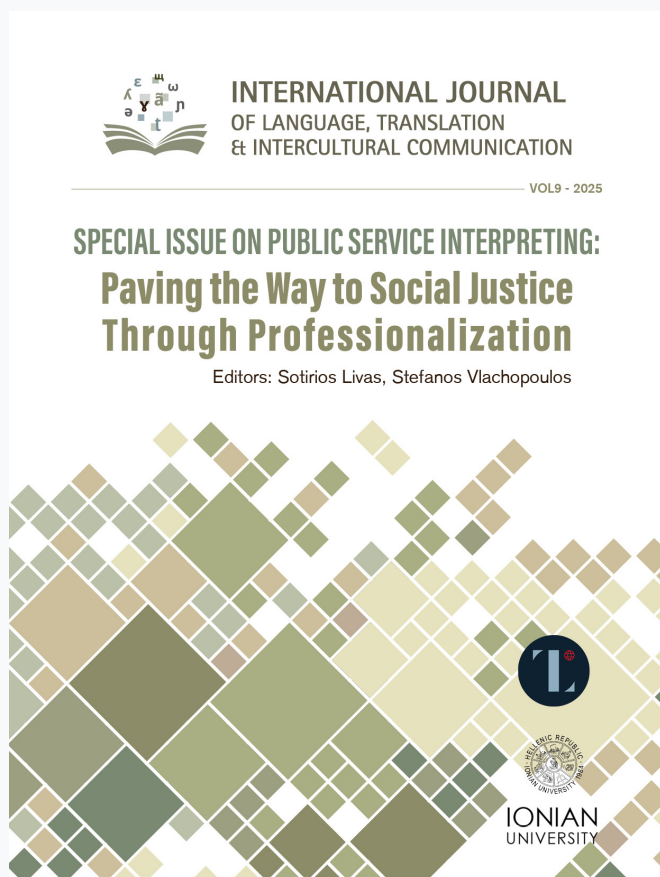INTERNATIONAL JOURNAL
OF LANGUAGE, TRANSLATION
& INTERCULTURAL COMMUNICATION

VOL9 - 2025

SPECIAL ISSUE ON PUBLIC SERVICE INTERPRETING:
Paving the Way to Social Justice Through Professionalization

Editors: Sotirios Livas, Stefanos Vlachopoulos

IONIAN UNIVERSITY

Multimodal Language Testing for Interpreter Training

Effrossyni Fragkou

doi: 10.12681/ijltic.43367

To cite this article:

Fragkou, E. (2025). Multimodal Language Testing for Interpreter Training: Designing and Implementing an AI-Supported Proficiency Test for Punjabi Public Service Interpreters in Greece. *International Journal of Language, Translation and Intercultural Communication*, *9*. https://doi.org/10.12681/ijltic.43367

Multimodal Language Testing for Interpreter Training: Designing and Implementing an AI-Supported Proficiency Test for Punjabi Public Service Interpreters in Greece

Effrossyni Fragkou
National and Kapodistrian University of Athens
effiefragkou@enl.uoa.gr

Abstract

This article documents the design and implementation of Greece's first multimodal, AI-supported proficiency test for Punjabi public service interpreters, developed in response to the March 2024 launch of the Public Service Interpreter Register. The test addresses the complex sociolinguistic and logistical realities faced in the Greek asylum and migration context. The paper reviews current frameworks for interpreter competence and language proficiency—drawing on Common European Framework of Reference (CEFR) standards—and critiques the limitations of traditional assessment models. The Punjabi test integrates audio-based tasks and visual cues, leveraging AI-generated prompts to accommodate dialectal and literacy variation, prioritizing fairness and scalability, thus offering a replicable model for rare and minority languages where written resources, local expertise, and institutional infrastructure are lacking. The paper concludes by evaluating initial implementation outcomes, highlighting both the potential and limitations of multimodal and AI-assisted testing approaches in professionalizing public service interpreting and safeguarding the communicative rights of vulnerable populations.

Περίληψη

Το άρθρο παρουσιάζει τον σχεδιασμό και την εφαρμογή του πρώτου πολυτροπικού, με υποστήριξη τεχνητής νοημοσύνης, τεστ γλωσσικής επάρκειας για Πουντζάμπι διερμηνείς δημοσίων υπηρεσιών στην Ελλάδα, αναπτυχθέντος μετά τη δημιουργία του Μητρώου Διερμηνέων Δημοσίων Υπηρεσιών (Μάρτιος 2024). Η εξέταση ανταποκρίνεται στις κοινωνιογλωσσικές και υλικοτεχνικές προκλήσεις του ελληνικού πλαισίου ασύλου και μετανάστευσης και στηρίζεται στα πρότυπα του CEFR, επισημαίνοντας τις αδυναμίες των παραδοσιακών μοντέλων αξιολόγησης. Το τεστ ενσωματώνει ακουστικές δραστηριότητες και οπτικά ερεθίσματα, αξιοποιώντας προτροπές παραγόμενες μέσω τεχνητής νοημοσύνης ώστε να ληφθούν υπόψη διαλεκτικές, πραγματολογικές και γραμματολογικές διαφοροποιήσεις, προτάσσοντας τη δικαιοσύνη και την επεκτασιμότητα. Το μοντέλο είναι αναπαραγώγιμο για σπάνιες ή μειονοτικές γλώσσες με περιορισμένους πόρους και θεσμική υποδομή. Το άρθρο ολοκληρώνεται με αξιολόγηση των πρώτων αποτελεσμάτων, αναδεικνύοντας δυνατότητες και περιορισμούς των πολυτροπικών, υποβοηθούμενων από τεχνητή νοημοσύνη μεθόδων αξιολόγησης στη διερμηνεία δημοσίων υπηρεσιών.

Keywords: AI-supported language testing; language proficiency; migration and asylum; multimodal assessment; Public Service Interpreting (PSI); Punjabi.

1 Introduction

Language testing is a key component in assessing the language proficiency of entry-level interpreters, yet it is rarely treated as distinct from assessing interpreting skills (Loiseau & Luchner, 2019: 4-15). In Public Service Interpreting (PSI), evaluating language proficiency is challenging, as test-takers are rarely homogeneous (Ginther et al., 1998; Gu, 2014; Manna et al., 2015). Their skills may vary due to prior testing

experience, literacy shaped by formal education, social and cultural backgrounds, and self-reported abilities (e.g., balanced or dominant bilinguals) (Flege et al., 2002: 568-569, 574; Oslon, 2023).

These elements must be understood in relation to the interpreter's A, B, and C languages, a distinction rooted in conference interpreting (CI) but only loosely applicable in PSI, where conference interpreting predominates. PSI's bidirectional nature means the notion of a C language may be redefined or deemed irrelevant, as interpreters must master both target languages (A and B) to a high level for productive and receptive purposes.

Thus, PSI language proficiency is often framed in terms of the native (A or L1) versus non-native (B or L2) active languages paradigm inherited from conference interpreting (Gile, 2005; Donovan, 2005; Lim, 2005) and how this dichotomy might influence interpreters' performance in terms of quality and audience perception (Kurz, 2009; Cheung, 2013). This classification assumes that (a) proficiency is measured comparatively, with the A language as the benchmark (Loiseau & Luchner, 2019: 6-7), and (b) A, B, and C distinctions determine permissible working combinations. However, criteria for assigning languages to these categories remain unclear, and labels such as "native", "quasi-native" or "non-native" are often sociolinguistically loaded and used inconsistently when describing proficiency either in translation or interpreting contexts (Pokorn, 2009). Despite "native-language benefits" in bilinguals, there are degrees of language proficiency and competence, which suggests that the native status exists on a spectrum relying heavily on subjectivities and variability of context rather than in binarity (Golestani et al., 2009; Lev-Ari, 2015). This is further supported by the lack of standardized, reliable methods for categorizing speakers as 'native' or 'non-native' (Wen et al., 2023).

Language proficiency is intrinsically linked to interpreting competence, as interpreting skills develop from the ability to function in interactive multilingual contexts. Assessing proficiency becomes more complex when testing multiple languages simultaneously, as in the present case. This requires tools which will ensure reliability, validity, standardization, and calibration across languages, regardless of dialectal variation, script, or standard variety. Robust assessment verifies bilingual capacity before interpreter training begins. Literature stresses the need for comprehensive evaluation of general abilities: Hale (2007) and Angelelli (2009) note that traditional tests often overlook contextually grounded performance. Credentialing systems in Australia, Canada, Norway, and the UK confirm that language assessment should precede or complement training (Hlavac, 2015) and distinguish between general proficiency and interpreting-specific skills.

Interpreting competence entails advanced listening comprehension, rapid lexical access, cognitive processing, and context-sensitive speech production. These require specialized methods. Angelelli (2007: 77-78) calls for separate, task-based testing that mirrors interpreting practice. In Norway, for instance, candidates face bidirectional testing and simulated consecutive interpreting (Skaaden & Wattne, 2009: 75-83); in Australia, multi-component assessments, including note-taking and pragmatic tasks, capture interpreting performance (Lai & Mulayim, 2010 & 2013). Such approaches reinforce a two-tiered model: general language proficiency and interpreting-specific competence.

The March 2024 launch of Greece's first Public Service Interpreter Register by the Ministry of Migration and Asylum highlighted the need for specialized proficiency

tests in critical languages, including Punjabi. Punjabi's dialectal diversity, contested scripts, and scarcity of local experts complicated test design and validation as will be shown in Part 3 of this article.

This paper documents the Punjabi proficiency test's development in the Greek PSI context and proposes a replicable model for languages with similar sociolinguistic and logistical constraints. The integration of multimodality and AI-supported solutions is central this project, bridging conventional language testing with evolving interpreter pre-training demands.

The paper first outlines the context leading to the Register, presenting theoretical considerations and project phases to situate the pre-entry exam's development. The framework, based on needs analysis within the Ministry of Migration and Asylum (MOMA), draws on the methodological principles of the National Foreign Language Exam System (KPG) (Dendrinos et al., 2013)[1], adapted to interpreter-focused assessment. A multimodal, communicative approach combines general language proficiency and interpreting requirements, forming a two-tiered bilingual model (Greek + other language[s]).

The author then examines the test's philosophy, format, and structure, with a focus on Punjabi's linguistic challenges, and addresses the targeted linguistic awareness—phonological, morphosyntactic, lexical, semantic, and pragmatic—across all versions. Emphasis is placed on validation, refinement, and initial implementation outcomes. The discussion then considers the broader implications of language competence testing for training and assessing interpreters in rare and minority languages, highlighting the potential of multimodal and AI-assisted testing to address sociolinguistic complexities such as dialectal variation and script diversity through innovative, non-language-based response formats.

The paper concludes by evaluating the test's contribution to Public Service Interpreting (PSI) training and assessment, outlining the benefits and limitations of multimodal and AI-supported testing in linguistically diverse contexts, and offering recommendations for future research and the scalability of the model to other rare language settings.

2 Background and Rational for Establishing the First Public Service Interpreter Register in Greece

The establishment of Greece's first Public Service Interpreter (PSI) Register emerged as a strategic response to longstanding gaps in the quality, coordination, and institutional oversight of interpreting services in the asylum and migration area. The initiative forms part of Sub-project 4: "Activities arising from the Cooperation Agreement with the Department of Foreign Languages, Translation and Interpreting of the Ionian University," under the Action "Strengthening and Development of National Capacity for Strategic Planning in the Fields of Asylum and Migration." The project

---

[1] The KPG Exam (State Certificate of Language Proficiency) is Greece's official standardized test for certifying proficiency in foreign languages. Established by the Ministry of Education, it assesses candidates' ability to understand and use a target language—such as English—at different levels, from beginner (A1/A2) to advanced (C2), according to the Common European Framework of Reference for Languages (CEFR). The exam tests reading, writing, listening, and speaking skills, and is recognized nationally and internationally for education and employment purposes.

was funded through the EEA Financial Mechanism 2014–2021, under the Program "Capacity Building of National Asylum and Migration Management Systems."

Spearheaded by the Ministry of Migration and Asylum (MOMA) in collaboration with the Department of Foreign Languages, Translation, and Interpreting of the Ionian University, and supported by Norway's Directorate of Integration and Diversity (IMDi) and Machas & Partners Law Firm, the project was designed to professionalize PSI in Greece. At its core lies a pressing concern: how to ensure that migrants, asylum seekers, and refugees—often vulnerable, traumatized, and disoriented—can exercise their rights and access essential services in a linguistically and culturally meaningful way.

2.1 Interpreters in the Context of Refugees and Asylum: The Greek State of Affairs

Interpreting within the refugee and asylum context is central to safeguarding fundamental rights. As Greece has shifted from a migration-neutral country to a frontline state in the European migration landscape—particularly since the 2015-2016 crisis and again in late 2025—the stakes of language mediation have risen sharply (Evangelinidis, 2016; Triandafyllidou et al., 2017; Dimitriadi et al., 2019; Kotoulas, 2024). The country has received large numbers of third-country nationals from Asia, Africa, and the Middle East, with little if any competence in Greek or English.

As a result, Greece can no longer claim to be facing mass migration for the first time. As Apostolou (2012: 213) observes, the influx of migrants has challenged "[…] the imaginary construct of the homogeneous ethnos/nation […]", eroding the notion of a monolingual, monocultural society. Linguistic mediation across multiple languages is therefore a structural necessity, not a contingency.

Interpreting in refugee contexts covers the entire asylum and integration process—border crossings, reception, registration, interviews, appeals, court proceedings, healthcare, child protection, school enrolment, etc. Work settings range from RICs, controlled access centers, hospitals, shelters, and courts to public services, faith institutions, and private homes. Interpreters act as communicators, institutional interlocutors, and cultural mediators—often the only link between refugees and the Greek state.

Our fieldwork showed that interpreting services are almost entirely outsourced to NGOs and, increasingly, private language service providers under EUAA mandates. Many of these services are delivered remotely, utilizing interpreters based abroad. While such a practice expands reach of service beneficiaries, safeguards privacy, and protects those at risk, it limits flexibility, contextual understanding, and user trust. Officials and service users stressed the need to prioritize Greece-based interpreters working to and from Greek for effective communication and institutional continuity.

Without an institutional recruitment, training, and certification framework, NGOs have stepped in with ad hoc programs, but training has been limited if existent (originally from 10-day intensive courses to as little as two days in some cases). This decline, amid rising demand, heightens risks of miscommunication and legal vulnerability for both interpreters and beneficiaries.

2.2   Integration, Language, and the Role of Interpreting

Interpreting must be viewed within the broader framework of refugee integration at the financial, social, and emotional level (Council of Europe, 2008 a & b); Isphording, 2015; Born et al., 2019; Daley, 2019; Montemitro et al., 2021; Schacht et al., 2022; Foged & van der Werf, 2023). Following Ager and Strang's (2008) model, language proficiency and access to information are among the key "facilitators" of integration. Yet, integration in the Greek context has often been approached through a neo-assimilationist lens, as Kondis (2013) argues, wherein migrants are granted rights but are simultaneously burdened with disproportionate responsibilities, particularly regarding language acquisition and civic participation.

The 2019 "National Strategy for Integration" explicitly recognized interpreting as a mechanism for fostering inclusion. Two policy actions are worth noting, namely Action 1.4.1 aimed to reinforce interpreting and intercultural mediation services for asylum seekers and unaccompanied minors, ensuring equitable communication and reducing cultural misunderstandings in RICs and asylum procedures (2019: 41); and Action 8.1.1, which targeted refugee and migrant women, offering training in interpreting and mediation to empower them and facilitate the reintegration of their families, particularly through school integration for children (2019: 73).

In these actions, interpreting was framed as both a support service and a vector of empowerment. However, this integrative vision weakened in a more recent 2021 strategy, where interpreting was referenced only vaguely. Terms like "appropriately trained and certified interpreters" appear in the strategy's wording without specifying certifying bodies, assessment criteria, or language coverage (2021: 10-11). Such ambiguity carries serious consequences in a multilingual asylum environment, especially when vital services hinge on accurate and impartial interpreting.

Bouroutis (2022: 72, 76) offers a sobering critique of this implementation gap: while national strategies may highlight the importance of language and interpreting, they often lack budgetary commitment, timelines, or institutional follow-through. For example, the 2019 strategy promoted the acquisition of Greek and English by migrants, yet did not specify funding or deadlines, rendering it aspirational rather than actionable.

2.3   Refugees' Own Perspectives and Interpreting as a Threshold

While the project described above did not directly investigate refugees' perceptions of integration, there is widespread evidence that Greece is not generally regarded as a final destination. As Naskou-Perraki, Papageorgiou, and Baxevanis (2017) observe, many recognized refugees in Greece apply for travel documents to continue their journey to Northern or Western Europe. Legal, social, and financial insecurities—often exacerbated by prolonged asylum processes and inconsistent interpreting services—contribute to this transient orientation.

In this context, interpreting operates at a critical threshold. It can either facilitate access to rights and services (Norström et al., 2011), thereby encouraging integration, or, when of poor quality or unavailable, it can reinforce exclusion and alienation (Heath et al., 2023). This dual potential also applies to the interpreters themselves, many of whom come from migrant backgrounds. Local interpreters who engage with state structures over time gain familiarity with institutional norms and linguistic registers, thus fostering their own integration into Greek society.

In short, the creation of a national PSI Register is not merely a technical intervention. It is a structural investment in the rights of migrants and the integrity of the Greek asylum system. It professionalizes crucial services, offers quality assurance, and enhances institutional trust. Ultimately, it recognizes that in a multilingual, multicultural migration landscape, interpreting is not a luxury but a prerequisite for justice and dignity (Gentile, Ozolins & Vasilakakos, 1996; Baixauli-Olmos, 2017; Moratto & Li, 2021; Vlachopoulos et al., 2023; Marianacci, 2024; Ioannidis, 2025; Vyzas, 2025).

2.4 Interpreter Profiles, Competencies, and the Challenges of Certification

One of the main outcomes of the fieldwork research conducted prior to and for the purposes of designing Greece's first Public Service Interpreter Register was to document the profile, qualifications, and working conditions of interpreters currently active in the asylum and migration sector. This data, collected through five semi-structured interviews with experienced Ministry officials and regional coordinators between February and March 2023 (caseworkers), reveals an interpreting services provision system that is nominally functional but fragmented, operating without consistent standards, institutional oversight, or a shared understanding of professional interpreting competencies.

2.5 Who Are the Interpreters in Asylum and Migration?

As of early 2023, all interpreters working within the Greek asylum and migration framework were engaged as external contractors. There were no interpreters formally employed as permanent Ministry of Migration and Asylum (MOMA) staff. The primary service providers were non-governmental organizations (NGOs) alternating or working in parallel with private language service companies subcontracted through the European Union Agency for Asylum (EUAA). These providers offered both on-site and remote interpreting services, although remote services—often delivered by interpreters based in other European countries—have become increasingly dominant, especially during and around the lockdowns. As of 2025, interpreting services are primarily provided on-site through contracts with local NGOs specializing in translation, interpreting, and intercultural mediation. However, these services have been fragmented due to issues of non-payment—interpreters have been reporting not being compensated, while NGOs often claim they have not received the funds owed to them by Ministry of Migration and Asylum (MOMA) under existing agreements.

While this shift to remote interpreting offers certain logistical advantages, such as wider language coverage and cost efficiency, it has introduced several challenges. According to the respondents, the disadvantages of remote services often outweigh the benefits. Remote sessions lack consistency, flexibility, and adaptability, especially in urgent or sensitive situations or recurrent meetings. Scheduling issues, cancellations, and a lack of familiarity with local context and institutions hinder communication, reducing trust between interpreters and service users.

By contrast, locally based interpreters are considered essential to the effective delivery of interpreting services. Their physical proximity allows for real-time availability in emergencies, greater flexibility in scheduling, and a deeper understanding of institutional procedures and workflows. They engage in continuous collaboration with case workers, social workers, and legal officers, and possess heightened cultural and contextual awareness of the migrant communities they serve.

This fosters their integration into institutional teams as trusted and consistent partners. Furthermore, many of these interpreters come from migrant backgrounds themselves, enabling them to develop an insider-outsider perspective that enhances their effectiveness. This dual positioning not only strengthens their ability to mediate between service providers and beneficiaries but also facilitates their own gradual integration into the professional and civic fabric of Greek society (Council of Europe, 2023; Urdal, 2024).

2.6   Current Interpreter Competence and Language Proficiency Levels

Despite their vital role, interpreters working in these contexts are not recruited or evaluated against a unified national standard. Instead, large NGOs and private providers apply their own internal systems for assessing interpreter competencies. One dominant NGO classifies interpreters into three broad levels based on language proficiency in both Greek and the interpreter's working language as shown in the table below.

Interpreter Level	Language Proficiency (Greek + L1) *	Assigned Tasks in the Asylum Context (Communicative + Interpreting-related)
Level 1	A1-A2	Basic Communication Tasks
Level 2	A2+	Registration of Asylum Seekers
Level 3	B1-B2	Conduction Asylum Interviews
* Levels designated according to the scale of the Common European Framework for Languages (CEFR).		

Table 1: PS interpreters language proficiency classification

Although some interpreters may hold university degrees or possess C2 or equivalent proficiency in their first language, very few actually reach beyond B2–C1 in Greek. Most fall within the A1 to B2 range in both languages, but that also depends on the language combination. This raises concerns about their ability to handle complex legal, medical, or administrative content without distortion or omission.

Furthermore, the levels depicted in Table 1 are provider-specific and, therefore, not aligned with any national or international certification system. With no central authority in Greece responsible for testing interpreter language proficiency, verifying credentials, or maintaining professional records, the absence of a standardized framework undermines the transparency, consistency, and perceived legitimacy of interpreting services across the asylum and migration infrastructure.

2.7   Ad Hoc Training Practices

Interpreter training represents another area characterized by significant inconsistency (Vlachopoulos, 2016: 162-164; Apostolou, 2015: 27). In the absence of formal state-

sponsored programs or a standardized national curriculum, responsibility for interpreter training has been delegated to non-governmental organizations. One prominent NGO previously implemented a ten-day foundational course structured around five core components. These included professional ethics and conduct, with a focus on confidentiality, neutrality, impartiality, and the management of conflicts of interest; legal and institutional knowledge, encompassing an understanding of asylum procedures, relevant documentation, and the broader institutional framework; interview techniques, which covered the roles of various stakeholders, interview typologies, questioning strategies, and common communicative challenges related to asylum procedures; language and terminology, with particular emphasis on legal and medical vocabulary in Greek and other working languages; and, finally, assessment and evaluation, involving written and oral examinations, role-play simulations, terminology testing, and ethical scenario analysis. However, due to financial constraints, combined with the need for high interpreter turnover, and growing demand during periods of increased migration, this training program has reportedly been shortened to just two days or under.

Such a reduced timeframe is insufficient even for the meaningful introduction of professional ethics, let alone for a comprehensive development of interpreting competencies. This decline in training quality and quantity reflects deeper structural vulnerabilities within the current system while underscoring the pressing need for a coherent, state-led intervention to ensure consistent and professional interpreter preparation.

2.8   The Case for a National Register and Certification Framework

The fragmented and inconsistent nature of interpreter recruitment, assessment, and training in Greece dictated the urgent need for establishing a unified Public Service Interpreter Register (Ioannidis, 2025). Such a register would serve multiple critical functions, namely, (a) define minimum thresholds for both language proficiency and professional competence; (b) establish transparent and standardized certification procedures, and (c) maintain a publicly accessible database of qualified interpreters. Furthermore, (d) it would guarantee ongoing training opportunities by supporting continuing professional development, while (e) safeguarding interpreters through clearly articulated codes of ethics and institutional protection mechanisms. Crucially, the creation of such a register would not only raise the quality and reliability of interpreting services but also reinforce the overall legal and procedural integrity of the asylum and migration system.

Implementing clear standards and formal verification of interpreter credentials are essential for upholding the rights of vulnerable individuals and for enabling public authorities to meet their obligations under European and international legal frameworks thus increasing the public's trust in the profession but also the trust of migrants and refugees in public authorities (Mikkelson, 2004). In addition, by formalizing and institutionalizing the role of public service interpreters, the register would contribute to the broader professionalization of the field and enhance the social recognition of interpreters, thereby supporting their long-term integration into Greece's public service infrastructure (Gentile, Ozolins and Vasilakakos, 1996).

3 The Construct of the Language Competence Test in the Context of the Greek PSI Register

3.1 General Considerations

The construct of interpreting competence is inherently linked to language competence. It encompasses the multidimensional skills and essential knowledge interpreters require to perform inherently complex tasks. Interpreting extends far beyond bilingualism (Kalina, 2000): interpreters must combine high-level proficiency in at least two languages (bilingual language competence) (Albl-Mikasa, 2013: 22) with an integrated set of abilities, typically categorized as: (a) interpreting-specific competence; (b) psycho-social and psychosomatic competence; (c) cognitive competence; (d) interpersonal and professional competence; (e) intercultural competence; and (f) strategic competence (Albl-Mikasa, 2012: 62, 64; Fragkou, 2023).
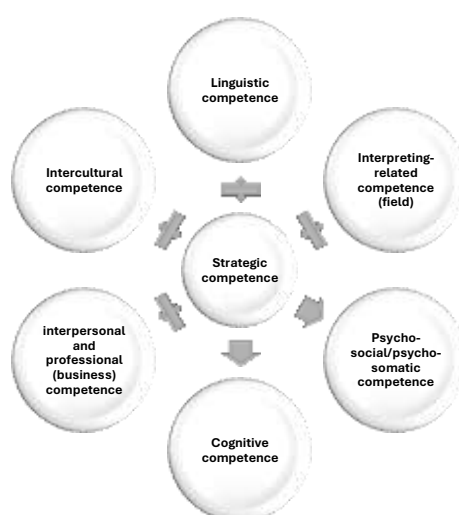

Table 2: Interconnection of competencies for PS interpreters

Although classifications vary, bilingual proficiency is the essential prerequisite—without it, interpreting cannot occur. Within this framework, proficiency in both source and target languages involves mastery of grammar, vocabulary, and discourse. Equally critical is cognitive processing, as interpreting is a real-time activity demanding intense memory, attention, and information management. Interpreters must comprehend, retain, and reformulate content almost instantaneously, requiring advanced linguistic and cognitive multitasking.

Central to most competence models is the notion of (*verbal*) *transfer*, theorized by Pöchhacker (2004 & 2009) as a *meme*, and by Setton & Dawrant (2016) as a complex procedure aimed to bridge cognitive and cultural gaps by making necessary adjustments to form and content. As such, transfer goes beyond verbatim rendition, aiming for pragmatic and contextual adequacy. Effective transfer depends on deep intercultural understanding and strategic competence, both indispensable to professional practice.

Another key component is extra-linguistic and subject-matter knowledge, often subsumed under interpreting-specific competence. Interpreters must not only perform simultaneous, consecutive, or sight translation, and use appropriate tools (e.g., remote platforms), but also draw on broad background and topical knowledge to comprehend the source material and produce coherent target renditions. According to ISO 23155,

this overlaps with research, information acquisition, and knowledge management, that is, the ability to conduct targeted research before an assignment, retrieve information from varied sources at short notice, and assess its quality.

This competence extends beyond general encyclopedic and cultural knowledge to include familiarity with precedent phenomena, realia, values, institutions, and everyday practices in the cultures involved. It also covers register awareness and broader cultural competence, as well as risk assessment and security when handling information before or during assignments (ISO 23155: 2022).

These skills intersect with interpersonal and professional competence, which entails understanding the practical, financial, and legal aspects of running a practice, participating in professional associations, and complying with legal, ethical, and deontological standards. As Setton & Dawrant (2016) note, the ability to work under stress falls within professional competence, though some frameworks (e.g., ISO 23155; Fragkou, 2023) classify it separately or as overlapping with psycho-social and psychosomatic competence.

3.2   Operationalizing the Construct of Interpreting in a Language Test for Entry-Level Interpreters

Now that the construct of interpreting has been clearly defined, this section focuses on designing and validating a language competence test for entry-level interpreters in the Greek public sector, particularly in immigration and asylum services. Developing such a test requires clear criteria and conditions for language tasks that are relevant and reflective of real-world demands. In Europe, the most recognized benchmarks are the Common European Framework of Reference for Languages (CEFR, 2020) "can-do statements," which, while adaptable, were not designed for pre-service interpreter language intake assessment.

As with any well-designed language test, assessment tools should have a clearly defined purpose, grounded in understanding both the test takers' abilities and the context of language use (Bachman & Palmer, 1996). Tests should engage a range of characteristics, or interactiveness, by activating:

a) Language ability, including linguistic knowledge and metacognitive or strategic competence, in other words, the ability to plan, monitor, and evaluate language use under test conditions;

b) Topical knowledge, or background knowledge of the subject matter; and,

c) Affective schemata, that is, personal attributes such as interests, attitudes, motivation, and emotional responses.

An effective test activates prior knowledge—both general and topical—while engaging affective involvement, strategic thinking, and core language ability. Recognizing that interpreting is both a process and a product rooted in discourse, various theoretical perspectives on discourse inform language competence assessment in interpreter training. Early linguistic views, such as Harris (1952), conceptualized discourse as coherent language beyond the sentence, focusing on structural patterns—insights valuable for designing tasks that assess cohesion and coherence. Austin (1962) and Searle (1969) shifted the focus to function, viewing language as social action; their speech act theory highlighted the performative nature of utterances, central to interpreting where language creates meaning in context-specific interactions.

Halliday's systemic-functional linguistics (1978) further emphasized language as a tool for social and functional communication, informing assessments of how interpreters manage interpersonal meaning, ideational content, and textual organization.

From a critical perspective, Foucault's (1969, 1971) post-structuralist analyses conceptualized discourse as constitutive of power, knowledge, and identity—highly relevant to interpreter-mediated encounters, where institutional discourse asymmetries shape communication. Building on this, Critical Discourse Analysis (CDA), developed by Fairclough (1989, 1992) and Fowler (1991), treats discourse as social practice that reflects and reproduces power relations, making interpreters' sensitivity to discursive positioning a crucial assessment dimension. Similarly, although originating in the study of environmental ideologies, Hajer (1995: 44) defines discourse as sets of ideas, narratives, and practices—concepts and categorizations—through which meaning is constructed, while Dryzek (1997) describes it as a shared framework for understanding the world, grounded in assumptions, judgments, and contentions that shape how issues are analyzed, debated, agreed upon, and disputed, and that are woven into coherent narratives.

A nuanced grasp of such frameworks allows the interpreter to recognize that communicative interactions are mediated by deeply rooted conceptual schemata, latent presuppositions, and culture-specific narratives which can align or diverge in interpreter-mediated interaction, and in the interpreter's own sociocultural and cognitive repertoire. Such metalinguistic and intercultural sensitivity is part of the complex of the interpreter's ability to effect pragmatic equivalence, preserve the illocutionary force of the original utterance, and negotiate the tensions brought about by divergent worldviews in complex multilingual communicative events.

3.3   Assessment Design: Key Preliminary Questions, Nature of the Test, Validity, Reliability and Test Authenticity

Bearing all this in mind, the task of designing language competence tests for all languages included in our Public Service Interpreter Register compelled us to confront a number of key questions. These questions can be grouped into two main categories: foundational considerations for language testing and essential criteria for test usefulness.

3.3.1 Foundational Considerations for Testing

Our tests were designed specifically to assess language competence for interpreting, prompting the question of which aspects of language should be evaluated. The goal was to establish clearly defined pass/fail criteria based on fixed, consistently difficult tasks reflecting progressively challenging communicative situations in the migration/asylum context. The tests were therefore criterion-referenced, targeting expected language competencies in these settings. Both macro and micro skills were included, with particular emphasis on reading comprehension for all languages and, in the case of Punjabi, listening comprehension, focusing on receptive rather than productive skills. The rationale for this deviation in the Punjabi test is discussed in a later section.

As far as micro skills were concerned, the focus was placed on lexis—with special attention to appropriate word use—syntax, and morphology, in order to assess accuracy and complexity in sentence structure. Orthography was also tested, particularly the correct spelling of homophones, pseudo-homophones, and homonyms,

as accuracy in meaning is essential for interpreters who must discern intended meaning from context, whether written or spoken. This ability becomes especially important when dealing with ambiguous utterances. In fact, the candidate's capacity to use context to disambiguate meaning was regarded as a core interpreting skill, particularly in real-time settings.

Despite being traditionally associated with written language (reading or writing), orthographic competence could serve as a predictor for error-prevention and self-correction in entry-level interpreters, who, as older individuals, ought to have more developed orthographic abilities (Martin et al., 2003; Zarić, Hasselhorn, & Nagler, 2020; Compton, Gilbert, Kearns, & Olson, 2020). These candidates are more susceptible to word confusion (linguistic interference) for several reasons: the problematic items appear in their 'B' language; it results from partial/erroneous cognitive inscription due to deviated acoustic-phonetic pattern perception (Kurz, 2008: 179-180, 183-154), or due to the pressure of performance in interpreting settings (which is also valid in an exam situation) (Cooper et al., 1982: 104). In either case, including such items assesses candidates' metacognitive abilities (planning, monitoring, and evaluating) (Kusiak, 2001), ensuring that only those who possess sufficient semantic awareness would pass.

Another micro skill assessed was sociolinguistic appropriacy—the ability to use language that is suitable to context, audience, and register. Pragmatic competence was also tested by simulating real-life situations and assessing the candidate's understanding of the socio-pragmatic features of the environments in which they are expected to interpret.

Speaking and writing were deliberately excluded from our test design. It was decided that the language assessment tool would be an objective, standardized instrument with predetermined answers and unambiguous scoring criteria across all tested languages. Such an approach removes the need for subjective judgment, enhances fairness, and facilitates the efficient administration and scoring of large cohorts, including in languages not mastered by the test administrators. This ensures the tool's long-term viability and operational autonomy beyond its initial development phase. Most importantly, it guarantees horizontal fairness for all test takers, regardless of the language being tested.

3.3.2 Modelling the Test and Assessing it Usefulness-Essential Criteria

A critical question in developing the assessment instruments was: Who would create them, how, and with what safeguards for validity? Such an operation requires a coordinated interdisciplinary action between "[…] content experts who have unique understanding of the context of interest, and applied linguists, who need to interpret these understandings within their own frame of reference for language teaching or testing purposes" (Elder & MacNamara, 2015: 2). Given the multilingual nature of the tool, native speaker language experts were essential to assume the role of content experts. They would have to operate under the guidance of an academic team in interpreting, linguistics, and language testing and assessment (applied linguists) to develop test instruments for their respective languages. The academic team was tasked with designing the initial format, structure, and objectives and presenting them to the language experts. The latter then followed a 10-hour introductory training session to the Greek-language prototype, its components, and guidelines for integrating authentic materials.

Following Greek public procurement requirements, the academic team issued a public call for language experts through the Special Account for Research Funds of the Ionian University. Candidates needed a degree from a recognized Greek or foreign university, documented competence in their native and/or working languages, and legal and tax status in Greece.

The assessment, offered in ten languages—Arabic, Dari, Farsi, Kurmanji, Punjabi, Somali, Sorani, Greek, English, and French—adhered to measurable benchmarks across all language pairs for fairness and comparability. For languages without standardized written forms, multimodal methods were used. Each version contained 50 closed-ended items with uniform structure and scoring (except for Punjabi and partly Somali), covering:

- Section I (Items 1–10): Assess vocabulary and grammar through multiple-choice questions that require candidates to choose the most contextually and grammatically appropriate word or phrase.

- Section II (Items 11–15): Test knowledge of synonymy, measuring the ability to understand and paraphrase synonymous vocabulary as well as register identification.

- Section III (Items 16–25): Assess comprehension of factual information, legal and civic knowledge, and language use in public service contexts.

- Section IV (Items 26–30): Focus on understanding text structure, legal reasoning, and cohesion through sequencing-based questions.

- Section V (Items 31–40): Cloze (fill-in-the-gap) items designed to evaluate vocabulary, syntax, institutional knowledge, and contextual awareness.

- Section VI (Items 41–45): Target literal and inferential reading comprehension skills.

- Section VII (Items 46–50): Assess understanding of Greek idiomatic expressions, particularly in relation to their literal and figurative meanings.

All sections scored medium-high in language ability and topical knowledge, low-medium in affective schemata (see Table 3).

Section	Language Ability	Topical Knowledge	Affective Schemata
I	High (vocabulary, grammar)	Medium (contextual)	Low
II	High (lexical)	Medium (contextual)	Low
III	Medium (comprehension)	High (facts, society)	Medium (beliefs and experiences)
IV	High (cohesion)	Medium (job and career)	Medium
V	High (contextual)	High (civic, culture)	Medium
VI	High (reading)	High (education)	Medium-High
VII	High (idioms)	Low	High (figurative language and cultural aspects of language)

Table 3: Test interactiveness according to Bachman and Palmer

Two experts worked on each language combination, independently producing five test versions and validating each other's work. Academic supervisors oversaw 3-4 language combinations each, monitored progress, and validated deliverables against a uniformed evaluation grid. Experts also validated the Greek-language test, assessing item quality, clarity, authenticity, and scalability. Their feedback was particularly relevant since most of them were active public service interpreters.

Test delivery parameters, such as location, scheduling, and frequency, were set by the commissioning body (MOMA) according to logistical constraints. The exam was designed as a paper-based, onsite test at MOMA venues. While an online version could expand access, it would require secure platforms, remote proctoring and related training as well as reliable candidate digital literacy. Remote testing would pose security and fraud-prevention challenges, and even controlled online administration would face venue and resource limitations (Manousou et al., 2024: 178). Other aspects, such as candidates' privacy concerns and exacerbated test anxiety would have to be factored in (Kuleva et al., 2024). It is worth noting, however, that the current test format lends itself seamlessly to a transition from paper-based to paperless administration.

A major challenge in our case was profiling the intended audience. Candidates come from diverse social, cultural, legal, and educational backgrounds, with significant variation in age, cognitive ability, and linguistic repertoire. Arabic best illustrated this complexity: alongside Classical and Modern Standard Arabic, numerous mutually unintelligible dialects exist. MOMA records group all Arabic speakers under one category, with no reliable dialectal classification, and self-reported data at EU entry is inconsistent.

Mapping interpreter profiles via surveys or interviews would also prove impractical due to interpreters' dispersion across Greece, absence of a national register, mistrust of data collection, and NGO reluctance to cooperate. Many interpreters—often former refugees—see their role as interpreters and their legal status as temporary, viewing Greece as a transit country. Consequently, the language experts involved in this project cannot be considered a representative sample, with the diversity of potential candidates remaining difficult to document.

Finally, it is important to note that our tests draw directly from the established principles of the State Certificate of Language Proficiency (KPG) examination format, recognized for its modular structure and emphasis on authenticity. Each test section corresponds to specific communicative tasks modeled on real-life scenarios. By adopting the KPG model—while restricting it to listening, reading, and pragmatic mediation—the tests ensure consistency in assessment and alignment with national certification standards, even though these were originally developed for foreign language assessment rather than interpreting. Furthermore, item construction draws, where feasible, on the rigorous validity and reliability criteria of the KPG framework, thus supporting transparent and objective marking procedures for large-scale candidate evaluation. The choice of the KPG paradigm also satisfies a fundamental theoretical requirement of test-making: *localization*, that is, the adaptation of assessment tools to the specific geographical, linguistic, social, and cultural environment of the test takers to ensure contextual appropriateness (O'Sullivan, 2012, p. 79).

Crucially, the assessment benchmarks of the tests are mapped onto the CEFR can-do statements, enabling a clear progression framework for candidates across B2 to C2 language proficiency levels. Tasks are designed to elicit functional language use, reflecting the CEFR's descriptors for comprehension and pragmatic awareness. For

example, in the case of the Punjabi test, candidates demonstrate their ability to understand main points in spoken messages, such as following health and safety announcements—mirroring the "can-do" statements for receptive abilities at each CEFR level. This integration not only anchors the test in European best practices but also facilitates recognition and comparability for candidates aiming to qualify for the Greek Public Service Interpreter Register.

4 A Multimodal Language Testing for Interpreter Training: The Punjabi Case

In developing the Punjabi language proficiency test for Greece's first Public Service Interpreter (PSI) Register, one of the most significant linguistic constraints was the dual-script system of Punjabi. According to Singh (2010), Punjabi is spoken by over 125 million people worldwide and is characterized by substantial dialectal diversity, the use of two distinct scripts—Gurmukhi in India and Shahmukhi in Pakistan—and a fragmentation along religious lines. Literacy rates among Punjabi speakers vary considerably (Singh, V., 2017; Pushkarna, M., 2017). In Pakistani Punjab, Punjabi literacy is markedly lower than literacy in Urdu or English. Many native speakers are fluent orally but can read and write only in Urdu or English, if literate at all. This disparity has direct consequences for the Greek PSI context, where the majority of Punjabi speakers originate from rural areas of Pakistani Punjab with comparatively low literacy rates. Moreover, intergenerational differences exacerbate the issue, as younger members of the diaspora—while maintaining strong oral competence—often exhibit limited literacy in either Gurmukhi or Shahmukhi, or even in Urdu, rendering conventional literacy-based assessment inappropriate.

The historical development of the script divide further complicates test construction. According to Murphy (2018), literary production in both scripts was vibrant during the colonial period. The *qissā* tradition, as well as religious and reformist literature, flourished in both systems, with Muslim poets producing highly regarded verse in Shahmukhi during the 18th and early 19th centuries, while significant Sikh literary activity occurred in Gurmukhi. The 1947 partition fundamentally "[…] altered the literary lives of Punjabi and enforced far stricter distinctions between Shahmukhi and Gurmukhi work" (Murphy, 2018: 5), leading to a close alignment between script choice and national borders. In Indian Punjab, Gurmukhi emerged as the dominant medium for modern literary production, particularly in narrative genres such as the short story and novel. As Murphy observes, "[…] modern Punjabi was the vehicle for modern literary creation meant to rework tradition and self-consciously utilize new forms, where Gurmukhi Punjabi dominated and for the most part Urdu supplanted Shahmukhi Punjabi, particularly for narrative forms like the short story and novel" (idem). This growth was supported by state patronage, institutional infrastructure, and a thriving modernist and progressive literary movement. In Pakistani Punjab, by contrast, Urdu displaced Shahmukhi Punjabi in much modern prose. While poetry in Shahmukhi retained some vitality and a large number of chapbooks continued to be produced, literary innovation and volume were restricted by the "alienation" of Muslim Punjabi writers from their language, the absence of state support, and the lack of orthographic standardization.

Derived from Perso-Arabic, the Shahmukhi script is neither a medium of formal instruction nor standardized, displaying considerable variation in spelling and style, which limits accessibility for readers accustomed to Indian Punjabi. As stated previously, its available written corpus is confined largely to classical poetry, a limited

selection of newspapers and online portals, unreliable social media content, and translated public awareness materials. In contrast, Gurmukhi benefits from a highly standardized orthography and a substantial written tradition encompassing newspapers, magazines, fiction and non-fiction, educational materials, religious scriptures, and public information texts. This asymmetry creates a structural imbalance in test design: any literacy-based assessment would privilege Indian Punjabi over Pakistani Punjabi, undermining fairness and validity in the PSI context.

Given these conditions, a conventional proficiency testing model—emphasizing reading and writing skills, and predicated on the assumption of a stable orthographic norm and widespread literacy—would have failed to reflect the communicative realities of PSI work in Greece, which is dominantly oral in nature. To address this, the design team adopted a multimodal assessment framework that foregrounded listening comprehension, rapid information processing, and pragmatic mediation skills. The resulting test was entirely audio-based, with visual prompts employed where appropriate, and comprised of thirteen listening comprehension tasks (50 items in total) built around authentic Punjabi audio excerpts, with stems and responses/options presented in Greek. Item types included image selection, factual information retrieval, and thematic inference, accommodating a wide range of literacy backgrounds and engaging alternative channels for meaning-making. Additional items targeted sociolinguistic and pragmatic competences pertinent to PSI scenarios, such as identifying communicative purposes, recognizing cultural practices, and interpreting public health messages. The test structure—featuring standardized recorded audio and closed-response formats—ensured scalability, security, and equity, while directly addressing the linguistic accessibility issues inherent in Punjabi's dual-script environment.

Beyond image selection, several listening items probe the candidate's ability to extract specific factual information. For instance, questions ask candidates to identify geographical references, count elements or retain numerical information mentioned in the narrative, or determine the thematic setting (urban, rural, or industrial) described in the audio. These tasks assess both literal and inferential comprehension, key to interpreting where information must often be inferred from context rather than explicitly stated.

The test also includes items designed to test sociolinguistic and pragmatic awareness. In some cases, the candidate must identify the communicative purpose of an audio text (e.g., distinguishing between a poem, a public announcement, or a charity message) or identify the intended audience and/or register. Others focus on cultural knowledge and domain-specific concepts relevant to PSI, such as recognizing traditional customs like *vaari* in Punjabi weddings, or understanding public health messages about polio vaccination. Such content mirrors real-life PSI scenarios, where interpreters frequently mediate communication in healthcare, legal, and community service contexts.

A further group of items addresses public health and safety communication, such as recognizing advice related to medical consultations, COVID-19 prevention, or the identification of vulnerable groups in emergencies. By including such material, the test embeds topical and domain-specific vocabulary alongside functional listening skills, ensuring that candidates demonstrate readiness for high-stakes communicative settings.

The structure of the Punjabi test embodies the multimodal principle at multiple levels. The use of recorded audios, sourced from reliable authentic sources and adapted to meet the objectives of each test item, ensures standardization across administrations while reducing examiner bias (Luoma, 2004). The integration of visual prompts provides alternative pathways to meaning-making for candidates with limited literacy, while also simulating the multimodal nature of real interpreting encounters, where meaning is often constructed through the interplay of verbal, visual, and contextual cues. Closed-response formats with predetermined correct answers allow for scalability and secure marking. Both are equally important given that many test administrators may not have proficiency in Punjabi.

AI-supported tools played a role in both the design and validation phases. They were used in generating images and in creating and vetting distractors for multiple-choice items with a view to ensuring plausible but clearly incorrect options, thereby enhancing discrimination power. AI-generated outputs were then submitted to human validation by our two language experts in coordination with the academic expert before administering the test.

From a validity perspective, the Punjabi test addresses Messick's (1996) tripartite concerns. Content validity is ensured by embedding authentic communicative content drawn from domains central to PSI—healthcare, legal procedures, public services, and community life. Construct validity is supported by the focus on listening comprehension, pragmatic judgment, and culturally situated inference, all of which are integral to interpreter competence at the pre-training stage. Consequential validity is addressed through multimodal design, which not only measures what is intended to measure but also mitigates bias against candidates with limited script literacy, thereby supporting fairness, equity and positive washback, as will be demonstrated in the discussion section.

The decision to exclude free oral or written production was deliberate. While this, in Bachman's conceptualization, reduces authenticity in terms of full-skill coverage (Bachman, 1991)[2], it aligns with the operational requirement for an objective, standardized, criterion-based, and easily administrable assessment tool across ten languages, some of which have no standard written form. Productive skills will have to be assessed later, during interpreter training, once candidates have been admitted to the Register and can be evaluated in controlled, simulation-based interpreting tasks.

The sequencing of items within the Punjabi section reflects a graduated approach, beginning with relatively straightforward image-audio matching and moving toward more complex tasks requiring inferential reasoning, recognition of communicative function, and domain-specific problem-solving. This progression not only scaffolds candidate performance: it attempts to mirror, be it partially, the cognitive demands of interpreting, where interpreters must first comprehend the literal message, then assess its pragmatic force, and finally act on it within institutional constraints, in this case the time limitations of an exam.

The Punjabi case thus offers a replicable model for other languages in the Greek PSI context that face similar challenges, such as Somali, certain Kurdish varieties, or other oral-dominant languages with fragmented literacy traditions. It demonstrates that

---

[2] The Punjabi test meets the criterion of *situational authenticity* as its tasks accurately represent language use within the PSI domain. However, *interactional authenticity* is only partly achieved because the lack of free oral and written production prevents an evaluation of the cognitive and metacognitive strategies that candidates would employ in such activities.

multimodal, AI-supported assessment can reconcile the competing demands of fairness, operational feasibility, and construct validity in a multilingual testing environment. While further research is needed to integrate secure, scalable oral production assessments, the current model provides a robust pre-entry filter that identifies candidates with the receptive and interpretive language competencies necessary for successful progression into interpreter-specific training.

5   Discussion of the Punjabi Test

The Punjabi language proficiency test, developed and implemented within the framework of Greece's Public Service Interpreter (PSI) Register, represents an innovative yet methodologically complex case of multimodal assessment in a low-resource, high-stakes linguistic environment. Drawing on the project's final deliverable, this section critically evaluates the test's design, administration, and broader implications. The appraisal considers both the advantages achieved and the limitations encountered, situating these findings within current research on language testing, assessment validation, and the integration of artificial intelligence (AI) in multilingual proficiency contexts.

5.1   Advantages

A key strength of the Punjabi test lies in its multimodal format, which combined written, aural (audio), visual, and AI-generated inputs. This approach enhanced accessibility for candidates with widely varying literacy levels, dialectal backgrounds, and script familiarity (Gurmukhi and Shahmukhi) by bypassing the latter. By embedding authentic materials—such as adapted YouTube transcriptions and contextually relevant video content—the test mirrored some of the relevant pragmatic demands of PSI but not all (e.g. asylum interviews). Visual mediation, supported by Greek as a bridge language, mitigated comprehension barriers and reflected the real-world conditions under which public service interpreters operate.

All languages in the PSI testing framework, including Punjabi, shared identical formats, item counts, difficulty ranges and progression, as well as marking criteria. Candidates were required to achieve minimum threshold scores (40 out of 50) in both Greek and their relevant native language(s) (candidates could sit more than one language combination), ensuring balanced bilingual/multilingual proficiency. Five parallel test forms were created for most languages (a set of three for Punjabi and Somali), each modular in design to facilitate repeated administration over time. This structure allowed for efficient rotation of items while controlling development costs, an important factor in sustaining assessments for low-resource languages.

AI tools were employed for lexical frequency analysis, phonetic variation, transcription, speech-to-text and text-to-speech verification, translation, and image generation to verify appropriateness of authentic materials and support listening comprehension prompts. In a context where proprietary, high-quality resources are scarce, this enabled the production of authentic, level-appropriate materials at scale. AI's capacity to adapt prompts and generate distractors provided a flexible means of addressing the diverse linguistic profiles of candidates.

5.2   Limitations of the Study

Despite its multimodal sophistication, the Punjabi test operated primarily as a summative, high-stakes examination. Candidates who sat the first exam were provided with a feedback form for each language test. Their comments in combination with success rate per language have been valuable, as they allow us to revisit the tests individually but also as a unified whole. More specifically, in the case of the Punjabi test, the only exclusively audio-and-image based test, the success rate was 89% (out of 9 participants in total) and the feedback was overwhelmingly positive: candidates found that the test was well-structured and reasonably timed as well as content and level appropriate.

However, the sole round of piloting, involving a small control group (consisting of the two language experts and the academic coordinator), enabled small-scale revisions before final deployment. Candidates did not receive detailed performance feedback or guidance for targeted skill development, nor were any re-take opportunities built into the system. This is mainly because the test was not associated with a follow-up training program. Consequently, the test offered limited scope for fostering ongoing competence growth, a significant gap given the dynamic demands of PSI work.

Methodological rigor was further undermined by the absence of a fully independent control group. The small pool of qualified Punjabi speakers in Greece meant that many individuals who participated in the pilot phase would later risk becoming candidates, thereby breaching standard validation protocols. Involving language experts based outside Greece was not an option because the legal terms of the agreement required team members to hold legal and professional status within the country; such involvement was also ruled out on confidentiality and financial grounds. This overlap introduced the risk of bias in item calibration, potentially inflating reliability estimates and limiting the generalizability of the results to other cohorts or contexts. In the absence of multiple piloting rounds, problematic or culturally biased items may have gone undetected, undermining both fairness and predictive validity.

While AI contributed to material development, its application was constrained by several factors. First, reliance on commercially licensed tools—rather than university-managed platforms—raised significant data security concerns. Test materials potentially processed on external servers without full institutional oversight or actual knowledge would pose risks to confidentiality and compliance with legal frameworks such as the EU General Data Protection Regulation (GDPR). Second, the lack of a dedicated, locally governed AI infrastructure limited customization for assessment-specific privacy needs, including anonymization protocols and restricted administrative access. Finally, the project's compressed timeline curtailed the possibility of more robust AI integration, iterative testing, and refinement.

5.3   Implications

The limitations identified above have important implications for future PSI assessment design. The absence of embedded feedback mechanisms weakens alignment with professional interpreting practice, where performance review and targeted skill development are central to quality assurance. Validation constraints point to the need for independent, sufficiently large control groups—even in low-resource language contexts—achievable through strategic synergies with diasporic communities or international academic partners. From a technological standpoint, the ethical and legal

challenges of relying on commercial AI tools highlight the urgency of investing in secure, institutionally controlled platforms that support both material generation and data governance in compliance with privacy regulations.

Despite these constraints, the Punjabi multimodal proficiency test achieved significant innovations in format, accessibility, and fairness within an exceptionally challenging sociolinguistic and logistical environment. Its standardized structure and authentic, context-driven materials set a valuable precedent for PSI assessment in other low-resource languages. Addressing current gaps—by integrating formative assessment elements, securing independent validation mechanisms, and building local technological capacity—will strengthen methodological robustness and replicability, aligning future PSI language testing with contemporary assessment scholarship and the professional realities of public service interpreting.

Conclusion

The development and deployment of a multimodal language proficiency test for Punjabi public service interpreters in Greece has highlighted both the urgent need and the inherent complexities involved in credentialing interpreters in migration and asylum contexts. By integrating authentic audio materials and AI-generated visual prompts, the test addresses critical concerns of fairness, reliability, and inclusivity, accommodating the diversity of Punjabi dialects and varying literacy levels among candidates in a time-efficient way. However, methodological challenges—including limited access to qualified test subjects, restrictions on independent validation, and constraints related to data privacy, and AI infrastructure—underscore the difficulties of implementing robust, adaptable assessment models under real-world conditions. Despite these limitations, the approach offers a promising blueprint for other rare and minority languages, demonstrating that multimodal and AI-supported solutions can enhance both the effectiveness and scalability of interpreter language proficiency testing. Ultimately, such innovations are instrumental in formalizing interpreter roles, protecting the rights of asylum seekers and refugees, and strengthening the institutional foundations of public service interpreting in Greece and beyond.

References

Ager, A., & Strang, A. (2008). Understanding integration: A conceptual framework. Journal of Refugee Studies, 21(2), 166–191. https://doi.org/10.1093/jrs/fen016

Albl-Mikasa, M. (2012). The importance of being not too earnest: a process- and experience-based model of interpreter competence. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies* (Vol. 3, pp. 62-67). John Benjamins.

——— (2013). Developing and cultivating expert interpreter competence. *The Interpreters' Newsletter*, 18, 17–34. https://core.ac.uk/download/pdf/149227171.pdf

——— (2024). English as a lingua franca in interpreting and translation: Professional insights and pedagogical implications. Frontiers in Communication, 9, 123456. https://doi.org/10.3389/fcomm.2024.123456

Angelelli, C. V. (2007). Assessing medical interpreters: The language and interpreting testing project. *The Translator, 13*(1), 63–82.

Angelelli, C. V., & Jacobson, H. E. (Eds.). (2009). Testing and assessment in translation and *interpreting studies: A call for dialogue between research and practice*. John Benjamins. https://doi.org/10.1075/btl.80

Apostolou, F. (2012). Interpreting services for immigrants: A new reality in Greece. *The Interpreters' Newsletter, 17*, 213–222.

——— (2015). *Translation and interpreting for public services in Greece* [Chapter]. In F. Apostolou, *Η μετάφραση και η διερμηνεία για δημόσιες υπηρεσίες στην Ελλάδα* [*Translation and interpreting for public services in Greece*] [Undergraduate textbook]. Kallipos, Open Academic Editions. https://hdl.handle.net/11419/960

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Bachman, L. (1991). What does language testing have to offer? TESOL Quarterly, 25, 671–704.

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

Baixauli-Olmos, L. (2017). Ethics codes as tools for change in public service interpreting. *The Journal of Specialised Translation*, 28. https://jostrans.soap2.ch/issue28/art_baixauli.php

Born, A. D., Oostrom, M. P., van Vugt, J. K., & Mark, N. (2019). Psychological individual differences as predictors of refugees' local language proficiency. International Journal of Intercultural Relations, 70, 162–175.

Bouroutis, A. (2022). Μετανάστευση και στρατηγικές για την ένταξη στην Ελλάδα. Η εμπειρία των ετών 2015–2022 [Migration and strategies for integration in Greece: The experience of the years 2015–2022]. *Ελληνική Επιθεώρηση Πολιτικής Επιστήμης* [Hellenic Review of Political Science], 47, 57–84.

Bradlow, A. R., & Bent, T. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. The Journal of the Acoustical Society of America, 106(2), 1023–1039. https://doi.org/10.1121/1.427115

Cheung, A. K. F. (2013). Non-native accents and simultaneous interpreting quality perceptions. *Interpreting, 15*(1), 25–47. https://doi.org/10.1075/INTP.15.1.02CHE

Compton, D. L., Gilbert, J. K., Kearns, D. M., & Olson, R. K. (2020). Using an item-specific predictor to test the dimensionality of the orthographic choice task. *Annals of Dyslexia, 70*(2), 243–258. https://doi.org/10.1007/s11881-020-00202-0

Cooper, G., Davies, R. and Tung, R.L. 1982. "Interpreting stress: Sources of job stress among conference interpreters." *Multilingua 1(2)*, 97–107.

Council of Europe. (2008a). *Raising young refugees' voices in Europe and beyond. Strasbourg: Council of Europe.* https://rm.coe.int/2008-raising-young-refugees-voices/1680702361

——— (2008b). *The role of languages in the integration of adult migrants. Strasbourg: Council of Europe.* https://rm.coe.int/16802fd54a

——— (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

——— (2023). Integration of migrants and refugees: Benefits for all parties involved. https://rm.coe.int/integration-of-migrants-and-refugees-benefits-for-all-parties-involved/1680aa9038

——— (2023). Languages and the integration of adult migrants. Council of Europe Publishing.

Daley, A. H. (2019). Language proficiency and immigrants' economic integration. IZA World of Labor, 457. https://doi.org/10.15185/izawol.457

Daley, C. (2019). Migrants and refugees: Exploring integration models in Canada and Scotland. *Churchill Fellowship Report*. https://centred.scot/wp-content/uploads/2022/05/Clare-Daly-Churchill-Report.pdf

Dendrinos, B., Karavas, K., (Eds.). (2013). The KPG Handbook: Performance Descriptors and Specifications. Athens: RCeL Publications, National and Kapodistrian University of Athens.

Dimitriadi, A., et al. (2019). Borders and the mobility of migrants in Greece. CEASEVAL Project Report. https://www.eliamep.gr/wp-content/uploads/2019/12/Borders-the-mobility-of-migrants_CeaSEVAL.pdf

Donovan, C. (2005). Teaching simultaneous interpretation into B: Training priorities. In M. Lambert & S. Moser-Mercer (Eds.), *Bridging the Gap: Empirical Research in Simultaneous Interpretation* (pp. 222–231). Amsterdam: John Benjamins.

Dryzek, J. S. (1997). *The politics of the earth: Environmental discourses* (1st ed.). Oxford University Press.

Elder, C., & McNamara, T. (2015). The hunt for "indigenous criteria" in assessing communication in the physiotherapy workplace. *Language Testing, 32*(1), 1–22. https://doi.org/10.1177/0265532215607398

Evangelinidis, A. (2016). The Greek State's Response to the Refugee Crisis and the Solidarity Movement. *Contemporary Southeastern Europe, 3*(1), 45-53. https://contemporarysee.org/allfiles/documents/45-the-greek-states-response-to-the-refugee-crisis-and-the-solidarity-movement.pdf

Fairclough, N. (1989). *Language and power*. Longman.

——— (1992). *Discourse and social change*. Polity Press.

Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics, 23*(4), 567–598.

Foged, M., & van der Werf, C. (2023). Access to language training and the local integration of refugees. Journal of Labor Economics, 41(1), 221–260.

Foucault, M. (1969). *L'archéologie du savoir* [*The archaeology of knowledge*]. Gallimard.

————— (1971). *L'ordre du discours* [*The order of discourse*]. Gallimard.

Fowler, R. (1991). *Language in the news: Discourse and ideology in the press*. Routledge.

Fragkou, E. (2023). *Public Service Interpreting in Greece and Around the World. The Profile of the Greek Public Service Interpreter*. In Migration and Asylum Interpreting Project - 3rd Predefined Action "Enhancing and Building-up national capacity of Migration and Asylum strategic planning" of "Programme G – Capacity Building of National Asylum and Migration Management Systems" (EEA 80629).

Gile, D. (2005). *Regards sur la recherche en interprétation*. Paris: Presses Universitaires de France.

Gentile, A., Ozolins, U., & Vasilakakos, M. (with Ko, L., & Ton-That, Q.-D.). (1996). *Liaison interpreting: A handbook*. Melbourne: Melbourne University Press.

Ginther, A., & Stevens, J. (1998). Language background, ethnicity, and the internal construct validity of the Advanced Placement Spanish Language Examination. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 169–194). Mahwah, NJ: Lawrence Erlbaum.

Golestani, N., Rosen, S., & Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: The contribution of semantics. Bilingualism: Language and Cognition, 12(3), 385–392. https://doi.org/10.1017/S1366728909990147

Gu, L. (2014). At the interface between language testing and second language acquisition: Language ability and context of learning. *Language Testing, 31*(1), 111–133.

Hajer, M. A. (1995). *The politics of environmental discourse: Ecological modernization and the policy process*. Oxford: Clarendon Press.

Hale, S. (2007). *Community interpreting*. Basingstoke: Palgrave Macmillan.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.

Harris, Z. S. (1952). Discourse analysis. *Language, 28*(1), 1–30.

Heath, M., Bainbridge, A., Chitah, B., et al. (2023). Interpreter services and effect on healthcare: A systematic review. *BMC Health Services Research, 23*, 26. https://pmc.ncbi.nlm.nih.gov/articles/PMC9932446/

Hellenic Republic, Ministry of Migration & Asylum. (2019). *National Integration Strategy* [Εθνική Στρατηγική για την Κοινωνική Ένταξη]. Athens: Ministry of Migration & Asylum. https://migration.gov.gr/wp-content/uploads/2022/09/NATIONAL-STRATEGY-FINAL.pdf

Hlavac, J. (2015). Formalizing community interpreting standards: A cross-national comparison of testing systems, certification conventions and recent ISO guidelines. *International Journal of Interpreter Education, 7*(2), 47-68. https://open.clemson.edu/ijie/vol7/iss2/4

International Organization for Standardization. (2022). *ISO 23155:2022—Interpreting services: Conference interpreting—Requirements and recommendations*. Geneva: ISO. https://www.iso.org/standard/74749.html

Ioannidis, A. (2025). Quality assurance in public service interpreting: An overview of the Greek Registry of Public Service Interpreters. *International Journal of Language, Translation and Intercultural Communication, 9*. https://doi.org/10.12681/ijltic.38695

Isphording, I. E. (2015). What drives the language proficiency of immigrants? IZA World of Labor, 177. https://doi.org/10.15185/izawol.177

Kalina, S. (2000). Interpreting competences as a basis and a goal for teaching. *The Interpreters' Newsletter*, 10, 3–32. University of Trieste. http://hdl.handle.net/10077/2440

Kontis, A. (Κόντης, Αντώνης). (2013). Elements of integration policy: Greek integration policy between Europeanism and ethno-culturalism [Στοιχεία Ενταξιακής πολιτικής: Η Ελληνική ενταξιακή πολιτική μεταξύ Ευρωπαϊσμού και Εθνοπολιτισμού]. In D. Balourdos & I. Tsiganou (Eds.), Meta-analysis of research on the social integration of immigrants in Greek society [Μετα-ανάλυση ερευνών για την κοινωνική ένταξη των μεταναστών στην ελληνική κοινωνία] (pp. 6–27). Athens: Papazisis [Αθήνα: Παπαζήσης].

Kotoulas, I.E. (2024). A new EU migration policy and Greece's perspective. *Weekly Briefing, Vol. 77, No. 4 (GR)*, China-CEE Institute. https://china-cee.eu/2024/12/19/greece-external-relations-briefing-a-new-eu-migration-policy-and-greeces-perspective/

Kuleva, M., & Miladinov, O. (2024). Exploring the efficacy of online proctoring in online examinations: A comprehensive review. *Environment. Technologies. Resources. Proceedings of the International Scientific and Practical Conference*, *2*, 27–33. https://doi.org/10.17770/etr2024vol2.8058

Kurz, I. (2008). The impact of non-native English on students' interpreting performance. In G. Hansen, A. Chesterman, & H. Gerzymisch-Arbogast (Eds.), *Efforts and models in interpreting and translation research: A tribute to Daniel Gile* (pp. 187–200). John Benjamins Publishing Company. https://doi.org/10.1075/btl.80.17kur

———— (2009). The impact of non-native English on students' interpreting performance. In I. Mason (Ed.), *Dialogue interpreting: A guide to interpreting in public services and community settings* (pp. 199–214). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.80.15kur

Kusiak, M. (2001). The effect of metacognitive strategy training on reading comprehension and metacognitive knowledge. *EUROSLA Yearbook, 1*(1), 255–274. https://doi.org/10.1075/eurosla.1.19kus

Laboratory for the Study of Migration and Diaspora (E.M.ME.DIA.). (2009). Issues of Social Integration of Migrants (A. Kontis, Ed.). Athens: Papazisis. (Special Studies on Migration and Diaspora Series, No. 4). ISBN: 978-960-02-2329-3

Lai, M., & Mulayim, S. (2010). Training refugees to become interpreters for refugees. *Translation & Interpreting*, 2(1), 48–65.

———— (2013). Training interpreters in rare and emerging languages: The problems of adjustment to a tertiary education setting. In C. Schäffner, K. Kredens, & Y. Fowler (Eds.), *Interpreting in a changing landscape* (pp. 287–303). Amsterdam & Philadelphia: John Benjamins.

Lev-Ari, S. (2015). Comprehending non-native speakers: theory and evidence for adjustment in manner of processing. *Frontiers in Psychology, 6*, 1546. https://doi.org/10.3389/fpsyg.2014.01546

Lim, H.-O. (2005). Working into the B language: The condoned taboo? *Meta: Translators' Journal*, 50(4), 153–167. https://doi.org/10.7202/019870ar

Lin, I.-I. (2013). The impact of non-native accented English on rendition accuracy in simultaneous interpreting. Translation & Interpreting, 5(2), 79–93. https://www.trans-int.org/index.php/transint/article/view/210

Loiseau, N., & Delgado Luchner, C. (2019). Contextualizing interpreter training in Africa: Two case studies from Kenya. *International Journal of Interpreter Education, 11*(2), 4–15.

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

Manna, V. F., & Yoo, H. (2015). Investigating the relationship between test-taker background characteristics and test performance in a heterogeneous English-as-a-second-language (ESL) test population: A factor analytic approach. *ETS Research Report Series, 2015*(1), 1–16. https://doi.org/10.1002/ets2.12072

Manousou, E., & Rassia, K. (2024). Online examinations, reliability and quality issues. *Art Human Open Access Journal, 6*(3), 173–179. https://doi.org/10.15406/ahoaj.2024.06.00259

Marianacci, A. (2024). "I know – that's the system. The system is wrong": Allyship and social justice in community interpreting from a service-user perspective. Linguistica Antverpiensia, New Series: Themes in Translation Studies, 23, 116–137.

Martin, F., Claydon, E., Morton, A., Binns, S., & Pratt, C. (2003). The development of orthographic and phonological strategies for the decoding of words in children. *Journal of Research in Reading, 26*(2), 151–163. https://doi.org/10.1111/1467-9817.00196

Messick, S. (1996). Validity and washback in language testing. Language Testing, 13(4), 241–256. https://doi.org/10.1177/026553229601300302

Mikkelson, H. (2004, October 25). The professionalization of community interpreting. *AIIC*. https://aiic.net/p/1546

Ministry of Migration and Asylum. (2021). National Strategy for Integration [Εθνική Στρατηγική για την Ένταξη]. Athens: Hellenic Republic [Αθήνα: Ελληνική Δημοκρατία].

Ministry of Migration Policy. (2019). National Strategy for Integration [Εθνική Στρατηγική για την Ένταξη].

Montemitro, C., Bellomo, A., Casucci, P., Di Fabio, F., Fulcheri, M., Cannavò, M., & Brasso, C. (2021). Language proficiency and mental disorders among migrants: A systematic review. European Psychiatry, 64(S1), S424–S425. https://doi.org/10.1192/j.eurpsy.2021.1128

Moratto, R., & Li, D. (Eds.). (2021). *Global insights into public service interpreting: Theory, practice and training*. Routledge. https://doi.org/10.4324/9781003197027

Murphy, A. (2018). Writing Punjabi across borders. *South Asian History and Culture, 9*(1), 68–91. https://doi.org/10.1080/19472498.2017.1411049

Naskou-Perraki, P., Papageorgiou, G., & Baxevanis, Ch. (2017). Πρόσφυγες και αιτούντες άσυλο: Παγκόσμια, ευρωπαϊκή και εθνική διάσταση [Refugees and asylum seekers: Global, European, and national dimension]. Athens–Thessaloniki: Sakkoulas. ISBN: 978-960-568-570-6.

Norström, E., Gustafsson, K., & Fioretos, I. (2011). Interpreters in Sweden—a tool for equal rights? In *Gramma: Journal of Theory and Criticism, 11*, 183–201. https://www.enl.auth.gr/gramma/gramma11/Norstrom_Gustafsson_Fioretos.pdf

O'Sullivan, B. (2012). Assessment issues in languages for specific purposes. *The Modern Language Journal*, *96*(S1), 71–88. https://doi.org/10.1111/j.1540-4781.2012.01298.x

Olson, D. J. (2023). A systematic review of proficiency assessment methods in bilingualism research. *International Journal of Bilingualism*, 1–36. https://doi.org/10.1177/13670069231153720

Pöchhacker, F. (2009). Issues in interpreting studies. In J. Munday (Ed.), *The Routledge companion to translation studies*). London: Routledge.

——— (2004). *Introducing interpreting studies*. London & New York: Routledge.

Pokorn, N. K. (2009). *Challenging the tradition: Translation into a non-mother tongue*. Amsterdam: John Benjamins.

——— (2009). Natives or Non-natives? That Is the Question. Translation Studies, 2(2), 185–198. https://doi.org/10.1080/14781700902943622

Pushkarna, M. (2017). Literacy patterns in Punjab: Rural–urban differentials. *IOSR Journal of Humanities and Social Science, 22*(7, Ver. 5), 32-38.

Schacht, D., Harris, S. M., et al. (2021). Syrian refugees' preferences and perceived barriers. *International Journal of Environmental Research and Public Health, 18*(23), 12413. https://pmc.ncbi.nlm.nih.gov/articles/PMC8669439/

Schacht, D., Martinovic, B., & Kappmeier, M. (2022). Language as a precondition for social integration of migrants in the educational context. In A. Haider, K. Mertens, & A. Demuth (Eds.), *The challenge of radicalization and extremism: Integrating research on education and citizenship in the context of migration* (pp. 153–181). Brill. https://doi.org/10.1163/9789004525658_007

Schmitz, J. R. (2013). The native speaker and nonnative speaker debate. *Calidoscópio*, 11(2), 135–152. https://doi.org/10.4013/cld.2013.112.07

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Setton, R., & Dawrant, A. (2016). *Conference interpreting: A complete course*. Amsterdam & Philadelphia: John Benjamins B.V.

Singh, G. (2010). Globalization and Change: Perspectives from Punjab. Rawat Publications.

Singh, V. (2017). Urban and rural literacy in Punjab and Haryana: A comparative current scenario and changes during the last four decades. *International Journal of Science and Research, 6*(3), 2242–2249.

Skaaden, H. & Wadensjö, C. (2014). Some considerations on the testing of interpreting skills. In C. Giambruno (ed.). *Assessing legal interpreter quality through testing and certification: the Qualitas project* (pp. 17-26). Universidad de Alicante.

Triandafyllidou, A., & Maroufof, M. (2017). *Migration in Greece: Recent Developments*. European University Institute. https://www.eliamep.gr/wp-content/uploads/2018/10/Greece-report-for-OECD_2-Nov-2017.pdf

Urdal, G. H. S. (2024). Making a difference: How does an interpreting course contribute to public service interpreters' professional identity? *FITISPos International Journal*, 11(2). https://doi.org/10.37536/FITISPos-IJ.2024.11.2.371

Vlachopoulos, S., & Ioannidis, A. (2023). Remote interpreting for the public service: From simple language service to an integrated framework of intercultural communication [Η εξ αποστάσεως διερμηνεία στις δημόσιες υπηρεσίες. Από την απλή γλωσσική υπηρεσία στο ολοκληρωμένο πλαίσιο διαπολιτισμικής επικοινωνίας]. Athens: Papayisi Publishing [Αθήνα: Εκδόσεις Παπαζήση].

——— (2016). *Κοινοτική διερμηνεία στην Ελλάδα: Κτίζοντας το μέλλον* [*Community interpreting in Greece: Building the future*]. TEI of Epirus. https://zenodo.org/records/161349

Vyzas, T. (2025). Interpreting for public services and the community: Towards conceptual accuracy and terminological stability. *International Journal of Language, Translation and Intercultural Communication, 9*. https://doi.org/10.12681/ijltic.39033

Wen, T. H., & Al-Bekai, W. (2023). Native vs. non-native: A study on simultaneous interpreting in the United Nations Security Council. In Corpus-based interpreting studies: UN testing ground (pp. 138–159). Routledge.

Wen, Y., Qiu, J., Leong, V., & van Heuven, W. J. B. (2023). LexCHI: A quick lexical test for estimating language proficiency. *PLoS ONE, 18*(7), e0287670. https://doi.org/10.1371/journal.pone.0287670

Zarić, J., Hasselhorn, M., & Nagler, T. (2020). Orthographic knowledge predicts reading and spelling skills over and above general intelligence and phonological awareness. *European Journal of Psychology of Education, 35*, 645–664. https://doi.org/10.1007/s10212-020-00464-7