Volume 35 Issue 1 2026

Applied Language Learning

A Practice-Based Journal for Foreign Language Educators

Special Issue: Language, Regional Expertise, & Culture (LREC)
with the U.S. Military Academy, West Point

Defense Language Institute
Foreign Language Center
Presidio of Monterey

Applied Language Learning

The mission of Professional Bulletin 65, ***Applied Language Learning*** (US ISSN 1041-679X and ISSN 2164-0912), is to provide a forum for the exchange of ideas and information on instructional methods and techniques, curriculum and materials development, assessment of needs within the profession, testing and evaluation, and implications and applications of research from related fields such as linguistics, education, communications, psychology, and the social sciences.

Applied Language Learning, published by the Defense Language Institute Foreign Language Center and Presidio of Monterey, California, presents professional information. The views expressed herein are those of the authors, not the Department of War or its elements. The content does not necessarily reflect the official U.S. Army position and does not change or supersede any information in official U.S. Army publications.

Further reproduction is not advisable. Whenever copyrighted materials are reproduced in this publication, copyright release has ordinarily been obtained only for use in this specific issue. Requests for reprints should be directed to the authors.

To access *Applied Language Learning* online:
https://www.dliflc.edu/ojs/ALL/index

To communicate with the Editor:
Send correspondence to the Editor at journals@dliflc.edu

Special Issue: Instructing Language Regional Expertise, and Culture (LREC)

This issue represents the partnership and collaboration between the Defense Language Institute Foreign Language Center (DLIFLC) and the U.S. Military Academy, West Point. Since the mid-2000s, the Department of Defense (DoD, now the Department of War, DoW) has elevated the importance of knowledge, skills, and attitudes pertaining to foreign language, intercultural, and regional competencies. To date, the LREC enterprise across the DoW has made strategic improvements to the way we teach languages and cultures in the context of America's military outreach across the globe. This special issue of the *Applied Language Learning* journal brings together scholarship from both the Defense Language Institute and the Service Academies to discuss best practices for instructing world languages, cultures, and regional dynamics at all levels of proficiency.

Special Issue Co-Editors:
Dawn Bikowski, Ph.D., Associate Professor, Defense Language Institute Foreign Language Center
Jeffrey Watson, Ph.D., Professor of Applied Linguistics, U.S. Military Academy

Editor-in-Chief

Dawn Bikowski

Editorial Board

Natalia Barley — *Defense Language Institute Foreign Language Center*

Christina Campbell — *Professor Emerita, Defense Language Institute Foreign Language Center*

Myoyoung Kim — *Defense Language Institute Foreign Language Center*

Pete Swanson — *U.S. Air Force Academy*

Jeffrey Watson — *U.S. Military Academy, West Point*

Rong Yuan — *Defense Language Institute Foreign Language Center*

Shen S. Zhu — *Defense Language Institute Foreign Language Center*

Contents

Introduction to the Special Issue

1 Foreword from the Defense Language Institute Foreign Language Center
~ *Robert Savukinas*

3 Foreword from the U.S. Military Academy, West Point
~ *E. John Gregory*

5 Instructing Language, Regional Expertise, and Culture
~ *Jeffrey Watson & Dawn Bikowski*

Technology

Research

13 Generative AI in Foreign Language Education: A Scoping Review of Pedagogical Applications and Challenges
~ *Yuliya Kozlova*

41 World Language Instructors' Perceptions of Using Generative AI Tools
~ *Pete Swanson & M. Ryan Richey*

Action Research

62 Revolutionizing Language Learning: AI-Assisted Learning as a Catalyst for High Proficiency Attainment
~ *Hanwei Tan*

79 An Artificial Intelligence-Enabled Tool for Real-Time Intercultural Competence Instruction
~ *Tara L. Schendel & Inna Kerlin*

Pedagogy

Research

96 The Impact of Linguistic Knowledge on Reading and Listening Comprehension Across ILR Proficiency Levels
~ *Sun-Kwang Bae*

Action Research

110 Integrating a Streaming Series into the Curriculum to Foster Listening Proficiency and Cultural Competence
~ *Jiayi Xu, Maria Wang, & Nina Bozeman*

Action Research, cont.

127 Integrating Cultural Competence into Army Special Operations Forces Basic Language Training

~ *Tong Sun, Kelly K. Lemmons, Inna Ruble, & Irina V. Emeliyanova*

Commentary

148 A Letter from Valley Forge and Words of the Developing World: Knowing the "Why" for Autonomous Language Learning

~ *Daniel Seibel*

DATA-DRIVEN TEACHING

Research

154 Language Aptitude as Predictor of Performance

~ *John Pendergast, Thomas Dean, & Robert Jannusch*

170 Beyond News and Documentaries: Developing a Corpus-Based Lexical Resource for Informal North Korean Speech

~ *Mi Hye Lee & Myoyoung Kim*

191 Motivational Dynamics among Persian Farsi Language Learners Across Two Military Institutions

~ *Olivia Holloway, Siyi Gao, Ra'ed Qasem, & Zachary Miller*

Commentary

209 Framing Intercultural Reflection in the Language Classroom

~ *Anjel Tozcu & Clarissa Eagle*

GENERAL INFORMATION

217 Call for Papers

218 Thank You Reviewers

Foreword from the Defense Language Institute Foreign Language Center

Robert Savukinas, Ed.D.
Provost, Defense Language Institute Foreign Language Center, Monterey, California

This special issue of *Applied Language Learning* arrives at a pivotal moment for Language, Regional Expertise, and Culture (LREC) across the Department of Defense and the broader language education community. Over the past two decades, LREC has evolved from a collection of complementary capabilities into a fully integrated framework essential to navigating today's complex, dynamic, and culturally mediated operational environments.

The articles in this issue reflect that evolution. Taken together, they demonstrate a clear shift from viewing language proficiency, regional knowledge, and cultural competence as discrete domains toward understanding them as mutually reinforcing dimensions of a single human capability. This integrated perspective underscores a central insight: success in contemporary missions depends not only on what individuals know, but on how they interpret, adapt, and act within unfamiliar sociocultural systems.

A defining theme that emerges throughout this volume is the importance of transformative instruction. Language learning is not conceptualized as the acquisition of vocabulary and grammar alone; it is a developmental process that challenges assumptions, reshapes perspectives, and builds adaptive leaders. Grounded in cultural relativism, effective LREC instruction enables learners to engage meaningfully with difference—cultivating not just proficiency, but empathy, flexibility, and ethical awareness. These qualities are indispensable in environments where influence, communication, and trust may be more decisive than force.

Equally prominent is the role of innovation in pedagogy, particularly through the integration of emerging technologies. Contributions examining generative artificial intelligence illustrate both its promise and its complexity. AI-enabled tools can expand opportunities for personalized learning, simulated interaction, and targeted feedback, offering powerful means to support high levels of proficiency and engagement. At the same time, their effective use depends on thoughtful implementation, informed instructors, and a clear alignment with pedagogical goals. Technology, as this issue emphasizes, is not a substitute for instruction but a force multiplier when grounded in sound educational principles.

The issue also illuminates enduring challenges within LREC education. Integrating culture-general and culture-specific competencies remains an ongoing tension, as does the difficulty of assessing intercultural development in meaningful ways. Similarly, while regional expertise is widely recognized as essential, it is not sufficient on its own; deep knowledge must be paired with the adaptability required to operate effectively across contexts. These challenges point to the inherent complexity of preparing individuals for real-world intercultural engagement.

Importantly, several studies included here highlight the human dimensions of learning—motivation, autonomy, and resilience. Findings on learner engagement, grit, and the importance of understanding the "why" and the "so what" behind language study reinforce a simple but powerful truth: transformative outcomes depend as much on the learner's internal drive as on instructional design. Instructors, therefore, play a critical role not only in teaching content, but in fostering purpose and sustaining commitment.

In aggregate, this special issue offers more than a collection of studies; it presents a coherent vision for the future of LREC. It calls for approaches that are integrative rather than fragmented, experiential rather than purely theoretical, and transformative rather than transactional. It affirms that developing linguistically capable, culturally agile, and regionally informed professionals is not merely an educational objective—it is a strategic necessity.

I hope this special issue of *Applied Language Learning* contributes meaningfully to ongoing discussions within the LREC community and beyond. More importantly, I hope it supports instructors, program designers, and leaders in their efforts to prepare learners to communicate effectively, adapt intelligently, and lead responsibly in an increasingly interconnected and complex world.

Foreword from the U.S. Military Academy, West Point

Colonel E. John Gregory
Professor of Chinese and Department Head, Department of English & World Languages
West Point, New York

This issue of *Applied Language Learning* reflects Charles Taylor's broader understanding of language learning that is captured in his description of the human being as fundamentally a "language animal." The contributions gathered here examine emerging technologies, language aptitude, cross-cultural competence, learner autonomy, and authentic language use. While diverse in method and focus, they share a common concern: how language educators can better prepare learners to navigate contemporary linguistic and cultural environments. The field of Foreign Language education indeed occupies a unique place among academic disciplines in the analysis of human affairs and the cultivation of human beings. I still remember the joy of learning to write my first Chinese character in eighth grade, of course to impress a fellow eighth grader (a girl). Twenty-five years later, that same joy had blossomed into deep socio-linguistic insight as I sat across from a PLA colonel in Beijing while I was serving as a working group lead on the *US-China Army-Army Dialog Mechanism*. Now, as the Head of West Point's Department of English and World Languages, I am involved daily in the profound endeavor of cultivating in Cadets that same joy of foreign language learning in order that they may also reap the benefits of mind and character that will make them, too, better commissioned officers. This transmutation of joy to insight is the guiding ethos of the foreign language teacher toiling in classrooms across the Republic, whether they be the ones in West Point's Washington Hall or at the Presidio of Monterey at the Defense Language Institute Foreign Language Center. I have taught in both classrooms and am certain of this: we rightfully keep the faith that our efforts will prove well spent at other places at other times, when the country will rely on the faculty of attention and nuance of judgment that can be the hallmark of cross socio-linguistic sophistication of the commissioned and non-commissioned officers, the cryptolinguists, the Foreign Area Officers and so many others that we are privileged to call our students.

Several articles in this volume explore the opportunities and challenges presented by generative artificial intelligence. These discussions arrive at a pivotal moment. Artificial intelligence promises unprecedented access to personalized instruction, feedback, and authentic interaction. At the same time, what aspects of the learning, production, and interpretation of language

should be delegated to technology, and what aspects remain irreducibly human? Language is not merely information exchange. It is *participation in shared forms of life*. The challenge before us is to harness technological innovation without losing sight of the relational, interpretive, and ethical dimensions of language education.

Other contributions to this special issue examine language aptitude and learner success. Such work reminds us that language acquisition is shaped by a complex interplay of cognitive, motivational, and contextual factors. At the same time, these studies challenge simplistic assumptions about talent and achievement. Aptitude matters, but so do persistence, purpose, instructional design, and opportunities for meaningful engagement. The most successful learners are often those who develop not only skill, but also a sustained sense of why language learning matters. That emphasis on purpose is especially evident in discussions of learner autonomy and intercultural reflection. The most enduring language education does not end when a course concludes, or a proficiency test is passed. Rather, it cultivates habits of inquiry that continue throughout a lifetime. Whether our students are Infantry Platoon Leaders, attachés or intelligence analysts, all become capable of asking new questions, seeking unfamiliar perspectives, engaging constructively with cultural complexity and providing better analysis.
Particularly noteworthy are the contributions that move beyond narrowly formal representations of language. Studies examining authentic discourse, informal speech, and intercultural engagement remind us that real language use rarely conforms to the tidy boundaries of textbooks. Human communication unfolds within relationships, communities, institutions, and traditions. Effective language instruction must therefore expose learners not only to linguistic structures but also to the realities that give those structures meaning. Understanding communication within its social, cultural, and historical context enables learners not only to interpret the outward manifestations of culture, but also to recognize the values, perspectives, and relationships that influence human behavior across communities and regions.

As someone deeply engaged with Chinese studies and broader regional dynamics, I have witnessed firsthand how language competence shapes understanding. Strategic competition, diplomacy, military cooperation, economic exchange, and cultural interaction all depend upon the ability to comprehend not merely what others say, but how they understand themselves and their world. The most consequential misunderstandings often arise not from inadequate vocabulary, but from unexamined assumptions about meaning, intention, and context. The articles in this volume contribute meaningfully to that mission. They remind us that language education stands at the intersection of technology and humanity, measurement and meaning, local practice and global engagement. Most importantly, they affirm that language learning remains, at its core, an encounter with other human beings and the worlds they inhabit.

I thank my longtime friend and colleague Dr. Jeff Watson and all the contributors to this edition for continuing to further this dialog. I invite readers to engage these contributions not simply as discussions of instructional techniques or educational policy, but as reflections on one of humanity's most remarkable capacities: the ability to create, share, and transform meaning through language.

Applied Language Learning

Introduction to the Special Issue

Instructing Language, Regional Expertise, and Culture

Jeffrey Watson
United States Military Academy, West Point

Dawn Bikowski
Defense Language Institute Foreign Language Center, Monterey, California

This special issue of *Applied Language Learning* is dedicated to innovative projects related to Instructing Language, Regional Expertise, and Culture (LREC). Since the passage of the Defense Language Transformation Roadmap (U.S. Department of Defense, 2005), the Department of Defense (now Department of War, DoW) has dedicated significant resources to building skillsets related to world language (WL) proficiency, cross-cultural competence (3C), and regional expertise. These resources led to new LREC programs across the DoD schools, training institutions, and service academies. The U.S. Military Academy at West Point, for example, began teaching world languages five days a week, opened its Center for Languages, Cultures, and Regional Studies (CLCRS), and developed an immersion program to send ~150 cadets abroad each semester to study languages in their native cultural contexts. All three service academies developed international program offices and increasingly embraced international exchange cadets/midshipmen from allied countries to "internationalize" their populations. The Air Force also opened its Air Force Culture & Language Center (AFCLC) and launched its Language-Enabled Airmen Program (LEAP). The Marine Corps also opened its Center for Advanced Operational Culture Learning (CAOCL) (Watson, 2010).

Building on the 2005 DoD Roadmap, "A Roadmap for Cultivating and Managing Skilled Language, Regional Expertise, and Culture Talent: Phase One" (U.S. Department of Defense, 2023) identified a need "to improve the management of LREC talent" as a "vital step forward to ensuring the Department has a LREC workforce ready to meet mission needs of today and the challenges of tomorrow" (p. 3). This 2023 Roadmap led to modernization initiatives at the Defense Language Institute Foreign Language Center (DLIFLC) in Monterey, California. Goals included increasing the percentage of students who graduate with at least an ILR level of 2+ in both Listening and Reading and also increasing the recruitment and retention rates of Cryptologic Language Analysts. In 2022, DLIFLC was granted the authorization to offer Bachelor of Arts Degrees and in 2023 the Intermediate and Advanced language programs were re-aligned under one directorate, thereby

increasing the accessibility of these courses, facilitating standardization, and maximizing talent utilization around the globe.

To introduce this special issue, it is important to first define the underlying tenets and skillsets that LREC instruction strives to produce in military personnel. Since the mid-2000s, the U.S. Department of Defense has increasingly recognized that success in contemporary military and security environments depends not only on technological superiority or kinetic capability, but on the human capacity to understand, interpret, and act within complex sociocultural systems. This recognition gave rise to the modern Language, Regional Expertise, and Culture enterprise, which integrates world language proficiency, cross-cultural competence, and regional understanding as mutually reinforcing capabilities rather than discrete and separate skill sets (Byram, 1997; Deardorff, 2006). Yet as the LREC enterprise has matured institutionally, an equally important question has emerged pedagogically: *How should LREC be defined and instructed in ways that genuinely prepare learners to operate across complex cultural landscapes?*

This special issue suggests that LREC is best understood through the lens of **cultural relativism** and that its instruction must therefore be **transformative** rather than merely informational. Cultural relativism is the principle that beliefs, values, and practices can only be meaningfully understood within their own cultural contexts (Byram, 1997; Deardorff, 2006). This principle provides the ethical and epistemological foundation for LREC. Without it, language instruction risks becoming mechanical, cultural instruction risks devolving into stereotyping, and regional expertise risks collapsing into static checklists of facts. With it, however, LREC becomes a developmental process that reshapes how learners perceive themselves, others, and the environments in which they operate.

Chapter 1 of *Language, Regional Expertise, and Culture in the Military: State of the Science* (Watson et al., 2025) situates LREC as a strategic response to the limitations exposed during military operations in the early 21st century, particularly in environments where U.S. forces encountered culturally embedded forms of resistance, influence, and legitimacy that could not be addressed through force alone. The resulting *Defense Language Transformation Roadmap* (2005) reframed language and regional expertise as warfighting competencies, while subsequent cultural initiatives sought to close persistent gaps in cross-cultural understanding. What emerges from this history is a clear insight: LREC is not additive. Language does not simply "support" culture (or vice versa), nor does culture merely contextualize regional knowledge. Rather, LREC functions as an integrated human capability aimed at navigating difference, ambiguity, and power across sociocultural systems.

If cultural relativism provides a foundation for LREC, **transformative instruction** defines *how* it should be taught. Watson and Leaver (2025) explicitly argue that transformative LREC instruction is a paradigm shift that moves beyond communicative competence toward personal and professional transformation. Drawing on the broader literature on transformative learning, they argue that effective LREC instruction develops bilingual and bicultural individuals capable of perspective-shifting, ethical reflection, and adaptive leadership in unfamiliar cultural environments.

TRANSFORMATIVE INSTRUCTION: BROADENING THE CONTEXT

Advanced language proficiency has long been the goal and focus of most WL programs across the DoD. The Defense Language Proficiency Test (DLPT) along with the Interagency Language Roundtable (ILR) proficiency scale serve as the goalpost and chief measure of this proficiency. However, in 2007, the Modern Language Association suggested that languages should be taught as a "core" element of cross-cultural competence (MLA, 2007). Building on this, in 2015, ACTFL published its *World Readiness Standards for Learning Languages*, the so-called "5 Cs:" Communication, Cultures, Connections, Comparisons, and Communities (ACTFL, 2015). These changes by our leading professional organizations reframed language education beyond proficiency alone and reorganized instructional principles around "interpretive, interpersonal, and presentational" modes of communication in order to emphasize learners' ability to use language within authentic sociocultural contexts (ACTFL, 2015). This shift signals a broader departure from earlier models focused primarily on developing language proficiency toward developing intercultural communicative competence (ICC) (Byram, 1997; Deardorff, 2006). This also aligns with the MLA's statement that language programs should produce "educated speakers who have deep translingual and transcultural competence" and situate language study within "cultural, historical, geographic, and cross-cultural frames" (MLA, 2007, p.4).

Moreover, transformative LREC instruction is grounded in the recognition that learning a language or culture – even if situated in its sociocultural complexity – is not a neutral cognitive activity. Rather, it is an encounter with difference that can unsettle deeply held assumptions about identity, values, and power. From this perspective, cultural relativism is not simply taught as a concept; it is *experienced* through carefully designed learning environments that confront learners with ambiguity, contradiction, and "disorienting dilemmas" that demand reflection and reassessment. These moments of cognitive and emotional disruption are not pedagogical failures but essential catalysts for transformation (Watson & Leaver, 2025).

Several articles in this special issue demonstrate this shift to a broader context for teaching language proficiency, specifically leveraging the unique capabilities of generative artificial intelligence (AI). For example, Kozlova at DLIFLC-Monterey in her article "**Generative AI in Foreign Language Education: A Scoping Review of Pedagogical Applications and Challenges"** as well as Swanson and Richey at the U.S. Air Force Academy in their article **"World Language Instructors' Perceptions of Using Generative AI Tools"** explore the potential benefits of using AI chatbots as a potentially "disruptive" technology that has the potential to challenge our current methodologies. In his article **"Revolutionizing Language Learning: AI-Assisted Learning as a Catalyst for High Proficiency Attainment,"** Tan demonstrates how an AI-Assisted Learning Framework can assist students with error identification and reduction. It is vital for the future of LREC instruction to explore how large language models can help teachers contextualize their teaching approaches and feedback mechanisms while affording students ways to explore simulated communication in their personalized language practice (Kozlova, this issue).

Other articles in this issue also explore meaningful ways to integrate language and culture through technology-enhanced instruction. Lee and Kim, for example, in **"Beyond News and Documentaries: Developing a Corpus-Based Lexical Resource for Informal North Korean Speech"** investigated lexical variation across formal and informal registers in contemporary North Korean discourse, illustrating the importance of utilizing various types of texts in preparing students not only for language exams, but also for understanding the culture and connections of a community. Xu, Wang, and Bozeman, in **"Integrating a Streaming Series into the Curriculum to Foster Listening Proficiency and Cultural Competence,"** explored the contributions of narrative media to students' development of listening skills and cultural understanding. In a joint study between West Point and DLIFLC-Monterey, Holloway, Gao, Qasem, and Miller investigate the highly relevant issue of motivational dynamics in the world language classroom. In their study, "**Motivational Dynamics among Persian Farsi Language Learners Across Two Military Institutions,"** they used guided reflective journals to explore how learners generate and maintain interest (through measures of grit and anxiety) to engage with their Persian language courses. This study highlights the importance of understanding individual differences of military learners and the "interplay between learner characteristics and motivational engagement" in military instructional environments (Holloway et al., this issue, p. 194). Similarly, Seibel in his commentary: **"A Letter from Valley Forge and Words of the Developing World: Knowing the 'Why' for Autonomous Language Learning"** discusses the importance of developing autonomous language learners. Seibel argues that learner motivation is vital for students at DLIFLC to attain proficiency in a short amount of time while also "fostering a love of the target language and culture" (p. 150). In this regard, instructors "must instill a sense of purpose" (p. 150) in their students in order to encourage the autonomous learning strategies that will increase "their survivability, adaptability, and potential to shape outcomes in the future conflicts that will unexpectedly arise in remote corners of the world" (p. 152).

Sun, Lemmons, Ruble, and Emeliyanova, in their study **"Integrating Cultural Competence into Army Special Operations Forces Basic Language Training,"** illustrate another important aspect of instructing LREC in this broader context. In their study, two interventions were used to instruct cultural competence in two groups of ARSOF students (Russian and Mandarin Chinese) at Fort Bragg: culture-general cross-cultural competence instruction and culture-specific intercultural communicative competence training. From the late 2000s to the early 2020s, the DoD invested in developing culture-general 3C models designed to cultivate core competencies across the general-purpose force.[1] As outlined in DoD Instruction 3126.01C and informed by Deardorff (2006), these competencies include "understanding culture, applying organizational awareness, cultural perspective taking, and cultural adaptability" (Chairman of the Joint Chiefs of Staff, 2023, p. F-1). Accordingly, the Marine Corps' *Culture General Guidebook* (Fosher & Mackenzie, 2021) emphasizes that cultural competence requires "a broad, multidimensional approach to culture training and education" grounded in "the interplay of intercultural concepts and skills" (p. 5). In principle, this reflects the integration of culture-general and culture-specific capabilities at the

[1] These training endeavors were led by organizations such as the Army's TRADOC Culture Center (TCC), the Air Force's Culture and Language Center (AFCLC), and the Marine Corps' Center for Advanced Operational Culture and Language Center (CAOCL). See also Fosher & Mackenzie, 2021, Mackenzie & Fosher, 2023, and Mackenzie & Henk, 2025.

heart of cross-cultural competence. Unfortunately, for mostly logistical reasons, the DoD explored this interplay by developing separate training paths (Watson, 2010). Culture-general skills, considered less time-intensive and applicable to all cultures, were developed as pre-deployment toolkits, while culture-specific knowledge remained the purview of WL institutions such as DLIFLC and the service academies (Mackenzie & Henk, 2025). This division allowed for practical training solutions but limited opportunities for meaningful integration of LREC capabilities.

As the DoD shifted its operational priorities, funding for many culture centers declined and culture-general programs across the DoD were reduced or eventually shuttered (Fosher & Mackenzie, 2021). As a result, this has shifted the responsibility to continue culture-general training to the better-funded and longer-standing WL organizations. While WL programs do not reach the majority of the force, they nonetheless represent an existing and institutionally viable venue for sustaining and advancing culture-general competencies. However, as demonstrated by Sun et al.'s findings, not all WL teachers are trained to incorporate culture-general 3C frameworks into their culture-specific WL classes. The result is a "teachability dilemma" in which there is a continued need for WL instructors to both appreciate the value of these 3C culture-general models and integrate them into the more familiar ICC frameworks. Similarly, the assessment of ICC skills has also been an ongoing challenge for WL instructors. In **"Framing Intercultural Reflection in the Language Classroom,"** Tozcu and Eagle outline a clear, pedagogically grounded way to embed ICC into language learning while meeting curricular demands for clarity and accountability. The authors demonstrate how to shift the focus from assessing ICC to documenting and supporting student engagement with foundational ICC concepts, such as perspective-taking, flexibility, and awareness of unspoken assumptions. The articles in this issue thus convey many of the complexities confronted by WL teachers both inside and outside of military contexts.

REGIONAL EXPERTISE

The RE of LREC is often the least discussed and perhaps the least understood. Historically, the DoD has defined regional expertise as a "multidimensional construct" representing knowledge "of a region's social, economic, political, and linguistic features" (Paletz et al., 2018, p. 528). Mackenzie & Henk (2025) further point out that regional experts are those with extensive knowledge of "U.S. interests and involvements, nation-states and their interests, international organizations, regional and local conflicts, regional histories, politics, societies, natural environments, [and] economies" (p. 45). As such, regional experts, such as Foreign Area Officers (FAOs), attachés, or Security Cooperation Officers (Chairman of the Joint Chiefs of Staff, 2023), develop in-depth culture- and region-specific knowledge, including significant foreign language proficiency, education, and sustained in-region experience. The skill sets of regional experts, however, do not inherently ensure the transferable, adaptive skills associated with cross-cultural competence, that is, "the ability to quickly understand and adjust to the cultural environment of new and unfamiliar operational areas" (Mackenzie & Henk, 2025, p. 47). As a result, instructing

regional proficiency must include training that promotes the cultural agility to set aside "cognitive biases and work within the patterns of thinking of host nation (or partner) actors" (p. 47).
To illustrate this, Schendel and Kerlin from DLI-Washington, in their article **"An Artificial Intelligence-Enabled Tool for Real-Time Intercultural Competence Instruction,"** explored the benefits of incorporating an AI-enabled cross-cultural training tool (Cross-cultural Interaction Real-time Assistant for Negotiators and Operations – CIRANO) within a 64-week Mandarin Chinese course through scenario-based interactive activities. Their findings illustrate how culture-specific and culture-general competencies interact in practice. This collaboration not only facilitated authentic communication and a better understanding of Chinese cultural norms (culture-specific) but also helped students practice cross-cultural skills such as perspective-taking, adaptability, and self-awareness (culture-general). Students became more aware of their posture, emotional tone, eye contact, and "their own communication patterns and linguistic choices in English" resulting in "adjustment of their behavior" to fit the scenario (Schendel & Kerlin, 2026, p. 87). This type of training aligns well with the "leader/influence" skillsets outlined in DoD Instruction 3126.01C pertaining to exerting cross-cultural influence while also at times utilizing an interpreter, or in this case, an interpretation tool (See Chairman of the Joint Chiefs of Staff, Enclosure F).

LANGUAGE TEACHING AND ASSESSMENT

Building and assessing students' language proficiency, the "L" of LREC, is fundamental to any WL program. This special issue thus includes two articles focusing on language development and assessment. **"The Impact of Linguistic Knowledge on Reading and Listening Comprehension Across ILR Proficiency Levels,"** by Sun-Kwang Bae, investigates the extent to which vocabulary, structural, and discourse knowledge predict reading and listening comprehension at various proficiency levels. Bae finds that learners gradually shift from a heavy reliance on lexical knowledge at lower proficiency levels toward a more coordinated use of lexical, structural, and discourse knowledge at higher levels, particularly in reading.

In addition to assessing language proficiency, assessment of learners' second language aptitude remains an important issue across the DoD. In 2020, West Point began using the Pre-Defense Language Aptitude Battery (Pre-DLAB) from the DLIFLC to help place cadets into world language courses, especially the more challenging Category III (Russian, Persian) and IV (Arabic, Chinese) languages. Pendergast, Dean, and Jannusch, in **"Language Aptitude as Predictor of Performance"** discuss the changing realities of student placement in WL courses, with cadets periodically being assigned to study languages they had not selected. In this context, they evaluate the effectiveness of different aptitude tests (and other variables) at West Point in predicting expected performance. Their finding on which aptitude assessments (i.e., the MLAT and Pre-DLAB) best predict performance provides useful information for any program facing similar challenges.

CONCLUSION

In closing, LREC skillsets remain vital to the DoW's ability to "exercise positive influence throughout the world" (Executive Office of the President, 2025). LREC is not only a supporting capability; it is an operational necessity. Building cooperative security partnerships, improving interoperability with allies, strengthening cryptologic and human intelligence gathering, and developing cross-cultural leaders that represent U.S. interests at home and abroad require these LREC skillsets and are a crucial part of our warfighting arsenal. Teachers who skillfully craft transformative lessons that leverage emerging technology, practice cross-cultural adaptability and perspective-taking, and strengthen language proficiency grounded in intercultural communicative competence and nuanced regional understanding play a critical role in developing these capabilities. We hope this special issue contributes to those efforts and advances the ongoing conversation about preparing military personnel to communicate, lead, and win in an increasingly complex cross-cultural environment.

Authors

Jeffrey R. Watson, Ph.D., is a professor of Applied Linguistics in the Department of English & World Languages (DEWL) and chair of linguistics and language acquisition in the Center for Languages, Cultures, and Regional Studies at the United States Military Academy (USMA). He holds a doctorate in Russian and Second Language Acquisition from Bryn Mawr College and teaches courses on applied linguistics, Russian, and English as a second language. Dr. Watson has published on topics related to second language acquisition and intercultural competence, and continues to pursue research in sociocultural theory, second language identity, language socialization, and curriculum design. Recent scholarship includes an edited book volume titled *Language, Regional Expertise, and Culture (LREC) in the Military: State of the Science* published by the West Point Press.

Dawn Bikowski, Ph.D., is an Associate Professor at the Defense Language Institute Foreign Language Center in Monterey, California. She serves as an advisor to the provost in the Office of Standardization and Academic Excellence and as Editor-in-Chief for DLIFLC's two academic journals. With a Ph.D. in Curriculum and Instruction and an M.A. in Applied Linguistics, her research and professional work focus on systems thinking, educational technology, and faculty development. In addition to many articles and book chapters, Dr. Bikowski co-authored the book *Teaching with a Global Perspective: Practical Strategies from Course Design to Assessment*, Routledge.

REFERENCES

American Council on the Teaching of Foreign Languages. (2012). *ACTFL proficiency guidelines 2012*. ACTFL. https://www.actfl.org/resources/actfl-proficiency-guidelines-2012

American Council on the Teaching of Foreign Languages. (2015). *World-readiness standards for learning languages* (4th ed.). ACTFL. https://www.actfl.org/educator-resources/world-readiness-standards-for-learning-languages

American Council on the Teaching of Foreign Languages. (2024). *ACTFL proficiency guidelines 2024*. ACTFL. https://www.actfl.org/resources/actfl-proficiency-guidelines-2024

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Chairman of the Joint Chiefs of Staff. (2023). *Language, regional expertise, and culture capability identification, planning, and sourcing (CJCSI 3126.01C)*. U.S. Department of Defense.

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education, 10*(3), 241–266.

Executive Office of the President. (2025). *National security strategy of the United States of America*. The White House.

Fosher, K. B., & Mackenzie, L. (Eds.). (2021). *The rise and decline of U.S. military culture programs, 2004–2020*. Marine Corps University Press.

Mackenzie, L., & Fosher, K. B. (2023). *Culture general guidebook*. Marine Corps University Press.

Mackenzie, L., & Henk, D. (2025). Protecting, preserving and maturing the culture component of language, regional expertise, and culture. In J. R. Watson, R. Wolfel, & A. Kalkstein (Eds.), *Language, regional expertise, and culture in the military: State of the science*. West Point Press.

Modern Language Association of America. (2007). *Foreign languages and higher education: New structures for a changed world*. Modern Language Association. https://eric.ed.gov/?id=ED500460

Paletz, S. B. F., Hughes, M. M., Sumer, A., Brugman, C. M., Bunting, M., Rytting, C. A., Buntain, C., & Ellis, D. (2018). A preliminary validation of the Regional Proficiency Assessment Tool (RPAT). *Military Psychology, 30*(6), 528–546. https://doi.org/10.1080/08995605.2018.1506196

Watson, J. R. (2010). Language and culture: Separate training paths. *Military Review*, 90(2), 81–87.

Watson, J. R., & Leaver, B. L. (2025). Transformative LREC instruction: Developing cross-cultural leadership. In J. R. Watson, R. Wolfel, & A. Kalkstein (Eds.), *Language, regional expertise, and culture in the military: State of the science* (pp. 279–296). West Point Press.

Watson, J. R., Wolfel, R., & Kalkstein, A. (Eds.). (2025). *Language, regional expertise, and culture in the military: State of the science*. West Point Press.

U.S. Department of Defense. (2005). *Defense language transformation roadmap*. https://archive.org/details/DTIC_ADB313370

U.S. Department of Defense. (2023). *A roadmap for cultivating and managing skilled language, regional expertise, and culture talent: Phase one*. https://www.dliflc.edu/wp-content/uploads/2023/09/Roadmap-for-Cultivating-and-Managing-Skilled-LREC-Talent-Phase-1.pdf

Applied Language Learning

Research Article

Generative AI in Foreign Language Education: A Scoping Review of Pedagogical Applications and Challenges

Yuliya Kozlova
Defense Language Institute Foreign Language Center, Fort Meade Language Detachment, Maryland

Generative artificial intelligence (AI) has quickly become an important force in education, especially in foreign language (FL) teaching and learning. While research on generative AI in education is growing, its value and the real challenges of using generative AI in language classrooms are not fully clear. This scoping review explores a dataset of 32 studies on generative AI's role in FL education, published between 2023 and January 2026. Specifically, it addresses: (1) how generative AI tools have been used to support language learning and teaching; (2) what instructional benefits and pedagogical affordances have been reported in relation to generative AI use; and (3) what pedagogical, technical, and ethical issues have been identified in their implementation. Following PRISMA-ScR guidelines, studies were systematically identified, screened, and coded based on their alignment with the research questions. The literature review findings reveal that generative AI tools are used in FL education as tutors that provide personalized feedback, as platforms for simulated communication through chatbots, and as a support for teachers in lesson planning and developing materials. Pedagogical affordances associated with the use of generative AI include increased teaching efficiency, improved learner engagement, more tailored feedback, and broader opportunities for personalized language practice. Finally, among challenges in incorporating generative AI in FL teaching and learning, the most common are uneven teacher preparedness, insufficient support from educational institutions, and concerns about academic integrity and data privacy. The gaps in cumulative knowledge and suggestions for future research are discussed at the end of the scoping review.

Keywords: *Generative AI, Foreign Language Education, Teacher Training, Pedagogy, Artificial Intelligence Literacy, Learner Autonomy, Scoping Review*

INTRODUCTION

Since the launch of ChatGPT in November 2022, the use of generative artificial intelligence (AI) has expanded rapidly in foreign language (FL) education. Unlike earlier technology tools, which were primarily used to deliver pre-programmed content, generative AI is able to create original language in context. Kasneci et al. (2023) describe this function as an ability that challenges conventional boundaries between learner, teacher, and tool. Research shows that generative AI can lower learners' anxiety when speaking in an FL (Wan & Moorhouse, 2025; Wang et al., 2024; Wang et al., 2025), improve skills development (Cakmak, 2025; Kohnke et al., 2023; Liao et al., 2023), and offer personalized help (Creely, 2024). Additionally, generative AI can foster learners' autonomy and reduce cognitive load (Feng, 2025). In instruction and assessment contexts, generative AI has been explored for the creation of materials and test items (Pack & Maloney, 2023; Perez-Nunez, 2023; Vigna-Taglianti, 2024), feedback (Roe et al., 2024; Zhang & Wang, 2025), and assisting with grading (Barrett & Pack, 2023; Zacharis & Papadakis, 2025). As empirical research on generative AI in FL education continues to grow, attention has shifted toward synthesizing the existing work to clarify predominant trends, benefits, and challenges of integrating this technology into practice.

Recent research has synthesized how generative AI can be incorporated into the FL classroom (Crosthwaite et al., 2025; Luo et al., 2025; Feng et al., 2025) with the prime emphasis on ChatGPT use (Hong, 2023; Li et al., 2024), English as a Foreign Language (EFL) context (Balcı, 2024; Lo et al., 2024), and narrow pedagogical areas of focus (Crosthwaite & Sun, 2025; Du & Daniel, 2024), leaving other languages, educational contexts, AI models, and theoretical perspectives underexplored. In response to this gap in understanding, this literature review takes a scoping review approach to systematically map how generative AI is situated within FL education as a pedagogical tool, including how it may be utilized in lesson planning, developing materials, and assessing learners. The review addresses the following research questions:

1. How have generative AI tools been applied to support foreign language teaching and learning?
2. What instructional benefits and pedagogical affordances have been reported in relation to generative AI use in foreign language teaching and learning?
3. What pedagogical, institutional, and ethical issues and challenges have emerged in the implementation of generative AI in foreign language teaching and learning?

By systematically mapping 32 empirical studies and coding their publication contexts, educational settings, types of generative AI tools, pedagogical applications, and reported benefits and challenges, this scoping review provides an evidence-based overview to inform future research and instructional practice in FL education.

METHODS

Operational Definitions

In this review, a foreign language refers to any language that is not the learner's native language and is not commonly used by the learner for daily communication, education, or official purposes (Richards & Schmidt, 2013). This definition does not distinguish between foreign and second language contexts and includes learning both inside and outside the target-language environment. Generative AI is understood as a subset of AI that can create original language outputs, typically through large language models used in tools like ChatGPT and Gemini. More broadly, AI refers to computer systems designed to complete tasks related to human intelligence, including earlier rule-based or automated writing evaluation tools that do not create new language. Large language models (LLMs) are data-driven neural network models trained on extensive textual data to produce human-like language. This study used a scoping review methodology to explore the breadth and nature of research on generative AI in FL education, following PRISMA-ScR guidelines (Tricco et al., 2018). Lastly, gray literature refers to non-peer-reviewed scholarly sources, such as reports or working papers, that were included to gather early insights on the initial stages of generative AI adoption in education.

Study Design

As generative AI began to gain traction in the research field only towards the end of November 2022, despite having been in development for years, there is insufficient literature to conduct a systematic literature review. This situation makes a scoping review a more appropriate choice. Unlike systematic reviews, scoping reviews are favorable for addressing broader questions, such as "What is the nature of the evidence for this intervention?" or "What is known about this concept?" (Delgado-Chaves, 2025). Given the distinct objectives of this study, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines (Tricco et al., 2018) were adopted to ensure methodological transparency and replicability. Following PRISMA-ScR, the process enabled a comprehensive mapping of generative AI applications, challenges, and pedagogical implications in FL education.

Eligibility Criteria

The documents were included based on the following criteria: a focus on generative AI-powered tools in FL teaching and learning contexts; involved teachers, students, or both as the target population; no language restrictions; and published between 2023 and January 2026. Eligible publication types included peer-reviewed empirical studies, conceptual or theoretical papers, review articles, and discussion papers. Gray literature was also included when it provided substantive scholarly insight into early applications or conceptual discussions of generative AI in language education.

Search Strategy

The process involved several steps: selecting databases, defining search terms, applying inclusion and exclusion criteria, and critically analyzing the selected literature. The databases were selected based on their relevance to education, applied linguistics, and educational technology research, as well as their coverage of peer-reviewed and interdisciplinary studies. ERIC and Education Source were used to find education-focused literature, ScienceDirect and JSTOR for foundational research, and Google Scholar to find new studies that the traditional databases do not always index. Major phrases were used to identify relevant studies: "generative AI", "Gen AI", "GAI", "Large Language Models", "LLM", "Chatbots", "Chatbot", "Chat GPT", "ChatGPT", "Gemini", "Copilot", coupled with "foreign language", "FL", "foreign language teaching", "foreign language learning", "foreign language instruction", "foreign language education", "impact on foreign language education", "future of foreign language education", "negative impact on foreign language education", "positive impact on foreign language education", or "integration in foreign language." The search strings were adjusted according to the search engine's requirements.

Selection of Sources of Evidence and Inclusion Criteria

The selection process included two phases. During the first phase, the researcher screened the titles and abstracts to assess their relevance to the study's topic. In the second phase, the full texts that had passed the previous screening were reviewed, with reasons for exclusion documented. The inclusion criteria were limited to empirical research studies that examined generative AI in foreign language teaching and learning and reported original data. The eligible studies used qualitative, quantitative, or mixed-methods designs and focused on the application of generative AI in foreign language education. When multiple publications reported on overlapping datasets or closely related analyses, the most comprehensive or recent study was included. Only studies focusing on instructional practices involving teachers, instructors, or educational institutions were considered. Studies that focused only on student-created work without any teaching context were excluded, as this review concentrated on how generative AI is integrated and implemented in foreign language education.

The flow diagram, which summarizes each step of the selection process, is presented in Figure 1. During the identification phase, an initial 214 records (100%) were retrieved from the scholarly search engine Google Scholar (170) and these selected databases: ERIC (10), Education Source (15), JSTOR (7), and ScienceDirect (12). No additional records were identified through other sources. All citation data were exported to Microsoft Excel for organization and duplicate removal. After this step, 154 unique records (≈72%) remained.

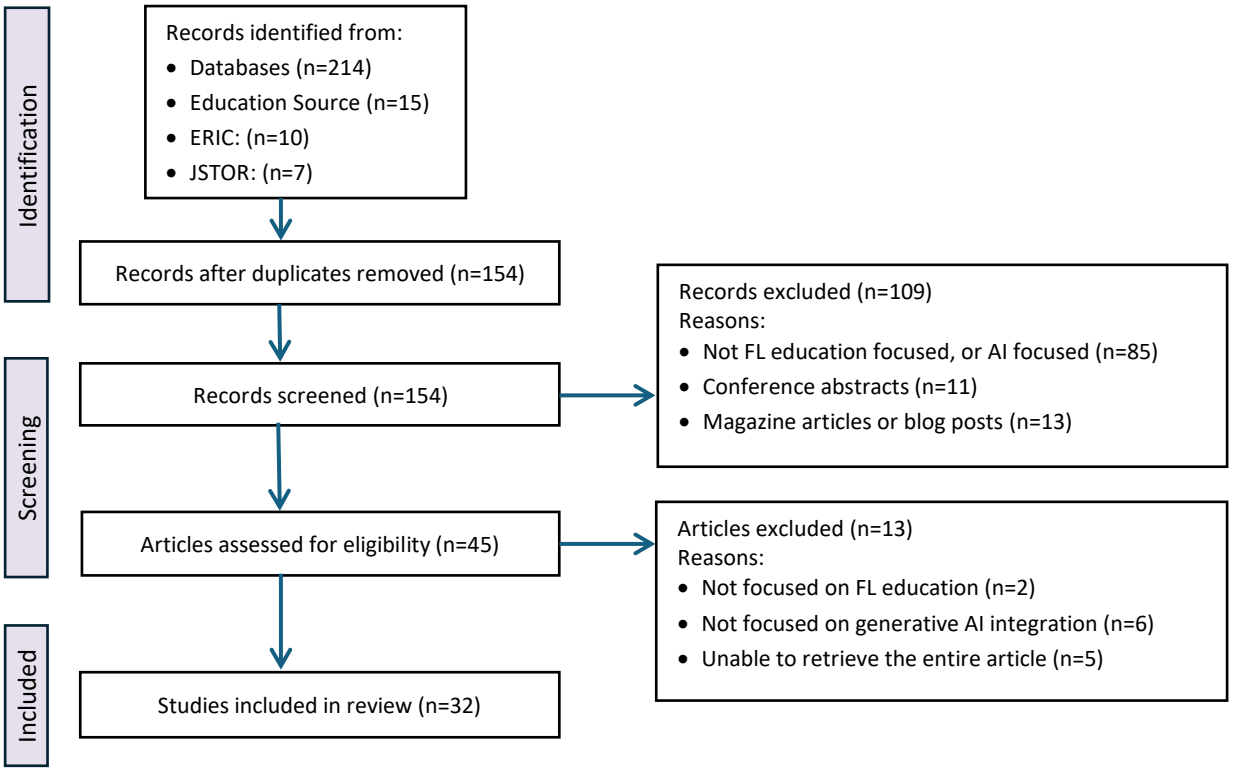
In the screening phase, these records were examined based on titles, abstracts, and relevance to the research focus. A total of 109 records were excluded because they were not related to foreign language education or generative AI, were duplicates of preprints, or originated from non-academic sources such as blogs, conference abstracts, or promotional materials. For example, studies that offered general overviews of generative AI in FL education without reporting original

empirical data (such as Mizumoto, 2025, which provides a synthesis of prior work rather than a study with clearly defined data sources and analytical methods) were excluded at this stage. The remaining 45 full-text articles (≈21%) were reviewed in detail for eligibility.

Of these, 13 articles were excluded due to insufficient methodological rigor, lack of a clear generative AI focus, or limited alignment with the review questions. For instance, Ulanova et al. (2025) was excluded from the review as it addresses AI broadly rather than specifically exploring generative AI applications. Also, the study does not provide insights into implementing or utilizing these methods in FL teaching and learning. In contrast, Chen (2025) was included because it offers an empirically grounded investigation of how generative AI can be integrated into FL education, with clearly defined research goals and direct relevance to the review questions. Finally, 32 peer-reviewed studies (≈15%) met all inclusion criteria and were included in the final literature review study.

Figure 1

Flowchart Summary of the Sources of Evidence



Data Analysis

The data analysis followed an iterative, descriptive-analytical approach consistent with scoping review methodology. After final study selection, all included articles were coded in a Microsoft Excel spreadsheet to enable transparent organization, comparison, and synthesis of findings.

Coding Process and Trustworthiness

A coding framework was developed deductively based on the study's research questions and refined inductively during initial familiarization with the literature. The final coding scheme was organized around three primary analytical categories aligned with the review questions: pedagogical applications of generative AI (RQ1), reported instructional benefits and affordances (RQ2), and pedagogical, institutional, and ethical challenges (RQ3). Additional contextual categories, including publication information, research design and methodology, and educational context and participant characteristics, were used to describe trends across the literature but were not analyzed as primary outcomes. The categories were organized into separate columns in the Microsoft Excel coding spreadsheet, which allowed a systematic mapping of how generative AI has been understood, used, and assessed in FL education.

Scoping reviews are explanatory and descriptive by nature, aiming to identify and map the literature on a given topic, clarify concepts, and examine how research has been conducted across different settings (Munn et al., 2022). This scoping review follows this approach, presenting the emerging characteristics and patterns of existing research on generative AI integration in foreign language education. Trustworthiness was achieved through qualitative synthesis methods, which included constant comparison across studies and an audit trail in the Excel dataset (Miles et al., 2014; Saldaña, 2011). For example, several publications described how generative AI can be used to enhance writing skills in FL, but the role of AI was not clearly stated, whether it served as support for teachers (e.g., generating prompts, lesson materials, or feedback models), or whether generative AI acted as a learner-facing tool used directly by students. In those cases, the literature was re-examined to determine how generative AI was used in practice and re-coded accordingly. Similarly, when studies reported multiple uses of generative AI in teaching (e.g., both feedback provision and content generation), coding decisions were compared across categories to ensure consistent classification. All revisions were documented in the Excel-based audit trail, which helped to maintain consistency and transparency throughout the review process.

Analytical Procedures

To analyze data, this scoping review summarized the frequencies and distributions of key study characteristics, such as research designs, types of generative AI applications, and educational contexts. For example, frequency counts were used to summarize study characteristics and were organized into frequency distributions for reporting.

Thematic analysis was used to identify recurring patterns related to benefits, challenges, pedagogical implications, and ethical considerations of generative AI in FL education. During analysis, findings sections of the included studies were examined to extract qualitative evidence describing instructional benefits, constraints, and pedagogical considerations. Themes were developed by repeatedly comparing studies and adjusted to reflect similar and differing viewpoints in the literature, using standard qualitative analysis methods (Braun & Clarke, 2006). The coding categories guiding this analysis and their alignment with the study's review questions

are outlined in Table 1. However, it is important to note that only categories directly aligned with the research questions (categories 4-7) were used in the thematic analysis. The remaining categories (1-3) were used to describe patterns in the literature.

Table 1

Coding Framework and Alignment with Review Questions

Number	Coding Category	Description	Related Research Question
1	Publication information	Author(s), year of publication, and journal outlet	Contextual descriptor (used to identify trends in the literature; not directly aligned with a specific RQ)
2	Research design and methodology	Study design, data collection methods, and analytical approaches	Contextual descriptor (used to characterize the evidence base; not directly aligned with a specific RQ)
3	Educational context and participants	Instructional setting, learner population, and language context	Contextual support for RQ1 (provides insight into where and how generative AI applications occur)
4	Generative AI tool or model	Type of generative AI system or large language model examined	RQ1: How have generative AI tools been applied to support language teaching and learning?
5	Pedagogical application	Instructional use (e.g., feedback, writing support, materials development, lesson planning)	RQ1: How have generative AI tools been applied to support language teaching and learning?
6	Reported benefits and affordances	Positive outcomes and instructional affordances associated with generative AI use	RQ2: What instructional benefits and pedagogical affordances have been reported in relation to generative AI use in foreign language teaching and learning?
7	Reported challenges and limitations	Pedagogical, technical, ethical, or institutional concerns	RQ3: What pedagogical, institutional, and ethical issues and challenges have emerged in the generative AI implementation in foreign language teaching and learning?

RESULTS

Characteristics of Sources

The review included 32 studies published from 2023 through January 2026. Most studies were quantitative (41%) or mixed-methods (31%), with fewer qualitative designs (28%). The majority were published in 2025, showing rapid growth in this area. Geographically, the studies came from various regions, with China and multi-regional contexts contributing the most.

ChatGPT was the most commonly examined tool at 56%, followed by studies looking at generative AI more generally or using multiple platforms. Most research focused on English as a foreign language, with limited representation of other languages, highlighting an imbalance in the current evidence base (see Table 2 for more details).

Table 2

Frequency of Various Article Characteristics

Characteristic Category	Subcategory	Frequency	Percentage
Publication Year	2023	2	6.3%
	2024	12	37.5%
	2025	17	53.1%
	2026	1	3.1%
Publication Type	Peer-reviewed journal/conference publications	32	100%
Study Type	Mixed Methods	10	31.3%
	Qualitative	9	28.1%
	Quantitative	13	40.6%
Country/ Region of Origin	China	6	18.8%
	Ukraine	3	9.4%
	Russia	3	9.4%
	Hong Kong	2	6.3%
	Indonesia	1	3.1%
	Saudi Arabia	1	3.1%
	Poland	1	3.1%
	Greece	1	3.1%
	Algeria	1	3.1%
	Bulgaria	1	3.1%
	Colombia	1	3.1%
	Italy	1	3.1%
	UK	1	3.1%
	Turkey	1	3.1%
	United States	1	3.1%
	Vietnam	1	3.1%

Characteristic Category	Subcategory	Frequency	Percentage
	Multi-regional/global*	6	18.8%
Generative AI Tool	ChatGPT (explicitly primary tool)	18	56.3%
	General Generative AI (unspecified category)	6	18.8%
	Multiple AI Tools (ChatGPT + others)	6	18.8%
	DeepSeek	1	3.1%
	AI Image Generators (text-to-image focus)	1	3.1%
FL Focus	English as an FL	19	59.4%
	Chinese as an FL	3	9.4%
	German as an FL	3	9.4%
	Multiple FLs	5	15.6%
	French & English (Teacher Education context)	1	3.1%
	Unspecified FL (general FL education)	1	3.1%

Note. Three studies were classified as "global" because they analyzed data or synthesized findings across multiple countries rather than focusing on a single national context.

Results of Individual Sources

The data collection methods, participant types, and research tools used in each publication were examined to provide an overview of the studies included in the final report. The review of 32 studies showed that most research on generative AI in FL education had been conducted in higher education settings, mainly involving university students and teachers. The initial analysis revealed that surveys, mixed-methods, and interviews were the most common instruments, which reflects an increasing interest in learners' perceptions of using generative AI in language learning and its classroom applications. A summary of the study characteristics is available in Appendix A. Also, the detailed findings and pedagogical applications are discussed in the thematic analysis section.

Results of Thematic Analysis

This scoping review identified the following overarching themes aligned with the research questions: (a) patterns of generative AI use in foreign language education, (b) instructional benefits and learning outcomes, and (c) pedagogical, technical, and institutional issues and challenges.

Review Question 1: How have generative AI tools been applied to support language teaching and learning?

Analysis of the 32 included studies revealed that generative AI is most commonly used to support writing-related tasks, followed by self-directed learning, instructional design, and conversational practice (see Table 3). Across studies, generative AI functioned primarily as a teacher-supported tool. Writing feedback and revision represented the dominant application. Learners used this technology to generate ideas, refine drafts, correct errors, and improve organization.

Conversational and simulation-based uses included dialogue practice and structured interaction tasks, often supporting speaking preparation.

In teacher-focused contexts, generative AI was used for lesson planning, material development, and assessment design. Additionally, several studies highlighted its role in independent learning within their courses, where students used AI tools for vocabulary development, grammar support, and exam preparation.

Table 3

Theme 1 Categories: Generative AI as an Instructional and Design Support Tool

Category	Description	Key Findings
Tool Distribution	Distribution of generative AI platforms studied across the dataset	• ChatGPT (*n*=18) was the most common tool • General AI and multi-tool studies (*n*=6 each) followed • Few studies examined DeepSeek (*n*=1) or Copilot comparisons (*n*=1)
Instructional activity types	Types of pedagogical tasks in which generative AI was integrated	• Writing feedback and revision (*n*=18) • Independent/self-directed learning (*n*=11) • Lesson planning and instructional design (*n*=10) • Conversational/simulation tasks (*n*=9)
Cross-study Trends	Key trends in AI integration across different settings	• Primarily used for writing and structured feedback • Less used for simulations and instructional design • Functioned as a teacher-supported (assistive) tool

Note. Here, *n* refers to the number of studies, not the number of participants.

Review Question 2: What instructional benefits and pedagogical affordances have been reported in relation to generative AI use in foreign language teaching and learning?

Two themes emerged after the careful analysis of the 32 studies, indicating that the positive pedagogical impact of generative AI in foreign language education can be understood through instructional affordances and differential skill gains in language learning. These themes reflect how AI reshapes instructional processes and the measurable learning outcomes that result from its integration. Table 4 presents the categories for Theme 2 and Theme 3, along with their key findings.

Table 4

Second and Third Themes' Categories

Theme	Category	Description	Key Findings
Theme 2: Instructional Affordances of Generative AI	Personalization and adaptive feedback	Generative AI supported personalized, self-paced learning	• Can support grammar and vocabulary development • Promotes learner independence • Particularly beneficial for lower-proficiency learners through tailored feedback
Theme 2: Instructional Affordances of Generative AI	Engagement and affective impact	Studies examined learner motivation, confidence, anxiety reduction, and behavioral engagement associated with generative AI use	• Increased motivation and reduced anxiety • Greater willingness to practice • Interactive tasks enhanced engagement
Theme 2: Instructional Affordances of Generative AI	Instructional efficiency and teacher support	Teachers' benefits of integrating generative AI in daily instructions	• Saved time on planning, feedback, and content creation • Improved workflow efficiency • Required ongoing pedagogical oversight
Theme 3: Writing-Dominant Learning Outcomes	Writing and vocabulary development	Measurable language skill improvements related to AI-supported instruction	• Writing quality • Lexical variety • Grammar accuracy • Vocabulary acquisition • Improved performance through AI-supported drafting and revision
Theme 3: Writing-Dominant Learning Outcomes	Reading, speaking, and other skills	Examination of non-writing skill outcomes	• Improved reading comprehension with structured AI tasks • Limited or inconsistent gains in speaking, listening, and intercultural competence

Generative AI supported personalized, self-paced learning through adaptive feedback and iterative revision (Alkolaly et al., 2025; Chen, 2025; Koka & Khan, 2024; Sun et al., 2025). These features particularly helped lower-proficiency learners, providing targeted support in grammar, vocabulary, and writing development (Karatas et al., 2024; Koka & Khan, 2024). Across studies, increases in learner motivation, confidence, and willingness to practice were observed when generative AI tools were integrated into instruction (Chukhno, 2024; Madison, 2025; Sehlaoui, 2024). Interactive tasks, immediate feedback, and low-pressure practice environments led to higher motivation and less anxiety. However, cognitive engagement relied on task design and instructional guidance (Guo, 2025; Michel et al., 2025; Moorhouse et al., 2024).

From a teaching perspective, generative AI improved efficiency by aiding lesson planning, material creation, and feedback processes (Dornburg & Davin, 2025; Evstigneev, 2024; Murcia et al., 2025). Despite these benefits, studies consistently noted the need for teacher oversight to ensure accuracy and proper teaching methods (Alkolaly et al., 2025; Moorhouse et al., 2024). Regarding learning outcomes, the biggest and most consistent improvements were seen in writing and vocabulary development (Dziemianko, 2025; Koka & Khan, 2024; Madison, 2025; Sun et al., 2025). Learners showed gains in organization, lexical diversity, and grammatical accuracy when working on AI-supported drafting and revision.

In contrast, improvements in speaking, listening, and intercultural competence were limited or inconsistent (Chen et al., 2026; Karatas et al., 2024; Vo & Nguyen, 2024). However, higher-level communication skills depend more on human interaction and contextual classroom practice (Michel et al., 2025; Zhang & Dong, 2024).

Review Question 3: What pedagogical, institutional, and ethical issues and challenges have emerged in the generative AI implementation in foreign language teaching and learning?

Even though the overall tone of research on generative AI integration is positive, the thematic synthesis identified several important pedagogical, institutional, and ethical challenges, which formed the fourth theme of this scoping review: ethical concerns, accuracy limitations, overreliance, and institutional constraints (see Table 5).

Table 5

Fourth Theme Categories

Theme	Categories	Description	Key Findings
Theme 4: Ethical, Technical, and Institutional Implementation Challenges	Ethical and academic integrity concerns	Concerns related to plagiarism, authorship ambiguity, originality, and responsible AI use in FL education	• Risks of overreliance and academic dishonesty • Unclear authorship boundaries • Teachers more concerned than students • Need for AI literacy and ethical guidelines
Theme 4: Ethical, Technical, and Institutional Implementation Challenges	Accuracy, reliability, and bias	Limitations of AI-generated content and feedback	• Misinformation and incorrect corrections • Outdated methods and cultural bias • Need for teacher validation of AI outputs
Theme 4: Ethical, Technical, and Institutional Implementation Challenges	Overreliance and surface-level learning	Risks of overdependence on AI and limited critical thinking	• Limited critical evaluation of AI outputs • Reduced ownership of writing • Hindered deeper engagement without structured tasks
Theme 4: Ethical, Technical, and Institutional Implementation Challenges	Institutional preparedness and policy gaps	Structural and systemic barriers to effective AI integration	• Uneven teacher preparedness and limited PD • Lack of clear policies and infrastructure gaps • Need for training, integration models, and institutional support

Ethical issues, including plagiarism and authorship ambiguity, were among the most frequently reported concerns (Guo, 2025; Karatas et al., 2024; Pei et al., 2025). Students generally had a positive view of AI (Alkolaly et al., 2025; Sehlaoui, 2024), but instructors were more cautious, stressing the need for clear ethical guidelines and for organized AI literacy instruction (Evstigneev, 2024; Moorhouse et al., 2024; Murcia et al., 2025).

Many studies also reported a challenge related to inaccuracies, biased outputs, and unreliable feedback, underscoring the importance of teacher validation (Dominique et al., 2025; Dziemianko, 2025; Tokmakova & Saenko, 2025; Vigna-Taglianti, 2024). The scholars highlighted

the risks of overreliance, in which learners accepted AI-generated responses without critical evaluation, leading to superficial engagement and reduced ownership of learning (Dominique et al., 2025; Guo, 2025; Huang et al., 2025).

Finally, institutional challenges included uneven teacher preparedness, lack of professional development, and unclear policies governing AI use (Evstigneev, 2024; Moorhouse et al., 2024; Suardika, 2024). Additionally, teacher attitudes toward AI influenced student adoption and classroom integration, stressing the need for coordinated institutional support (Alkolaly et al., 2025; Sun et al., 2025). The literature suggests that successful adoption relies on institutional frameworks that support training, ethical policy development, and effective integration models.

DISCUSSION

Significance of the Findings

A synthesis of 32 empirical studies published between January 2023 and January 2026 shows that generative AI offers multiple advantages for FL education. Enhanced personalization, increased learner engagement, and improved instructional efficiency were the most frequent benefits reported across the reviewed studies. Generative AI features, including adaptive feedback, iterative revision, and self-paced practice, supported differentiated instruction and were especially helpful for lower-proficiency learners (Alkolaly et al., 2025; Karatas et al., 2024; Pei et al., 2025; Sun et al., 2025).

The writing assessment showed the strongest and long-lasting learning achievement across all skills. The use of generative AI to assist students with drafting and revision activities resulted in improved organization, lexical diversity, grammatical accuracy, and vocabulary development (Athanassopoulos et al., 2023; Guo, 2025; Huang et al., 2025; Nykyporets et al., 2025). The integration of structured reading tasks improved students' reading proficiency and enhanced their motivation to study the language (Shang et al., 2025). However, generative AI's effects on spontaneous speaking, listening, and intercultural competence were limited or inconsistent (Chen et al., 2026; Vo & Nguyen, 2024).

The literature review also indicates shifts in instructional practice. Teachers who used generative AI to assist them with lesson planning and feedback reported increased efficiency (Dornburg & Davin, 2025; Evstigneev, 2024). However, successful implementation required active teacher involvement and careful assessment of AI outputs (Michel et al., 2025; Zhang & Dong, 2024).

At the same time, recurring concerns related to academic integrity, reliability, overreliance, and institutional preparedness underscore the need for structured integration and clear policy guidance (Chukhno, 2024; Rudnik et al., 2024; Sun et al., 2025). Mainly, the scholars agree that generative AI can support writing development, personalization, and instructional efficiency when thoughtfully integrated into language teaching practice.

Implications for Research

The findings of this scoping review reveal five directions for future research on generative AI in FL education. The literature published between 2023 and January 2026 shows rapid growth, but the studies remain uneven in scope, method quality, and theoretical foundation.

First, many studies focus on short-term or exploratory outcomes, underscoring the need for longer-term research examining generative AI's lasting effects on language skills, learner independence, and motivation. A significant portion of the existing research has concentrated on immediate outcomes, such as quiz scores and participants' perceived engagement. However, there is a noticeable gap in studies exploring long-term retention and knowledge transfer across contexts.

Second, research must extend beyond outcomes focused mainly on writing. Improvements in writing were consistent, but there is limited evidence for spontaneous speaking, listening comprehension, intercultural skills, and pragmatic development. This gap may exist because writing tasks are easier to design, implement, and assess in controlled research settings. In contrast, speaking and interaction skills are harder to measure and require real-time, dynamic environments. Also, many generative AI tools are currently better suited for text-based output, which may bias research toward writing-focused outcomes. Studies should examine how AI-supported dialogue and real-time interaction can improve communication skills in real-life situations.

Third, research is uneven across languages and regions. Most studies focus on English as an FL and are primarily located in Asia and Eastern Europe, leaving questions about how generative AI impacts less commonly taught languages or multilingual learning settings.

Fourth, more attention should be given to cognitive engagement and AI literacy. Although risks of overreliance were frequently noted, few studies directly measured metacognitive development or critical evaluation skills. Future work should define AI literacy in FL education and explore how structured training and reflective tasks affect learning outcomes. For example, researchers could create interventions that include prompt evaluation, output comparison, and guided reflection, and assess their impact using performance tasks and metacognitive measures. Lastly, research should address teacher professional development and institutional frameworks. Empirical studies on the design of teacher training, policy implementation, and ethical administration remain limited. Comparisons and studies across different contexts would improve the field's theoretical and practical foundations.

Implications for Practice

The findings of this review suggest that generative AI can be effectively integrated into FL instruction through strategic educational applications. Across studies, the strongest and most consistent benefits were reported in structured writing contexts. Instructors may find it most

effective to use generative AI to develop students' writing skills, with learners using AI to draft their papers, revise them, and analyze provided feedback. However, the evidence also shows that unstructured generative AI use often leads to surface-level revisions that may or may not be necessary or accurate. Writing tasks should include guided comparisons, reflections on AI suggestions, and clear reasons for changes to keep learners cognitively engaged.

At the same time, the evidence does not support replacing face-to-face interaction with AI-mediated dialogue. The studies show that chatbot simulations increased students' willingness to practice and reduced anxiety in low-pressure settings. However, the improvements in spontaneous speaking were limited. In practice, generative AI tools were most effective during rehearsals and useful for preparation and generating ideas. Live communication and instructor feedback remain essential for higher-order proficiency.

Generative AI also supports differentiated instruction through adaptive feedback and self-paced practice. Teachers can use AI to provide targeted grammar clarification and contextualized vocabulary support. Tasks that ask students to evaluate AI outputs, spot inaccuracies, or compare different responses can support the development of critical thinking skills.

Given ongoing concerns about academic integrity and authorship, clear classroom guidelines and institutional policies are essential. Most scholars suggest that banning AI tools might not be a meaningful solution. It is recommended that educators focus on promoting AI literacy and teaching students how to use these tools appropriately in their coursework and assessments.

Finally, effective implementation depends on teacher preparation and institutional alignment. The research shows that the successful use of generative AI depends heavily on how well instructors understand these tools' capabilities and role in FL education. Professional development should go beyond simply showing how to use generative AI. It should focus on designing tasks with clear learning objectives in mind, framing prompts, evaluating AI-generated content, and addressing inaccuracies and bias. Institutional support, including clear guidelines, access to infrastructure, and open discussions across departments, can help reduce uncertainty and encourage effective implementation.

Limitations

Multiple limitations should be considered when interpreting this scoping review. Although 32 empirical studies published between 2023 and January 2026 were analyzed, the field remains at an early stage. Many studies used short-term interventions, survey designs, or relatively small samples, which limit conclusions about long-term language development and instructional impact. The research base is also uneven in scope. Most studies focused on English as an FL and were conducted mainly in Asia or Europe, with limited representation from less commonly taught languages and other countries. This focus restricts the generalizability of the findings across different learner populations and teaching environments. In addition, the selection of databases used for this review may have limited the inclusion of relevant studies, as relevant research

published in other databases, regional journals, or non-English languages might not have been considered.

The rapid development of generative AI technologies presents an additional challenge. Several studies examined specific tool versions that have changed considerably (e.g., ChatGPT 3, ChatGPT 3.5). As these systems continue to evolve, findings from earlier versions may no longer accurately reflect current capabilities. Methodological diversity makes synthesis more complex. The studies reviewed used different designs, measures, and analytical methods, which made direct comparisons difficult. Lastly, the review included only English-language publications, potentially omitting important research from other languages and regions. Taken together, these limitations show that the findings provide a structured overview of a developing area rather than conclusive evidence of effectiveness. More broad, methodologically rigorous, and diverse research is needed to strengthen the evidence base for generative AI in FL education.

CONCLUSION

This scoping review of 32 empirical studies suggests that generative AI is most useful in FL education when it is used as a teacher-supported tool, particularly for writing and instructional support. Most studies have linked the use of generative AI to improvements in writing, but its effects on speaking and listening are limited or context-dependent (Athanassopoulos et al., 2023; Guo, 2025; Huang et al., 2025; Chen et al., 2026). At the same time, the findings highlight that generative AI cannot be used effectively without teacher support, as concerns about accuracy, overreliance, and academic integrity remain (Dziemianko, 2025; Murcia et al., 2025; Rudnik et al., 2024). Future research should continue to explore its use across different languages and learning contexts, examine its long-term impact, and identify ways to address the challenges associated with this technology use.

Author

Yuliya Kozlova, Ed.D., is an Associate Professor and serves as a Cross-Training Branch Chief in the Continuing Education Language, Regional Expertise, and Culture Extension Programs, Defense Language Institute Foreign Language Center. Her areas of interest include the integration of educational technology in foreign language instruction, faculty development, and technology-enhanced learning.

REFERENCES

Alkolaly, M., Zeid, F., Al-Shamali, N., Khasawneh, M., & Tashtoush, M. (2025). Comparing lecturers and students attitude towards the role of generative artificial intelligence systems in foreign language teaching and learning. *Qubahan Academic Journal*, *5*(3), 1-15. https://doi.org/10.48161/qaj.v5n3a1583

Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural

classroom. *Advances in Mobile Learning Educational Research*, *3*(2), 818-824. https://doi.org/10.25082/AMLER.2023.02.009

Balcı, Ö. (2024). The role of ChatGPT in English as a foreign language (EFL) learning and teaching: A systematic review. *International Journal of Current Educational Studies*, *3*(1), 66-82. https://doi.org/10.46328/ijces.107

Barrett, A., & Pack, A. (2023). Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education, 20*(1), Article 59*.* https://doi.org/10.1186/s41239-023-00427-0

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Cakmak, F. (2025). Generative AI for learning English as a foreign language (EFL). In *AI applications for English language learning* (pp. 293-312). IGI Global Scientific Publishing. doi:10.4018/979-8-3693-9077-1.ch012

Chan, C. K. Y., & Tsi, L. H. (2024). Will generative AI replace teachers in higher education? A study of teacher and student perceptions. *Studies in Educational Evaluation*, *83*, Article 101395. https://doi.org/10.1016/j.stueduc.2024.101395

Chen, M. (2025, May). Generative AI in foreign language inquiry-based learning: A qualitative study of student experiences. In *Proceedings of the 2025 2nd International Conference on Digital Society and Artificial Intelligence* (pp. 151-158). https://doi.org/10.1145/3748825.3748851

Chen, Z., Wei, W., & Zou, D. (2025). Generative AI technology and language learning: Global language learners' responses to ChatGPT videos in social media. *Interactive Learning Environments*, 1-14. https://doi.org/10.1080/10494820.2025.2511248

Chen, X., Zhou, X., & Goh, Y. S. (2026). How Chinese as a foreign language learners use generative AI for oral script-writing: A qualitative perspective on cognitive scaffolding in project-based learning. *Acta Psychologica*, *262*, N.PAG. https://doi.org/10.1016/j.actpsy.2025.106071

Chukhno, O. (2024). Teachers and learners' perspectives on the use of generative AI in foreign language education. *Modern Information Technologies and Innovation Methodologies of Education in Professional Training: Methodology, Theory, Experience, Problems, 73*, 53–60. doi:10.31652/2412-1142-2024-73-53-60

Creely, E. (2024). Exploring the role of generative AI in enhancing language learning: Opportunities and challenges. *International Journal of Changes in Education*, *1*(3), 158-167. https://doi.org/10.47852/bonviewIJCE42022495

Crosthwaite, P., & Sun, S. (2025). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal*, *57*(1). https://doi.org/10.1177/00336882251386530

Delgado-Chaves, F. M., Jennings, M. J., Atalaia, A., Wolff, J., Horvath, R., Mamdouh, Z. M., ... & Baumbach, L. (2025). Transforming literature screening: The emerging role of large language models in systematic reviews. *Proceedings of the National Academy of Sciences*, *122*(2), e2411962122. https://doi.org/10.1073/pnas.2411962122

Dominique, C. V., Huang, J., Polo-Perez, N., & Zaher, A. (2025). Developing writing skills and feedback in foreign language education with ChatGPT: A multilingual

perspective. *Language Learning in Higher Education, 15*(2), 463-483. https://doi.org/10.1515/cercles-2024-0102

Dornburg, A., & Davin, K. J. (2025). ChatGPT in foreign language lesson plan creation: Trends, variability, and historical biases. *ReCALL*, *37*(3), 332–347. doi:10.1017/S0958344024000272

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, *6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Evstigneev, M. N. (2024). Planning a foreign language lesson using generative artificial intelligence technologies. *Tambov University Review. Series: Humanities, 29*(3), 617-634. https://doi.org/10.20310/1810-0201-2024-29-3-617-634

Feng, H., Li, K., & Zhang, L. J. (2025). What does AI bring to second language writing? A systematic review (2014-2024). *Language Learning & Technology*, *29*(1). https://doi.org/10.64152/10125/73629

Feng, L. (2025). Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: a comparative study. *Journal of Educational Computing Research*, *62*(8), 1741-1774. https://doi.org/10.1177/07356331241268349

Guo, X. (2025). Facilitator or thinking inhibitor: Understanding the role of ChatGPT-generated written corrective feedback in language learning. *Interactive Learning Environments*, *33*(5), 3526–3544. https://doi.org/10.1080/10494820.2024.2445177

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, *5*(1). https://doi.org/10.61414/jeti.v5i1.103

Huang, J., Mizumoto, A., & Huang, Y. L. (2025). Comparing peer feedback and generative artificial intelligence feedback in Japanese English as a foreign language speaking context: Impacts on motivation, engagement, and writing self-efficacy. *International Journal of Applied Linguistics*, 1. https://doi.org/10.1111/ijal.70020

Hubbard, P., & Schulze, M. (2025). AI and the future of language teaching. *International Journal of Computer-Assisted Language Learning and Teaching*, *15*(1). https://doi.org/10.4018/IJCALLT.378304

Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, *29*(15), 19343-19366. https://doi.org/10.1007/s10639-024-12574-6

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537-550. https://doi.org/10.1177/00336882231162868

Koka, N. A., & Khan, M. R. (2024). Exploring insights from undergraduates on the use of artificial intelligence (AI) tools to enhance proficiency in tense and aspect in foreign language

instruction. *Pakistan Journal of Life & Social Sciences*, *22*(2), 19633–19656. https://doi.org/10.57239/PJLSS-2024-22.2.001440

Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, *7*, 100266. https://doi.org/10.1016/j.caeai.2024.100266

Liao, H., Xiao, H., & Hu, B. (2023). Revolutionizing ESL teaching with generative artificial intelligence—Take ChatGPT as an example. *International Journal of New Developments in Education*, *5*(20), 39-46. doi:10.25236/IJNDE.2023.052008

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, *11*(1), 50. https://doi.org/10.1186/s40561-024-00342-5

Luo, Y., Xia, Y. & Lu, X. (2025). A systematic review of the role of artificial intelligence in second language writing education. *Digital Studies in Language and Literature*, *2*(2), 302-325. https://doi.org/10.1515/dsll-2025-0001

Madison Jr, A. (2025). The impact of generative AI on language learning: Transforming pedagogical approaches and learner experiences in the digital age. *TLEP–International Journal of Multidiscipline*, *2*(4), 90-99. https://tlepub.org/index.php/1/article/view/209

Martynyuk, O. (2024). AI tools in foreign language teaching: Pros and cons of Ukrainian and foreign practices. *Comparative Professional Pedagogy, 14*(2), 65–73. https://doi.org/10.31891/2308-4081/2024-14(2)-7

Michel, M., Bazhutkina, I., Abel, N., & Strobl, C. (2025). Collaborative writing based on generative AI models: Revision and deliberation processes in German as a foreign language. *Journal of Second Language Writing*, *67*, N.PAG. https://doi.org/10.1016/j.jslw.2025.101185

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook.* (3rd ed). Sage Publications.

Mizumoto, A. (2025). Generative AI in foreign language learning and teaching: A summary and insights from our empirical investigations. *Information and Technology in Education and Learning*, *5*(1), Inv-p001. https://doi.org/10.12937/itel.5.1.Inv.p001

Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y., & Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, *125*, 103399. https://doi.org/10.1016/j.system.2024.103399

Munn, Z., Pollock, D., Khalil, H., Alexander, L., Mclnerney, P., Godfrey, C. M., ... & Tricco, A. C. (2022). What are scoping reviews? Providing a formal definition of scoping reviews as a type of evidence synthesis. *JBI Evidence Synthesis*, *20*(4), 950-952. doi*:* 10.11124/JBIES-21-00483

Murcia, D., Osorio, M. F. J., Calderon, L. F. J., & Pipicano, Y. A. H. (2025). GenAI in language teaching, learning, and assessment: Stakeholders insights from two undergraduate language programs. *Lenguaje, 53*(1S), e20314387-e20314387. https://doi.org/10.25100/lenguaje.v53i1S.14387

Nykyporets, S. S., Melnyk, O. D., Piddubchak, S. Y., Melnyk, M. B., & Krutchenko, O. O. (2025). The efficacy of AI-driven pedagogical strategies for foreign language acquisition in higher education. *Society and National Interests, 2*(10), 24-36. https://doi.org/10.52058/3041-1572-2025-2(10)-24-36

Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *TESOL Quarterly*, *57*(4), 1571-1582. https://doi.org/10.1002/tesq.3253

Perez-Nunez, A. (2023). Exploring the potential of generative AI (ChatGPT) for foreign language instruction: Applications and challenges. *Hispania*, *106*(3), 355-362. https://dx.doi.org/10.1353/hpn.2023.a906568.

Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. Routledge. https://doi.org/10.4324/9781315833835

Rudnik, Y., Petryk, L., Kosharna, N., Dzhurylo, A., & Popova, L. (2024). The students' awareness considering the academic integrity of artificial intelligence use in terms of foreign language acquisition. *Cadernos de Educação Tecnologia e Sociedade, 17*(3), 50-60. https://doi.org/10.14571/brajets.v17.nse3.50-60

Saldaña, J. (2011). *Fundamentals of qualitative research*. Oxford University Press.

Sehlaoui, F. Z. (2024). Integrating AI in foreign language teaching and learning: Learner autonomy and tool utilization in an Algerian university. *Passerelle, 13*(2), 116-139. https://revue.univ-oran2.dz/Revue/Passerelle/index.php/Passerelle/article/view/361

Shang, J., Huang, Y., Xu, M., Huang, Y., Shen, X., Wang, G., Wang, Y., & Zhang, L. (2025). Competition between human learners and ChatGPT: enhancing university EFL students' reading comprehension and critical thinking through competitive questioning. *Computer Assisted Language Learning*, 1–35. https://doi.org/10.1080/09588221.2025.2577356

Sofronieva, E., Beleva, C., & Georgieva, G. (2025). Use of artificial intelligence in foreign language teaching. *Pedagogy/Pedagogika,* 3961-3982. https://doi.org/10.53656/ped2025-1s.03

Suardika, I. P. G. (2024). Exploring the application of generative artificial intelligence (Gen-AI) at EFL primary class in Kuta District. *Indonesian Journal of Foreign Language Studies*, *1*(1), 10-15. https://doi.org/10.65576/indofes.v1i1.3

Sun, J., Wang, Y., & Qian, Z. (2025). Examining the influence of individual-level cultural values on CFL learners' acceptance of ChatGPT for Chinese learning. *Interactive Learning Environments*, *33*(5), 3393–3407. https://doi.org/10.1080/10494820.2024.2443785

Tokmakova, Y. V., & Saenko, E. S. (2025). The use of corrective feedback from generative artificial intelligence in teaching a professional foreign language to students of an agricultural university. *Tambov University Review. Series: Humanities*, *30*(1), 50-66. https://doi.org/10.20310/1810-0201-2025-30-1-50-66

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467-473. https://doi.org/10.7326/M18-0850

Ulanova, A. E., Buniakova, A. V., & Lazko, A. Y. (2025). Artificial intelligence and foreign language learning in higher education: Analysis of research methods. *Professional Discourse & Communication, 7*(3), 54–68. https://doi.org/10.24833/2687-0126-2025-7-3-54-68

Vigna-Taglianti, J. (2024). AI-Generated images as a teaching tool in foreign language acquisition. *Technology and language*, *5*(3), 85–105. doi:10.48417/technolang.2024.03.07

Vo, A., & Nguyen, H. (2024). Generative artificial intelligence and ChatGPT in language learning: EFL students' perceptions of technology acceptance. *Journal of University Teaching and Learning Practice*, *21*(6), 199-218. doi/10.3316/informit.T202409290000404091943901204

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, *127*, 103533. https://doi.org/10.1016/j.system.2024.103533

Wang, C., Li, X., & Zou, B. (2025). Revisiting integrated model of technology acceptance among the generative AI-Powered foreign language speaking practice: Through the lens of positive psychology and intrinsic motivation. *European Journal of Education*, *60*(1), e70054. https://doi.org/10.1111/ejed.70054

Wan, Y., & Moorhouse, B. L. (2025). Challenges of AI in language education. *Insights into AI and language teaching and learning*, 32-52. doi:10.29140/9781763711600-03

Young, J. C., & Shishido, M. (2023). Investigating OpenAI's ChatGPT potentials in generating chatbot's dialogue for English as a foreign language learning. *International journal of advanced computer science and applications*, *14*(6). https://doi.org/10.14569/IJACSA.2023.0140607

Zacharis, G., & Papadakis, S. (2025). Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment. *Educational Process: International Journal*, *19*, e2025591. https://doi.org/10.22521/edupij.2025.19.591

Zhang, J., & Wang, J. (2025). Student perceptions of hybrid feedback: Using gen-AI to enhance engagement with EAP writing feedback. *The JALT CALL Journal*, *21*(2), 2175-2175. https://doi.org/10.29140/jaltcall.v21n2.2175

Zhang, Y., & Dong, C. (2024). Exploring the digital transformation of generative AI-assisted foreign language education: A socio-technical systems perspective based on mixed-methods. *Systems*, *12*(11), 462. https://doi.org/10.3390/systems12110462

Zhang, Y., & Miao, M. (2025). The application of generative artificial intelligence in POA guided foreign language teaching. In *Proceedings of the 2024 4th International Conference on Education, Language and Art (ICELA 2024)* (Vol. 907, p. 442). Springer Nature. 10.2991/978-2-38476-364-1_55

Appendix A

An Overview of the Studies and Findings of Research Included in the Analysis

Author	Year	Educational Context	Type of Generative AI Tool	Research Methods and Participants	Findings
1. Alkolaly et al.	2025	Higher education; FL education (postgraduate level); universities in Saudi Arabia	Generative AI systems (general reference to generative)	Quantitative survey study; 45 FL lecturers and 493 postgraduate FL students	Students showed higher acceptance than instructors; benefits included efficiency and personalization.
2. Athanassopoulos et al.	2023	Secondary education (Junior High School); multilingual and multicultural FL classroom in Greece; learners with refugee/migrant background learning German as a foreign language (A2 level)	ChatGPT (GPT-3.5–based generative AI chatbot)	Empirical classroom-based intervention (pilot study); pre–post writing task design; 8 students aged 15; analysis of lexical and syntactic features	Improved lexical diversity and sentence complexity in writing.
3. Chen	2025	Higher education; non-English-major undergraduate students in an FL course; university in western China	Generative AI tools (general reference to generative)	Empirical qualitative study; semi-structured interviews; 23 undergraduate students	Enhanced vocabulary, writing, and critical thinking through personalized support.
4. Chen et al.	2026	Higher education; Chinese as an FL instruction	Generative AI tools (general reference to generative)	Empirical qualitative design; participants were university-level Chinese as FL learners	Improved writing quality; limited transfer to spontaneous speaking.
5. Chen et al.	2025	Informal and self-directed FL learning; global FL learners using with ChatGPT-related language learning content on social media (YouTube)	ChatGPT	Sequential mixed-methods study. Data consisted of 1,088 first-level YouTube comments collected from 19 highly viewed ChatGPT-for-language-learning videos (Jan–Apr 2023). Participants were global FL learners commenting on social media (no demographic data reported)	Positive perceptions; issues with feedback depth and communication quality.

Author	Year	Educational Context	Type of Generative AI Tool	Research Methods and Participants	Findings
6. Chukhno	2024	FL education, English, secondary and higher education; the Ukrainian context shaped by COVID-19 and the challenges of remote learning during wartime	Generative AI tools (general category; includes AI chatbots and content-generating systems)	Empirical quantitative survey study. 683 participants (110 English language teachers, 573 senior school and university students) from multiple Ukrainian regions	Positive attitudes; concerns about ethics, reliability, and overreliance.
7. Dominique et al.	2025	Higher education; FL instruction at Durham University (UK)	ChatGPT	Qualitative empirical design. 18 university students with (four Italian, two per other language: Arabic, Chinese, French, German, Japanese, Russian, Spanish)	Increased motivation; concerns about accuracy and authorship.
8. Dziemianko	2025	University-level EFL instruction, Poland	ChatGPT-3.5 and Microsoft Copilot, compared with online dictionaries (COBUILD, LDOCE) and Google Dictionary	Experimental, within-subject design, quantitative analysis of learning outcomes; participants were 128 Polish university students enrolled in English language courses	Strong for immediate comprehension; weaker for long-term retention.
9. Dornburg & Davin	2025	FL teacher education and instructional design in higher education in U.S.	ChatGPT (GPT-4)	Quantitative design with document analysis. No human participants were involved. The dataset consisted of 50 AI-generated FL lesson plans	Effective lesson drafts; limited cultural depth and interaction.
10. Evstigneev	2024	Higher education - FL teacher education (pre-service FL	Generative AI tools, including chatbots (Mistral AI, ChatGPT variants) and multimodal AI tools	Empirical mixed-methods study; 17 undergraduate pre-service FL teachers in FL education	Positive attitudes; outputs required teacher correction.

Author	Year	Educational Context	Type of Generative AI Tool	Research Methods and Participants	Findings
		teachers); university-level FL methodology course in Russia	(Twee, Kandinsky, Suno) used for lesson planning and instructional design		
11. Guo	2025	University-level German as an FL; undergraduate students in China	ChatGPT	Mixed method design; 50 undergraduate German FL students (aged 18–21)	More feedback generated than teachers; lower accuracy and acceptance.
12. Huang et al.	2025	Higher education; EFL instruction in Japanese universities	ChatGPT	Longitudinal mixed-methods design; participants were 147 first- and second-year undergraduate EFL students	Improved engagement and writing support; concerns about dependence.
13. Karatas et al.	2024	Higher education (preparatory FL program); university in Turkey; online distance education	ChatGPT (GPT-3.5-based generative AI chatbot)	Empirical qualitative case study; 13 preparatory-level English FL students	Improved writing and motivation; limited impact on speaking/listening.
14. Koka & Khan	2024	Multiple educational contexts in FL learning, (higher education and institutional language programs)	ChatGPT	Mixed-methods design. The data were collected from approximately 450 language learners across multiple institutions	Improved vocabulary, fluency, and engagement; risks of overreliance.
15. Madison	2025	Foreign/second language learning across multiple settings (universities, community colleges, and private language schools); multilingual global learner population	ChatGPT, GPT-4, and specialized generative AI language learning applications	Empirical mixed-methods study; 450 language learners (ages 18–65) studying English, Spanish, French, or Mandarin	Improved multiple skills; increased engagement and reduced anxiety.
16. Michel et al.	2025	Higher education; German as an FL	ChatGPT	Qualitative, classroom-based design; eight participants (4 pairs), aged 18–23	Supported collaborative writing and deeper revision processes.

Author	Year	Educational Context	Type of Generative AI Tool	Research Methods and Participants	Findings
17. Moorhouse et al.	2024	Initial Language Teacher Education (ILTE); master's program in English Language Teaching (Hong Kong university).	Primarily ChatGPT; limited exposure to image generators (e.g., Hotpot.ai), writing assistants (Quill Bot), and other generative AI tools	Mixed-methods intervention study; 54 pre-service language teachers (M.Ed. students)	Improved teacher readiness and prompt design skills.
18. Murcia et al.	2025	Higher education; two undergraduate English Language Teaching programs in Colombia	ChatGPT and similar generative AI platforms used for text generation, feedback, summarization, and assessment support	Qualitative exploratory case study; 195 undergraduate students, 35 professors, and 2 program chairs	Widespread use; concerns about academic integrity and AI literacy.
19. Nykyporets et al.	2025	Higher education; FL instruction at a university in Ukraine	AI-driven instructional tools, including NLP-based automated writing feedback systems and AI chatbots	Mixed-method, multi-case design combining experimental and quasi-experimental interventions; Three cohorts of university English FL learners	Improved grammar, fluency, and engagement; need for teacher guidance.
20. Pei et al.	2025	Secondary education; Grade 10 EFL writing instruction in a Hong Kong secondary school	ChatGPT	Quasi-experimental design; 99 EFL 10 grade students	Strong writing gains, especially for lower-level learners.
21. Rudnik et al.	2024	Higher education; FL education; Ukrainian university context	Generative AI tools (general category, primarily focused on ChatGPT)	Empirical quantitative survey study; 37 undergraduate students enrolled in pedagogical specialties	Positive perceptions; strong ethical concerns and need for policies.
22. Sehlaoui	2024	Higher education; FL teaching and learning (French and English); Algerian university (teacher education context)	AI-powered language learning tools, including Duolingo, ChatGPT, DataBot (generative and adaptive AI tools)	Empirical quantitative study; 120 undergraduate students enrolled in French and English teacher education programs	Increased motivation, autonomy, and learning effectiveness.

Author	Year	Educational Context	Type of Generative AI Tool	Research Methods and Participants	Findings
23. Shang et al.	2025	University-level EFL; undergraduate students	ChatGPT	Quasi-experimental design; participants were 87 Chinese university freshmen	Improved reading comprehension and critical thinking.
24. Sofronieva et al.	2024	Higher education; FL teacher education and pedagogy programs; Sofia University "St. Kliment Ohridski," Bulgaria	ChatGPT-3.5	Empirical quantitative study; 134 undergraduate and postgraduate students in language teaching and pedagogy programs	Difficulty distinguishing AI vs human texts; ethical implications.
25. Suardika	2024	Primary education; English as an FL (EFL) instruction; Kuta District, Bali, Indonesia	Generative AI tools for text (ChatGPT), image (Stable Diffusion and DALL·E), audio (AIVA, Soundful, and Murf.ai), and video (Invideo AI) creation	Empirical descriptive qualitative study; 40 EFL primary school teachers from 22 schools	Improved engagement; barriers included infrastructure and training.
26. Sun et al.	2025	Chinese as an FL; international university students in China learning Chinese as an FL	ChatGPT	Explanatory mixed-methods design; participants were132 Chinese FL students at the university in China	Positive attitudes linked to usefulness and ease of use.
27. Tokmakova & Saenko	2025	Higher education; professional FL instruction at Voronezh State Agrarian University, Russia	DeepSeek	Empirical quasi-experimental study; 86 FL learners	Improved grammar, vocabulary, and professional language use.
28. Vigna-Taglianti	2024	Higher education; professional English for specific purposes (ESP); Landscape Architecture students at a non-linguistic university in Russia	AI image generators (e.g., Craiyon v3, Freepik Pikaso, Pixlr, DeepAI, Kandinsky 3.0, NeiroPlod, Shedevrum); text-to-image generative AI tools	Empirical design-based study; comparative analysis of seven open-source AI image generators	Visual tools improved vocabulary and engagement.

Author	Year	Educational Context	Type of Generative AI Tool	Research Methods and Participants	Findings
29. Vo & Nguyen	2024	Higher education; English-majored undergraduate students at universities in Vietnam	ChatGPT	Empirical quantitative study; questionnaire survey with 369 English-major students	Useful for reading/writing; concerns about accuracy and misuse.
30. Young & Shishido	2023	EFL; learners using a mobile, voice-enabled chatbot application for speaking and dialogue practice	ChatGPT (GPT-3.5)	Quantitative study design; 450 ChatGPT-generated dialogue samples were analyzed. No direct learner participants were involved	Effective for beginner/intermediate dialogue practice.
31. Zhang & Dong	2024	Higher education; FL programs in Chinese universities	Generative AI (general category, primarily focused on ChatGPT)	Mixed-method design; participants included FL teachers working in university-level language programs, university students enrolled in FL courses, and institutional stakeholders.	Improved personalization; concerns about bias and policy gaps.
32. Zhang & Miao	2025	Higher education (vocational college FL instruction) in China; POA (Production-Oriented Approach) - guided classroom teaching	Generative AI tools, including ChatGPT and KIMI (Chinese generative AI chatbot)	Empirical practice-based study (instructional design implementation with teacher-led classroom application); no student sample size explicitly reported	Improved efficiency and differentiation; required teacher mediation.

Applied Language Learning

Research

World Language Instructors' Perceptions of Using Generative AI Tools

Pete Swanson
United States Air Force Academy, Colorado Springs

M. Ryan Richey
United States Air Force Academy, Colorado Springs

The paradigm of second language teaching and learning has transformed significantly over the past several decades with a focus on standards and student proficiency in the target language. Emerging digital technologies continue to be practical tools for both instructors and students, among which artificial intelligence has gained substantial interest. Large language models such as ChatGPT can be described as a disruptive technology—"an innovation that significantly alters or even revolutionizes an existing industry, often introducing new and more efficient approaches, challenging methodologies, or rendering older methods obsolete" (Zimotti et al., 2024, p. 2). Following IRB approval, the researchers examined world language instructors' perceptions of using artificial intelligence for a variety of purposes framed in Su and Yang's (2023) IDEE framework. Results indicated that there are advantages and concerns when integrating artificial intelligence tools in the world language classroom. The study has implications for world language teachers, world language program coordinators, and educationalists.

Keywords: *Artificial Intelligence, World Language Teaching And Learning, Instructor Perceptions, ChatGPT, IDEE framework*

BACKGROUND

Large language models like those used in ChatGPT can be described as a disruptive technology—"an innovation that significantly alters or even revolutionizes an existing industry, often introducing new and more efficient approaches, challenging methodologies, or rendering older

methods obsolete" (Zimotti et al., 2024, p. 2). These artificial intelligence (AI) models are trained on massive amounts of textual data and can generate human-like text, answer questions, and complete other language-related tasks with incredible accuracy. Furthermore, they can predict the probability of word sequences or even generate new text based on a given input, which can advance world language (WL) teaching and learning. Such technology has the capability to personalize learning, develop curricula and learning exercises, and provide feedback to learners on their progress up the proficiency ladder. New technology requires educators to adapt, which can be rather intimidating (Zimotti et al., 2024). Thus, it is important to understand WL teachers' attitudes towards this innovative technology. The purpose of this survey research was to explore WL instructors' perceptions of using AI for instructional purposes.

LITERATURE REVIEW

AI in World Language Teaching and Learning

Benefits of Using AI in World Language Classes

By integrating AI into today's paradigm of second language (L2) teaching and learning, researchers suggest a landscape of cautious optimism tempered by skepticism, where students are able to comprehend and produce the new language while developing both linguistic proficiency and intercultural competence (de Vicente Yagüe Jara et al., 2023; Guo & Wang, 2023; Mohamed, 2023). Research indicates that generative artificial intelligence (GenAI) can be effective in improving student outcomes but only in the context of their "extreme specialization in computable tasks" (Fuchs, 2022, p. 255). These tools have been found to be particularly helpful in the following areas for language teaching and learning.

AI for Curriculum and Materials Development. When learning a new language, practice and focus on function (Hildebrandt & Swanson, 2016) play a crucial and decisive role. However, with large class sizes and varying levels of student proficiency in classes, the development of traditional teacher-created materials poses distinct challenges for language teachers (e.g., students' varying linguistic backgrounds, learning styles, and levels of prior knowledge). Employing large language models like ChatGPT can create engaging classroom materials, lesson plans, activities, and formative assessments (Çobanoğullari, 2024; Lee et al., 2023), as well as level-specific writing prompts (Baidoo-Anu & Owusu Ansah, 2023). Specifically, large language models can assist language teachers in the teaching of grammatical principles and even facilitate conversational practice (Kasneci et al., 2023) and have shown promise for enhancing core language learning skills for courses that utilize a learning management system, such as Canvas or Blackboard (Ayotunde et al., 2023).

AI for Student Feedback. Research has shown that students find teacher feedback helpful and motivating (Fong & Schallert, 2023; Graham et al., 2015). Nevertheless, providing teacher feedback can be challenging due to large class sizes, time constraints, and language learners' need for more personalized and actionable feedback (Williams, 2024). In addition to providing

immediate and comprehensive feedback that saves teachers time, AI tools offer several feedback-related benefits: (1) offering increased adaptability, interactivity, and practicality for the L2 student (Kurt & Kurt, 2024); (2) supporting students at different proficiency levels or those who have different learning needs (Guo & Wang, 2023); (3) offering language learners explanations, personalized solutions, and suggestions for improving their L2 writing (Zhang, 2024); and (4) challenging students to achieve specific L2 learning goals in a self-directed manner (Kurt & Kurt, 2024).

These benefits can lead to improved writing skills (Baidoo-Anu & Owusu Ansah, 2023). Specifically, research shows that students receiving such feedback improved not only their grammar, lexical density, syntax, and vocabulary but also the overall quality of their writing (Yang et al., 2023; Zou & Huang, 2023). Students have reported that such feedback enhances their comprehension because it is comprehensive and explained well (Zhang, 2024) while enhancing their engagement in writing in the target language (Cheng et al., 2023). Some researchers have reported that students find such feedback helpful, motivating, and even preferable over other feedback types (Fong & Schallert, 2023). Yang et al. (2023) mapped students' interactions with AI tools for feedback and found they engaged in a "dynamic process, in which students' responses changed from the initial mechanical responses at the discrete language level to more considered approaches in response to machine feedback" (p. 3837).

AI as a Vehicle to Personalize Learning and Build Engagement. AI can also personalize student learning (Su & Yang, 2022). Baskara and Mukarto (2023) reported that ChatGPT's personalization capability can help WL instructors create practice exercises that align with the L2 learners' proficiency levels, interests, and teachers' objectives. They suggested that these targeted practice activities can improve learners' L2 abilities and motivation. These tools have been shown to give instructors the ability to develop a more engaging learning experience for students (Su & Yang, 2023). Such large language models may also be able to positively influence L2 learners' motivation with respect to writing (Marzuki et al., 2023). Ali et al.'s (2023) findings suggest that ChatGPT-based teaching is motivational for language learners and that teachers should use it as a learning tool instead of avoiding it out of fear.

Concerns with AI in Education

While AI tools such as ChatGPT can serve as a valuable assistant in education, many concerns have arisen in the literature, beyond the reality that these tools currently *hallucinate*, or create inaccurate statements or feedback (Fuchs, 2022; OpenAI, 2022). These concerns include: (1) bias, data security, and inaccuracies (Almansor & Hussain, 2020; Perkins, 2023); (2) academic integrity breaches such as plagiarism, cheating, and academic fraud (Lund & Wang, 2023; Ray, 2023); (3) user over-reliance, which can hinder critical thinking and writing abilities (Bishop, 2023; García-Peñalvo, 2023); and (4) disagreements between learners and teachers with respect to acceptable generative AI usage (Barrett & Pack, 2023). Researchers have called for explicit guidelines and teacher professional development on integrating generative AI into educational contexts (Barrett & Pack, 2023) and have noted that additional safety guidelines are needed to ensure the proper use of this technology (Castro, 2023). Barrett and Pack (2023) pointed out the need for increased

professional development and policy guidelines at the institutional level. Karataş et al. (2024) noted that while AI has led to an increase in curiosity about the influence of AI (notably ChatGPT) on education and on L2 teaching and learning, they stress the need to recognize "the indispensable role of human expertise" (p. 19344). Large language models like ChatGPT may grapple to comprehend the complexities and nuances of human language, which can lead to incorrect responses as well as misunderstandings.

Given the current landscape, AI's use in L2 teaching and learning requires further investigation to ensure responsible integration, mitigate pitfalls, and maximize benefits for enriching teaching and learning WLs.

Instructors' Perceptions of Using AI

While the literature suggests that AI can be used to teach WLs, there is currently a dearth of research on the topic of WL educators' perceptions of using it (Espartinez, 2024). Findings from Iqbal et al.'s (2023) interviews with 20 higher education faculty members indicated that most faculty members want more information about AI and more professional development about how to implement it to make informed decisions. In terms of L2 writing development, Barrett and Pack's (2023) study on perceptions regarding the appropriate use of generative AI in the writing process revealed a need for explicit guidelines and teacher training on how to integrate AI into their writing-oriented lessons.

Two AI Choices

This study focuses on two AI chatbots: ChatGPT 5.1 and Microsoft Copilot. ChatGPT—Generative Pre-trained Transformer—is an advanced online AI chatbot created by developers at OpenAI that uses a large language model called a Generative Pre-trained Transformer. It allows users to enter prompts to receive humanlike text, images, or even videos that are created by AI (OpenAI, 2022). Much like other large language models, ChatGPT was trained on a much larger dataset (i.e., texts from a very large web corpus) and has shown state-of-the-art performance on an impressive range of natural-language tasks from translation to question answering and composing coherent essays at amazing speed. It can read a variety of WL scripts and contains additional features (e.g., attachment upload and voice production).

Copilot is a chat-based AI generative assistant that has natural language processing ability and machine learning algorithms that analyze user behaviors, preferences, and patterns (Microsoft, 2025). Copilots, as they are known, "have specialized functionality based on their users and use cases" (p. 1). Some Copilots are experienced standalone applications, while others are integrated into Microsoft's proprietary devices, products, and services. For example, Copilot can be used with *Microsoft Word*, *Excel*, *PowerPoint*, and *Teams* for tasks such as document creation, data analysis, and project management. Via Outlook, Microsoft 365 Copilot can help summarize key action items such as lengthy email conversations. Additionally, it can be used within Microsoft Azure to help automate infrastructure management and optimize cloud workloads. Microsoft

offers both free and paid versions of Copilot. Both ChatGPT and Copilot can communicate in different languages.

Conceptual Framework: Identify-Determine-Ensure-Evaluate (IDEE)

In an effort to align our research with best practices in curricular development, the Identify-Determine-Ensure-Evaluate (IDEE) framework (Su & Yang, 2023), a backward design framework for educational curriculum planning (Wiggins & McTighe, 2005), was used as a lens for the present exploratory study from the perspective of WL instructors at the United States Air Force Academy (USAFA). This approach allowed the researchers to (1) Identify the desired learning outcomes, (2) Determine the appropriate level of automation, (3) Ensure ethical concerns, and (4) Evaluate the effectiveness of using ChatGPT. Although Su and Yang (2023) developed this four-part framework specifically for ChatGPT use for educational purposes, its underlying emphasis on intentional instructional alignment and evaluation reflects broader curriculum design principles in general (e.g., Chapter 1 in Wiggins & McTighe, 2005, pp. 13–34) and can be extended to other generative AI applications in educational settings (Su & Yang, 2023).

***Identify Desired Outcomes*.** As a first step in the educational process, the identification of the objectives is important. Wiggins and McTighe (2005) advanced the notion that it is vital to know what knowledge is worth being familiar with, what knowledge and skills should be mastered, and which understandings can be considered enduring. That is, information that is important to remember once students have completed a course. Deng et al. (2025) caution that "a significant knowledge gap exists regarding its impact across different learning outcomes" (p. 105224). Jensen et al. (2025) warn that a failure to examine and understand such impacts can lessen the quality of learning as well as raise concerns about academic integrity. Therefore, it is important to align desired curricular outcomes with any technology used. With respect to the present study, the overall objective was to gain insight into WL instructors' perceptions of using AI for instructional purposes.

***Determine the Appropriate Level of Automation*.** Su and Yang (2023) noted that "depending on the objectives, it may be appropriate to fully automate the teaching or learning experience using educative AI or to use it as a supplement to traditional teaching methods" (p. 4). Thus, instructors can use AI for multiple purposes (e.g., creation of practice exercises, authentic reading materials) and such creations can be customized to the individual students' needs. In the case of the present study, AI can provide a more personalized learning experience for language learners whereby the technology can serve to "encourage teachers to reflect on educational content and trust between teachers and students" (Sun & Yang, p. 7). By doing so, language learners can receive increased targeted instruction based on their individual needs. Instructors can determine the level of automation on a continuum ranging from fully automated AI teaching and learning to occasional use to meet course outcomes and learning objectives.

Ensure Ethical Considerations. Su and Yang (2023) go on to state, "The ethical implications of using educative AI must be carefully considered" (p. 4) as not all technologies impact all users in a similar fashion. With respect to the large consumption of energy needed to operate large

language models like ChatGPT, the International Energy Agency (2025) noted that US power consumption is "on course to account for almost half of the growth in electricity demand between now and 2030" (p. 1). The economy is predicted to devour more electricity in 2030 for processing data than for manufacturing all energy-intensive goods combined (e.g., aluminum, cement, chemicals, and steel). As a second ethical dilemma, Gašević et al. (2023) cautioned that some student populations might be at greater risk of harm than others because the data used to train AI large language models have bias embedded within them or within the process of creating algorithms. Such bias can produce problematic results that can be harmful to certain populations. Thus, ethics, bias, and fairness are central to AI's growing influence. Su and Yang (2023) highlight the importance of ensuring that ChatGPT's feedback is free from bias and based on objective observations. Additionally, they posit that ChatGPT should not replace the important role of human coaches in providing support and guidance to instructors. With respect to the present study, the researchers acknowledge that as AI models continue to mature, they can create bias and must be monitored carefully regardless of the level of automation selected by the instructor.

Evaluate Effectiveness. Finally, researchers must also evaluate the effectiveness of ChatGPT in achieving the desired outcomes. Effectiveness can be measured by analyzing the impact of feedback from instructors using AI for educational purposes. Su and Yang (2023) advanced the notion that effectiveness can be achieved via teacher evaluations, feedback surveys, and other assessment tools. To that end, the researchers developed a participant questionnaire to gauge participants' perceptions of using AI to teach WLs.

The purpose of the present study was to explore instructors' perspectives of using GenAI tools in the WL classroom, noting that "to date no consensus has arisen regarding what constitutes appropriate use of GenAI in higher education" (Barrett & Pack, 2023, p. 2). The following research questions guided this study:

The present study is guided by the following research questions:

1. With respect to the IDEE framework:
    a. Did the participants feel that using AI in their instruction helped meet the desired course outcomes?
    b. Did the participants feel that the artificial intelligence they employed in their instruction was at the appropriate level of automation?
    c. Did the participants feel that it was ethically appropriate to use AI in their instruction given possible bias and accuracy?
    d. Did the participants feel that the use of AI in their instruction was effective for student learning?
2. What were the advantages and disadvantages of using AI for instructional purposes?
3. Under which conditions did WL instructors choose not to use AI for instructional purposes?
4. How did WL instructors employ AI pedagogically?

METHODS

Rationale

The researchers recognize the potential of GenAI tools in the field of WL teaching and learning. The purpose of this study was to explore if instructors teaching WLs were using AI for educational purposes at USAFA given that the Department of Defense (United States Air Force, 2024) places a premium on L2 learning, what AI platforms were being used, perceptions of the (dis)advantages of using AI for teaching WLs, and their use of AI to teach WLs as framed by the IDEE framework.

Procedures

Following IRB approval in July 2025, instructors (*N*=41) in the Department of Languages and Cultures at the US Air Force Academy received an email inviting them to participate in the present study with a link to a two-part questionnaire about (1) their perceptions and their use of GenAI tools for instructional and assessment purposes along with (2) a section requesting participant demographics. Following data collection, the researchers used IBM SPSS Statistics 27.0 to analyze the participants' ratings to the statements on the questionnaire as well as the demographic variables.

Questionnaire

In order to create a questionnaire to elicit participants' perceptions about using GenAI tools in the field of WL teaching and learning, the researchers reviewed the literature and developed a 10-item questionnaire using a 6-point Likert scale (1 — Strongly Disagree to 6 —Strongly Agree) with a participant demographic sheet and several open-ended questions (see Appendix A). The researchers field-tested the questionnaire with four colleagues using *Google Forms* to assess its functionality as well as its content validity and reliability. The reliability coefficient for the Likert scale questionnaire statements was high (Cronbach's alpha was .88), indicating that (1) the respondents' answers to the statements were consistent and that (2) the coefficient shows satisfactory consistency for research purposes (Henson, 2001). In addition to the Likert scale statements, the researchers asked several open-ended questions about the advantages and disadvantages of using GenAI tools for instructional purposes, if they do not use GenAI tools for instructional purposes, why they choose not to use it, and finally how those that use GenAI tools for instructional purposes employ it pedagogically. Data collection concluded in August 2025.

Participants

Of the 41 WL instructors at USAFA, 19 chose to complete the questionnaire. Eleven (58%) self-reported as male. Thirteen participants self-reported their ethnicity as White/Caucasian with two reporting as Latino/a, three as Asian, and one as multiracial. Six reported being in the US Air Force while the remaining participants reported being civilians. All of the participants held graduate degrees (masters, *n*=7; doctorate, *n*=12) and served for an average of 17.87 years (*SD*=10.85) as

an instructor of their WL. The languages they teach include Arabic (*n*=1), Chinese (*n*=3), French (*n*=3), German (*n*=2), Japanese (*n*=3), Portuguese (*n*=2), Russian (*n*=2), and Spanish (*n*=3).

Participants' mean age was 48.61 (*SD*=9.44). Thirteen reported using GenAI tools in their instruction. Of those individuals, when asked which tool was used, 47% of the participants (*n*= 9) reported using ChatGPT while 16% reported using Microsoft Copilot (*n*=3) and one using a military application.

FINDINGS

Given the small sample size of this exploratory study, data are reported with respect to Su and Yang's (2023) IDEE conceptual framework. Participant responses from the 6-point Likert scale were collapsed and reported here to reflect only agreement or disagreement with the 10 questionnaire statements.

With respect to the first of four research questions representing the four pillars of the IDEE framework (Su & Yang, 2023), 84.6% of the participants (*n*=11) agreed that using GenAI tools in their instruction helped meet course outcomes. However, when asked if it is appropriate to use AI platforms for the grading of language learner work, the 13 participants were almost equally divided. Fifty-four percent (*n*=7) disagreed while 46.2% (*n*=6) expressed agreement. The second of four IDEE research questions was addressed by asking participants about whether the GenAI tools they employed in their instruction were at the appropriate level of automation. Almost all of the respondents (84.6%) agreed that not only did the tool as a supplement in instruction help improve students' L2 acquisition, it also increased student motivation to acquire the new language, suggesting that the use of GenAI tools in the WL classroom helped personalize teaching and learning (Su & Yang, 2023).

For the third IDEE research question, regarding whether it was ethically appropriate to use GenAI tools in their instruction given possible bias and accuracy, 92.3% of the participants (*n*=12) agreed that instructors should be aware of such biases. Almost three quarters of the participants (69.2%) expressed that student data (e.g., student essays) should not be stored and shared as part of the AI model data set. However, almost all the participants (84.6%) agreed that considering its impact on the environment, the large amount of energy used to power AI applications is worth its use for WL learning and teaching. Furthermore, 92.3% (*n*=12) expressed agreement that the Department of Languages and Cultures should incorporate more GenAI tools with respect to student learning. Regarding the fourth IDEE question, about the evaluation of the effectiveness of using AI tools for achieving desired outcomes for student learning, most of the 13 participants (*n*=11) felt that their use of GenAI tools was effective and that using the tools increases their motivation to teach WLs (*n*=10).

Next, turning to research question two regarding participants' perceptions of the advantages and disadvantages of using GenAI tools for instructional purposes, five of the 13 respondents mentioned that GenAI tools can save instructors precious time preparing lessons, can help

evaluate WL written artifacts (both formative and summative), and can simplify authentic literature for elementary- and intermediate-level learners. A 27-year-old first-year instructor with a master's degree stated that "using ChatGPT saves me a lot of time while juggling other duties." Three others reported that ChatGPT rapidly helps improve target-language writing yet caution that the instructor-created prompt to evaluate learners' written artifacts must be developed carefully so as not to provide learners' inaccurate feedback. A third-year Spanish instructor with a master's degree wrote: "When I develop a prompt for providing corrective written feedback for my students, I have to tell ChatGPT exactly what I want it to review the student writing pieces for and for which grammatical features I do not want it [ChatGPT] to provide feedback."

Other responses about the advantages of using GenAI tools in the WL classroom revolved around the creation of rubrics for the assessment of productive skills (i.e., speaking, writing), searching for creative ideas to teach grammar and reading, and the creation of lesson plans. With respect to the disadvantages of using GenAI tools for instructional purposes, participants mentioned the inaccuracy of responses from ChatGPT and Copilot, an overdependence on technology, poor translation ability from English to the target language, security, privacy, and environmental concerns. A 60-year-old male instructor with a doctoral degree stated that "language teachers still need to read through AI-generated output to verify that the material is suitable [for learners] (both content- and level-wise)."

Turning to the third research question about why the participants chose not to use AI, the notions of security and privacy were mentioned. Additionally, several general philosophical, ethical, and environmental concerns were raised. Specifically, a male instructor holding a doctorate with almost 20 years of experience in the classroom pointed to "the massive amounts of water and energy used in places where water is scarce for crops and people." Others reported that they needed professional development opportunities because they were not sure how the chatbots worked or how they could be applied pedagogically.

Finally, with respect to the final research question about the participants employing GenAI tools pedagogically, much like responses about the advantages of using these tools in the WL classroom, the participants stated that they use them for multiple purposes. For beginning and intermediate learners, instructors cited using GenAI tools to create practice exercises (e.g., contextualized cloze activities) that are similar to those in their textbooks and online homework. Additionally, the participants reported that ChatGPT is adept at the provision of corrective written feedback on short written assignments. For learners at the more advanced levels, instructor responses varied from using GenAI tools when teaching literature to simple excerpts from popular literary pieces to interpreting literary critiques from multiple perspectives to summarizing lengthy news stories about current events. One response regarding languages using non-Roman script was mentioned by a Chinese instructor. She stated that the language learning model for AI platforms created in English (e.g., ChatGPT) tends to lack specific data and proficiency in WLs, which may provide less trustworthy output. This shortcoming made it difficult to use GenAI tools for her classes.

In sum, we found that participants perceived that using GenAI tools to teach WLs meets course objectives, requires effort to develop a quality prompt for the AI tool to deliver meaningful feedback, is effective for multiple purposes, and has certain distinct advantages and disadvantages.

DISCUSSION

The purpose of this study was to examine WL instructors' perceptions of using GenAI tools for a variety of purposes. The Department of Defense continues to place a premium on language and culture-enabled military personnel (United States Air Force, 2024). Findings from this study have important implications for the field with respect to WL instructors' perceptions of using AI for instructional purposes. As mentioned earlier, growing fascination with AI along with its utility is indisputable (Pazzanese, 2020). AI is advancing at unprecedented speed, and global AI spending is projected to reach $1.5 trillion in 2025 and to exceed $2 trillion by 2026 (World Economic Forum, 2025). With such investment, education may be transformed dramatically. Yet challenges remain, not the least of which is teacher perceptions. We will discuss our findings as they are framed in Su and Yang's (2023) IDEE framework.

Identify Desired Outcomes

When viewing the findings through the IDEE framework, most of the participants agreed that the GenAI tools they use, primarily ChatGPT and Microsoft Copilot, helped meet desired course outcomes and learning objectives even though the respondents were divided on AI's ability to evaluate student work. Nevertheless, such findings are encouraging and begin to fill the gap of research on AI's ability to meet educational objectives (Deng et al., 2025). Clearly articulated learning objectives—the purpose of instruction—are critical, especially when integrating AI-oriented teaching and learning activities. Objectives should describe the purpose using an action verb and express the expected performance and conditions under which the performance should occur (Hildebrandt & Swanson, 2016). When written appropriately, AI-oriented learning objectives convey instructors' expectations in terms of what students should know and be able to do after completing a course of study. Additionally, these carefully worded objectives should help improve student performance on assessments (Orr et al., 2022). According to Paesani (Glisan & Donato, 2017), via the process of designing the curriculum and setting course objectives, "what mattered more to me was not coverage of content or transmission of information, but the opportunity for students to apply what they were learning in practical, yet theoretically grounded ways" (p. vii). Integrating AI into instruction in the WL classroom in a theoretically grounded way can be beneficial for both L2 learners and their instructors.

Determine the Appropriate Level of Automation

In addition to meeting learning objectives, the participants agreed that instructors can use GenAI tools at an appropriate level of automation in order to achieve satisfactory results, even though these tools can be fallible and can provide inaccurate responses. Su and Yang (2023) included in their framework the notion that there is a continuum for AI use, stating, "depending on the objectives, it may be appropriate to fully automate the teaching or learning experience using educative AI or to use it as a supplement to traditional teaching methods" (p. 4). While researchers (Almansor & Hussain, 2020; Perkins, 2023) have identified potential concerns with AI's use for educational purposes (e.g., bias, data security, inaccuracies), the determination of an appropriate level of automation becomes necessary. For example, most of the participants (84.6%) agreed that using GenAI tools as a supplement to their instruction helped improve students' L2 acquisition. The supplemental use of GenAI tools also increased student motivation to acquire the new language with a limitation that it should not be allowed to store students' personal data. By addressing such concerns, instructors can limit risks to academic integrity (e.g., plagiarism, cheating) and personalize learning for the students. Participants believe that shaping the AI learning experience to the students, which includes content, goals, and pace of learning, can enhance student academic achievement, engagement, and motivation. Technology tools have the potential to meet students where they are and foster a deeper understanding of the material (Shrum & Glisan, 2015); with the growth and development of AI, these capabilities have only increased. Moreover, in our study, we found that participants view AI-oriented learning as being able to impact student persistence in that motivated students tend to learn more deeply, produce higher quality work, and have a stronger sense of efficacy in the face of challenges. As always, increasing student motivation holds serious implications for teaching and learning as it is the driving force behind students' desire to learn, engage, and succeed in both the classroom and in future endeavors (Hulleman & Hulleman, 2018).

Ensure Ethical Considerations

Along with determining the appropriate level of automation, there are ethical considerations that must be taken into consideration when using GenAI tools for educational purposes. First, the large quantity of electricity needed to power AI is extraordinary even though almost all of the participants (84.6%) reported that this impact on the environment was worth its use for WL learning and teaching. Second, from an educational perspective, some student populations might be at greater risk of harm than others due to embedded bias from the process of creating algorithms (Gašević et al., 2023). With respect to the present study, the researchers acknowledge that as AI models continue to mature, they must be monitored carefully regardless of the level of automation selected by the instructor. Findings from the present study align with earlier scholarship emphasizing the importance of human oversight and maintaining human guidance in technology-enhanced learning environments (Bayne, 2015; Roll & Wylie, 2016). Within the context of the Air Force Academy, the participants reported that student data (e.g., student essays) should not be stored and shared as part of the AI model data set. This aligns with Su and Yang (2023) who warned that there is potential for such technology to collect sensitive

information from students without their prior knowledge or consent. Similarly, as Su and Yang (2023) also noted, AI should be monitored so that large language models such as ChatGPT should not diminish the important role of human coaches in providing support and guidance. The authors strongly advocate that large language models such as ChatGPT should not replace the vital role of human instructors.

Evaluate Effectiveness

While such ethical issues are clearly important, results from the present study showed that 92.3% of the participants believed that the departmental faculty should integrate more GenAI tools into WL teaching and learning. The participating instructors reported using these tools to create contextualized practice exercises, provide corrective written feedback on writing assignments, and build interpretive mode exercises allowing students to experience multiple perspectives from literary excerpts. Additionally, most of the participants expressed that their use of GenAI tools was effective in saving them preparatory time, providing assistance when evaluating student-created written artifacts, and even simplifying authentic literature for elementary- and intermediate-level learners. Such findings align with other researchers who suggest that AI can be used to develop intriguing lesson plans, such as those involving gamification (Çobanoğullari, 2024). AI can help instructors create writing prompts and find authentic L2 materials such as dialogues and reading passages (Baidoo-Anu & Owusu Ansah, 2023). Additionally, results of the current study align with Krashen's (1981) seminal principle about comprehensible input. Enhancing and personalizing input in the target language so that it is slightly above a learner's current level can lead to meaningful acquisition of vocabulary and grammatical principles (Krashen, 1981); AI tools can help students and teachers in this area. However, the participants noted that ChatGPT and Co-Pilot sometimes provide inaccurate responses [input], suggesting that instructors need to examine the output from these two AI platforms to verify its accuracy. As noted by one of the participants teaching Chinese, GenAI tools trained on one language tend to lack specific data that can make their output in a different language less trustworthy. Thus, the selection of which GenAI tool to use for teaching WLs is an important consideration. Finally, participants in this study reported that they need more professional development opportunities to develop and maintain skills in these areas, similar to previous researchers (Barrett & Pack, 2023; Iqbal et al., 2023).

CONCLUSION

AI clearly holds potential for innovation in the field of WL teaching and learning as demonstrated by the findings from the present study. Given the wide range of implications from the present research, this study has elucidated certain limitations. Since USAFA is a unique institution of higher education, a replication study from a public institution of higher education would provide additional insights. Findings from this study should also be interpreted with caution due to its small sample size. Even though almost half of the department's instructors chose to participate in the study, further research involving a larger sample size is warranted. Additionally, self-reported data are not verifiable, and inaccuracies can be present such as social desirability bias

(i.e., responding untruthfully to appear favorable) and recall bias. As a result of these limitations, the authors call for further research on the use of AI for teaching WLs.

Lastly, qualitative interviews with instructors would also be insightful to provide a deeper understanding of their perceptions of using AI for educational purposes. Avenues for further research might include overviews of successful teaching practices using large language models for specific purposes such as the development of productive skill activities (i.e., speaking and writing exercises) at different proficiency levels (e.g., Novice-High, Intermediate-Mid). Furthermore, studies regarding L2 learners' perceptions of using AI to accelerate their proficiency in the target language would also provide important understandings.

Authors

Pete Swanson, Ph.D., is Professor of Spanish at the United States Air Force Academy and serves as the Division Chief for Spanish and Portuguese in the Department of Languages and Cultures. His research focuses on the identity and practices of highly effective world language teachers. He is the author of six books, scores of peer-reviewed articles and book chapters, and is Past President of the American Council on the Teaching of Foreign Languages.

Major M. Ryan Richey is a senior instructor of Spanish at the United States Air Force Academy in Colorado Springs, Colorado. He serves as the program manager for the Cadet Summer Language Immersion Program and the Cadet Semester Study Abroad program. His research focuses on non-native and native teacher and heritage speaker student language teaching abilities as well as the use of generative AI in the language classroom.

REFERENCES

Ali, J.K.M., Shamsan, M.A.A., Hezam, T.A., & Mohammed, A.A.Q. (2023). Impact of ChatGPT on learning motivation. *Journal of English Studies in Arabia Felix,2*(1), 41–49. https://doi.org/10.56540/jesaf.v2i1.51

Almansor, E. H., & Hussain, F. K. (2020). Survey on intelligent chatbots: State-of-the-art and future research directions. In L. Barolli, F. K. Hussain, & M. Ikeda (Eds.), *Complex, intelligent, and software Intensive systems* (pp. 534–543). Springer.

Ayotunde, O. O., Jamil, D. I., & Çavuş, N. (2023). The impact of artificial intelligence in foreign language learning using learning management systems: A systematic literature review. *Information Technologies and Learning Tools, 95*(3), 215–228. https://doi.org/10.3407/itlt.v95i3.5233

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4337484

Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of*

Educational Technology in Higher Education, 20(1), 59. https://doi.org/10.1186/s41239-023-00427-0

Baskara, R., & Mukarto, M. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics, 7*(2), 343–358.

Bayne, S. (2015). Posthumanism and research in digital education. In C. Haythornthwaite, R. Andrews, J. Fransman, & E.M. Meyers (Eds.), *SAGE handbook of e-learning research* (pp. 82-89). Sage.

Bishop, L. (2023). *A computer wrote this paper: What ChatGPT means for education, research, and writing* (SSRN Scholarly Paper 4338981). https://doi.org/10.2139/ssrn.4338981

Castro, C. A. (2023). Discussion about the impact of ChatGPT in education: Benefits and concerns. J*ournal of Business Theory and Practice, 11*(2). 28–34. http://dx.doi.org/10.22158/jbtp.v11n2p28

Cheng, X., Liu, Y., & Wang, C. (2023). Understanding student engagement with teacher and peer feedback in L2 writing. *System, 119*, 103176. https://doi.org/10.1016/j.system.2023.103176

Çobanoğullari, F. (2024). Learning and teaching with ChatGPT: Potentials and applications in education. *The EuroCALL Review, 31*(1), 4–15. https://doi.org/10.4995/eurocall.2024.19957

De Vicente Yagüe Jara, M., Martínez, O.L., Navarro-Navarro, V., & Cuéllar-Santiago, F. (2023). Writing, creativity, and artificial intelligence. ChatGPT in the university context. *Comunicar, 31*(77), 47–57. https://doi.org/10.3916/c77-2023-04

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies*, Computers & Education, 227*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Espartinez, A.S. (2024). Exploring student and teacher perceptions of ChatGPT use in higher education: A q-methodology study. *Computers and Education: Artificial Intelligence, 7*, 100264. https://doi.org/10.1016/j.caeai.2024.100264

Fong, C.J., & Schallert, D.L. (2023). Feedback to the future: Advancing motivational and emotional perspectives in feedback research. *Educational Psychologist, 58*(3), 146–161. https://doi.org/10.1080/00461520.2022.2134135

Fuchs, T. (2022). Human and artificial intelligence: A critical comparison. In R. M. Holm-Hadulla, J. Funke, & M. Wink (Eds.), *Intelligence - Theories and applications* (pp. 249-259). Springer.

García-Peñalvo, F.J. (2023). The perception of artificial intelligence in educational contexts after the launch of ChatGPT: Disruption or panic? *Education in the Knowledge Society, 24.* https://doi.org/10.14201/eks.31279

Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 4*, 100130. https://doi.org/10.1016/j.caeai.2023.100130

Glisan, E.W., & Donato, R. (2017). *Enacting the work of language instruction: High-leverage teaching practices*. ACTFL.

Graham, S., Hebert, M., & Harris, K.R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115*(4), 523–547. https://doi.org/10.1086/681947

Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*, 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*(3), 177–189.

Hildebrandt, S.A., & Swanson, P. (2016). *Understanding the world language edTPA: Research-based policy and practice*. Information Age Publishing.

Hulleman, C.S., & Hulleman, T. (2018, January 10). An important piece of the student motivation puzzle. *FutureEd*. https://www.future-ed.org/reversing-the-decline-in-student-motivation

International Energy Agency. (2025, April 10). AI is set to drive surging electricity demand from data centres while offering the potential to transform how the energy sector works. https://www.iea.org/news/ai-is-set-to-drive-surging-electricity-demand-from-data-centres-while-offering-the-potential-to-transform-how-the-energy-sector-works

Iqbal, N., Ahmed, H., & Azhar, K. (2023). Exploring teachers' attitudes towards using chat GPT. *Global Journal for Management and Administrative Sciences, 3*. https://doi.org/10.46568/gjmas.v3i4.163

Jensen, L.X., Buhl, A., Sharma, A., & Bearman, M. (2025). Generative AI and higher education: A review of claims from the first months of ChatGPT. *Higher Education, 89*, 1145–1161 https://doi.org/10.1007/s10734-024-01265-3

Karataş, F., Faramarz Y.A., Filiz, O.G., Karadeniz, D., & Yasemin, K. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies, 29*, 19343–19366.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F.,… & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Krashen, S.D. (1981). *Second language acquisition and second language learning*. Pergamon Press Inc.

Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education, 23*(24), 2–17. https://doi.org/10.28945/5370

Lee, U., Jung, H. C., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2023). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies, 29,* 11483–11515. https://doi.org/10.1007/s10639-023-12249-8

Lund, B.D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*. https://doi.org/10.1108/LHTN-01-2023-0009

Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2), 1–17. https://doi.org/10.1080/2331186X.2023.2236469

Microsoft. (2025). *What is a copilot?* https://www.microsoft.com/en-us/microsoft-copilot/copilot-101/what-is-copilot

Mohamed, A.M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies, 29,* 3195–3217. https://doi.org/10.1007/s10639-023-11917-z

OpenAI. (2022). *What is ChatGPT? Commonly asked questions about ChatGPT.* https://help.openai.com/en/articles/6783457-what-is-chatgpt

Orr, R.B., Csikari, M.M., Freeman, S., & Rodriguez, M.C. (2022). Writing and using learning objectives. *CBE—Life Sciences Education, 21*(3), 1-6. https://doi.org/10.1187/cbe.22-04-0073

Pazzanese, C. (2020, October 26). Ethical concerns mount as AI takes bigger decision-making role in more industries. *The Harvard Gazette*. https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-biggerdecision-making-role/

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice, 20*(2), 6–24. https://doi.org/10.53761/1.20.02.07

Ray, P.P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence Education, 26*, 582–599. https://doi.org/10.1007/s40593-016-0110-3

Shrum, J.W., & Glisan, E.W. (2015). *Teacher's handbook: Contextualized language instruction* (5th ed.). Cengage.

Su, J., & Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. *Computers and Education: Artificial Intelligence, 3*, 1–13. https://doi.org/10.1016/j.caeai.2022.100049

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, 1–12. https://doi.org/10.1177/20965311231168423

United States Air Force. (2024). *Language-enabled airmen bridge cultures, elevate missions*. https://www.aetc.af.mil/News/Article-Display/Article/3818845/language-enabled-airmen-bridge-cultures-elevate-missions/

Wiggins, G., & McTighe, J. (2005) *Understanding by design* (2nd ed.). Association for Supervision and Curriculum Development.

Williams, A. (2024). Delivering effective student feedback in higher education: An evaluation of the challenges and best practice. *International Journal of Research in Education and Science, 10*(2), 473–501. https://doi.org/10.46328/ijres.3404

World Economic Forum. (2025, December 1). *The AI-energy nexus will determine AI's impact. We must account for it better*. https://www.weforum.org/stories/2025/12/ai-energy-nexus-ai-future

Yang, H., Gao, C., & Shen, H. (2023). Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study. *Education and Information Technologies, 29*, 3837–3858. https://doi.org/10.1007/s10639-023-11991-3

Zhang, Y. (2024). Incorporating ChatGPT as an automated written corrective feedback tool into L2 writing class. *Journal of Language Teaching, 4*(4), 22–34. https://doi.org/10.54475/jlt.2024.024

Zimotti, G., Frances, C., & Whitaker, L. (2024). The future of language education: Teachers' perceptions about the surge of large language models like ChatGPT. *Technology in Language Teaching & Learning, 6*(2), 1–24. https://doi.org/10.29140/tltl.v6n2.1136

Zou, M., & Huang, L. (2023). The impact of ChatGPT on L2 writing and expected responses: Voice from doctoral students. *Education and Information Technologies, 29*(11), 13201–13219. https://doi.org/10.1007/s10639-023-12397-x

APPENDIX A

Questionnaire and IDEE Framework

The IDEE framework for using ChatGPT in education. The number in parentheses after each statement reflects the domains of the framework guiding this study.

1. **Identify the Desired Outcomes:** Before using ChatGPT or other generative AI in education (or "educative AI"), it is important to identify the objectives of the application. This ensures that the use of technology aligns with desired outcomes.
2. **Determine the Appropriate Level of Automation:** Depending on the objectives, it may be appropriate to fully automate the teaching or learning experience using educative AI or to use it as a supplement to traditional teaching methods.
3. **Ensure Ethical Considerations:** The ethical implications of using educational AI must be carefully considered, including potential biases, and their impact on teachers and students.
4. **Evaluate the Effectiveness:** It is important to evaluate the effectiveness of educative AI in achieving the desired outcomes.

Instructions

Part 1. **Please answer the following questions about using Artificial Intelligence.**		
1. Do you or have you used artificial intelligence for teaching the language(s) you teach at USAFA?	**Yes**	**No**
a. If you answered Yes to the first question above, please elaborate on what you use artificial intelligence for (e.g., creating lesson plans, assessments, practice exercises?		
b. If you answered Yes to the first question above, what platform(s) do you use (e.g., ChatGPT)?		
c. If you answered Yes to the first question above, what do you feel are the advantages of using artificial intelligence for teaching the language(s) you teach at USAFA?		

d. If you answered Yes to the first question above, what do you feel are the disadvantages of using artificial intelligence for teaching the language(s) you teach at USAFA?	
2. If you answered No to the first question above, why have you chosen not to use artificial intelligence as an instructor?	

Part 2. If you answered Yes to the first question above, please answer the following questions by rating your level of agreement from Strongly Disagree to Strongly Agree.

	Strongly Disagree	**Disagree**	**Slightly Disagree**	**Slightly Agree**	**Agree**	**Strongly Agree**
	1	2	3	4	5	6
1. Do you feel that using artificial intelligence for teaching the language(s) you teach at USAFA is effective for meeting course outcomes? (1)						
2. Do you feel using artificial intelligence for teaching the language(s) you teach at USAFA helps improve student acquisition of the language(s) you teach at USAFA? (1)						
3. Do you feel that the Department of Languages and Cultures should incorporate more use of artificial intelligence for cadet learning? (2)						
4. Do you feel the large amount of energy used to power artificial intelligence applications such as ChatGPT is worth its use for world language teaching and learning purposes considering its impact on the environment? (3)						
5. With respect to one's privacy, do you feel it is appropriate that cadets' data (e.g., short essays the cadets						

write in class, practice exercises) should be stored and shared as part of the artificial intelligence language model dataset? (3)						
6. Given the quantity of data input into an artificial intelligence model receives (e.g., language, demographics), do you feel that there is an inherent bias in these systems that cadets and instructors should be aware of because these models can provide inaccurate, misleading, and unethical information? (3)						
7. Do you feel that it is appropriate to use artificial intelligence platforms (e.g., ChatGPT) as an aid for grading cadet work? (3)						
8. Given the advantages artificial intelligence offers for instructors (e.g., saving time grading, preparing exercises), do you think using artificial intelligence in your instruction increases your motivation to teach the language(s) you teach at USAFA? (4)						
9. Given the advantages artificial intelligence offers for instructors (e.g., saving time grading, preparing exercises), do you feel the use of artificial intelligence for teaching world languages increases cadet confidence to acquire the target language? (4)						
10. Do you think that using artificial intelligence to aid your instruction helps you to meet course outcomes for cadets? (4)						

Comments: _____

Please respond:

Gender: Male(1) ____ Female(2) _____

Ethnicity: White/Caucasian(1) _____
Latino/a(2) _____
African American(3) _____
Asian(4) _____
Pacific Islander(5) _____
Multiracial(6) _____

Age _____

Language(s) you teach at USAFA _____
Are you a member of the military? Yes ____ No _____
How many years of teaching the language(s) you teach do you have in total? _____

Highest Degree Earned: B.A. / B.S. _____ Masters _____ Doctorate _____

Applied Language Learning

Action Research

Revolutionizing Language Learning: AI-Assisted Learning as a Catalyst for High Proficiency Attainment

Hanwei Tan
Defense Language Institute Foreign Language Center, Monterey, California

This action research project investigates whether an AI-Assisted Learning Framework can help instructors and learners identify, track, and reduce recurring errors and non-native structures among intermediate and advanced Chinese learners at the Defense Language Institute Foreign Language Center. Recurring grammatical, lexical, and discourse patterns were coded as key forms when they met three criteria, distinguishing them from fatigue slips or one-off mistakes. Over a 19-week cycle, five learners engaged in an AI-enhanced intervention integrating five coordinated strategies: Error-Pattern Detection, Micro-Targeted Adaptive Exercises, Multimodal Feedback, AI-Mediated Conversational Tutoring, and Adaptive Review through Spaced Repetition. Data sources included pre- and post-test scores, logs, and learner journals. AI tools were used to assist instructors in identifying key forms across student output and generating targeted instructional materials, while all pedagogical decisions and validation remained instructor-mediated. Learners engaged in approximately 4.2 hours per week of AI-supported activities in addition to full-time classroom instruction. Findings indicate that three of the five learners demonstrated measurable ILR gains, while the frequency of key form errors declined by approximately 40%. Learner reflections suggest increased awareness of habitual error patterns and greater engagement with feedback processes. Although improvements cannot be attributed to AI use alone, the results suggest that structured, continuous diagnostic cycles may enhance the salience and treatment of persistent difficulties within high-intensity language programs. The study proposes an AI-assisted learning framework as a practical, instructor-guided approach for integrating diagnostic feedback, targeted practice, and longitudinal monitoring into intensive language instruction.

Keywords: *AI-Assisted Learning, Recurring Learner Patterns, Learner Autonomy, ILR Proficiency, Adaptive Feedback, Military Language Education*

BACKGROUND

Student Learning and the Role of Noticing

Language learners make errors for a variety of reasons. Some recurring learner errors may become stabilized over time despite repeated exposure and corrective feedback, meaning the learner has been taught an accurate form or pronunciation, yet continues to produce errors persistently despite continued exposure. Han (2004) conceptualized fossilization as developmental arrest, "a process whereby learning manifests a strong tendency toward cessation" that is both persistent and resistant to normal input cycles (p. 23). These patterns may become habitualized, persistent, and resistant to change. Such patterns frequently become automatized through repeated communicative success and may be reinforced by L1 transfer, early instructional routines, or strategic simplification under cognitive load (DeKeyser, 2007; Long, 1996). Other errors are developmental and a normal step in the learning process; these errors arise as learners test hypotheses about the target language system while restructuring their evolving grammatical representations (Corder, 1967; Selinker, 1972). Such errors are transitional and typically diminish as input, feedback, and restructuring mechanisms take effect.

Causes of Language Learning Errors and Non-Native Constructions

There are many reasons why learners produce errors. A main one is the influence of the native language. For example, English-speaking learners of Chinese often overgeneralize possessive constructions (e.g., 我的父亲 *my father* instead of 父亲 *father* in neutral contexts) or insert unnecessary determiners due to the influence of their L1. L1 influence can also manifest not as overt grammatical inaccuracy but as discourse-level or rhetorical deviation. Consider the learner production: 昨天我和我女朋友坐火车去旧金山**为了**吃中餐 (literal translation as: *Yesterday I and my girlfriend taking a train to San Francisco in order to eat Chinese food*). Although grammatically interpretable, the insertion of 为了 (*in order to*) mirrors English purposive logic rather than preferred Chinese verb-chaining sequencing: 昨天我和女朋友坐火车去旧金山吃中餐 (literal translation as: *Yesterday I and my girlfriend taking a train to San Francisco to eat Chinese food*). The former construction is grammatically interpretable but less target-like in natural Chinese discourse flow. These non-native forms may impede further language development and limit learners' ability to achieve higher levels of proficiency.

Helping Students Learn and Minimize Errors

DeKeyser (2007) argues that practice strengthens existing mental representations; therefore, if learners repeatedly practice inaccurate forms, those forms may become further proceduralized rather than corrected (p. 2). Without mechanisms that destabilize entrenched routines, repeated communicative success may consolidate non-target-like forms, increasing their resistance to change. These errors require more than corrective feedback; they necessitate diagnostic

identification and structured destabilization to reopen the interlanguage system to further development.

Autonomous Language Learning

A key strategy for helping students minimize persistent errors is the development of learner autonomy and self-directed learning. When learners become more aware of their own recurring weaknesses, actively monitor their language use, and take responsibility for reviewing and correcting their own output, they are more likely to notice non-target-like patterns and engage in sustained improvement. Benson (2011) associates learner autonomy with learners' capacity to take control of their learning through planning, monitoring, and evaluating their progress (pp. 58–63). Similarly, Little (1991) defines learner autonomy as a capacity for "detachment, critical reflection, decision-making, and independent action" (p. 4). In second language acquisition, Schmidt's (1990) Noticing Hypothesis proposes that conscious attention to linguistic forms is a necessary condition for learning, suggesting that learners must notice problematic forms before they can modify their language use (pp. 140–141). Therefore, fostering autonomy and self-directed learning is not only a motivational strategy, but also an essential condition for reducing recurring stabilized learner errors and supporting sustained proficiency growth.

AI Tools in Language Teaching and Learning

Much has been written recently on the ways in which AI tools can benefit the language teaching and learning process. In her article focusing on DLIFLC students, Amini Harsin (2026) showcases case studies on how teachers can use these tools to create interactive, relevant materials tailored to individual learning needs, noting the importance of "crafting precise and pedagogically sound prompts" (p. 32). She concludes that these "tools offer a valuable partnership for language educators at DLIFLC, where innovative instructional methods are essential due to the program's fast-paced nature" (p. 49). Wallace and Lima (2025) also note the importance of well-written prompts in using these tools. They propose the *CTPSC Prompting Framework* (Context, Task, Purpose, Sample, Clarification) and offer sample prompts that would be applicable to teaching any language. Students and teachers alike must learn how to use AI tools, and not trust them blindly (Slamet, 2024); training and practice in addition to the critical review of output are crucial.

DLIFLC Teaching and Learning Challenges and Opportunities

DLIFLC instructors in Intermediate and Advanced courses must provide fast-paced instruction in a short period of time, with students being required to take a high-stakes test (the DLPT) at the end of 19 weeks of study. To support students' learning, instructors in these team-teaching courses use data from a number of pre-course assessments (termed Diagnostic Assessment, DA) in order to plan individualized instruction for each learner. These include online and in-person proficiency tests to pinpoint areas of weakness and strength and also various inventories (i.e., the Motivated Strategies for Learning Questionnaire, Barsch Learning Style Inventory, and Myers-Briggs Type Indicator). These tools are not intended to categorize learners rigidly; rather, they provide instructors with contextual information about processing tendencies, learning strategies

(such as reliance on memorized patterns, avoidance of complex structures, or overdependence on translation), and affective factors such as confidence, motivation, and anxiety that may influence performance. Instructors use this information to create instructional plans and materials, but implementing these plans and providing students with sufficient feedback can be challenging due to time limitations and workload expectations.

As an instructor of Chinese at DLIFLC, I was interested in how AI might be able to help students identify and correct their errors and non-native forms. I developed a framework integrating continuous error-pattern detection, micro-targeted adaptive practice, multimodal reinforcement, guided AI-mediated production, and instructor-managed spaced review. Rather than replacing instructor expertise, this framework is designed to enhance diagnostic precision and feedback consistency within existing instructional structures, thereby creating conditions more conducive to noticing and learning, in addition to allowing students to develop as lifelong autonomous learners.

Research Questions

Consistent with the exploratory and practice-based orientation of action research, this action research project addresses the following questions:

1. In what ways can AI tools help teachers and students as they move through the steps of error-pattern detection, micro-targeted drills, multimodal feedback, conversational tutoring, and adaptive spaced review as related to key forms?
2. How do learners describe their experience of AI-supported diagnostic and feedback processes in terms of clarity, usefulness, engagement, and their awareness of key forms?

ACTION PLAN

Participants and Context

This action research was conducted over a 19-week instructional cycle in two Chinese courses at DLIFLC—one Intermediate (entry approximately ILR 2 range) and one Advanced (entry approximately ILR 2+ range). Five learners (three Intermediate and two Advanced) were selected as participants. On average, participants engaged in AI-supported tasks for approximately 4.2 hours per week (both in- and out-of-class). Engagement time was recorded weekly to maintain relative consistency across participants and to contextualize performance trends.

Data Sources

Data collection included each learner's:

- Most recent DLPT score

- Online Proficiency Test administered during Week 1
- In-Person Proficiency Test conducted in Week 2
- Online Proficiency Test administered in Week 12
- Final DLPT score at course completion
- Reflective Journals

The pre- and post-DLPT and OPI scores were used to indicate overall proficiency development and whether each participant's listening and reading scores increased, decreased, or remained stable during the 19-week instructional cycle. The Week 1 Online Proficiency Test and Week 2 In-Person Proficiency Test served as baseline diagnostic measures. These assessments were reviewed by the instructor-researcher to identify recurring grammatical, lexical, and discourse-level patterns across learner output. Reflective journals offered insight into learners' awareness of their own recurring patterns and their perceptions of AI-mediated feedback; these reflections were reviewed during scheduled one-on-one conferences and were used to refine individualized intervention plans throughout the course.

A linguistic structure was classified as a key form only if it met three criteria: (1) it recurred across multiple tasks and modalities, (2) it persisted despite prior instruction or corrective feedback, and (3) it was present in both the Week 1 Online Proficiency Test and the Week 2 In-Person Proficiency Test. Isolated mistakes, fatigue errors, and newly introduced developmental errors were excluded. To determine the number of key form occurrences, the number of confirmed key-form occurrences was divided by the total number of words produced and multiplied by 100 to calculate a rate per 100 words. This allowed comparison of recurring error density before and after the intervention across syntactic, lexical, and discourse domains.

The five strategies described below were designed to operationalize the research questions within the constraints of action research, integrating detection, targeted intervention, learner reflection, and performance monitoring into a unified instructional cycle.

The 5-Step AI-Assisted Learning Framework

To meet students' needs and streamline instruction, I created the 5-step AI-Assisted Learning Framework: Error-Pattern Detection**,** Micro-Targeted Adaptive Exercises**,** Multimodal Feedback**,** AI-Mediated Conversational Tutoring**,** and Adaptive Review through Spaced Repetition. I used this framework throughout the course. These components operated as an integrated diagnostic–instructional loop rather than isolated interventions. Each phase generated data that informed the next. A wide range of learner production—including daily quizzes, translation assignments, reflective journals, weekly research-based presentations, debate performances, and transcribed speaking samples—served as the continuous data stream informing this cycle. All AI tool use was conducted in accordance with DLIFLC and Department of War policy. Microsoft Copilot served as the primary institutionally approved tool used on secure systems; no proprietary or sensitive material was input into AI tools.

Step 1. Detection of Key Forms and Patterns (Diagnostic Phase)

Identifying Key Forms

The researcher used Copilot to streamline the process for identifying key forms. I input learner-created texts (essays, translations, and speaking transcripts) into Copilot. The analysis targeted aspect-marker usage (such as omission or misplacement of 了, 过, 着), word-order deviations (including adverbial positioning and 把/被 constructions), discourse cohesion patterns (for example, overuse or omission of connectors such as 所以 *so*, 因此 *therefore*, 但是 *but*), and lexical precision in formal contexts. I used standardized prompt templates to ensure methodological consistency.

- Sample prompt for specific key forms, learner texts were submitted with instructions such as: "Analyze the following Chinese text and identify recurring grammatical or structural errors. Focus on aspect markers (了, 过, 着), word order, connective devices, and register. Count the frequency of each error type and indicate whether the pattern recurs across clauses."
- Sample prompt for spoken transcripts: "Review this transcript and flag repeated omissions or misuses of aspect markers or connectors. Highlight patterns that occur more than once and may indicate systematic rather than incidental errors."

AI-generated outputs provided frequency summaries and example sentences. All outputs were manually verified, ensuring that AI functioned as a co-analyst rather than an autonomous evaluator.

Identifying Patterns of Key Forms

I categorized recurring patterns into syntactic, lexical, or discourse domains. For instance, repeated omission of 了 was categorized as a syntactic aspect-marking pattern; overreliance on 想 (*want*) in formal argumentative contexts was categorized as lexical overgeneralization; and fragmented clause sequencing without cohesive markers was classified as discourse-level instability. Only patterns that appeared in at least two task types and aligned with Week 2 findings were retained as key forms.

Utilize Contrastive Modeling

To support instructional planning and targeted revision, instructors used AI tools to generate contrastive comparisons between learner-produced sentences (from assignments or assessments) and more target-like forms. These comparisons helped both instructors and learners see exactly how meaning, grammar, and discourse structure differed, with all AI-generated suggestions reviewed and validated by the instructor before being used in class.

Instructors and students used prompts such as: "Compare this learner-created text with a more target-like Chinese version. Identify any grammatical or discourse-level differences, explain why the original form is less natural or has errors, and provide a short teaching explanation." As an example, one learner had produced the following sentence in an assignment, which was flagged by AI: 我吃饭以后去图书馆 (*I eat and then go to the library.*). Copilot created the revised form: 我吃了饭以后去图书馆 (*I ate and then went to the library.*) and generated an explanation that the aspect marker 了 indicates a completed action in past-time narration, and that without 了, the sentence sounds less complete and less natural in this context. The contrast helped the student notice that the issue was not vocabulary, but aspect marking and temporal clarity; the instructor was available to provide follow-up information as needed.

For discourse-level revision, instructors and students used prompts such as: "Revise this learner paragraph to improve logical cohesion and formal register. Focus on connectors, sentence flow, and argument structure. Explain the specific changes made." As an example, one student produced fragmented logical connections in this sentence, flagged by Copilot: 经济发展很快，所以政策变化很大 (*The economy developed quickly, so policy changes were large.*) Copilot suggested the following revision: 由于经济发展很快，因此政策变化也很大 (*Because the economy developed quickly, policy changes also became significant.*). The revised version strengthened logical cohesion by using a more formal cause-and-effect structure (由于 *because*… 因此 *so*…) that is preferred in advanced academic and professional Chinese.

Through this structured sequence—identification, categorization, and contrastive modeling—the process remained systematic, transparent, and pedagogically controlled. With this information, I created a Learner Profile for each student. These individualized profiles directly informed subsequent planning. Using Copilot for analysis and explanations streamlined the pedagogical process, saving me time to focus on other tasks.

Advanced Group Examples: Students K and S

- **Student K**: This student exhibited discourse cohesion errors characterized by repeated clause fragmentation, omission of logical connectors such as 因此, and incomplete syntactic closure. Across a three-week period, 27 instances of such fragmentation were documented across writing and speaking modalities. Because this pattern appeared longitudinally and aligned with Week 2 In-Person Proficiency Test findings, it was classified as a key form.

- **Student S**: This student demonstrated aspect-marker errors, with a 43% recurrence rate of 了 omission in completed events. This syntactic instability co-occurred with lexical overgeneralization in formal contexts, including frequent reliance on 想 (*want*) and 在 (*at*). These patterns were classified as key forms.

Step 2. Micro-Targeted Exercises (Adaptive Practice Phase)

Grounded by Schmidt's (1990) Noticing Hypothesis, I designed Micro-Targeted Adaptive Exercises focusing on each learner's key forms and Learner Profile. This phase included two steps:

1) Task Creation by Instructor: I used AI tools to produce narrowly focused drills targeting a single domain at a time (e.g., aspect marking, discourse connectors, modifier placement). Prompts were standardized to enhance replicability (e.g., "Generate five sentence-ordering exercises requiring appropriate use of 因此 (*therefore*), 然而 (*however*), 不过 (*but*), with model answers" or "Create six fill-in-the-blank items requiring correct use of 了 in past-time narration"). All AI-generated materials were reviewed and refined before implementation to ensure linguistic accuracy and alignment with instructional goals.
2) Student Participation: Learners actively completed drills and revised responses; they also orally rehearsed corrected forms with their peers or the instructor. The instructor curated AI-generated explanations and provided targeted clarification. Key forms that learners produced accurately and consistently in subsequent communicative tasks were entered into instructor-managed review logs for later reinforcement. Key forms that remained unstable re-entered the micro-targeting cycle.

Advanced Group Examples: Students K and S

Advanced-level cases illustrate the process.

- **Student K**: Challenges: discourse fragmentation involved missing connectors and illogical sequencing. AI Exercises: drills emphasizing paragraph reconstruction and connector selection. When errors persisted, prompt parameters were adjusted to reduce lexical load and isolate cohesion variables.

- **Student S**: Challenges aspect omission (particularly missing 了) and lexical overgeneralization recurred across tasks. AI Exercises: structured practice progressed from recognition to controlled production and then to transfer tasks requiring time-sequenced narratives.

Learner reflections indicated that isolating one feature per cycle reduced cognitive overload and increased metacognitive noticing of previously automatized forms. By narrowing focus to entrenched patterns rather than addressing all weaknesses simultaneously, the exercises supported targeted interlanguage destabilization while preserving communicative fluency.

Step 3. Multimodal Feedback (Reinforcement Phase)

In the third stage of the AI-Assisted Learning Framework, students received multimodal feedback to reinforce corrected language forms through synchronized text, audio, and visual input. The purpose of this stage was to help students move beyond simply recognizing errors toward

accurately producing corrected forms in listening, speaking, and writing. Students processed the same target form through multiple channels—seeing it, hearing it, and producing it themselves. This approach aligns with multimedia learning theory (Mayer & Moreno, 2003), which suggests that coordinated visual and auditory input improves retention, deeper processing, and long-term procedural control. Learners engaged in structured comparison tasks. For example, when using LTEA Transcript Trainer 5.0 (a DLIFLC-supported learning platform), students first listened to the audio and transcribed the sentence exactly as spoken. They then compared their original utterances with instructor-approved reformulations, listened to native-speaker recordings of corrected forms at adjustable speeds, and reviewed color-coded overlays indicating omissions, misplacements, or structural deviations. Afterwards, they were asked to explain why the corrected version was more natural, particularly in terms of aspect markers, discourse connectors, collocations, or word order differences. Forms that students produced accurately and consistently across multiple subsequent tasks were marked as stabilized. These stabilized forms were then entered into the instructor-managed spaced review system for periodic reactivation in future lessons. Forms that remained unstable were returned to the earlier micro-targeted practice stage for additional focused work.

Step 4. AI-Mediated Tutoring and Role Play (Guided Production Phase)

AI-Mediated Tutoring extended the restructuring process into real-time text-based production. The goal was to help students transfer stabilized language patterns from controlled practice into functional communication. This stage was conducted twice a week through instructor-guided AI interaction sessions. Instructors first designed prompts based on each learner's diagnostic profile. AI tools functioned as simulated interlocutors, generating immediate responses and follow-up questions that sustained interaction and forced learners to negotiate meaning in real time.

Instructors gave students prompts that required the use of targeted structures. For example, for a student who frequently omitted aspect markers, the prompt for Copilot might be: "I am an advanced student studying Chinese Mandarin. I want your help with practicing grammar. I will describe a major decision I made in my military career and explain how that experience has changed my current professional goals. I will use completed actions and resulting states. I want you to identify any errors and explain them." After this practice activity with AI, students were ready to give oral presentations in class.

If students attempted to avoid a targeted structure during interaction, instructors redirected the conversation through follow-up prompts that required its use. This selective intervention prevented cognitive overload while ensuring that practice remained focused, measurable, and transferable to future spontaneous speaking performance.

Step 5. Adaptive Review (Instructor-Guided Spaced-Repetition Phase)

The final component, Adaptive Review through Spaced Repetition, is intended to help learners with long-term retention of key forms. While the earlier phases emphasized identifying errors,

correcting them, and practicing them in guided production, this final stage ensured that improvements became stable habits rather than temporary performance gains.

In this phase, instructors—not students—used AI tools to track learner progress across time. After students completed speaking tasks, writing assignments, translation exercises, and classroom discussions, instructors used AI-assisted analysis to review recurring error patterns and compare current performance with earlier samples. This process allowed instructors to identify whether previously corrected forms had become stable or whether they were still likely to reappear. In this sense, "longitudinal aggregation" refers to collecting language samples from different weeks and across multiple skill areas (listening, reading, speaking, and writing) in order to monitor long-term progress rather than judging performance from a single assignment.

The principle behind this stage was spaced repetition: important language forms should be reviewed repeatedly over time rather than intensively only once. Ebbinghaus (1913) demonstrated that forgetting occurs rapidly without distributed review, and later research confirmed that spaced retrieval strengthens long-term retention more effectively than massed practice (Cepeda et al., 2006). In practical teaching, this meant that instructors intentionally brought back previously corrected forms in later lessons, discussions, and assessments to test whether learners could still use them accurately under new conditions.

Instructors monitored four practical indicators of stabilization. First, they checked recurrence frequency across modalities—whether the same error appeared repeatedly in speaking, writing, listening, or translation tasks. Second, they considered learner self-reported confidence during regular student–teacher conferences, asking students whether they felt the form had become natural or still required conscious attention. Third, they observed response latency in spontaneous production—whether students could use the form quickly and automatically or whether they paused noticeably before producing it. Fourth, they examined cross-modal consistency to see whether students could use the corrected form accurately not only in writing but also in spontaneous speaking and listening interpretation. Only forms that remained accurate and automatic across time and context were removed from intensive review cycles. This prevented students from repeatedly practicing structures they had already mastered while allowing instructors to focus attention on forms most likely to re-stabilize as recurring errors.

This final phase completed the AI-Assisted Learning Framework. Diagnostic identification led to targeted intervention; intervention outcomes were monitored across time, and reactivation occurred whenever instability reappeared. In this way, short-term correction was systematically transformed into sustained interlanguage restructuring rather than temporary performance adjustment.

FINDINGS AND DISCUSSION

Three of the five learners (students M, B, K) improved in their DLPT/OPI scores from the beginning to the end of the 19-week course (see Table 1). One student (J) dropped from 2 to 1+ in their

reading DLPT score and another student (S) stayed at the same scores. Student S remained at ILR 2+/2+ in Listening and Reading despite substantial improvement in structural accuracy, discourse organization, and control of recurring patterns throughout the instructional cycle. Similarly, Student J achieved ILR 2+ in Listening but did not reach ILR 2+ in Reading. Classroom observations and diagnostic assessments suggested that student J's listening comprehension developed more rapidly than reading fluency and processing speed, particularly when interpreting complex written texts under time constraints. Diagnostic records nevertheless documented reductions in recurring error frequency for both learners across writing samples, guided speaking tasks, and reflective journals. Several students also demonstrated improved self-monitoring and greater awareness of their own recurring patterns, even when these changes were not fully reflected in final DLPT scores. These findings suggest that standardized proficiency outcomes may not fully capture ongoing developmental progress occurring during intensive language instruction.

Table 1 also shows the scores students received on the other proficiency tests they took at the beginning and end of the course. These scores reveal some of the challenges and successes students experienced as they moved through the coursework; dips in scores were used by the instructors to inform instruction. While the AI-assisted learning framework may have contributed to any gains made in DLPT/OPI scores, it should be noted that many other classroom-based and learner-based factors undoubtedly affected learners' scores as well.

Table 1

Pre- and Post-Intervention Scores for Intermediate and Advanced Participants (N=5)

		Pre-Intervention Scores			**Post-Intervention Scores**	
Group	**Student**	**Entry DLPT/OPI (L/R/S)**	**Week 1 Test (L/R/S)**	**Week 2 Test (L/R/S)**	**Week 12 Test (L/R)**	**Final DLPT/OPI (L/R/S)**
INT	**M**	2+/2/1+	2+/3/1+	2/2/1+	1+/3	2+/2+/2
	J	2/2/1+	2/1/1+	2/1+/1+	2+/2+	2+/1+/2
	B	2+/2/2	1+/2/1+	2/2+/2	2/2+	3/2+/2
ADV	**S**	2+/2+/2	1+/2+	2 Low/2 Low	–	2+/2+/2
	K	2+/2+/2	1+/2+	2/2/2	–	3/3/2+

Research Question 1

The first research question asked how AI tools can assist teachers and students with error detection and correction. Table 2 shows the number of key form errors at the beginning and the ending of the 19-week program.

Table 2

Number of Key Form Errors, Before and After Use of AI-Assisted Learning Framework

Error Domain	Pre-Intervention Key Error Rate (/100 words)	Post-Intervention Key Error Rate (/100 words)	% Reduction
Syntactic	0.41	0.24	−41%
Lexical	0.33	0.19	−42%
Discourse	0.28	0.15	−46%
Mean Reduction	-	-	≈ 43%

The approximately 43% mean reduction reflects longitudinal tracking of previously confirmed recurring key-form occurrences. Because learners concurrently received full-time instruction, the reduction cannot be attributed to AI tools in isolation. Instead, the pattern is consistent with the hypothesis that structured diagnostic review cycles—more frequent identification, targeted activation, and scheduled re-checking of stabilized forms—can increase the likelihood that previously persistent routines are destabilized and replaced by more target-like representations. The largest reductions were observed in aspect marking, discourse connector usage, and formal-register precision. Students who participated more consistently in targeted review and feedback activities generally demonstrated stronger gains in structural accuracy and control of recurring patterns.

Table 3 synthesizes representative individual cases and aggregate trends emerging from this cycle.

Table 3

Representative AI-Assisted Learning Framework Interventions on Recurring Learner Errors

Student / Focus Area	Diagnostic Findings (from DA / AI logs)	AI-Generated Micro-Tasks	Post-Intervention Improvement
Student S — Aspect-Marker & Lexical Precision	• 43 % recurrence of aspect-marker omission (了) • Overuse of high-frequency verbs (想 *want,* 在 *at*) in formal writing • Connector overuse (但是 *but*)	**Micro-drill sequence:** 1. Recognition → choose correct aspect marker 2. Controlled → fill-in or translation drills 3. Contextual → summarize short news text inserting appropriate markers 4. Transfer → compose 150-character paragraph with time sequencing	• Aspect-marker accuracy +44 %. • Reduction in 想 *(want)* / 在 (*at*) misuse across 3 weeks. • Student journal (Week 6): "I could see where '在' (*at*) doesn't belong — it finally clicked."
Student K — Discourse Cohesion & Word-Order Accuracy	• 27 instances of fragmented discourse; missing connectors (所以 *so*, 因此 *therefore*) • Modifier/adverbial misplacement; 把 /被 errors • Limited cohesion in paragraph summaries	**Adaptive sequence:** • Sentence ordering to enforce logical flow • Connector selection tasks (然而 *however*, 因此 *therefore*, 不过 *but*) • Sentence combining using 把/被 patterns • Inference-cue simplification when AI detected repeated error	• Discourse cohesion errors −38 %. • Word-order errors −41 %. • Listening summary scores +0.5 ILR.

Students S and K (Table 3) exemplify how iterative targeting of aspect marking and discourse cohesion, respectively, corresponded with measurable declines in recurrence and parallel improvement in performance indicators. These associations do not establish causation but illustrate how focused destabilization of entrenched patterns may align with upward proficiency movement. Taken together, the findings suggest that when key forms are systematically identified, repeatedly activated under communicative demands, and revisited through spaced monitoring, learners may demonstrate both reduced recurrence and improved performance stability. Within high-intensity military language programs, AI-assisted diagnostic support may enhance targeted learning and feedback processes.

Research Question 2

The second research question explored students' perceptions of the use of these AI tools. Weekly reflective journals provided insight into learners' perceptions of the intervention and their

evolving awareness of recurrent error patterns. Journal prompts invited learners to identify persistent difficulties and explain how feedback informed their revisions and subsequent performance.

Overall, students reported that they valued the AI-assisted activities and intended to continue using them independently. Many described incorporating AI into an iterative revision cycle. For example, one student wrote, "This time, I want to focus on connectors, collocations, transitional expressions, and contrastive structures." Such comments indicate increasing metacognitive awareness and intentional control over learning targets and that they use the tools for different goals. Also, learners consistently described AI-mediated feedback as pattern-oriented and cumulative across tasks. Unlike live instructor feedback—which is necessarily selective due to time constraints—AI-assisted analysis repeatedly highlighted the same forms across multiple drafts and transcripts. This recurrence appeared to increase salience. One learner observed, "I didn't realize how often I overuse '在' (*zai*) until it kept appearing in the corrections. I finally saw the pattern" (Student S, Week 6).

Importantly, learners distinguished between unfamiliar grammar and entrenched usage habits. Several reported that the intervention was most effective for forms they "already knew but kept using incorrectly," suggesting that AI-assisted feedback was particularly useful for addressing recurring patterns rather than introducing new structures. One student commented, "It's great at pointing out when you have a habit of using grammar wrong" (Student, Week 11, P55LJY).

Multimodal reinforcement also emerged as a salient theme. Learners reported that synchronized text–audio comparison, color-coded overlays, and adjustable playback speeds facilitated more precise self-monitoring. In particular, students described how comparing their own output with corrected versions across visual and auditory channels helped them identify subtle differences in word order, collocation, and discourse flow. This aligns with the design of the intervention, in which learners repeatedly processed corrected forms through multiple modalities to strengthen retention and proceduralization.

Finally, several learners indicated that they planned to continue using AI tools beyond the course, demonstrating a shift toward greater learner autonomy, with AI functioning not only as a feedback mechanism but also as a self-directed learning partner. More importantly, their reflections suggest that AI was not used passively, but rather as a tool for critical evaluation and informed, self-directed decision-making. One student explicitly described accepting and rejecting AI feedback based on communicative goals: "The words highlighted in yellow are corrections made by the AI; they are more accurate than my original expressions. The words highlighted in green are suggestions I chose to reject. I rejected them for two reasons: first, my original expression was also acceptable; second, although the AI's version was more formal, my goal was not formality but correctness." The student further provided examples of rejected suggestions, such as replacing 但是 (*but*) with 然而 (*however*), and 不管 (*no matter*) with 无论 (*regardless of*), indicating an awareness of stylistic nuance and register choice.

This type of reflection demonstrates that learners were not only internalizing corrected forms but also developing the ability to evaluate language choices based on context, purpose, and audience. In this sense, AI functioned not merely as a corrective tool but as a catalyst for higher-level metalinguistic awareness and autonomous control over language use. Taken together, the journal data indicate that the intervention fostered not only pattern recognition and accuracy, but also independent, strategic engagement with language learning beyond the classroom.

Learner reflections suggested increased awareness of recurring language patterns and greater ability to explain corrections metalinguistically. Several learners distinguished between unfamiliar grammar and long-standing habitual errors, indicating increased self-monitoring and attention to form. These observations are consistent with Schmidt's (1990) Noticing Hypothesis, which emphasizes the importance of conscious attention in language development.

Learners also responded positively to multimodal feedback activities, including synchronized text-audio comparison, color-coded visual feedback, and slowed audio playback. These activities appeared to support more precise self-monitoring and reinforced corrected forms through multiple channels of input. This finding aligns with multimedia learning research (Mayer & Moreno, 2003), which suggests that coordinated visual and auditory input may improve retention and procedural control. The integration of AI-supported comparison and feedback tools helped instructors provide these multimodal learning experiences more consistently within the constraints of an intensive instructional schedule.

Pedagogical Implications

Although this study was conducted within a DLIFLC cohort, the framework may also be relevant for other language programs characterized by intensive schedules, rotating instructors, and performance-based assessment systems. In such environments, instructors often face challenges tracking recurring learner patterns across multiple tasks, modalities, and instructional cycles. Without structured support, feedback may become fragmented or inconsistent across instructors and over time.

The AI-Assisted Learning Framework provided a structured method for organizing recurring learner patterns, monitoring changes over time, and supporting more consistent instructional follow-up. By clustering recurring forms and identifying frequency patterns, instructors were better able to focus attention on persistent weaknesses rather than isolated errors. AI functioned as a diagnostic support tool that assisted instructors in identifying patterns, preparing targeted materials, and organizing feedback, while instructional decisions and final validation remained instructor-mediated throughout the process.

Although instructor workload increased during the initial setup phase because of prompt development, calibration, and validation of AI outputs, the framework later reduced repetitive manual tracking and improved continuity across rotating teaching teams. Early investment in diagnostic organization was therefore partially offset by improved efficiency and consistency of feedback.

Reflection and Future Directions

Several limitations should be acknowledged. First, the implementation relied primarily on written production and transcribed speech because of institutional AI tool policies. Future research incorporating real-time oral analysis and prosodic tracking may provide additional insight into spontaneous speaking development. Second, although several learners demonstrated substantial gains, not all participants reached target proficiency levels during the 19-week cycle. Future studies may benefit from delayed post-tests, controlled production tasks, or comparison groups to examine the durability of observed improvements more precisely. Finally, the small sample size (n = 5) and absence of a control group limit generalizability. This study should therefore be understood as exploratory classroom-based action research demonstrating feasibility and instructional potential rather than definitive evidence of causal impact.

CONCLUSION

This action research proposed an AI-Assisted Learning Framework designed to support learner development and autonomy among intermediate and advanced Chinese learners in an intensive instructional setting. Over the 19-week instructional cycle, several participants demonstrated measurable proficiency gains, while recurring grammatical, lexical, and discourse-level patterns declined across multiple task types and modalities. The primary contribution of this study is not the claim that AI independently improves proficiency. Rather, the study demonstrates how AI-assisted diagnostic processes may support identification, reinforcement, review, and monitoring of recurring learner patterns within intensive instruction. Rather than replacing instructor expertise, the framework supported greater visibility of learner patterns and more consistent coordination within a team-taught environment.

Author

Hanwei Tan, PhD, is an Associate Professor at the Defense Language Institute Foreign Language Center (DLIFLC). He earned his PhD in Comparative Literature from Indiana University and has taught and administered language programs in both the US and China, including serving as Chair of the Chinese Department in the School of Resident Education at DLIFLC and Associate Dean of Foreign Languages at Hebei Normal University. His research interests include diagnostic assessment, learner autonomy, AI-assisted language learning, visual poetry, and comparative literary studies. Dr. Tan is the author of *Chinese Visual Poetry: The Path of a Picturesque Literature* (2018) and *A Course in American Literature* (1995), and has published internationally on language teaching, assessment, intercultural learning, literary criticism, and visual poetics.

Ethics and Consent

This action research followed DLIFLC and Department of War guidelines for minimal-risk educational studies. Students were informed of the project's purpose, the types of data collected (test scores, classroom artifacts, AI logs, journals), and their right to opt out at any time without academic consequence. All data were anonymized using pseudonyms, and identifying

information was removed from excerpts and tables. AI tools were used only for instructional diagnostics, not for high-stakes evaluation. Data were stored on secure, password-protected institutional systems and accessed only by the instructor-researcher. All procedures adhered to DLIFLC data-security and confidentiality requirements.

REFERENCES

Amini Harsin, N. (2026). Leveraging AI-driven ChatGPT for pedagogical personalization. *Dialog on Language Instruction, 36(*1), 33–49. https://www.dliflc.edu/ojs/dialog/article/view/44/38

Benson, P. (2011). *Teaching and researching autonomy in language learning* (2nd ed.). Routledge.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching, 5*(1–4), 161–170.

DeKeyser, R. M. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). Teachers College, Columbia University. (Original work published 1885).

Han, Z.-H. (2004). *Fossilization in adult second language acquisition*. Multilingual Matters.

Holec, H. (1981). *Autonomy and foreign language learning*. Pergamon Press.

Little, D. (1991). *Learner autonomy 1: definitions, issues and problems*. Authentik.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

Schmidt, R. (1990). The role of consciousness in second language learning. A*pplied Linguistics, 11*(2), 129–158.

Selinker, L. (1972). Interlanguage. I*nternational Review of Applied Linguistics in Language Teaching, 10*(3), 209–231.

Slamet, J. (2024). Potential of ChatGPT as a digital language learning assistant: EFL teachers' and students' perceptions. *Discover Artificial Intelligence, 4*, Article 46. https://doi.org/10.1007/s44163-024-00143-2

Wallace, L., & Lima, E. (2025). Using ChatGPT to elevate pronunciation pedagogy with the CTPSC prompting framework. *Proceedings of the 15th pronunciation in second language learning and teaching conference.* Iowa State University Digital Press. https://doi.org/10.31274/psllt.18671

Applied Language Learning

Action Research

An Artificial Intelligence-Enabled Tool for Real-Time Intercultural Competence Instruction

Tara L. Schendel
Defense Language Institute–Washington, D.C.

Inna Kerlin
Defense Language Institute–Washington, D.C.

This paper reports on a collaborative action research project between the Defense Advanced Research Projects Agency (DARPA) and the Defense Language Institute, Washington Office (DLI-W). The project examined the utility of using an existing interpretation system, not originally created for instructional purposes, for teaching and assessing intercultural competence. The system studied was the Cross-cultural Interaction Real-time Assistant for Negotiators and Operators (CIRANO), an artificial intelligence (AI)-enabled, consecutive interpretation and cultural norm violation detection system. CIRANO extends the ability of normal digital interpretation tools by adding capabilities to identify cultural norm violations and to detect emotional cues, and provides real-time rephrasing suggestions to ease interpersonal communication. CIRANO's utility for culturally-based foreign language instruction had not been explored. DARPA and DLI-W thus collaborated to pilot CIRANO with 19 U.S. Air Force students and four native-speaking instructors participating in a 64-week Mandarin Chinese language program. Students and instructors incorporated the tool into a scenario-based lesson. Analysis of student and instructor feedback and research assistants' observation notes indicated that CIRANO enhanced awareness of cultural differences, supported the practical application of intercultural competence principles, and generated valuable instructional interactions. However, serious limitations included the absence of monolingual support, the lack of context-sensitive feedback, and a heavy reliance on reading screen prompts. CIRANO shows potential for classroom integration, contingent upon further refinement to align with DLI-W's pedagogical environment, and to advance broader Department of War objectives in language and culture education.

Keywords: *Action Research, Artificial Intelligence (AI), Intercultural Competence, Foreign Language Instruction, Human Language Technology (HLT), Real-Time Interpretation*

BACKGROUND

This paper describes the findings of an action research project by the Defense Language Institute, Washington Office (DLI-W), which piloted a prototype of the language and cultural application CIRANO (Cross-cultural Interaction Real-time Assistant for Negotiators and Operators), developed by the Defense Advanced Research Projects Agency (DARPA) and its partners (Defense Advanced Research Projects Agency, n.d.; Defense Advanced Research Projects Agency, 2021). The pilot's main purpose was to explore CIRANO's potential to enhance and enrich DLI-W's training programs related to its main mission, which is to provide regional and culturally-based foreign language education globally in support of the Department of Defense.

DLI-W is seeking instructional tools to enhance its programs for intercultural competency, which applies to all students regardless of their military occupational specialty (per CJCSI 3126.01C Joint Chiefs of Staff, 2023 and DoDI 5160.70 (Office of the Under Secretary of Defense for Personnel and Readiness, 2016). Additionally, DLI-W is working with its vendor schools to implement its Final Learning Objectives for its Foreign Area Officers (FAO FLOs). FAO FLOs outline the minimum language skills, intercultural competence, and regional knowledge that FAOs must possess by the end of their full basic language course at DLI-Washington to effectively meet their key mission, which is to build relationships and alliances. These FLOs operationalize both CJCSI 3126.01C and DoDi 5160.70 at the ground level. However, selecting the most appropriate instructional approach and formative assessment of intercultural competence continues to be an ongoing challenge in the foreign language instructional environment (Al-Afifi et al., 2025; Jin et al., 2025). Additionally, the understanding of the concept of intercultural competence and the knowledge and skills associated with it continue to evolve (Byram et al., 2002; Deardorff, 2006; Luo & Chan, 2022). For example, Deardorff's (2004, p. 194) definition focuses on the "the ability to communicate effectively and appropriately in intercultural situations based on one's intercultural knowledge, skills, and attitudes" (as cited in Deardorff, 2006), while Luo and Chan (2022) recontextualize IC as an adaptive toolkit for a world where people simultaneously inhabit physical spaces and borderless digital landscapes. Preparing foreign language learners in the military for this range of intercultural interactions and competencies is an ongoing challenge. This action research project provided DLI-W with the opportunity to evaluate CIRANO and determine its potential for future adoption, adaptation, and maturation in the DLI-W foreign language program. It provided DARPA and its associates with real-world user input on the performance and utility of the application in a military foreign language training environment.

Understanding CIRANO

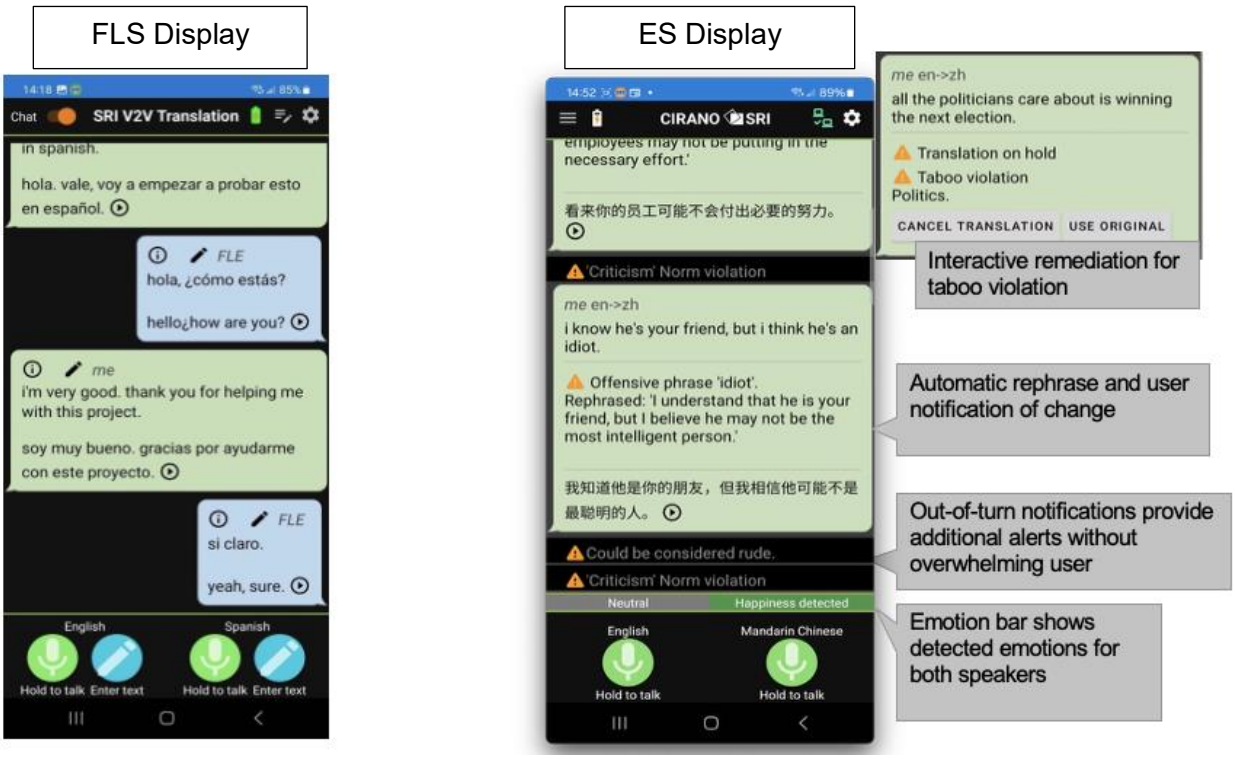
CIRANO is an updated version of previously developed an AI-assisted consecutive interpretation system currently used in the field by Special Operations Command forces when human interpreters are unavailable. These AI supported systems merely interpret, but produce no insight into non-verbal communications or possible cultural missteps that might be degrading communication and putting operators at risk.

DARPA took the existing interpretation system and added newly developed algorithms for emotion detection and detection of cultural norm violations in interactions between an English Speaker (ES) and a Foreign Language Speaker (FLS). DARPA funded and developed these algorithms as part of Cross-Cultural Understanding (CCU) Program. These algorithms use word choice, tone, and volume to detect violations of cultural norms and some speaker emotions (Defense Advanced Research Projects Agency, n.d.). Plutchik's (1980) general psychoevolutionary theory of emotion served as the foundational theoretical framework for the DARPA project's emotion detection. For the development of the CIRANO algorithms, Wang et al. (2023) note that a unique framework was created which drew from multiple other theories, including research on descriptive cultural norms (Cialdini et al., 1990) and situated reasoning (Emelin et al., 2021).

CIRANO was originally developed for use with an Oculus VR headset. For this pilot, DLI-Washington requested that it be re-designed as a cell phone application, which would be more appropriate for a training environment and more easily accessible to students and teachers. Each participant had a cellphone with CIRANO installed, one for ES and one for the FLS. The ES cellphone displays the most comprehensive information while the FLS cellphone has only limited information.

When either an ES or FLS speak, CIRANO determines the gender and approximate age of the speakers which impacts cultural interactions and language use in many cultures. When the ES or FLS speak, CIRANO transcribes what is said, but does not interpret it. Interpretation only occurs once the ES or FS signal CIRANO to do so. In addition to the transcription, the ES cellphone displays warnings about possible culture and taboo violations that could adversely impact communication (see Figure 1). In some cases, the system offers the ES alternative phrasing recommendations along with the cues to help mitigate potential rudeness. CIRANO also provides warnings to the ES about negative emotions such as anger or tiredness that either they or their interlocutors may be transmitting. Armed with this information, the ES can make adjustments by rephrasing a communication before allowing CIRANO to interpret or go forward without changes. The tool is designed to flag violations, but it is ultimately up to the user to determine the proper course of action and when an interpretation will occur.

Figure 1
CIRANO Cellphone Displays



The FLS cellphone displays only the transcript and translation without providing any cues or rephrasing options. This is by design as CIRANO is intended to aid the ES field operator. At the end of a dialog, CIRANO produces a full transcript of the exchange in .pdf format which includes translations, transcriptions, cues provided, remediation and rephrasing suggestions, and English-speaker choices.

The transcripts provided by CIRANO are a key feature to this tool. Teacher-student in-person interactions are undoubtedly crucial for learning both the language and intercultural communication, but they lack any physical artifact that can be used for future learning or reflection.

The decision was made to have students speak English rather than Chinese-Mandarin in this pilot for several reasons: 1.) CIRANO was not designed or optimized to support mono-lingual conversations, 2.) CIRANO was not optimized to process students' spoken target language, and 3.) the focus of the pilot was on intercultural competence vs. foreign language proficiency. Allowing students to be themselves in an interaction unhindered by language limitations would allow for more natural interactions and more norm violations.

ACTION PLAN

This section describes the project goal and objectives, participants, design, implementation, and data analysis for this pilot. The overarching goal was to explore the practical application of the CIRANO tool within a Department of War foreign language training environment.

> Research Question: What were the perceptions of the students, teachers, and research assistants with their interactions with CIRANO, including their perceptions of the tool's strengths and weaknesses for use as a training tool?

This project offered the opportunity to evaluate CIRANO and consider these new technologies for additional finding for future adoption, adaptation, and maturation; as well as giving DARPA the opportunity to receive feedback from DLI-W on how CIRANO might be improved to better serve DLI-W and DLIFLC's training needs.

The DLI-W team developed the pilot design and procedures in close collaboration with DARPA and their partners, as well as the selected DLI-W vendor school. Prior to conducting the pilot, DLI-W presented its pilot design and plan to DLIFLC's Human Research Protections Program (HRPP) for review. The project was approved by DLIFLC HRPP in January 2025 as it met the ethical review requirements for a project focusing on instructional technology applications.

Participants

DLI-W selected the participants using a purposeful and convenient sampling method based on student and teacher availability at the time of the pilot. The students were 19 U.S. military personnel training as cryptologic language analysts, four native Chinese-Mandarin teachers, and five research assistants (see Table 1). All participants had varying degrees of familiarity with instructional technology and AI-enabled tools based on their self-descriptions at the start of the pilot. The students and the teachers had no prior knowledge or experience utilizing CIRANO.

Table 1
Participants Overview

Category	Number	Details
Students	19	• 18 U.S. military personnel (cryptologic language analysts) • 1 military spouse • Enrolled in DLI-W full basic course for Chinese-Mandarin • First half of 64-week course • Estimated ILR proficiency: 1 to 1+ (speaking and listening)
Teachers/FLS	4	• Native Mandarin Chinese speakers • Assigned to teach the student group throughout course
Research assistants	5	• Representatives from DARPA, DLI-W, and vendor school

To further highlight the specifics relevant to the study's participants, Table 2 provides a description of the DLI-W instructional environment.

Table 2
DLI-W Learning Environment

Aspect	Description
Class Size	Maximum 5 students per section
Daily In-Person Instruction	6 hours
Daily Homework	Up to 2 hours
Instructional Format	Group instruction

Pilot Design and Implementation

The pilot design consisted of extensive development of the instructional scenarios and pre-briefing of the participants. The team designed the scenarios to align with the students' syllabus and learning objectives for their progress in their course. Scenarios were focused on situations that involved known differences between Chinese and American cultures to generate dialog and possible norm violations detectable by CIRANO The team consulted the Chinese instructors to devise the two final scenarios as represented in Table 3.

Table 3
Instructional Scenarios

Day 1: Formal Scenario	Day 2: Informal Scenario
You are unhappy with your teacher's feedback. Talk to your teacher. (Convince the teacher to change feedback).	You must decline an invitation from your Chinese friend. Find a way to do it without breaking the relationship.

Approximately two weeks before the pilot, vendor personnel, DLI-W, and DARPA representatives met with the students and teachers to introduce and discuss pilot goals and design, as well as to familiarize them with the CIRANO application for the first time. This was a hands-on exploration before the actual pilot. Teachers were given the role-plays a few days before the pilot to prepare their strategies and approaches in the interactions. Students received their role-play scenarios on pilot days in order to encourage more natural, impromptu interactions from students.

DLI-W and DARPA representatives conducted the pilot on March 4-6, 2025, at the DLI-W vendor facility. During the pilot, the students played the roles of English speakers while the teachers played the roles of Chinese speakers. This allowed participants to focus on intercultural competency and their personal identity without potential interference of the students' nascent foreign language ability (Byram et al., 2002).

The participants were broken into two groups:

- Group A: nine students and two teachers worked through a formal scenario. This scenario was used on the first day in both morning and afternoon sessions.
- Group B: ten students and two different teachers played through an informal scenario. This scenario was used on the second day in both morning and afternoon sessions.

During the pilot, morning and afternoon sessions were held with a lunch break between the two sessions. Four to five students used CIRANO during each session. Each session consisted of one student and teacher interacting with CIRANO for approximately 10 minutes. Two official research assistants monitored each exchange in order to fill out standardized observation checklists. Up to three additional research assistants were present at any given time but did not fill out the official observer survey. Research assistants changed at the end of each morning session.

After their individual sessions with CIRANO, students filled out a written feedback form in Microsoft Forms intended to capture their first impressions of the tool without the influence of others (see Appendix A). Four to five students then spoke with each teacher before the teacher also filled out a written feedback form to capture their first impressions (see Appendix B). The students received their own transcripts as well as the transcripts of their fellow students in the same session. After-action review (AAR) sessions were conducted at the end of each session with teachers, students, and research assistants using non-structured open-ended questions to facilitate group discussion. These sessions provided teachers and students with the opportunity to discuss their interactions while reviewing the transcripts, to ask questions, and to get more in-depth information about the various violations and cues CIRANO provided.

Several days after the pilot phase, the team conducted two separate focus group sessions (one with students, one with teachers) using semi-structured protocols. Each focus group session lasted about one hour. The purpose of these sessions was: (a) to capture any additional comments or ideas that arose after a period of reflection; and (b) to compare the focus group feedback to the individual feedback forms to identify any significant differences.

Data Collection and Analysis

The pilot design allowed for collection and analysis of both quantitative and qualitative feedback from the students, teachers, and research assistants. The quantitative and qualitative data were collected from the written feedback forms mentioned above. The forms contained a mix of Likert Scale questions (i.e., Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree) and open-ended prompts. The team gathered additional qualitative data from the focus groups and the research assistants' written notes. Overall, the quantitative and qualitative data supported each other. There was a high degree of agreement and consistency among participants across the various feedback instruments. Comments on the individual feedback forms aligned closely with those made during the focus groups and the AARs. Research assistants' comments also aligned well with those of students and teachers.

Our thematic analysis of the qualitative data produced 81 initial codes and 13 preliminary categories, which resulted in two final themes—Technology Performance and Learning and Instruction—and 8 final categories (see Table 4).

Table 4

Final Themes and Categories

Themes	Categories
Technology Performance	• Technical Features • Translation Quality • Emotion Detection, Culture Cues, Taboo Cues • Authenticity of Conversation
Learning and Instruction	• Learning • Instruction • Intercultural Competence • Tool Applicability

Limitations

Several limitations were observed during the pilot. The first was the student sample chosen for the project. Initially, DLI-W intended to recruit participants from among students who serve as Foreign Area Officers or Defense Attachés. Their jobs require them to develop strong speaking and participatory listening skills as well as intercultural competence. However, at the time of the pilot planning and implementation DLI-W did not have any FAOs enrolled in Mandarin Chinese language training. The 19 students who were ultimately selected to participate in the pilot were all Air Force Cryptologic Language Analysts. Cryptologic Language Analysts predominately work in secure, compartmentalized environments utilizing their reading and listening skills. Triaging material and completing tasks associated with transcription and translation work take up much of their time. They do not use their speaking skills for their work and are passive listeners. However, they learn speaking skills and intercultural competence as part of their language training with DLI-W. Therefore, the students were very receptive to the opportunity to interact with a native speaker, even in English, and to engage with the target language culture.

A second limitation was the potential for personal bias from both participants and research assistants. To address this, the team used several methods to collect qualitative and quantitative information, such as individual written surveys, focus groups discussions, and observer notes. All these steps helped strengthen the reliability and trustworthiness of the team's analysis and findings.

FINDINGS AND DISCUSSION

Overall, the results indicate that the students, teachers, and research assistants perceived the CIRANO tool as facilitating a successful instructional interaction that resulted in culturally-based language learning. It also provided insight into how such technology could be integrated within a foreign language training context related to the guidance in CJCSI 3126.01C (Joint Chiefs of Staff, 2023). However, the results clearly indicate that the tool requires further development to be applicable within the DLI-W instructional environment.

Intercultural Competence

Both teachers and students agreed that the tool helped facilitate intercultural communication, with 42% of students strongly agreeing with that statement. Multiple teachers and students reported that alternative phrasings and suggestions provided by CIRANO helped improve cultural appropriateness, which helped facilitate communication. The research assistants noted that the communication was smooth and without major interruptions. Additionally, a majority of students (80%) reported that they adjusted their communicative approach based on the suggestions provided by the tool. These results align with the qualitative feedback about the value of language phrasing options in choosing the most appropriate speaking strategy. This also aligns with the concept of the intercultural competence skills of *discovery* and *interaction* described by Byram et al. (2002).

Students also reported that using the tool raised their overall self-awareness, making them more thoughtful about their posture and lack of eye contact with their dialog partner. This resulted in adjustment of their behavior, aligning with what Deardorff (2006) describes as the desired external outcome at the top of her pyramid model, i.e., effective communication and behavior in intercultural contexts. Additionally, students commented that CIRANO raised their awareness of their own communication patterns and linguistic choices in English, such as using "like" frequently as a filler word. By uncovering their automated linguistic habits, the students engaged in what Byram et al. (2002) define as *savoir s'engager* (critical cultural awareness)—the ability to critically evaluate features of one's own culture and communication practices rather than treating them as the universal standard. Furthermore, this sudden hyper-awareness of internal speech patterns aligns with Deardorff's (2006) process model, which asserts that true intercultural competence begins with an inward-facing baseline of cultural self-awareness. Students reported that the emotion cues raised awareness of their own voice, tone, lexicon, and intonation. Students reported changing their tone to achieve a more "neutral" emotion indication from CIRANO.

Teachers had mixed feelings about how CIRANO represented their speech. Some felt it made them sound either more abrupt or more polite than intended. During the focus group sessions, both students and teachers suggested that CIRANO's segmentation of Chinese sentences may have contributed to this discrepancy. This is consistent with the research literature on the subject of rendition of oral speech in written language. Unlike structured written language,

natural, spontaneous speech is inherently characterized by disfluencies, fragmented phrases, unfinished sentences, and continuous run-on sentences. This organic oral communication diverges sharply from the neatly bounded grammatical rules governing written discourse. As a result, the automated evaluation software struggled with the strict application of written language rules—such as rigid syntactic structures and punctuation mapping—when forced to process authentic spoken language. This fundamental structural tension is heavily documented in natural language processing research; for instance, Fraser et al. (2015) stress that the definitive concept of a "sentence" as structured in written typography does not naturally exist in spoken language. Because automated pipelines struggle to superimpose written syntactic boundaries onto the messy, continuous flow of oral delivery, segmentation errors frequently compound. When evaluating languages that lack explicit orthographic word boundaries like Chinese, these architectural constraints become even more acute. As noted by Tsang et al. (2025), Chinese characters are processed without standard visual spacing, making word segmentation and subsequent sentence boundary detection inherently ambiguous and highly prone to error transmission in automated software pipelines. By trying to fit spontaneous speech into pre-programmed written standards, CIRANO introduced punctuation and structural segmentation errors that artificially penalized the students' authentic oral productions.

Additionally, both students and teachers appeared to enjoy the interactions via CIRANO, as noted by the research assistants in their observation checklists. Students and teachers used idiomatic expressions to intentionally test the system's responses. In some instances, students and teachers attempted to come across as rude or forceful in order to trigger a warning from CIRANO. The research assistants interpreted this behavior as curiosity about the technology and as maintaining motivation and engagement, rather than as detracting from the main goal of the activity. Periodically, the system would reveal an awkward moment eliciting laughter from the group. In one scenario, to avoid attending a party, the student used his dog as an excuse. Immediately realizing the inappropriateness of this excuse (even without a cue from CIRANO), he and his teacher laughed, and he tried again. This instance came up again in a focus group with the student still embarrassed by his tactic. This finding aligns with the literature which suggests that attitude is a fundamental requirement for the successful development of intercultural competence (Byram et al., 2002; Deardorff, 2006).

In another scenario, one or two students planned ahead and took time to prepare to see if they could successfully navigate the interaction with a minimum of assistance from CIRANO. In the scenario, the student, originally from the Philippines, attempted to convince her teacher (the interlocutor) to call her father about her progress instead of her mother. When the teacher responded that it was only appropriate for her to contact the mother, the student called upon her knowledge of Filipino culture and invited the teacher to her home for lunch on the weekend. This behavior was not flagged by CIRANO because, it turned out, this was an appropriate response in Chinese culture as well.

Taboo and Culture Violations

Overall, students found the culture violation warnings and taboo indicators useful, even if sometimes inaccurate. Over half of the students (68%) reported that the cues provided by CIRANO helped them to understand the Chinese culture norms (Appendix A) However, students indicated that taboo and culture cues from CIRANO were rather formulaic and confusing at times, as they lacked clear labeling, explanation or guidance on how to repair the violation. They also noted that a better understanding of what triggered the taboo and cultural warnings would be helpful as would more on-screen guidance on how to repair a violation.

Non-verbal Cues

When asked whether using the tool during the interaction increased their awareness of non-verbal cues, only 37% of the students agreed with the statement, while more than half either provided a neutral rating (53%) or disagreed (11%). (Appendix A). The research assistants, students, and teachers noted that because there was a lot of reading involved while using the tool, they spent most of the time looking at the screen of the phone rather than in more natural conversation with eye contact.

CIRANO is designed to facilitate oral communication by removing the language barrier while supporting better cross-cultural awareness through written feedback on the phones. In a foreign language teaching environment, the amount of attention required for reading CIRANO prompts interferes with the authenticity of oral communication and distracts from focusing on speaking and possibly listening. This is an important observation, as it clearly indicates the tool's limitations.

Authenticity of Conversation

Teachers and students commented that the system speed was impressive and maintained a good conversational flow. While the system requires turn taking, the gaps between conversation turns were minimal, with the system able to transcribe, translate, and voice quickly.

Teachers reported that the translations produced by CIRANO were generally accurate but overly literal, especially with idiomatic expressions and colloquial language. This resulted in odd translations and misunderstandings between speakers (Appendix B).

In general, teachers commented that having students speak English allowed them to see them as their more natural selves. They noted that the low levels of speaking proficiency limit students' ability to authentically represent themselves. They often lack the ability to express emotions, humor, and use complex grammar and vocabulary to express themselves naturally in the foreign language at this early stage of their study. This aligns with the literature on multiple authentic identities that people who study and interact with other cultures develop while engaging in intercultural interactions (Byram et al., 2002).

Just over half (63%) of the students agreed with the statement that they would like to use the tool in their training, while 11% of the students disagreed with this statement (Appendix A). In their comments, these students in the minority indicated that the information received from the tool was readily available to them already from their teacher, and therefore, the tool was redundant. According to participants' comments, enabling the tool to perform in a monolingual setting would make it more relevant to their learning program.

Additionally. multiple participants and research assistants commented that receiving printed transcripts, including the transcripts of others, was extremely beneficial for their learning. It allowed them to see the different options for expressing similar ideas in Mandarin Chinese, as well as introduce them to the novel communication strategies employed by their fellow students as they negotiated a culturally sensitive situation. Reviewing other students' cultural violations and taboos was very beneficial as it facilitated interesting discussions about new concepts related to cultural norms. These results mirror the intercultural competence frameworks of Deardorff (2006) and Byram et al. (2002), specifically regarding openness, adaptability, and the ability to see things from others' perspectives.

Most of the teachers (75%) either agreed (50%) or strongly agreed (25%) that they would like to use the tool in the future to teach intercultural competence (Appendix B). The teachers mentioned a couple of teaching approaches for using this tool in the future, such as making it a game to recognize cultural violations. They acknowledged that the tool enhanced role-play, a common and widely used instructional activity, during which students and instructors play out a scenario to accomplish a specific language task. Many teachers might be uncomfortable reacting to a cultural taboo or explaining why a statement is rude, thereby raising the usefulness of using a tool such as CIRANO as an intermediary.

Lastly, CIRANO presents a novel and intriguing tool in our training environment. A 64-week intensive course in Mandarin Chinese is long and difficult. Maintaining motivation can be challenging for both students and teachers. Offering novel tools presents opportunities to exercise different modes of critical thinking, break up monotony, and maintain motivation and drive (Cabrera-Solano, 2022). However, all the participants and research assistants agreed that the tool needs further development to be useful within the DLI-W foreign language training environment. For instance, CIRANO needs to support monolingual interactions with the culture cues to promote not only the acquisition of cultural knowledge and skills, but also the development of strong communication skills in the target language.

Recommendations

The team provided several recommendations for using CIRANO and similar AI tools to build intercultural competence in foreign language training. These suggestions align with CJCSI 3126.01C (Joint Chiefs of Staff, 2023) guidance on expanding military language, regional

expertise, and culture (LREC) capabilities. Specifically, the team focused on identifying multi-language technologies that support these mission goals:

- **Cues for Both Users:** For instructional purposes, it is recommended that the users have equal access to all the prompts, violation cues, and other information to facilitate real-time corrective action and feedback and allow for more instructional flexibility.
- **Flagging a Wider Array of Cues and Emotions:** CIRANO is designed to flag a limited number of norm violations and is tuned to focus on emotions and violations that may negatively impact communication. Widening the variety of cues and including positive feedback would help students understand what works when interacting with native speakers.
- **Highlighting What Caused a Violation:** Providing any information on what triggered a cultural violation (e.g., a word, intonation, or topic) would significantly enhance acquiring the knowledge and skills associated with intercultural competence (Deardorff, 2006) and provide an understanding of why rephrasing was suggested.
- **Etiquette Rules or Guidelines for Using AI-enabled Communication Devices:** Like in the past, when people adjusted to the peculiarities of talking over the phone or through a video-conferencing platform, or when using a human interpreter, it would be valuable to develop etiquette rules for the users of devices and tools like CIRANO. These rules would address proper eye contact, how to disclose device use, where and when it is appropriate to use the device, or what to do when the device fails or makes a mistake. Use of such devices and the digital social norms associated with them are culturally laden in and of themselves (Heitmayer & Schimmelpfennig, 2023; Walton, 2024). In the DLI-W context, this would be taught as part of the initial orientation to learning tools.
- **Monolingual Input as an Option:** The goal of DLIFLC is to teach foreign languages; therefore, the support of monolingual dialogues conducted in the same language by both speakers, e.g. either all in English or all in the target language, would be most beneficial for teaching the principles of IC. Moreover, the resulting transcript would align with the widely used instructional approach of integrating a variety of modalities into a single learning activity (Dai et al., 2024; Joohoon Kang & Kim, 2024). Given the usefulness of the option of beginning students speaking English, this option should still be retained.
- **Natural Oral Speech Accuracy:** It is recommended that the tool developers continue to improve the rendering of natural oral speech that differs from written speech. For instance, it is normal and acceptable to have filler words, pauses, unfinished and "run-on" sentences (Siepmann, 2023). However, CIRANO currently adheres to written language norms, for example eliminating filler words and truncating long sentences. It also puts periods on unfinished sentences. Additionally, because CIRANO uses a generic mechanical voice for all interactions, it eliminates pauses and other aspects of spoken language that are key carriers of emotion. CIRANO needs to accommodate different voice options for its vocalizations that are more refined.

- **Explore Options for Different Delivery Platforms:** For this pilot, CIRANO was delivered as a cellphone application. However, with the amount of information it provides to students, development of an alternative application that can be integrated into a learning management system and used on an interactive whiteboard is recommended. This would allow for both individual practice with an instructor as well as more group activity work. It would reduce the need for extra equipment and fit more naturally into the typical learning environment that enables intercultural competence development (Al-Afifi et al., 2025).

CONCLUSION

CIRANO shows strong potential to support the teaching and assessment of intercultural competence, especially when used as a springboard for deeper discussion and reflection, which are foundational in acquiring intercultural competence (Byram et al., 2002; Deardorff, 2006). Key recommendations for improving CIRANO's suitability for DLI-W's training environment include adding options monolingual interactions, improving context sensitivity and voice quality, integrating CCU features into a more versatile platform, providing information about cues and rephrasing prompts, improving natural oral speech rendering, and developing etiquette guidelines for using AI-enabled communication devices. Since this CIRANO pilot, there have been major advances in AI technology and more to come that could improve many aspects of CIRANO provided there is funding to continue development.

CIRANO is designed to facilitate oral communication through written feedback on the phones. In a foreign language teaching environment, the amount of attention required for reading CIRANO prompts interferes with the authenticity of oral communication and distracts from focusing on speaking and possibly listening. However, DLI-W's instructional approach aligns with educational research on multimodal learning, which supports integrating reading, listening, and speaking into training activities to maximize training results (Dai et al., 2024; Joohoon Kang & Kim, 2024). For educators, this highlights the importance of diverse learning models. For CIRANO's developers, it points to the need for further exploration of how the tool's current features support this approach, or what additional features might be necessary to support LREC requirements outlined in CJCSI 3126.01C (Joint Chiefs of Staff, 2023).

This pilot represents a successful model of collaboration and innovation between DARPA and DLI-W, advancing DoD training capabilities while providing DARPA with vital user-centered feedback to refine its tools for both operational and instructional environments.

Acknowledgement

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under the Cross-Cultural Understanding (CCU) program. The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Authors

Tara L. Schendel, M.A., is an Associate Professor serving as Academic Advisor for Educational and Information Technology at DLI Washington. Her current academic focus is on developing new technology tools to better gather and analyze student progress reporting data.

Inna Kerlin, Ph.D., is an Associate Professor and serves as an Academic Advisor at DLI-Washington. Her professional areas of expertise encompass academic leadership, language testing and assessment, and the application of adult educational principles in foreign language education.

REFERENCES

Al-Afifi, M., Nygren, T., Rasmusson, M., & Tväråna, M. (2025). Intercultural competence in the classroom: A systematic review of classroom-based research on teaching and learning practices for developing intercultural competence in different educational contexts. *Intercultural Education, 37*(2), 217-250. https://doi.org/10.1080/14675986.2025.2555782

Byram, M., Gribkova, B., & Starkey, H. (2002). *Developing the intercultural dimension in language teaching: A practical introduction for teachers*. Council of Europe. https://rm.coe.int/16802fc1c3

Cabrera-Solano, P. (2022). Game-based learning in higher education: The pedagogical effect of Genially games in English as a foreign language instruction. *International Journal of Educational Methodology, 8*(4), 719–729. https://doi.org/10.12973/ijem.8.4.719

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026.

Dai, Z., Thipatdee, G., & Metjiranont, M. (2024). The development of a task-based Chinese speaking instructional model for Chinese as a foreign language learners in Thailand. *Journal of Education and Learning, 13*(1), 162. https://doi.org/10.5539/jel.v13n1p162

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, *10*(3), 241-266. https://doi.org/10.1177/1028315306287002

Defense Advanced Research Projects Agency. (n.d.). *Computational cultural understanding*. Retrieved April 25, 2025, from https://www.darpa.mil/research/programs/computational-cultural-understanding

Defense Advanced Research Projects Agency. (2021, June 3). *Broad agency announcement computational cultural understanding (CCU) information innovation office HR001121S0024 amendment 2*. https://www.darpa.mil/sites/default/files/HR001121S0024-Amendment02.pdf

Emelin, D., Le Bras, R., Hwang, J. D., Forbes, M., & Choi, Y. (2021). Moral stories: Situated reasoning about norms, intents, actions, and their consequences. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 698–718).

Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.55

Fraser, K. C., Ben-David, N., Hirst, G., Graham, N. L., & Rochon, E. (2015). Sentence segmentation of aphasic speech. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 862–871). Association for Computational Linguistics. https://aclanthology.org/N15-1087/

Heitmayer, M., & Schimmelpfennig, R. (2023). Netiquette as digital social norms. *International Journal of Human–Computer Interaction*, 40(13), 3334–3354. https://doi.org/10.1080/10447318.2023.2188534

Jin, Y., Sercu, L., Zhou, J., & Chen, Y. (2025). Assessing intercultural competence in the digital age: A Chinese scale development and validation. *Journal of Studies in International Education*, *0*(0). https://doi.org/10.1177/10283153251371030

Joint Chiefs of Staff. (2023, March 8). *Language, regional expertise, and culture capability identification, planning, and sourcing* (CJCSI 3126.01C). https://www.jcs.mil/Portals/36/Documents/Library/Instructions/CJCSI%20312601C.pdf

Joohoon Kang, & Kim, H. G. (2024). Multimodality in English language teaching: A review of L2 multimodal composition. *Multimedia-Assisted Language Learning,* 27(4), 160–178. https://doi.org/10.15702/mall.2024.27.4.160

Luo, J., & Chan, C. K. Y. (2022). Qualitative methods to assess intercultural competence in higher education research: A systematic review with practical implications. *Educational Research Review, 37*. 100476. https://doi.org/10.1016/j.edurev.2022.100476

Office of the Under Secretary of Defense for Personnel and Readiness. (2016, December 30). *DoD instruction 5160.70: Management of the defense language, regional expertise, and culture (LREC) program*. Retrieved April 25, 2025, from https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/516070_dodi_2016.pdf

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *R. Plutchik & H. Kellerman (Eds.), Theories of Emotion* (pp. 3–33). Academic Press. https://doi.org/10.1016/B978-0-12-558701-3.50007-7

Siepmann, D. (2023). This deserves a brief mention: A multi-corpus comparison of written and spoken academic discourse in English and French with implications for pedagogy and lexicography. *Languages in Contrast*, 23(1), 60–92. https://doi.org/10.1075/lic.21004.sie

Tsang, Y. K., Chen, X., & Jin, Z. (2025). *Parsing through boundaries in Chinese word segmentation*. arXiv. https://doi.org/10.48550/arXiv.2503.23091

Walton, D. J. (2024). Culturally responsive computing: An introduction into computer science, security, and technology. *Rotel*. Retrieved June 6, 2025, from https://rotel.pressbooks.pub/culturally-responsive-computing/

Wang, S. C.-H., Saakyan, A., Li, O., Yu, Z., & Muresan, S. (2023). Sociocultural norm similarities and differences via situational alignment and explainable textual entailment. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 3548–3564). Association for Computational Linguistics. https://aclanthology.org/2023.emnlp-main.215.pdf

APPENDIX A

CIRANO CCU Student Feedback Form

1. Please indicate your level of agreement with each statement about using the CCU CIRANO Application, choosing one of the following for each: *Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree*
    a. The application helped me communicate more effectively with my instructor.
    b. The cues provided helped me better understand cultural norms.
    c. The application increased my awareness of non-verbal communication cues.
    d. I am comfortable using AI powered tools like this application.
    e. I would like to use this application more as my training progresses.
2. Describe any specific instances where the application was particularly helpful or challenging.
3. I adjusted my approach to the conversation based on the suggestions provided by the application: *Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree*
4. What did you like most about using the application?
5. What did you like least about using this application?
6. If you have any concerns about using this application, please share them with us.

APPENDIX B

CIRANO CCU Instructor Feedback Form

1. Please indicate your level of agreement or disagreement with each statement: *Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree*.
    a. The tool's Chinese translations were accurate.
    b. The tool helped facilitate culturally appropriate communication.
    c. I would like to use this tool in the future for teaching intercultural competence.
    d. The tool was easy to use and navigate.
2. What did you like best about this application?
3. What did you like least about using this application?
4. Please share any suggestions you have for improving the tool for possible future use in the DLI-W foreign language program.

Applied Language Learning

Action Research

The Impact of Linguistic Knowledge on Reading and Listening Comprehension Across ILR Proficiency Levels

Sun-Kwang Bae
Defense Language Institute Foreign Language Center, Monterey, California

The ability to comprehend written and spoken language is the cornerstone of second language (L2) acquisition. This study investigates the extent to which three linguistic knowledge components—vocabulary, structural, and discourse knowledge—predict reading and listening comprehension in L2 Korean learners across proficiency levels. It further examines how the relative predictive power of these components differs between modalities. This study analyzed a large-scale dataset drawn from the Online Diagnostic Assessment (ODA) system, comprising over 10,000 data points across five Interagency Language Roundtable (ILR) proficiency levels (1, 1+, 2, 2+, and 3). Multiple regression analyses revealed a significant cross-modal asymmetry. In reading, a developmental shift occurred: vocabulary was the primary predictor of reading ODA scores at lower levels, but grammatical knowledge (structural and discourse) became dominant at higher levels. In listening, however, vocabulary remained the strongest score predictor across all proficiency levels. These findings support a developmental interactive model of L2 comprehension, in which learners gradually shift from heavy reliance on lexical knowledge at lower proficiency levels toward a more coordinated use of lexical, structural, and discourse knowledge at higher levels, particularly in reading. The manuscript discusses important implications for proficiency-sensitive pedagogy and diagnostic assessment in L2 Korean.

Keywords: *Action Research, Linguistic correlates, Vocabulary, Syntax, Discourse, Online Diagnostic Assessment*

INTRODUCTION

The ability to comprehend written and spoken language is the cornerstone of second language (L2) acquisition and communicative competence (Bachman, 1990; Canale & Swain, 1980; Krashen, 1985). Understanding written and spoken language is not monolithic; rather, it is constructed from the integration of multiple linguistic subsystems, including knowledge of vocabulary (lexis), grammar (syntax), and text organization (discourse) (Pae, 2018).

It is well-established within the field of second language acquisition (SLA) that all of these linguistic components are essential for successful comprehension (Jeon & Yamashita, 2014; Joyce, 2019; Mlakar, 2022; Susoy & Tanyer, 2019; Vafaee & Suzuki, 2020). Rather than debating which component is universally "most important," contemporary SLA research seeks to understand how the contributions of these components interact and shift dynamically as learners develop (Sarbazi et al., 2021). This study is built upon the understanding that the architecture of comprehension is not static but dynamic, and that the relative contributions of its constituent linguistic components shift systematically as a function of two critical moderating variables: the modality of input (reading versus listening) and the learner's proficiency level. Such a perspective reframes the central question from a static "what is more important?" to a more nuanced, developmental "when, why, and under what conditions is a particular component more important?"

This study focused on Korean as the target language. For many learners, especially those with an L1 such as English, Korean presents a distinct set of linguistic challenges. Its agglutinative morphology, the use of case-marking particles to encode grammatical relations, and a relatively flexible word order create a processing environment where both lexical and syntactic analyses can be demanding. For instance, L2 learners of Korean frequently misinterpret case-marking particles during reading and listening tasks, which can significantly impede accurate syntactic parsing (Chun & Kim, 2021). These differences make Korean an ideal test case for examining how learners build comprehension when faced with a linguistic system that may differ substantially from their L1. The specific difficulties Korean L2 learners encounter in both reading and listening underscore the need for a fine-grained analysis of the underlying linguistic skills that support comprehension in this context.

LITERATURE REVIEW

To situate the present study, this section reviews theoretical models and empirical evidence that inform our understanding of how L2 learners comprehend written and spoken language. It briefly outlines interactive psycholinguistic frameworks for reading and listening, and concludes by examining the established roles of the core linguistic components under investigation.

Interactive Processing and Psycholinguistic Constraints

Early conceptualizations of comprehension were often framed as unidirectional processes. Bottom-up models describe comprehension as a linear, data-driven process that begins with the smallest units of input—letters on a page or phonemes in an auditory stream—and sequentially builds them into words, phrases, sentences, and finally, a representation of the text's meaning. This perspective emphasizes the centrality of decoding and linguistic analysis. In contrast, top-down models, often associated with psycholinguistic approaches, posit that comprehension is a concept-driven process where the comprehender's prior knowledge, expectations, and schemata guide the interpretation of the input. The reader or listener forms hypotheses about the meaning and uses the incoming linguistic data to confirm or disconfirm them.

Skilled comprehension is currently understood through the lens of interactive models, which posit that comprehension involves a complex, simultaneous interplay between bottom-up (data-driven decoding) and top-down (concept-driven) processes (Eskey & Grabe, 1988; Wang, 2023). For example, strong top-down processing can sometimes assist a learner when bottom-up processing is strained, though relying strictly on context to guess unknown words has recognized limitations (Grabe & Yamashita, 2022). This interactive framework is fundamental to the present study, as it provides the theoretical basis for investigating how multiple knowledge sources—lexical, syntactic, and discourse—contribute concurrently to the construction of meaning.

Psycholinguistic and Cognitive Frameworks

While reading and listening both draw on a common pool of linguistic knowledge, listening comprehension operates under a unique and demanding set of psycholinguistic constraints (Hogan et al., 2014). The auditory signal is ephemeral; it unfolds in real-time and cannot be reviewed once it has passed (Joyce, 2019). This *now or never* quality places an immense cognitive load on the listener, requiring rapid and efficient online processing (Fang, 2011). Cognitive models of listening, such as Anderson's (1995) three-stage model of perception, parsing, and utilization, underscore the critical importance of foundational, bottom-up processes. A failure at the early stages of perception or parsing can cause a catastrophic breakdown in comprehension, as the listener cannot simply go back and listen again in the way a reader can go back and re-read a sentence. This fundamental asymmetry between the modalities suggests that the linguistic skills supporting rapid and accurate bottom-up processing, particularly vocabulary recognition, may play a disproportionately critical role in listening.

The Role of Core Linguistic Components in L2 Comprehension

This study focuses on three core components: vocabulary, syntax, and discourse.

Vocabulary Knowledge

An extensive body of research has firmly established vocabulary knowledge as a cornerstone of L2 proficiency and a powerful predictor of comprehension in both reading and listening (Zhang, 2025). Studies consistently demonstrate a strong positive correlation between vocabulary size and comprehension ability, with some research suggesting that vocabulary knowledge can account for 40-50% of the variance in reading and listening performance (Alderson, 2005; Qian, 2002; Stæhr, 2009). While vital for both modalities, vocabulary's role is often found to be especially pronounced in listening comprehension (Stæhr, 2009).

Syntactic Knowledge

The specific weight of syntactic knowledge, or grammar, in L2 comprehension varies depending on factors such as text complexity and learner proficiency. Studies involving adult L2 learners indicate that syntactic knowledge becomes increasingly vital for reading comprehension as texts become more structurally complex (Morvay, 2012; Susoy & Tanyer, 2019). In L2 listening, research indicates that while both vocabulary and syntactic knowledge are significant predictors, vocabulary often emerges as the stronger of the two overall (Vafaee & Suzuki, 2020).

Discourse Knowledge

Discourse knowledge represents a higher level of linguistic competence, involving the understanding of how sentences are connected to form a coherent and meaningful whole (Grabe & Yamashita, 2022). While foundational vocabulary and syntactic skills are necessary for processing individual sentences, discourse knowledge is what enables the comprehender to move beyond the sentence level to grasp the overarching message (Grabe & Yamashita, 2022). As such, the ability to effectively utilize discourse knowledge is often associated with more advanced stages of language proficiency.

Research Questions

Based on the existing literature concerning L2 processing constraints and developmental trajectories, it is hypothesized that vocabulary knowledge will serve as a stronger predictor for listening comprehension than for reading comprehension, particularly at lower proficiency levels where foundational decoding is paramount. It is further hypothesized that the predictive power of structural and discourse knowledge will increase commensurately with proficiency in both modalities, as learners' processing becomes more automatized, freeing cognitive resources for higher-order analysis.

This study leverages a large-scale dataset to systematically investigate these dynamics. The primary research questions guiding this inquiry are as follows:

1. To what extent do vocabulary, structural, and discourse knowledge uniquely predict reading comprehension in L2 Korean learners at Interagency Language Roundtable (ILR) proficiency levels 1, 1+, 2, 2+, and 3?
2. To what extent do vocabulary, structural, and discourse knowledge uniquely predict listening comprehension in the same population and at the same proficiency levels?
3. How does the relative predictive power of these linguistic components differ between reading and listening comprehension across the proficiency spectrum?

METHODOLOGY

Data Corpus

The study utilizes a large, anonymized dataset sourced from the Online Diagnostic Assessment (ODA) system, a platform designed for learners of foreign languages at a major US government language institute. The dataset consists of Korean learner data collected between December 2008 and February 2023. The final corpus for analysis was compiled after standard data cleaning procedures, resulting in 5,465 valid data points for reading comprehension and 5,318 valid data points for listening comprehension.

The data were drawn from a database of assessment sessions collected from Korean language learners from diverse backgrounds. For the purposes of this study, learners in the database were categorized into five distinct proficiency levels of their Target Level Performance based on the ILR scale (ILR, n.d.). Target Level Performance refers to the proficiency level the ODA is intended to assess—the level at which a user does not yet demonstrate sufficient mastery (i.e., the failed level). This is in contrast to the Current Level Performance, which reflects the learner's actual performance, the level successfully attained and passed. The Target Level Performance was chosen to use for analysis because it is supported by a substantially larger amount of response data—twice as much as that available for the Current Level Performance—resulting in more stable estimates and greater reliability for analytical purposes. The five levels analyzed were: ILR 1 (Elementary), ILR 1+ (Elementary Plus), ILR 2 (Limited Working), ILR 2+ (Limited Working Plus), and ILR 3 (Professional Working). The distribution of learner records across these proficiency levels is detailed in the Descriptive Statistics section below (Table 1).

Instrumentation and Procedures

The primary instrument for data collection was the ODA (DLIFLC, n.d.), a web-based, computer-adaptive style assessment designed for proficiency maintenance and enhancement. It utilizes authentic source materials (both written texts and audio recordings) and a series of interactive activities to assess comprehension. The dataset includes records of learners who accessed the ODA platform and completed diagnostic assessments aligned with their designated target proficiency level. The assessment generated scores for discrete linguistic competencies, which served as the independent variables in this study:

- **Vocabulary Knowledge:** Measured by performance on tasks targeting the recognition and comprehension of lexical items within the context of the authentic source material. The assessment utilized constructed-response test (CRT) prompts (i.e., student generated).
- **Structural Knowledge**: Measured by performance on tasks in ODA that require learners to process and understand grammatical relationships, word order, and other syntactic features. The assessment incorporated multiple formats, including CRT, multiple-choice, and matching questions.
- **Discourse Knowledge**: Measured by performance on tasks assessing the ability to understand the relationships between sentences and identify the overall organizational structure. The assessment incorporated multiple formats, including CRT, multiple-choice, and matching questions.

The dependent variables were the L2 comprehension scores, operationalized as the percentage of correctly answered questions following each reading or listening passage. These questions consisted exclusively of CRT items. To allow for a more nuanced analysis, comprehension questions were categorized into two types:

- **Main Idea/Topic Comprehension**: Identifying the gist or primary message.
- **Supporting Idea/Detail Comprehension**: Identifying specific information or examples.

DATA ANALYSIS

To investigate the unique contribution of each linguistic knowledge component to reading and listening comprehension across the five proficiency levels, a series of multiple linear regression analyses was conducted. A total of 10 primary regression models were calculated: one for each of the five ILR proficiency levels (1, 1+, 2, 2+, 3) for reading comprehension, and one for each of the five levels for listening comprehension.

Prior to analysis, standard assumptions for multiple regression were evaluated. Linearity and homoscedasticity were checked via residual plots. Multicollinearity among the linguistic predictors (Vocabulary, Structure, Discourse) was assessed using Variance Inflation Factors (VIF); while correlations existed, VIF values remained within acceptable limits (typically < 5), allowing for the interpretation of unique contributions. To control for Type I errors (i.e., false positives) across the multiple models, the alpha level was maintained at $p < .05$, but exact p-values are reported to provide transparency regarding the strength of evidence.

For each model, the standardized beta coefficient is reported to compare the relative predictive strength of the independent variables. The coefficient of determination (R-squared) is reported to indicate the proportion of total variance explained.

RESULTS

Descriptive Statistics

Table 1 provides the descriptive statistics for all predictor and outcome variables, disaggregated by modality and ILR proficiency level. An analysis of the mean scores reveals a non-linear trend as proficiency increases, with scores often peaking at ILR Level 1+ before decreasing, reflecting the increasing difficulty of the authentic materials at higher levels. Notably, data for the Discourse Score was not available for ILR Level 1 in either modality due to the design of the assessment instrument.

Table 1

Descriptive Statistics for All Variables by Proficiency Level

		Reading			Listening		
Target ILR Level	**Variable Scores**	***N***	**Mean**	**SD**	***N***	**Mean**	**SD**
1	Main Proposition	613	0.279	0.223	384	0.355	0.229
	Supporting Proposition	613	0.283	0.207	384	0.257	0.201
	Vocabulary	613	0.505	0.276	384	0.462	0.257
	Structural	613	0.586	0.258	384	0.521	0.260
1+	Main Proposition	779	0.445	0.210	936	0.469	0.206
	Supporting Proposition	779	0.389	0.193	936	0.366	0.195
	Vocabulary	779	0.633	0.199	936	0.581	0.195
	Structural	779	0.609	0.240	936	0.640	0.225
	Discourse	779	0.637	0.241	936	0.580	0.244
2	Main Proposition	1626	0.432	0.217	2283	0.399	0.221
	Supporting Proposition	1626	0.380	0.164	2283	0.350	0.162
	Vocabulary	1626	0.524	0.215	2283	0.484	0.190
	Structural	1626	0.557	0.242	2283	0.538	0.228
	Discourse	1626	0.514	0.244	2283	0.489	0.226
2+	Main Proposition	1854	0.359	0.219	1363	0.338	0.210
	Supporting Proposition	1854	0.279	0.173	1363	0.259	0.158
	Vocabulary	1854	0.422	0.208	1363	0.283	0.168
	Structural	1854	0.399	0.233	1363	0.503	0.239
	Discourse	1854	0.408	0.234	1363	0.439	0.228
3	Main Proposition	593	0.294	0.218	352	0.347	0.235
	Supporting Proposition	593	0.262	0.178	352	0.241	0.170
	Vocabulary	593	0.376	0.235	352	0.336	0.204
	Structural	593	0.349	0.245	352	0.514	0.251
	Discourse	593	0.349	0.245	352	0.488	0.226

Note: Discourse Score was not available per assessment design for ILR Level 1.

RQ1: Predictors of Reading Comprehension

Multiple regression analyses were conducted to assess the predictive power of the various linguistic knowledge components on reading comprehension (Table 2).

Main Propositions (Gist)

At the lower proficiency levels (ILR 1 and 1+), vocabulary knowledge was the dominant predictor (beta = .475 and .464, respectively, $p < .001$). A clear shift begins at ILR Level 2, where all three components were highly significant. By ILR Level 3, discourse knowledge emerged as the strongest predictor (beta = .280), followed by structural knowledge (beta = .196), while vocabulary's influence diminished (beta = .169).

Supporting Propositions (Details)

A similar, yet more pronounced developmental trend was observed for detail comprehension. At ILR 1, vocabulary was the strongest predictor. However, by ILR Level 2, the combined predictive power of structural and discourse knowledge surpassed that of vocabulary.

At ILR 3, the shift was most dramatic: Discourse (beta = .313) and structural (beta = .294) knowledge were the strongest predictors, with the combined influence (beta = .607) substantially outweighing vocabulary (beta = .152).

Table 2

Multiple Regression Models Predicting Reading Comprehension Across ILR Levels

		Main Proposition as Outcome Variable				Supporting Proposition as Outcome Variable			
Target ILR Level	**Predictor**	**B**	**SE**	**beta**	***p***	**B**	**SE**	**beta**	***p***
1	**(intercept)**	0.036	0.020		0.070	0.049	0.018		0.008
	Vocabulary	0.384	0.033	0.475	<.001	0.324	0.031	0.432	<.001
	Structural	0.086	0.036	0.099	0.016	0.120	0.033	0.150	<.001
		R^2 = 0.287, F(2, 612) = 123, p <.001				R^2 = 0.281, F(2, 612) = 119, p < .001			
1+	**(intercept)**	0.080	0.024		<.001	0.103	0.023		
	Vocabulary	0.488	0.040	0.464	<.001	0.302	0.039	0.312	<.001
	Structural	0.058	0.031	0.067	0.059	0.099	0.030	0.123	0.001
	Discourse	0.033	0.032	0.037	0.312	0.054	0.032	0.067	0.088
		R^2 = 0.269, F(3, 776) = 95.2, p <.001				R^2 = 0.180, F(3, 776) = 56.7, p < .001			
2	**(intercept)**	0.111	0.015		<.001	0.091	0.010		<.001
	Vocabulary	0.388	0.024	0.385	<.001	0.258	0.017	0.339	<.001
	Structural	0.077	0.021	0.086	<.001	0.136	0.015	0.200	<.001
	Discourse	0.145	0.021	0.163	<.001	0.153	0.015	0.226	<.001
		R^2 = 0.268, F(3, 1623) = 198, p <.001				R^2 = 0.352, F(3, 1623) = 293, p < .001			
2+	**(intercept)**	0.107	0.012		<.001	0.030	0.008		<.001
	Vocabulary	0.304	0.024	0.289	<.001	0.324	0.017	0.391	<.001
	Structural	0.165	0.022	0.176	<.001	0.143	0.015	0.194	<.001
	Discourse	0.141	0.022	0.150	<.001	0.135	0.016	0.183	<.001
		R^2 = 0.230, F(3, 1851) = 184, p <.001				R^2 = 0.366, F(3, 1851) = 355, p < .001			
3	**(intercept)**	0.081	0.016		<.001	0.057	0.012		<.001
	Vocabulary	0.157	0.040	0.169	<.001	0.115	0.030	0.152	<.001
	Structural	0.170	0.037	0.196	<.001	0.209	0.028	0.294	<.001
	Discourse	0.249	0.039	0.280	<.001	0.228	0.029	0.313	<.001
		R^2 = 0.283, F(3, 590)= 77.6, p <.001				R^2 = 0.394, F(3, 590) = 128, p < .001			

RQ2: Predictors of Listening Comprehension

Parallel regression analyses for listening comprehension revealed a markedly different pattern (Table 3).

Main Propositions (Gist)

Vocabulary knowledge remained the dominant predictor across *all* five proficiency levels, with consistently high predictive strength (e.g., ILR Level 3 beta = .493, p < .001). In contrast, structural knowledge was a small predictor at lower levels and became non-significant at ILR Level 3 (p = .076). Discourse knowledge was largely a weak or non-significant predictor for listening gist.

Supporting Propositions (Details)

Vocabulary knowledge was also the strongest predictor for detail comprehension at every level (e.g., ILR Level 3 beta = .406). However, unlike in the main proposition models, both structural and discourse knowledge were consistently significant predictors. Discourse knowledge peaked at ILR Level 3 (beta = .226), yet even at this peak, it remained secondary to vocabulary.

Table 3

Multiple Regression Models Predicting Listening Comprehension Across ILR Levels

		Main Proposition as Outcome Variable				Supporting Proposition as Outcome Variable			
Target ILR Level	**Predictor**	**B**	**SE**	**beta**	***p***	**B**	**SE**	**beta**	**p**
1	**(intercept)**	0.094	0.023		<.001	0.003	0.020		0.892
	Vocabulary	0.404	0.046	0.453	<.001	0.307	0.039	0.392	<.001
	Structural	0.143	0.045	0.163	0.002	0.215	0.039	0.278	<.001
		R^2 = 0.315, F(2, 382) = 87.8, p < .001				R^2 = 0.355, F(2, 382) = 105, p < .001			
1+	**(intercept)**	0.092	0.021		<.001	0.032	0.020		
	Vocabulary	0.504	0.039	0.476	<.001	0.302	0.035	0.302	<.001
	Structural	0.107	0.028	0.116	<.001	0.142	0.027	0.143	<.001
	Discourse	0.027	0.028	0.032	0.338	0.135	0.028	0.169	<.001
		R^2 = 0.305, F(3, 933) = 136, p < .001				R^2 = 0.249, F(3, 933) = 103, p < .001			
2	**(intercept)**	0.059	0.013		<.001	0.057	0.009		<.001
	Vocabulary	0.576	0.024	0.494	<.001	0.363	0.017	0.426	<.001
	Structural	0.077	0.019	0.079	<.001	0.120	0.013	0.169	<.001
	Discourse	0.041	0.019	0.042	0.030	0.108	0.013	0.151	<.001
		R^2 = 0.305, F(3, 2283) = 333, p < .001				R^2 = 0.363, F(3, 2283) = 433, p < .001			
2+	**(intercept)**	0.127	0.013		<.001	0.055	0.009		<.001
	Vocabulary	0.476	0.035	0.379	<.001	0.423	0.024	0.448	<.001
	Structural	0.061	0.024	0.069	<.001	0.083	0.002	0.125	<.001
	Discourse	0.105	0.025	0.114	<.001	0.098	0.017	0.141	<.001
		R^2 = 0.228, F(3, 1360) = 134, p < .001				R^2 = 0.354, F(3, 1360) = 248, p < .001			
3	**(intercept)**	0.090	0.028		0.001	0.008	0.019		0.670
	Vocabulary	0.568	0.066	0.493	<.001	0.339	0.045	0.406	<.001
	Structural	0.086	0.049	0.092	0.076	0.069	0.033	0.102	0.038
	Discourse	0.043	0.058	0.041	0.457	0.170	0.040	0.226	<.001
		R^2 = 0.325, F(3, 349) = 55.8, p < .001				R^2 = 0.394, F(3, 349) = 75.3, p < .001			

RQ3: Cross-Modal Differences

Synthesizing the results reveals a stark cross-modal asymmetry. In Reading, a clear shift occurs where vocabulary dominance at low levels is replaced by grammatical knowledge (structural and

discourse) at high levels. In Listening, this shift does not occur; vocabulary remains the primary predictor of comprehension for both gist and details from ILR 1 through ILR 3.

DISCUSSION

This study investigated the dynamic contributions of various linguistic knowledge components to L2 Korean comprehension. The findings reveal a clear developmental trajectory and a significant cross-modal asymmetry.

The Primacy of Vocabulary in Listening

The most robust finding is the persistently dominant role of vocabulary knowledge in listening comprehension across all proficiency levels. This aligns with psycholinguistic models characterizing listening as a now-or-never process (Joyce, 2019). The listener must perceive and parse input in real-time; in this high-pressure environment, rapid lexical access is the primary gatekeeper to meaning (Fang, 2011). If a key word is not recognized instantly, the listener cannot look back to use context as a reader might. Consequently, vocabulary knowledge explains the greatest amount of variance among the predictors examined, suggesting that a strong lexicon is the non-negotiable foundation of auditory comprehension.

Developmental Shifts and Joint Contributions

The results for reading comprehension support a developmental interactive model. At lower levels (ILR 1, 1+), comprehension is constrained by bottom-up processing (decoding). As learners progress to intermediate and advanced levels (ILR 2+), the predictive power of structural and discourse knowledge increases significantly. This suggests that as basic lexical decoding becomes automatized, cognitive resources are reallocated to higher-order analysis (Verhoeven et al., 2011).

It is important to note a limitation regarding ILR Level 1 data. The absence of discourse scores at this level means that the dominance of vocabulary at ILR 1 must be interpreted with caution, as the potential contribution of discourse knowledge was not mathematically controlled. However, the trend remains consistent at ILR 1+, where discourse was measured but remained a weaker predictor than vocabulary.

The data also highlight a distinction between comprehending the *gist* versus *supporting details*. For listening, while vocabulary drives gist comprehension almost exclusively, the comprehension of details requires a significant joint contribution from structural and discourse knowledge (peaking at ILR Level 3). This indicates that while learners may grasp the general topic of a spoken text via keywords, a precise understanding of specific details requires the ability to parse syntax and follow discourse markers.

Pedagogical Implications

The results advocate for a differentiated pedagogical approach. For lower-proficiency learners, the overwhelming importance of vocabulary suggests that instruction should prioritize building a robust, high-frequency lexical foundation. For listening, this includes connecting spoken forms to meanings through contextualized vocabulary learning, rather than isolated memorization. While grammar is essential, instruction prioritizing complex grammar rules at the expense of lexical accumulation may be less effective at these early stages.

For higher-proficiency learners (ILR Level 2+ and 3), instruction should shift to encompass complex linguistic features. The growing importance of structural and discourse knowledge in the regression models suggests these learners need to engage with the intricacies of grammar—such as cohesive devices and rhetorical structures—to facilitate advanced comprehension (Grabe & Yamashita, 2022).

Limitations and Future Directions

Several limitations should be acknowledged. First, as a cross-sectional correlational study, these results identify predictive relationships but cannot establish causality or confirm longitudinal developmental trajectories within individuals. Future longitudinal research is needed to track how these component contributions shift over time for the same learners. Second, the absence of discourse measures at ILR Level 1 limits the conclusions drawn for the lowest proficiency level. Finally, this study's focus on L2 Korean—a language with distinct agglutinative morphology—may influence the results; cross-linguistic comparisons would help determine if these patterns are universal.

CONCLUSION

This study demonstrates that the architecture of L2 comprehension is not fixed but varies significantly by modality and proficiency. While reading comprehension evolves from a lexically driven process to one heavily reliant on grammatical knowledge, listening comprehension remains consistently anchored in vocabulary knowledge. These findings underscore the need for proficiency-sensitive and modality-specific approaches in both L2 assessment and instruction.

Author

Sun-Kwang Bae (Ph.D., The Australian National University) is an Associate Professor at the Defense Language Institute Foreign Language Center. He serves as the Program Manager for the Online Diagnostic Assessment and the Global Language Online Support System projects within the Technology Integration Division of Educational Technology and Development Directorate.

REFERENCES

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.

Anderson, J. R. (1995). *Learning and memory: An integrated approach*. Wiley.

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

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1–47.

Chun, J., & Kim, M. H. (2021). Corpus-informed application based on Korean learners' corpus: Substitution errors of topic and nominative markers. *Asian-Pacific Journal of Second and Foreign Language Education, 6*(13). https://doi.org/10.1186/s40862-021-00112-7

Eskey, D. E., & Grabe, W. (1988). Interactive models for second language reading: perspectives on instruction. In P. L. Carrell, J. Devine, & D. E. Eskey (Eds.*), Interactive approaches to second language reading* (pp. 223–238). Cambridge University Press.

Fang, X. (2011). The priority of listening comprehension over speaking in the language acquisition process. *International Education Studies, 4*(1), 161–167. Doi: 10.5539/ies.v4n1p161

Grabe, W., & Yamashita, J. (2022). Building awareness of discourse structure. In *Reading in a second language: Moving from theory to practice* (pp. 338–368). Cambridge University Press.

Hogan, T., Adlof, S., & Alonzo, C. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology, 16*(3), 199–207. https://doi.org/10.3109/17549507.2014.904441

Interagency Language Roundtable (ILR). (n.d.). *Interagency Language Roundtable (ILR)*. https://www.govtilr.org/

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis*. Language Learning, 64*(1), 160–212. https://doi.org/10.1111/lang.12034

Joyce, P. (2019). The relationship between L2 listening proficiency and L2 aural language processing. *PASAA, 57,* 9–32. https://digital.car.chula.ac.th/pasaa/vol57/iss1/2/

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.

Morvay, G. (2012). The relationship between syntactic knowledge and reading comprehension in EFL learners. *Studies in Second Language Learning and Teaching, 2*(3), 415–432. https://doi.org/10.14746/ssllt.2012.2.3.8

Mlakar, H. (2022). Linguistic factors influencing L2 reading comprehension. In *Literacy without borders: Engaging literacy learners in diverse settings* (pp. 166–185). Eden Print Ltd.

Pae, T.-I. (2018). Effects of task type and L2 proficiency on the relationship between L1 and L2 in reading and writing: An SEM approach. *Studies in Second Language Acquisition, 40*(1), 63–90. https://doi.org/10.1017/S0272263116000462

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513–536. https://doi.org/10.1111/1467-9922.00193

Sarbazi, M., Khany, R., & Shoja, L. (2021). The predictive power of vocabulary, syntax and metacognitive strategies for L2 reading comprehension. *Southern African Linguistics and*

Applied Language Studies, *39*(3), 244–258. https://doi.org/10.2989/16073614.2021.1939076

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577–607. https://doi.org/10.1017/S0272263109990039

Susoy, Z., & Tanyer, S. (2019). The role of vocabulary vs. syntactic knowledge in L2 reading comprehension. *Eurasian Journal of Applied Linguistics, 5*(1), 113–130. https://doi.org/10.32601/ejal.543787

Vafaee, P., & Suzuki, Y. (2020). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition, 42*(2), 383–410. https://doi.org/10.1017/S0272263119000676

Verhoeven, L., Reitsma, P., & Siegel, L. S. (2011). Cognitive and linguistic factors in reading acquisition. *Reading and Writing, 24*(4), 387–394. https://doi.org/10.1007/s11145-010-9232-4

Wang, J. (2023). Interactive approaches in ESL teaching and its application. *Sino-US English Teaching, 20*(3), 85–89. Doi: 10.17265/1539-8072/2023.03.001

Zhang, X. (2025). Vocabulary knowledge towards L2 reading and listening performance. *Reading in a Foreign Language, 37*(2), 109–126. https://doi.org/10.64152/10125/67500

Applied Language Learning

Action Research

Integrating a Streaming Series into the Curriculum to Foster Listening Proficiency and Cultural Competence

Jiayi Xu
Defense Language Institute Foreign Language Center, Hawaii Language Training Detachment

Maria Wang
Defense Language Institute Foreign Language Center, Hawaii Language Training Detachment

Nina Bozeman
Defense Language Institute Foreign Language Center, Hawaii Language Training Detachment

This classroom-based action research project investigates the use of a Chinese streaming series to foster listening proficiency and cultural competence in a 19-week Intermediate Chinese Program. The authors integrated narrative media (e.g., films or television dramas) within the course's Language, Regional Expertise, and Culture (LREC) curriculum. For 12 weeks, eight students participated in weekly episode viewings with Chinese subtitles, followed by guided analysis of excerpts from the novel the series was based on and instructor-facilitated discussion. The module included culture-based discussions and culminated in a 600-word student reflection. Data included Defense Language Proficiency Test (DLPT5) scores, student reflections, and post-project student feedback. Seven of eight students achieved Interagency Language Roundtable (ILR) Level 2+ in listening. Students perceived the integration positively, reporting increased confidence in processing fast-paced spoken Chinese and enhanced cultural competence through intercultural comparison and reflective engagement with sociocultural themes. Findings indicate that this structured integration of narrative media can support listening proficiency development, cultural competence, and positive learner perceptions in intensive, military-oriented language education.

Keywords: *Action Research, Chinese Language, Cultural Competence, DLPT5, Listening Proficiency, LREC, Media-Based Instruction, Narrative Media*

INTRODUCTION

Media-based instruction has been widely recognized for its role in improving vocabulary acquisition (Chen, 2012; Wang, Y.C., 2012), listening proficiency (Fujita, 2019; Wang, D., 2012), and cultural competence in second language learning (Fujita, 2019). Cultural competence is a multidimensional construct that goes beyond factual knowledge of a target culture (Byram et al., 2002; Deardorff, 2006). Byram et al. (2002) describe it as comprising five interrelated capacities: attitudes of curiosity and openness toward otherness; knowledge of social groups and intercultural processes; skills of interpreting and relating across cultural contexts; skills of discovery and interaction in real-time communication; and critical cultural awareness, the ability to evaluate practices and perspectives in both one's own and other cultures. Rather than positioning the learner as a native-speaker imitator, this framework aims to develop an intercultural mediator capable of engaging with complexity, multiple identities, and unfamiliar perspectives.

Building on this foundation, Deardorff (2006) empirically validated a process model in which attitude, specifically respect, openness, and curiosity, serves as the essential starting point, leading toward an internal shift in frame of reference and ultimately an external outcome of behaving and communicating effectively and appropriately in intercultural situations. Deardorff further emphasizes that cultural competence exists on a continuum and is best assessed through multiple qualitative methods rather than any single instrument.

In this study, cultural competence is defined as the ability to interpret cultural practices, compare perspectives across sociocultural contexts, and reflect on one's own cultural assumptions (Byram et al., 2002; Deardorff, 2006). Grounded in the interpretive and critical dimensions of Byram et al.'s framework along with Deardorff's emphasis on perspective-taking and self-reflective awareness, this paper reports on a 12-week action research project that incorporated *The Long Night*, a Chinese-language streaming series (based on a novel) about the justice system in China, into an Intermediate Chinese Program.

BACKGROUND

The Efficacy of Media-Based Instruction and Activities

The effectiveness of media-based instruction in second language acquisition (SLA) has been widely explored, particularly in vocabulary acquisition and listening comprehension. Chen (2012) conducted a corpus-based lexical richness study analyzing the sitcom *Friends*, concluding that the show's scripts offer "a high possibility to be an appropriate resource for English as an additional language (EAL) learners' lexicon instruction" due to their alignment with high-frequency words in English corpora. This aligns with Y. C. Wang (2012), who examined American TV dramas' impact on L2 vocabulary learning and emphasized that learners benefit from "authentic dialogues and contextualized expressions" that reinforce vocabulary retention. These findings suggest that narrative media (e.g., films or television dramas) may support vocabulary development by

providing contextualized input. However, much of the existing research has focused primarily on vocabulary outcomes rather than broader discourse-level analysis (Chen, 2012; Y. C. Wang, 2012). Fewer studies have provided detailed descriptions of how narrative media can be systematically integrated into classroom instruction.

Listening comprehension also improves through exposure to TV dramas. While some learners struggle with rapid speech and phonological reductions, research suggests that structured scaffolding can enhance decoding ability and retention. Fujita (2019) found that students identified fast speech as the primary obstacle to comprehension, with higher-proficiency learners demonstrating greater awareness of phonological features such as linking and reduced forms. The study further suggested that structured scaffolding and guided analysis help learners decode these features more effectively. These findings support the use of layered instructional tasks, such as guided analysis and structured follow-up activities, in classroom media-based instruction.

Media-Based Instruction and Cultural Competence

In addition to its efficacy, media-based instruction plays a vital role in authentic linguistic and cultural competence. In the context of China, where many learners lack immersive English-speaking environments, D. Wang (2012) emphasized that watching TV dramas enables students to "gain exposure to natural speech patterns and pragmatic competence that traditional textbooks fail to provide" (p. 2). This perspective highlights the role of narrative media as a form that allows learners to access authentic linguistic input beyond formal classroom settings. Media also supports cultural engagement and sociocultural awareness. Fujita (2019) reported that over 70% of students expressed greater interest in other cultures after watching dramas that "depicted daily life overseas" (p. 26). In addition, Richardson (2010) argues that television drama operates as a complex semiotic system in which meaning is constructed through multiple simultaneous channels, making it a uniquely rich resource for language learners. The research indicates that narrative media offers an integrated context for developing both linguistic and cultural competence by combining authentic linguistic input, cultural representation, and multimodal meaning-making.

Research Questions

These studies shaped the planning and implementation of the present classroom-based project. Fujita's (2019) findings regarding rapid speech and phonological complexity informed the decision to embed guided analysis and the combination of the streaming series and corresponding excerpts from the novel into weekly lessons. While prior research on media-based language learning has primarily examined vocabulary acquisition and learner motivation (Chen, 2012; Y. C. Wang, 2012; Fujita, 2019), fewer studies have provided detailed pedagogical models demonstrating how narrative media can be systematically integrated into classroom instruction through structured tasks and instructor-facilitated activities. The present project addresses this gap by implementing an instructional sequence centered on a Chinese streaming series, incorporating guided analysis of excerpts from the novel, and instructor-facilitated discussions to

support both listening proficiency and cultural competence, and examines how students perceive this integration.

Informed by this literature and grounded in the needs of Chinese learners of Interagency Language Roundtable scale (ILR) Level 2 (Intermediate High proficiency), the study was guided by the following research questions:

1. What are students' perceptions of integrating narrative media into a Chinese language course, in terms of development of their listening proficiency and cultural competence?
2. In what ways might including narrative media in a language course impact students' listening proficiency development?

ACTIONS TAKEN

Student Profile

The study was conducted in a 19-week Chinese BA Degree Program with eight students enrolled at the Defense Language Institute Foreign Language Center (DLIFLC). All participants were preparing to take the Defense Language Proficiency Test (DLPT5) at the conclusion of the course, with an institutional target of achieving at least ILR Level 2+ (described as Advanced Plus) in listening and reading. While their professional background is omitted here for anonymity, it is worth noting that they shared similar career goals requiring strong linguistic proficiency to perform their military duties. Their prior DLPT5 score histories (see Appendix A) indicated variability in listening and reading performance across testing cycles.

At the beginning of the course, all eight students had achieved at least ILR Level 2 in listening and reading, with entry scores ranging from 2 to 2+ in listening and from 2 to 3 in reading (see Appendix A). Speaking proficiency, as measured by Oral Proficiency Interview (OPI) ratings, showed greater variability, with scores ranging from 1+ to 2. For example, one student entered with a 2+/3/2 profile (listening/reading/speaking), while another entered with a 2/2/1+ profile, reflecting stronger receptive skills relative to productive ability. This pattern of initial inconsistency followed by improvement was seen across several learners and strongly informed the design of the instructional intervention. The aim was to consolidate recent gains, address lingering weak areas (especially listening and oral fluency), and maintain forward momentum.

Although students reported that they had engaged with authentic materials in earlier classes (e.g., news articles, podcasts), none had previously participated in a long-term, curriculum-integrated media-based project. Their experience with audiovisual materials had been limited to isolated clips or test prep items. As such, this project served not only to develop linguistic and cultural skills, but also to introduce learners to a structured mode of learning, which required clear scaffolding, pacing, and reflection tools to help learners transition from familiar formats.

Instructional Design and Teaching Steps

Implementing a 12-week media-based project within an intensive, schedule-driven language curriculum required both administrative support and instructional flexibility. The program in which the project was conducted follows a tightly structured weekly plan, with nearly every instructional hour accounted for in the standardized syllabus. With approval from program leadership, time was made available each week to run this project without interfering with the standardized curriculum content or required objectives.

The instructional content was centered around the Chinese streaming series *The Long Night* (《沉默的真相》). The series, which was based on a novel, provided rich cultural content on the legal system in China. In addition to the series, instructional activities included excerpts from the original novel. Students did not read the full novel independently; instead, short instructor-selected excerpts were introduced during guided review sessions in class to support analysis of key scenes. The careful selection and pairing of episodes and excerpts from the novel were undertaken by the instructional team, whose deep cultural and literary insight ensured that the materials were not only thematically appropriate but also pedagogically rich. Weekly implementation included guided classroom discussions, vocabulary review, and concept mapping activities designed to support comprehension and critical reflection.

Each week, students viewed one full episode of *The Long Night*, which aligned with a 50-minute instructional period. Due to the length of the episodes and time constraints, there was no pausing or rewatching, and the show was screened with Chinese subtitles only, without English glossing or scene breakdowns. While some pedagogical approaches advocate limiting subtitle use to promote inferencing strategies, the use of Chinese subtitles throughout viewing was intentional to provide lexical support appropriate for ILR Level 2 learners while maintaining exposure to authentic spoken language. This format challenged students to follow naturalistic, fast-paced dialogue in real time, replicating the kind of immersive listening scenario often encountered in real-world or testing contexts.

Vocabulary and grammar were not pre-taught prior to viewing. Instead, a structured 50-minute review session was conducted during the class period on the following day after each episode was viewed. During these review sessions, instructors guided students through key vocabulary lists, thematic prompts, and character mapping linked to the corresponding episode and excerpts from the novel. This process allowed students to consolidate their understanding of the language while also reflecting on the story's progression and moral dilemmas.

Scaffolding of the justice system and related culture was an essential part of the instructional design. Every week during the 50-minute review session of each episode, instructors posed guiding questions such as: "How is evidence handled in this case?" "What role do prosecutors and defense attorneys play?" and "How might this situation be handled differently in the United States?" These prompts prepared students for discussions that would happen in the next class sessions. Meanwhile, the instructors maintained a concept map on the classroom whiteboard to help students track characters, motives, and events within the evolving crime narrative. This

visual framework became a shared interpretive tool, linking plot elements to linguistic features and sociocultural themes (e.g., legal terminology, formal register). Though photographs of the map could not be retained due to security protocol, the concept map remained a central artifact of the project's development and classroom engagement.

After they watched the show, students discussed the reflective topics they had been prompted to consider, such as comparing the Chinese justice system depicted in the streaming series with the U.S. system based on their prior knowledge or real-life experience. These conversations extended beyond language practice into meaningful cultural interpretation, enhancing the development of students' cultural competence in the process.

Instruction was delivered using a hybrid of physical and digital tools. Episodes were streamed from YouTube, projected in-class. Instructional prompts, vocabulary notes, and reading materials were compiled in Microsoft OneNote, which students accessed regularly. The classroom whiteboard was used for concept maps and group contributions, to reinforce the interactive nature of the instructional design.

In summary, the instructional workflow followed a consistent weekly structure:

Step 1: Episode Viewing on Monday

Students viewed one full episode of *The Long Night* during a 50-minute session. Episodes were screened with Chinese subtitles only, without pausing or English glossing, requiring real-time processing of naturalistic dialogue.

Step 2: Guided Discourse Analysis on Tuesday

Instructors facilitated structured analysis comparing key scenes from Monday's episode with corresponding excerpts from the original novel. These excerpts were introduced and read collaboratively in class rather than assigned as independent homework reading. Activities included examination of dialogue structure, narrative perspective, legal terminology, and thematic interpretation.

Step 3: Vocabulary Consolidation and Cultural Discussion

Students reviewed targeted vocabulary lists derived from the episode and excerpts from the novel. Discussions incorporated comparative discussion of Chinese and U.S. legal systems, guided by instructor framing questions.

Step 4: Concept Mapping and Narrative Tracking

A collaborative concept map was updated weekly to track characters, relationships, motivations, and key events, supporting comprehension of the complex narrative structure.

Step 5: Reflective Engagement

Students participated in group discussions and ongoing reflective interpretation, culminating in a final written reflection synthesizing linguistic and cultural insights.

Student Engagement and Participation

Given the intensity of the course and the already substantial volume of homework associated with the standardized curriculum, no additional assignments were given outside of class for this media-based project. The design was intentional: the goal was to enhance in-class learning experiences without increasing the cognitive load on students. Instead, participation was emphasized during class time, with special attention to the review sessions on Tuesday, which followed each episode viewing and focused on the analysis of the televised episode alongside the excerpt from the novel, vocabulary review, and cultural interpretation.

Students were expected to contribute actively to group discussions each Tuesday. These discussions were led by the instructors to ensure that students could follow the complex storyline and navigate advanced vocabulary, given their approximate ILR Level 2 proficiency. Because *The Long Night* presents multiple timelines and dense legal terminology, students often relied on the instructor's framing questions, narrative summaries, and conceptual cues (e.g., character maps or cause-effect prompts) to make sense of the story's progression. While students did not lead the discussion independently, their verbal contributions, questions, and reflections were critical in creating a shared understanding of both language and narrative content.

The only formal written assignment required of students was an individual 600-word final reflection, submitted by each student at the end of the 12-week module. This reflection asked students to synthesize their understanding of the show's narrative, compare elements of Chinese and American justice systems, and express personal insights about characters, plot developments, and broader cultural themes. Though this was not formally graded, it provided a capstone opportunity to demonstrate comprehension growth.

Students were not assessed through quizzes or comprehension checks during the weekly reviews or discussions. Instead, the real-time evaluation in class emphasized formative participation and qualitative indicators of engagement, including meaningful contributions to thematic discussions and reflective analysis. Changes in proficiency were indicated through DLPT5 scores (see Appendix A). However, these scores are presented as contextual indicators rather than direct causal outcomes of the intervention.

Data Collection and Analysis

The data sources included in this study were standardized proficiency measures, students' qualitative written reflections, and students' post-project feedback, providing both quantitative and qualitative perspectives on language development and learner experience.

The primary performance indicator was DLPT5, which all students completed near the end of the course. Though not designed specifically to evaluate the project itself, the DLPT5 offered a standardized benchmark for assessing changes in listening proficiency over the duration of the course. DLPT5 scores were interpreted as contextual indicators of proficiency development that served as a valuable reference point for correlating instructional exposure to receptive language outcomes.

To explore the cultural and interpretive dimensions of learning, we analyzed students' 600-word final reflection in Chinese (the written reflections are not authorized to publish due to the security protocol). This reflection required students to:

- Compare aspects of Chinese and American justice systems,
- Reflect on the cultural themes presented, and
- Comment on their personal experience engaging with the story.

While not formally scored, the reflections were reviewed through repeated readings to identify recurring themes related to justice system comparison and cultural interpretation. Notes were taken to document patterns across the eight submissions, and common themes were summarized descriptively rather than coded using formal qualitative software. This descriptive thematic approach allowed the researchers to identify shared interpretive tendencies while maintaining the exploratory nature of the classroom-based project.

In addition, students completed a post-project feedback form that consisted of Likert-scale items and open-ended response fields (See Appendix B). The form gathered feedback on:

- Perceived improvements in listening proficiency and vocabulary recognition,
- Confidence following Chinese narrative media without English support,
- Cultural learning and engagement with the sociocultural themes presented in the series,
- Preferences for media-based versus traditional instructional materials.

Together, these data sources provided complementary perspectives on the instructional intervention. Student reflections and feedback addressed students' perceptions of narrative media integration and cultural competence development, corresponding to RQ1, while DLPT5 scores provided contextual indicators of listening proficiency development in relation to RQ2. This combination of performance data and supporting qualitative evidence allowed the study to examine measurable listening outcomes within an intensive, military-oriented language learning context.

RESULTS

The findings suggest that the integration of narrative media and structured cultural discussion was associated with the improvements in listening proficiency and cultural competence. Three major outcomes emerged from the data: (1) the majority of students demonstrated improved listening proficiency as reflected in DLPT5 scores (see Appendix A), (2) students' cultural competence deepened, evidenced through students' written reflections, and (3) students reported strong personal engagement with and positive perceptions of the narrative media approach, evidenced through post-project feedback. It should be noted that all students reported being Very to Extremely Satisfied with the program.

Listening Proficiency Outcomes

The most measurable outcome was observed in the DLPT5 scores (See Appendix A). A closer examination of score movement shows that five students (S2, S4, S5, S6 and S8) advanced from Level 2 to Level 2+ in listening. Students who entered at Level 2+ in listening generally maintained that level at the end of the course. One student (S3) remained at Level 2 in listening. DLPT5 scores were merely interpreted as a contextual indicator of proficiency development instead of a causal indicator of instructional effectiveness. Student feedback sheds light on these gains. Seven out of eight students reported overall feeling that this program greatly helped their listening improve, while one student reported feeling the program helped a neutral amount. Open-ended responses indicated that they felt increased confidence in processing fast-paced spoken Chinese (see Appendix B). Several noted that viewing episodes with Chinese subtitles, without English translation, required sustained attention to real-time speech.

Cultural Competence

Across the reflections, students demonstrated evidence of cultural competence, particularly regarding themes of corruption, justice, sacrifice, and legal complexity. Many were able to articulate intercultural comparisons between Chinese and American justice systems. For example, one student wrote, "在中国，因为有很多腐败，江阳花了很长时间找证据，还付出了很大代价。但在美国，新闻自由，法院公平，可能让真相更快被发现。" ("In China, because there's a lot of corruption, Jiang Yang spent a long time collecting evidence and paid a great price. But in the U.S., freedom of the press and fair courts might allow the truth to be revealed more quickly.")

Another student reflected on the moral nature of the justice system in China, observing, "即使'好人'犯法也会受到惩罚。" ("Even if 'good people' break the law, they will still be punished.")

This student highlighted the show's emphasis on the rule of law as interpreted in Chinese cultural narratives. Multiple students also referenced their surprise that such a critical depiction of

Chinese institutions had passed censorship, indicating deeper engagement with questions of media production, institutional context, and cultural interpretation.

Students responded actively to the series' portrayal of the justice system, despite having limited prior exposure to Chinese legal language. Many reflections discussed the dilemma of ethical sacrifice and the role of whistleblowers. One student wrote, "《沉默的真相》让我思考：为了真相，你愿意付出什么？" ("The Long Night made me ask: what would you be willing to sacrifice for the truth?") Rather than reflecting solely on character motivation, such responses demonstrated perspective-taking and self-reflection across sociocultural contexts, which were key components of the study's operational definition of cultural competence. Through iterative qualitative analysis of reflection content, these patterns were identified as evidence of students' increasing ability to interpret cultural practices and evaluate them in relation to their own assumptions.

Increased Motivation and Personal Engagement

Nearly all students expressed strong personal engagement with the material. Survey results showed that students found the show emotionally impactful, intellectually stimulating, and linguistically challenging, which contributed to a greater sense of relevance in their learning. One student reflected, "我肯定会向任何学习中文的人推荐这个电视剧……可以提高听力和生词知识。" ("I would definitely recommend this series to anyone learning Chinese… it can really improve your listening and vocabulary.")

Another added, "这部剧不仅是一个谋杀悬疑的故事，更是关于权力、腐败和正义代价的深度探索。" ("This series isn't just a murder mystery—it's a deep exploration of power, corruption, and the cost of justice.")

Several students noted they had never completed a media project like this before, and some requested similar units in future courses. Survey responses and reflective comments collectively suggest increased motivation and perceived personal relevance of language learning within a narrative-driven instructional context. Although formal speaking output was not a major focus of the project, students did note that the in-class discussions helped them "find their voice" when explaining complex ideas in Chinese.

DISCUSSION

The findings suggest that media-based instruction can support the development and consolidation of listening proficiency and cultural competence in an LREC-focused classroom. While media-based instruction is not new to second language acquisition, the structured integration of *The Long Night*, paired with instructor-facilitated discussions and analysis of

excerpts from the novel, demonstrated a structured pedagogical approach for Chinese learners of ILR Level 2.

Listening Development

In terms of language development, the project's outcomes show alignment with prior scholarship. Student feedback indicated perceived growth in vocabulary recognition during listening and greater comfort processing authentic spoken input following sustained exposure to high-frequency lexical items in the media. These perceptions align conceptually with findings reported by Chen (2012) and Y. C. Wang (2012), who also observed vocabulary and listening benefits associated with media-based instruction. The DLPT5 results reflect a similar pattern: seven reached ILR Level 2+ in listening, with five students moving from Level 2 to Level 2+. While gains were not uniform across all participants, these findings are broadly consistent with Fujita's (2019) observation that higher-proficiency learners may benefit from media-based instruction when accompanied by structured scaffolding. By embedding weekly episode viewing combined with subtitle-supported listening, guided analysis, and post-viewing vocabulary review rather than repeated scene playback, the project appears to have supported students' efforts to decode complex dialogue.

Cultural Competence Development

The findings also suggest evidence of cultural competence development as defined in this study. Rather than merely engaging with narrative themes, students' reflections demonstrated interpretive engagement with culturally embedded practices and institutional structures. For example, several students analyzed representations of judicial corruption and civil responsibility within the sociopolitical context of contemporary China and contrasted these depictions with their understandings of the justice system in the United States.

These reflections went beyond plot comprehension to include intercultural comparison and perspective-taking, two core components of the operational definition of cultural competence guiding this study. In examining differences in media portrayal, institutional trust, and legal accountability, students articulated both cultural distinctions and the assumptions underlying their own perspectives. Such reflective comparison aligns with Richardson's (2010) emphasis on multimodality as a vehicle for cultural meaning-making.

Student Perceptions and Engagement

Beyond proficiency outcomes and cultural competence development, students' feedback revealed strong personal engagement with the instructional approach, which addresses the perceptual dimension of RQ1. Nearly all students reported that the program meaningfully contributed to their listening development and expressed a preference for narrative media over traditional instructional materials. Students particularly noted the value of sustained, real-time exposure to naturalistic spoken Chinese, a feature of the instructional design that traditional

curricula rarely replicate. This perceived confidence in processing fast-paced authentic input is consistent with Fujita's (2019) finding that structured scaffolding enhances learners' awareness of their own listening strategies. Importantly, students' positive perceptions of the approach suggest that narrative media integration need not come at the expense of learner motivation or program satisfaction, a meaningful consideration for instructors weighing the feasibility of media-based projects within standards-driven, LREC-focused curricula.

Limitations

Several limitations should be acknowledged. First, while most students demonstrated stabilization or advancement within the ILR Level 2 to 2+ range in listening, gains were not uniform across all participants. Two out of eight students did not demonstrate measurable improvement in listening but remained at Level 2+ and one remained at Level 2. While it is possible that these students' listening proficiency did not increase to the level needed for a higher score, it is also the case that score fluctuation across administrations is not uncommon within mid-level proficiency bands of the DLPT5. Also, the DLPT5 was not designed to isolate the effects of a single instructional unit. Because both assessments capture performance at a specific testing moment, factors such as test-day conditions, interactional dynamics, and individual variability may influence outcomes. Future research employing longitudinal designs or multiple assessment points could provide a more stable picture of instructional impact.

Second, the small sample size limits generalizability. Replication across multiple cohorts and institutional contexts would strengthen claims regarding the broader applicability of narrative media-based instruction for supporting listening development and cultural competence.

CONCLUSION

This project suggests that the structured integration of narrative media within an intermediate-level Chinese curriculum may support the development and consolidation of listening proficiency and cultural competence. Through a 12-week module built around the streaming series *The Long Night*, students engaged with complex language, culturally embedded themes, and authentic discourse practices in a format aligned with DLPT-oriented demands.

Triangulated evidence from DLPT5 scores (Appendix A) indicates upward movement within the ILR Level 2 to 2+ proficiency range for several learners, alongside qualitative evidence of intercultural comparison and interpretive reflection (student feedback shown in Appendix B and student reflections). While gains were not uniform across all participants, the findings suggest that guided analysis and reflective discussion can create opportunities for learners to interpret culturally situated practices and examine their own assumptions, which are core components of the study's operational definition of cultural competence.

Student feedback further indicated strong personal engagement with the narrative media approach, with nearly all participants reporting perceived improvements in listening confidence

and a preference for this instructional format over traditional materials. These perceptions suggest that structured media integration can support learner motivation and satisfaction alongside measurable proficiency outcomes, an important consideration for programs weighing the feasibility of similar interventions within demanding, standards-driven curricula.

The findings indicate that narrative media can function as a structured context for engaging learners with linguistically complex and culturally embedded content within an LREC-focused curriculum. Rather than displacing core instructional goals, the module operated alongside established proficiency targets and assessment benchmarks.

Use of AI Tools

The authors used generative AI tools (OpenAI, ChatGPT, and Microsoft Copilot) solely for limited language editing support, including grammar and phrasing refinement, as none of the authors are native English speakers. The authors drafted all substantive content independently, including research design, analysis, interpretation, and conclusions. AI tools were not used to generate research ideas, arguments, data, analyses, references, or sources. All final content was reviewed and approved by the authors in accordance with journal policy.

Authors

Jiayi Xu, Ph.D., is an Assistant Professor and Chinese language instructor at the Defense Language Institute Foreign Language Center's Hawaii Language Training Detachment. He holds a Ph.D. in Bilingual Education from the University of Florida. His research interests include translanguaging, cross-cultural communication, and language ideology. Committed to mission-driven language education, he is dedicated to helping students develop the linguistic proficiency and sociocultural knowledge needed to fulfill their national defense responsibilities.

Maria Wang is an Assistant Professor at the Defense Language Institute Foreign Language Center's Hawaii Language Training Detachment. She holds a Master of Public Administration (MPA) from the University of Hawaii. Her research interests include task-based language teaching, communicative language teaching, content-based instruction, and project-based learning. She focuses on fostering learners' linguistic proficiency and intercultural competence using scaffolded instruction for complex texts and interactions, differentiated instruction for diverse proficiency levels, strategy-based learning, and performance-based tasks and simulations.

Nina Bozeman is an Associate Professor of Chinese at DLIFLC Hawaii Language Training Detachment. Her work focuses on curriculum design, developing advanced proficiency, gamification, and the integration of authentic multimedia and streaming content to strengthen learners' proficiency and cultural competence. A technology-savvy educator, she is particularly interested in gamifying instructional content to increase engagement, motivation, and meaningful interaction with real-world language and cultural perspectives.

REFERENCES

Byram, M., Gribkova, B., & Starkey, H. (2002). *Developing the intercultural dimension in language teaching: A practical introduction for teachers*. Council of Europe. https://rm.coe.int/16802fc1c3

Chen, S. (2012). TV drama as an effective method for lexical instruction: a corpus-based lexical richness study. In *Proceedings of the 2012 annual conference of the Canadian Linguistic Association* (pp. 1-10). https://cla-acl.ca/pdfs/actes-2012/Chen_2012.pdf

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, *10*(3), 241-266. https://doi.org/10.1177/1028315306287002

Fujita, R. (2019). How do TV-drama-based materials affect the listening abilities of EFL learners with different proficiency levels?. *Teaching English through Multimedia: ATEM Journal*, *24*, 17-30. https://doi.org/10.24499/atemnew.24.0_17

Richardson, K. (2010). Multimodality and the study of popular drama. *Language and Literature*, *19*(4), 378-395. https://doi.org/10.1177/0963947010373873

Wang, D. (2012). Self-directed English language learning through watching English television drama in China. *Changing English*, *19*(3), 339-348. https://doi.org/10.1080/1358684X.2012.704584

Wang, Y. C. (2012). Learning L2 vocabulary with American TV drama: From the learner's perspective. *English Language Teaching*, *5*(8), 217-225. https://doi.org/10.5539/elt.v5n8p217

APPENDIX A

DLPT5 Scores of the Students in this Project

#	Entry DLPT & OPI	DLPT L	DLPT R	OPI
S1	2+/3/2	2+	2+	2
S2	2/2+/2	2+	2+	2
S3	2/2+/2	2	2+	1+
S4	2/2/1+	2+	2+	2
S5	2/2+/2	2+	2+	2
S6	2/2+/1+	2+	2+	2
S7	2+/2/2	2+	2+	1+
S8	2/2+/2	2+	2+	2

DLPT5 Scores of the Students in this Project

#	Entry DLPT Listening	Exit DLPT Listening	Change from Entry to Exit
S1	2+	2+	No Change
S2	2	2+	Increase
S3	2	2	No Change
S4	2	2+	Increase
S5	2	2+	Increase
S6	2	2+	Increase
S7	2+	2+	No Change
S8	2	2+	Increase

APPENDIX B

Post-project Student Feedback Gathered in Online Form
Scale from 1 (lowest) to 5 (highest)

Questions	S1	S2	S3	S4	S5	S6	S7	S8
How satisfied were you with the overall Media-Enhanced Linguistic and Cultural Studies (MELCS) program?	5	5	5	5	5	4	5	5
How much did you feel that your listening comprehension improved as a result of this program?	3	5	5	4	5	5	5	5
How much did your vocabulary expand through the TV-based content?	3	4	5	4	5	4	5	4
How effective were the after-show activities (e.g., discussions, reading) in reinforcing your language learning?	2	4	5	4	5	4	5	4
How effective was the instructor's guidance in connecting language and cultural aspects?	4	5	5	4	5	5	5	4
How much did the program enhance your ability to interpret cultural nuances?	4	5	5	4	4	4	5	5
How confident do you feel in using the language in real-life situations after the program?	3	5	5	3	4	4	4	5
How much would you recommend using the same teaching program for future students?	3	5	5	4	5	5	5	5

	Do you think the TV show *The Long Night* is a valuable addition to the CHIN3950 Course?	**In what ways did *The Long Night* contribute to your advanced Chinese language learning in this course?**
S1	I did enjoy the show. The part that I enjoyed the most was picking up parts of the language I had studied previous in a more authentic environment.	I learned a lot of crime vocab and learned how to use it situationally. I will say that I wish I would have done more exercises with these words to better grasp them, but I definitely know them better now with the show than I would have if we had not watched the show.
S2	Yes. It is a good way to relax but also take in new and focused vocabulary as well as colloquial terms. However, i think that it would have been better if we had come up with vocab lists for the episodes the day after and ran through the timeline of the previous episode. I feel like i didn't	I liked discussing what had happened in the previous episode and going over the character charts. I wish we had been able to come up with vocabulary words as a class for each episode and maybe did a basic timeline together so that if we got a little lost, we could get un-lost.

	really learn much from reading excerpts from the script.	
S3	Yes, I was able to learn more crime and science related vocabulary in a less stressful way	improved reading, listening, and speaking skills through analysis of the show and studying vocabulary used.
S4	Yes, but I feel like it should be moved to a Thursday/Friday, Monday and Tuesday, I feel we should focus on other things, and it'd be a nice way to end the week.	we heard more colloquial spoken language and read colloquial text.
S5	yes, it shows us how the language gets used outside of the classroom setting	it allowed us to use our listening skills with more than a couple minutes of context, also it used more colloquial language than we are normally exposed to
S6	Yes-- the material is more recent/modern than a lot of other curriculum used	The show was full of realistic modern speaking patterns/structure
S7	I do believe it was helpful. I believe it helped with rapid understanding and trying to understand through context.	I think the show helped to expand our exposure to language/colloquialisms and culture. Having the novel analysis the next day was also helpful in cementing understand while also showing the differences in spoken versus written language.
S8	Yes, because adding a more passive way to learn gave me a way to slowly digest information that I was being given without being overwhelmed. That being said I would have preferred this to be an every other day type of class instead of just once a week so we could have explored different topics.	I think the show aided in acquisition of rare vocab and giving a unique perspective on the history/ development of police departments in China.

Applied Language Learning

Action Research

Integrating Cultural Competence into Army Special Operations Forces Basic Language Training

Tong Sun
Defense Language Institute Foreign Language Center, Fort Bragg, North Carolina

Kelly K. Lemmons
United States Air Force Academy, Colorado Springs, Colorado

Inna Ruble
Defense Language Institute Foreign Language Center, Fort Bragg, North Carolina

Irina V. Emelyanova
Defense Language Institute Foreign Language Center, Fort Bragg, North Carolina

The Army Special Operations Forces (ARSOF) Basic Special Operations Language Training (BSOLT) program has demonstrated strong outcomes in language proficiency development, with students exceeding established standards by 43% at the 2/2 level and higher and achieving a 100% pass rate at the 1+/1+ level in FY 2025 (Language, Regional Education, and Culture Directorate, 2025). However, global ARSOF missions require language training that integrates cultural competence to mirror complex mission realities. This action research examines the effectiveness of cultural interventions in extensive language training in the ARSOF context and highlights potential challenges of conducting action research in such settings. In this study, a cross-cultural competence (3C) intervention was implemented with two groups of students, with one of those groups also receiving an intervention related to intercultural communicative competence (ICC). Data include pre- and post-intervention Intercultural Development Inventory (IDI) surveys to measure changes in 3C, open-ended surveys to explore students' perceptions of the intervention, and observations of differences in implementation. Data shows that the group receiving only the 3C intervention showed a decline in 3C scores, as measured by the IDI. This group also expressed that the 3C intervention did not align with their language training, whereas the

group receiving both the 3C and ICC interventions showed an increase in 3C and expressed positive attitudes towards both cultural interventions, including increased self-efficacy in intercultural communication and mission readiness. This finding emphasizes that superficial instruction of culture-general models in language training might be detrimental, while a dynamic approach that integrates culture-general and culture-specific elements and promotes active cultural meaning-making can substantially enhance 3C.

Keywords: *Cross-Cultural Competence; Intercultural Communicative Competence; Basic Language Training; Army Special Operations Forces (ARSOF); Action Research*

BACKGROUND

This action research paper first seeks to understand the effectiveness of implementing specific cultural "interventions" within the Army Special Operations Forces (ARSOF) Basic Special Operations Language Training (BSOLT). BSOLT is an intensive program administered in support of the Language, Regional Expertise, and Culture (LREC) guidance from the Department of Defense (DoD Instructions 5160.70, 2016, and 3126.01C, 2023) that support partner-force-centric, culturally embedded missions. Second, it seeks to serve as an example of the potential difficulties encountered when executing action research within these respective programs.

The American Council on the Teaching of Foreign Languages (ACTFL) established two important proficiency benchmarks and performance indicators for the teaching and learning of foreign languages (NCSSFL & ACTFL, 2021). The first proficiency benchmark and performance indicator was communication, focusing on interpretive, interpersonal, and presentational modes of communication. The second, which is the focus of this action research, was intercultural communicative competence (ICC). ICC can generally be defined as "the ability to communicate in another language and behave appropriately in that culture" (NCSSFL & ACTFL, 2021, p. 2). The establishment of ICC as a primary benchmark and performance indicator within ACTFL demonstrates the importance of ICC in language learning. However, despite the increased emphasis, the field of foreign language education (FLE) continues to struggle to integrate ICC into classroom practices. Hennebry (2014) states that it "has been argued that culture is the marginalized sister of language" (p. 135). Studies indicate that teachers generally hold positive attitudes towards cultivating ICC within language instruction; however, in practice, ICC receives limited prioritization compared to proficiency benchmark of communication, with a primary focus on grammar and vocabulary (Griffith & Lim, 2024; Sercu, 2006; Young & Sachdev, 2011).

The challenge of incorporating ICC is further compounded by reductive and static conceptualization of culture within FLE (Byram, 2021). Culture, within FLE, and within education in general, is often reduced to the transmission of cultural facts, products, and behavioral norms, such as food, festivals, and appropriate ways of greeting and interacting with others (see Gao, 2020). This common interpretation of culture ignores the deeper meaning of culture that focuses

on underlying values and belief systems that drive those outward symbols and behaviors (Byram et al., 2002; Deardorff, 2009). Due, in part, to this interpretation, culture is seen as separate from language and that taking time to teach cultural knowledge is taking time away from teaching language (Biebricher et al., 2019; Gao, 2020). The primary source of this tension is the debate over the cognitive and social orientations of FLE. FLE tends to focus on cognitive orientation, which focuses on ACTFL's first proficiency benchmark and performance indicator, and views language learning as a universal cognitive process emphasizing linguistic functions, memory, and grammar rules. However, this largely ignores the culture-general aspects of intercultural communication, a social-oriented perspective at the heart of cross-cultural competence (3C) treating intercultural interactions as a dynamic social practice through which culture influences meaning, negotiation, and interpretation (Liddicoat & Scarino, 2013). 3C is defined as "a culture-general skill set that includes awareness of one's 'self' in the context of culture, an open mind towards and appreciation of diversity, and the ability to apply 'culture analytical models' to any region" (Watson, 2007, p. 2). ICC focuses on communication and differs from 3C in that it "takes into account language teaching and focuses on [communicating] in a foreign language" (López-Rocha, 2016, p. 107). 3C, however, refers to the broader set of knowledge, attitudes, and behaviors that enable individuals to act effectively and appropriately across cultural contexts regardless of the language used. Greater ICC helps learners understand the cultural meanings embedded in communication, allowing them to interact more appropriately and effectively. Greater 3C helps learners to act in culturally appropriate ways within that target language and culture or others.

To exemplify this tension, this action research study focuses on a specific case study within the ARSOF LREC program that examines the divide between cognitive and social orientations, and seeks to effectively teach both ICC and 3C in the BSOLT curriculum. ARSOF soldiers must build rapport, gain trust, and maintain relationships with foreign partners to accomplish global missions in high-stakes environments (Department of the Army, 2025; Joint Chiefs of Staff, 2014). Their mission requires not only language proficiency but also integrated intercultural effectiveness (i.e., 3C) grounded in ICC. Early on, an Army research report on the ICC requirements for Special Forces teams suggested language training should play a vital role in providing both culture-specific knowledge and "culture-related abilities" such as empathy, adaptability, and tolerance for ambiguity (Russell et al., 1995, p. 3). Two decades later, the *ARSOF NEXT: Return to First Principles* (US Army, 2015) publication reaffirmed that ARSOF LREC training should incorporate ICC/3C along with language training. At the policy level, this emphasis became even more explicit: The Department of Defense (DoD) instruction 5160.70 (2016) mandates the integration of two, and if possible, all three LREC competencies–language, regional expertise, and cultural competence. Complementing this directive, DoD instruction 3126.01C (2023) further reinforces the operational necessity of cultural competence. However, an internal ARSOF LREC program review concluded that BSOLT still primarily focuses on the cognitive aspect of FLE and largely ignores the social orientation—"curricula are focused on building language skills and unsystematically introduce tidbits of information in the form of culture facts that are not contextualized in terms of ARSOF mission demands" (Duran et al., 2018, p. 11).

To address this gap, in 2018, BSOLT implemented a pilot Spanish curriculum program integrating the 12 competencies of Adaptive Readiness for Culture (ARC) listed in the DoD instruction 5160.70 (2016). However, this initiative was discontinued. No debrief report exists, and anecdotal evidence from those involved in the Spanish language pilot study suggests it failed for several reasons. First, the 12 competencies are prescriptive and not descriptive, providing no methodology on how they might be taught, which therefore made it too difficult for language practitioners to adopt these effectively into the curriculum. Second, teachers weren't fully trained to teach the aspects of cross-cultural competence implied by ARC. Research consistently shows that foreign language (FL) educators tend to teach culture as transmission of knowledge and facts of the target culture (Gong et al, 2018; Griffith & Lim, 2024) a set of traits (culture-specific), such as the visible aspects of culture, as opposed to culture as a dynamic meaning-making process (culture-general). As intercultural scholars have noted, when culture is framed primarily as a set of traits rather than a dynamic process of interpretation, reflection, and critical engagement, implementation in language classrooms becomes particularly challenging (Byram et al., 2002; Liddicoat & Scarino, 2013). However, a number of studies indicate that FL instructors are generally unprepared for teaching language in an intercultural and dynamic way (Petosi & Karras, 2020; Wang et al., 2023). This second reason is mirrored in the findings of this action research, which will be elaborated on in the discussion section, namely the dilemma of teaching culture-general vs. teaching culture-specific principles across various instructors with differing backgrounds (Kramsch, 2013; Lemmons, 2015; Liddicoat & Scarino, 2013).

Third, the ARC initiative in BSOLT failed because of curricular choices made as a result of its standalone final assessment instrument, the Oral Proficiency Interview (OPI), which does not specifically measure cultural knowledge or competence and offers only limited insight into learners' intercultural development. Although at the ILR 1-2 levels, the OPI does include culturally situated role-play tasks, its primary evaluation criteria remain linguistic: communicative functions, grammatical control, and lexical range. Consequently, at the end of BSOLT, the OPI provides relatively little information about learners' cultural understanding or their ability to navigate broader social and intercultural contexts. With the OPI as the standalone assessment for success, there is no incentive to incorporate aspects of culture-general. Conversely, a disincentive to focus on culture might be perceived if instructors believe that time spent on culture comes at the cost of time spent on language proficiency. While BSOLT effectively cultivates language proficiency (Language, Regional Education, and Culture Directorate, 2025), developing ICC and 3C is essential to mirror the complex communicative realities of ARSOF missions worldwide.

This action research aims to explore effective methods for integrating cultural competence into BSOLT, with the goal of enhancing learners' ICC to align with ARSOF missions without compromising (and hopefully improving) language proficiency outcomes. The study consists of two methods: an ICC-focused intervention and a 3C-focused intervention. For the ICC-focused intervention, a set of ICC-integrative materials comprised of 17 lessons is implemented alongside the core curriculum. Each lesson's material encompasses a mission-oriented scenario, sample dialogue in the target language, vocabulary lists, grammar notes, cultural notes, and applied practice tasks. A performance-based approach was adopted to implement these materials,

requiring students to perform tasks with language accuracy and cultural appropriateness for simulated real-world interactions. The 3C-focused intervention consisted of six lectures, with the first delivered in-person and the remaining five through video recordings. Students watched the videos and engaged in discussions with their language instructors.

The research questions guiding this action study were:

1. Does integrating cultural competence training (both culture-general and culture-specific interventions) into BSOLT enhance intercultural development within foreign language classes?
2. How do students perceive the integration of these two cultural interventions and their language proficiency development?

ACTION PLAN

Context

The study was conducted within the ARSOF LREC BSOLT program. Within BSOLT, class sizes range from two to six, depending on mission needs. Each instructor is responsible for one small group of students, allowing for intensive and personalized instruction. Instructors teach six hours per day, four to five days per week, depending on the training calendar. The final OPI assessment is evaluated according to the Interagency Language Roundtable (ILR) scale, which ranges from 0 to 5 with "plus" ("+") designations between levels 0 and 4 to indicate sub-levels of proficiency. To graduate from BSOLT and complete the military Qualification Course, learners must achieve a minimum rating of ILR 1+ in both speaking and participatory listening, measured by the OPI. This level reflects proficiency beyond the basic survival abilities of ILR 1 and enables learners to perform many ILR 2 tasks. However, recurring individualized deficiencies prevent them from demonstrating the consistency required for an ILR 2 rating. Learners at the ILR 1+ level can generally handle routine social interactions and basic professional exchanges, though their linguistic control, range, and sociolinguistic awareness remain limited (ILR, n.d.).

The study was conducted during a six-month instructional cycle of BSOLT in 2025. A total of ten students participated in this study, with six studying Chinese Mandarin (referred to as the Chinese group for this article) and four studying Russian. The students were originally divided into three groups: two students of Chinese were in the control group which received no intervention; four Russian students received only the 3C-focused intervention; and four students of Chinese received both the ICC-focused and 3C-focused interventions. Each group had a different instructor. However, due to a staff shortage during the training, the two Chinese classes were combined into one after four months. This adjustment left two groups remaining for the study: the 3C-focused intervention group of four Russian students and a 3C- and ICC-focused intervention group of four Chinese students.

Actions Taken

3C-focused Intervention

The 3C-focused intervention was composed of six lectures specifically designed for ARSOF to foster cultural competence and cultural introspection. The lectures were prepared by a professor at the US Air Force Academy (USAFA) who specializes in teaching 3C to US Air Force Foreign Area Officers (FAO) and USAFA study abroad participants. The lectures catered to the ARSOF mission but generally followed the same instruction recently published in a West Point Press volume on LREC (Lemmons & Schell, 2025). The 3C lectures were taught as *culture-general* 3C, meaning that lectures were not specific to the students' target language culture. Students were taught in this manner so that they might implement the topics covered in the lectures across the various cultures in which they might operate. The six lecture topics included in the 3C intervention are summarized in Table 1. This method of instruction has shown to be an effective means of presenting 3C material both in-person (Lemmons, 2023), and as a recorded asynchronous course (Lemmons & Schell, 2025) in the past. Both studies showed a significant increase in student intercultural competence after participating in a similar course. The lectures were recorded in such a fashion as to elicit engagement from the students and facilitate prompts by the instructor/facilitator who played the recorded lectures.

Table 1

Topics of the Intentional 3C Lectures

Lecture #	Title
1	What Is Culture?
2	Plausibility Structure and Cognitive Dissonance
3	My Plausibility Structure
4	Iceberg Model and Introspection
5	Culture Shock
6	Overcoming Barriers and Conclusion

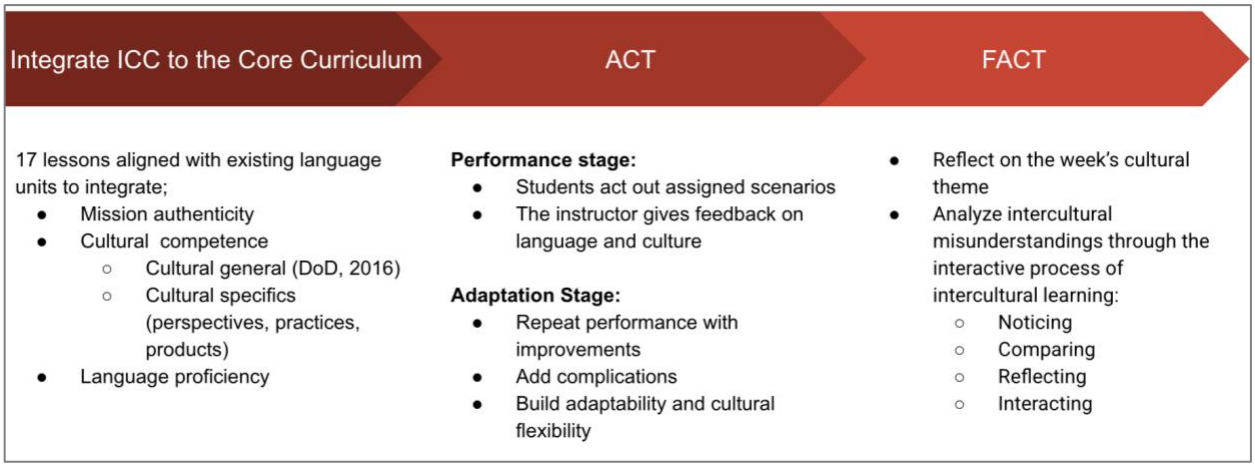
In an effort to establish rapport with the students, the first lecture was delivered in person to the students at Ft. Bragg, North Carolina. The remaining five lectures, ranging from 30 to 50 minutes in length, were provided via pre-recorded video lectures, which BSOLT language faculty from Defense Language Institute Foreign Language Center (DLIFLC) incorporated into class every four to five weeks. Before each lecture, faculty conducted pre-viewing discussions to activate prior knowledge. While watching the videos, faculty also paused at key moments to clarify ideas, explain cultural nuances, and compare target cultures with students' own experiences. When the videos were shorter than 50 minutes, they were followed by a discussion led by the faculty to connect the lecture to class-specific cultural contexts. For both Chinese and Russian groups, the videos were shown during the last hour of Fridays or after module assessments.

ICC-focused Intervention

The ICC-focused intervention included three parts: Integrating ICC into the current core curriculum, ACT classes, and FACT classes[1], as shown in Figure 1.

Figure 1

ICC Intervention Flowchart



Integrating ICC into the Current Core Curriculum. The ICC-focused intervention was implemented using a set of ICC-integrated materials consisting of 17 lessons. Each lesson began with a mission-oriented scenario. The scenarios highlighted operational authenticity, general cultural competency, specific cultural products, practices, and perspectives, as well as communication competence. For example, one scenario titled "Visiting a Taiwanese Counterpart's House" situates learning in a realistic social setting:

> "Scenario: Having recently joined the Taiwan team, you've quickly become acquainted with your Taiwanese counterpart, Captain Wang. After a week of collaboration, he extends an invitation to his home for a weekend lunch. Captain Wang lives with his wife and child, along with his mother—a traditional family arrangement in Chinese culture. You are now at the dining room table. Lunch is about to start."

All ICC-focused materials are designed to fit within the language core curriculum. For instance, this scenario is introduced directly after students receive the "Chinese Food" lesson from the core curriculum, in order to "translate" the curriculum into an authentic communicative context. Table 2 demonstrates the rationale and target objectives that guided the design of the lesson containing the above scenario:

[1] ACT ("ACTing") classes involve performative intercultural practice and FACT (fact-based) classes involve explicit cultural explanations and reflection (Yu, 2020).

Table 2
Example of Rationale and Objectives for Designing ICC-integrative Materials

Target Area	Rationale and Objectives
Operational Authenticity	Interaction reflects a realistic social situation that ARSOF members may encounter during partner-nation engagement or liaison assignments. The importance of cultural understanding in fostering rapport, trust, and long-term cooperation extends beyond formal settings.
Intercultural Effectiveness	Cultural general (DoD, 2016):
	Coping with cultural surprises Taking perspective of others Diplomatic mindset
	Cultural specifics:
	Perspectives: Hierarchy, face, filial piety *Practices:* Waiting for the elders to eat first Responding appropriately when the host serves food or pours drinks Avoiding declining directly *Products:* Shared family meal Signature dishes and drinks
Communicative Competence	Sufficient linguistic proficiency to engage in small talk in a lunch party Using an indirect communication style to show politeness

After the scenario, an example dialogue was provided to demonstrate the desired level of linguistic accuracy and cultural appropriateness. Vocabulary lists, grammar notes, and cultural notes were provided to assist students' independent study before the class.

Instruction through the ACT/FACT Model of the Performed Culture Approach. The Performed Culture Approach (PCA) is a pedagogical framework that integrates language proficiency and cultural appropriateness to develop learners' ICC and enable effective, appropriate communication with members of the target-language community (Yu, 2020). By utilizing performance as the basic unit for both teaching and learning, the instructor situated intercultural communications within specific social contexts, integrating "language and culture as a unitary whole" (Wang & Jia, 2022, p. 4). The PCA utilizes two types of classes: ACT classes and FACT classes. In the ACT classes, the instructor provided context through the five key elements of a performance: time, location, roles, audience, and script, as exemplified in Table 3. Students also communicated in the target language while performing culturally appropriate behaviors. The FACT lessons were designed to allow students to use English when needed, focusing on declarative knowledge and bridging the language–culture gap to foster cognitive understanding (Christensen & Warnick, 2006).

Table 3
Example Five Elements of a Performance

Element	Example	
Time	A weekend lunch following a week of collaboration with Taiwanese counterparts	
Location	Taiwanese counterpart's family dining room	
Roles	American soldier and Taiwanese counterpart	
Audience	Other American soldiers, Taiwanese counterpart's family	
Script	Captain Wang: 大卫，来，你坐这儿。你喝什么？ David, come, sit here. What would you like to drink?	
	David: 随便，喝什么都行。Anything's fine; I'll have whatever you're having.	Guest-host relationship: • Follow the host's convenience (culture-specific) • Take perspectives of others (culture-general)
	Captain Wang: 那尝尝台湾的高山茶吧。快开始吃吧！ Then try some Taiwanese high-mountain tea. Let's start eating!	
	David: 阿姨先吃。Auntie, you eat first.	• Hierarchy and filial piety (culture-specific) • Take perspectives of others (culture-general)
	Captain Wang: 你尝尝这道菜，这是"臭豆腐"。你一定没吃过。Try this dish. It's called "stinky tofu." I bet you've never had it before.	
	David: 哇，闻起来真的很臭，但是吃起来味道还不错。是阿姨做的吗？Wow, it really smells strong, but it tastes surprisingly good! Did your mom make it?	• Giving face by balancing honesty with politeness (culture-specific) • Managing attitudes towards culture (culture-general)
	Captain Wang: 臭豆腐是买的，其他菜是我妈妈做的。 The stinky tofu is store-bought, but the other dishes are my mom's cooking.	
	David: 阿姨的手艺真好。Your mom's cooking is amazing!	•Giving face by complimenting (culture-specific)
	Captain Wang: 哎呀，就是家常菜。对了，我们下个星期什么时候开会？ Oh, it's just simple home-style food. By the way, when's our meeting next week?	• Saving face by avoiding saying "no" directly (culture-specific) • Diplomatic Mindset (culture-general)
	David: 时间和活动都在电邮里。我会发给你。The time and activities are all in the email. I'll send it to you.	

This material serves as self-study homework content. Students review the scenario, dialogue, and accompanying cultural notes before class to become familiar with the communicative context, relevant language forms, and culturally appropriate behaviors associated with the situation. The example dialogue is intended to model one possible way of interacting appropriately in the given context and serves as a foundation for subsequent classroom practice and application.

Since the students had already previewed the ICC material required for their performance, the instructor showed a background picture to set the scene for the scenario. The instructor assigned roles to two students, who performed the scenario from the memory. Longer dialogues were divided into sections to reduce cognitive load for performance. For example, the above dialogue was divided into three sections: pre-meal interaction, trying the food and expressing appreciation, and discussing business during the meal. After students demonstrated their ability to navigate the scenario, the instructor introduced contextual complications to enhance adaptability and flexibility in intercultural communication. For example, the American soldier was allergic to seafood, did not want to drink tea, or felt uncomfortable when served a chicken feet dish.

The ACT lessons were conducted in an immersive environment, focusing on performance and procedural knowledge. Students learned through enactment—performing scenes with linguistic accuracy and cultural appropriateness. By engaging in manageable performances from memory, students built their second cultural worldview by compiling stories about the target culture (Walker & Noda, 2000).

The FACT lessons were conducted during the last hour on Fridays, when the 3C lecture was not scheduled. Each FACT lesson was designed to highlight a cultural theme that emerged from the ACT classes (e.g., in the example from Table 3, themes include giving and saving face and respecting social hierarchy). Each lesson began with students recalling the cultural theme from that week's ACT class. This was followed by an analysis of cultural misunderstandings using guiding questions that promoted the interactive process of intercultural learning—namely noticing, comparing, reflecting, and interacting (Liddicoat & Scarino, 2013). Table 4 presents an example of cultural misunderstanding analysis.

Table 4

Example: Cultural Misunderstanding Analysis–Theme is Face (面子, Mianzi)

Scenario	
Captain Smith had recently arrived in a Chinese-speaking target region for his first overseas assignment. During a multinational humanitarian assistance exercise, he worked closely with local counterparts to coordinate the logistics and transportation of relief supplies. At the conclusion of the exercise, personnel from both sides gathered for an after-action review meeting. During the discussion, a colonel from the counterpart force publicly praised Captain Smith for his professionalism and contributions to the mission. The colonel stated that Smith's efforts had been instrumental to the success of the operation and expressed appreciation on behalf of the counterpart unit. Captain Smith felt somewhat uncomfortable. As a newly assigned officer, he believed that many of the key decisions had been made by more experienced personnel and that the operation's success was the result of a collective effort. Unsure how to respond, he briefly thanked the colonel and attempted to redirect the recognition to the rest of his team.	
Questions	
Noticing	Why did the colonel from the counterpart force praise Captain Smith in front of all personnel? What Chinese cultural perspectives are at play?
Comparing	How would such an incident be comprehended in American culture?
Reflecting	Why does Captain Smith feel uncomfortable?
Interacting	If you were Captain Smith, how would you respond to the praise in front of their leader (in Chinese)?

Predetermined "correct" answers were intentionally not provided, as intercultural interpretation and cultural reasoning rarely lend themselves to a single authoritative explanation. Instead, the activity was designed to encourage students to consider multiple contextual, relational, and cultural factors when interpreting the interaction. In practice, students frequently generated thoughtful and plausible interpretations that differed from the instructor's initial assumptions, reflecting the interpretive and dialogic nature of intercultural learning.

More importantly, the instructional goal was not for students to memorize fixed cultural rules specific to Chinese Mandarin-speaking societies, but rather to develop broader ICC applicable to future military contexts. For example, in this lesson, students connected the cultural concept of *face* to the conduct of after-action reviews (AARs) in Asian contexts. One student with operational experience in Asia noted that he was often unable to elicit direct "areas for improvement" during formal AARs, but later received more candid feedback privately from his partners. Through discussions such as these, students were encouraged to connect cultural perspectives to real-world military interactions and consider how communication strategies may need to be adapted across cultural contexts.

Together, the ACT and FACT classes provided a complementary balance: ACT classes offered an immersive environment where students learned inductively, whereas FACT classes offered

opportunities to explicitly develop a second-culture worldview and apply the target language to mission-oriented intercultural communication in dynamic contexts.

Data Collection

Data collected included a pre- and post-course survey that measured students' 3C, an open-ended survey of students' feedback on the 3C and ICC interventions, and teacher observations. Three groups were originally part of this research experiment with a total of ten students: two students in the Chinese control group, four students in the Chinese experiment group involving both ICC and 3C interventions, and four students in the Russian group involved in the 3C-only intervention. Four months into the 6-month program, the Chinese control group, for reasons outside of this experiment's control, was merged into the Chinese experiment group. For this reason, those two students no longer served as a control group and were therefore removed from the statistical analysis of the survey.

3C was measured using the Intercultural Development Inventory (IDI). The IDI is a 50-question survey that has proven to be statistically reliable and valid (Hammer et al., 2003). The IDI has become the gold standard for measuring 3C pre- and post-cultural training/experiences/ interventions. For this study, the IDI was used to measure the effect of ICC and 3C training on students' 3C in the context of language instruction.

The five-question open-ended survey (See Appendix A) was conducted at the end of the course, allowing students to provide anonymous written feedback about the respective interventions. The first three questions were derived from Byram's ICC model (2021): 1. Attitudes (Curiosity and Openness), 2. Knowledge (Understanding of Self and Others), 3. Skills (Interpreting and Responding Appropriately). Question 4 was aimed at understanding the impact of ICC activities, and question 5 was used for overall feedback and mission effectiveness.

The qualitative method of inductive content analysis (Vears & Gillam, 2022) was used to discern themes and overall opinions from the open-ended survey for both the Russian and Chinese groups. Although classroom observations were not originally planned as a formal data source, one faculty member expressed interest in examining how the 3C-interventions were enacted in practice. As a result, she conducted informal observations of the same intervention sessions in both the Russian and Chinese Mandarin classes. Notes were taken to document instructional strategies, teacher-student interaction, and student response to the cultural components of the lessons. While those notes were not systematic across all sections, they provided contextual insights that helped triangulate findings from the IDI and open-ended survey data. Therefore, observation notes are included as a supplementary data source to support interpretation of the primary data.

FINDINGS AND DISCUSSION

3C Findings

The Russian group received only the 3C intervention. On the pre-course IDI assessment, the students' average score was 90.75. In the post-course IDI, the average score was 84.25, a decrease of 6.5 points. The Chinese group received both the 3C and ICC interventions. The Chinese group had an average pre-course IDI score of 101.25 and an average post-course score of 110.75, an increase of 9.5 points. The two groups had too small a population to run a paired sample t-test to determine if the mean difference between the two sets of paired data was statistically significant. The IDI publishes statistically significant levels for individuals taking pre- and post-tests to measure for change. The standard error of measurement for the IDI is 3.66, which means that if an individual takes the IDI twice within a short period of time with no specific interventions, then their score would be +/- 3.66 from their original score. This means that a difference in score of more than 3.66 is statistically significant (IDI, 2025), showing that the Chinese group significantly improved their 3C, and the Russian group showed a statistically significant decrease in 3C as measured by the IDI.

Open-ended Survey Findings

The open-ended survey results show that the two groups stand in stark contrast to one another. For the Russian group, one theme noticed in the responses from Question 1 was that students perceived the 3C lectures as having low utility, preferring practical language instruction, with respondent #1 stating "I've learned more about the culture from talking to my teacher…" Similarly, a theme noticed from Question 2 was skepticism toward the theory-heavy lectures on 3C, with one respondent stating, "to be honest, some guy speaking in theory had no effect on how I view culture." One theme noticed from Question 3 responses was that there was little perceived application of 3C, and that only the lecture on high vs low cultural context was somewhat helpful. One theme observed in the responses to Question 4 was that the "opportunity cost" of teaching 3C was too high, with one respondent even stating that the 3C "power point presentations hindered my language learning process." Lastly, a theme from Question 5 suggested the 3C lectures needed to be translated into the target language culture being studied and that they were too theoretical and abstract to be useful. Respondent #3 stated "I think that understanding other cultures is a critical part of the ARSOF mission. Unfortunately, the lectures didn't feel applicable to our future needs." Overall, the Russian group perceived 3C as a hindrance to their language learning and was too theoretical when not applied to their target language culture.

For the Chinese group, who received both 3C and ICC interventions, one theme noted in the responses to Question 1 was that these interventions gave them confidence and more willingness to engage, i.e., the feeling of being more able to transfer language skills to real life interactions. Respondent #5, for example, stated that "I now feel more confident and excited to interact with Chinese speakers, both in and outside the classroom." For Question 2 responses reflected a

theme of perspective-taking and adaptability with respondent #6 stating "by understanding these cultural characteristics I have been able to look into my own culture and understand how my world understanding is unique in its own way." From Question 3, one observed theme was personal behavioral shifts and a greater awareness of culture with respondent #8 calling 3C/ICC, "extremely helpful. I can't imagine I would understand much at all about communication/culture strategy without the… activities. Not only am I more aware of critical behaviors in building personal relationships but military/professional as well." One theme observed from Question 4 noted that cultural understanding had been enhanced from understanding the meaning and values that underpin behaviors. In this regard, respondent #5 stated that "understanding the 'why' behind expressions made them easier to remember and use appropriately." From Question 5, one observed theme pertained to strong mission relevance and professional growth with respondent #5 stating that "overall, the 3C/ICC component was a vital part of my growth as a language learner and Special Forces soldier." Overall, the Chinese students stated that 3C and ICC enhanced their language learning and that the cultural aspects improved confidence, language retention, appropriate usage, and mission readiness.

DISCUSSION

For the Russian group, 3C was taught in a "culture general" manner, i.e., not specific to any one culture but rather focused on the foundational principles of cross-cultural competence. The experiment was set up in such a way to measure the effectiveness of teaching 3C only versus teaching 3C in conjunction with ICC. In reviewing the findings above, an obvious distinction was found between the Russian and Chinese groups. According to the open-ended survey results, the Russian group felt that the time spent on 3C lectures came at the detriment of their language learning. The Chinese group expressed the opposite sentiment, stating that 3C principles (culture-general) combined with ICC (culture-specific) had a dramatic positive effect on how confident they felt learning their target language. This same sentiment is also reflected in their 3C score change as measured by the IDI. The Russian group decreased by 6.5 points, while the Chinese group increased by 9.5 points. In contrasting the responses between the two groups, it is apparent that the Russian group felt that the lessons on 3C were too far removed from the language curriculum, too abstract, and took time away from learning basic language skills. In contrast, the Chinese group, through 3C and ICC integration, learned language not as separate from culture, but as a part of culture, stating that this level of understanding gave them greater confidence in their ability to communicate in Mandarin Chinese.

As evidenced by the classroom observations conducted by the faculty member mentioned above, the significant discrepancies in the IDI results and survey responses may be partly explained by learners' affective responses and the differing instructional approaches. The observations showed the Russian group experienced the 3C training in a more static, lecture-heavy format, where students read PowerPoint slides aloud in English before responding to discussion prompts. When no one volunteered to respond, the teacher supplied the answers, resulting in a subdued and disengaged classroom atmosphere. The recorded lectures were not used in the manner in which they were intended in that the instructor did not facilitate follow-up discussions and at

times dismissed them because they did not pertain to aspects of Russian culture specifically. This was at no fault of the instructor. As mentioned above, FL educators often teach culture-specific aspects but do not have the background/training to teach culture-general/3C models. Respondent #3 from the Russian group provides further context stating, "the majority of the cultural understanding I got from this course came from the teacher telling stories about how life was in Russia, the 3C studies were not very useful." The ability of instructors to incorporate culture-general 3C principles into the FL classroom was not one of the original questions driving this action research but became obvious as we analyzed the data.

In contrast to the Russian group, the instructor of the Chinese group had a background in ICC and was more able to incorporate 3C principles via ICC, i.e., more able to teach culture-general models in conjunction with culture-specifics. Additionally, the Chinese group exhibited an emotionally engaging learning environment, where personalized prompts, role-plays, and reflective tasks connected to the students' first culture and therefore reflected a more dynamic approach to teaching and learning culture. These findings support the argument that intercultural competence should be actively developed through dynamic practices which allow learners to construct cultural meanings, rather than through passive transmission of cultural-specific knowledge (Kramsch, 2013; Liddicoat & Scarino, 2013).

The elimination from the control group can serve as an example of the difficulties encountered during action research within these respective programs. This project was originally a pilot study to explore whether or not these interventions merited further scrutiny in a larger experiment with more participants. While it was not the intention to publish these data with such a small *n*, halfway through the research project, priorities surrounding this program shifted and the focus on teaching culture seemed to fall out of favor. As a result, the two Chinese groups were combined, eliminating our control group. After reviewing the initial data and identifying the significant findings, we decided to publish despite the small *n.* Due to unforeseen administrative constraints and the fact that this study might not be continued in the future, we provide these data as a baseline for other researchers. Other difficulties encountered were the government shutdown in late 2025, which kept us from accessing the OPI results in a timely manner, which could have provided a significant data point in the context of this study.

As described above, a major limitation to this study is the small *n*. To reiterate, despite the small *n* we hope that this research can serve as a baseline for future research. Other limitations include the fact that there is no pre-course language evaluation which could allow us to measure the effect of cultural interventions on language acquisition before and after the interventions. Additionally, future studies should be conducted within the same language to allow more valid comparisons between control and intervention groups.

CONCLUSION

The two questions driving this research were first, does integrating cultural competence training (both culture-general and culture-specific interventions) into BSOLT enhance intercultural

development within foreign language classes and second, how do students perceive the integration of these two cultural interventions and their language proficiency development? The findings suggest that merely incorporating culture into language to "check a box" may have a detrimental effect, as demonstrated by the participants in the Russian group whose IDI scores (3C) decreased and who described their cultural instruction as hindering their language instruction. However, the group that purposefully and conspicuously tied culture-general and culture-specific knowledge to language in a participatory and dynamic manner showed a significant increase in 3C scores and reported a greater confidence in their ability to communicate in the foreign language. Furthermore, this research highlighted the consistent teaching dilemma of incorporating cultural competency training into military language programs. While policies emphasize the importance of transferable and adaptive culture-general competencies, classroom operationalization appeared to be anchored in cultural specifics. This tension does not reflect the lack of commitment of instructors but rather foregrounds the structural gap between policy-level expectations and pedagogical preparation. Without explicit training on how to operationalize culture-general competencies through language training, 3C risks being perceived as abstract and disconnected from language training.

Although explicit pedagogical training on integrating 3C into language instruction is important, it alone does not explain the variation in classroom enactment. Future studies should also investigate instructors' attitudes about the value and relevance of cultural competence integration, alongside institutional constraints, such as limited instructional time and assessment washback, to better understand the cognitive and structural factors shaping the teaching dilemma.

Authors

Tong Sun, Ed.D., is a Chinese language Assistant Professor with the Defense Language Institute Foreign Language Center's Language Training Detachment at Fort Bragg, North Carolina. Her professional interests include intercultural communicative competence, intercultural language teaching and learning, and cognitive science-informed AI integration in adult learning environments. She recently completed her Doctor of Education (Ed.D.) studies at Texas A&M University.

Kelly Kristopher Lemmons, Ph.D., is the director of the Geospatial Science program and Associate Professor of Geospatial Science at the U.S. Air Force Academy. As a human/cultural geographer, his research focuses on cultural geography, the development of intercultural competence, and travel. Dr. Lemmons is a graduate of Brigham Young University (BS), University of Massachusetts (MS), and Texas A&M University (PhD).

Inna Ruble is an Associate Professor at the Defense Language Institute Foreign Language Center. She has served with the DLIFLC since 2001 and currently leads the Language Training Detachment within the Language, Regional Expertise, and Culture (LREC) - Operational Programs, Continuing Education, at Fort Bragg, North Carolina. She previously led curriculum development projects for

Russian language programs and has been officially recognized for her contributions to Special Operations language training.

Irina Emelyanova is a Russian language Assistant Professor with the Defense Language Institute Foreign Language Center's Language Training Detachment at Fort Bragg, North Carolina. As a language teacher, her professional interests include communicative language teaching, intercultural communicative competence, and creating engaging and supportive learning environments for students.

Statement on the Use of Generative AI

I, Tong Sun, used generative AI tools (ChatGPT) to support the editing stage of my writing. The AI's influence was minimal, limited to correcting grammar, spelling, formatting, and APA style, and occasionally suggesting alternative phrasing to reduce redundancy or clarify meaning. It did not generate or alter the original content or ideas. I critically evaluated all AI-generated suggestions and incorporated only those that improved readability or accuracy. The core ideas, analysis, and synthesis of literature are entirely my own. https://chatgpt.com/share/691cb50d-c7c0-8000-82ae-05eed487ddd4

REFERENCES

American Council on the Teaching of Foreign Languages (ACTFL). (1996). Executive summary of standards for foreign language learning: preparing for the twenty-first century. https://www.actfl.org/uploads/files/general/1996-National-Standards-for-FL-L-Exec-Summary.pdf

American Council on the Teaching of Foreign Languages (ACTFL). (2017). *Intercultural communication proficiency benchmarks and performance indicators*, p. 5. ACTFL. https://www.actfl.org/uploads/files/general/Resources-Publications/Can-Do_Benchmarks_Indicators-wide.pdf

Biebricher, C., East, M., Howard, J., & Tolosa, C. (2019). Navigating intercultural language teaching in New Zealand classrooms. *Cambridge Journal of Education, 49*(5), 605–621. https://doi.org/10.1080/0305764X.2019.1581137

Byram, M., Gribkova, B., & Starkey, H. (2002). *Developing the intercultural dimension in language teaching: A practical introduction for teachers*. Council of Europe.

Byram, M. (2021). *Teaching and assessing intercultural communicative competence revisited*. Multilingual Matters.

Christensen, M. B., & Warnick, P. (2006). *Performed culture: An approach to East Asian language pedagogy*. National East Asian Language Resource Center at The Ohio State University.

Deardorff, D. K. (Ed.). (2009). *The SAGE handbook of intercultural competence*. Sage Publications.

Department of the Army. (2025). *FM 3-05: Army special operations*. Headquarters, Department of the Army. https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN44116-FM_3-05-000-WEB-1.pdf

Department of Defense. (2016). *DoD Instruction 5160.70: Management of the Defense Language, Regional Expertise, and Culture (LREC) Program*. U.S. Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/516070_dodi_2016.pdf

Department of Defense. (2023). *DoD Instruction 3126.01C: Language, regional expertise, and culture capability identification, planning, and sourcing.* https://www.jcs.mil/Portals/36/Documents/Library/Instructions/CJCSI%203126.01C.pdf

Duran, J. L., Rasmussen, L. J., & Sieck, W. R. (2018, April). *Review and analysis of the SWEG (A) LREC program of instruction* [Unpublished internal report]. Global Cognition. Prepared for the Language, Regional Expertise, and Culture Program, John F. Kennedy Special Warfare Center and School, Special Warfare Education Group (Airborne), Fort Bragg, NC.

Gao, X. (2020). Teachers' perceptions of effective strategies for developing intercultural competence. *Global Chinese*, *6*(2), 333–385. https://doi.org/10.1515/glochi-2020-0017

Gong, Y., Hu, X., & Lai, C. (2018). Chinese as a second language teachers' cognition in teaching intercultural communicative competence. *System, 78*, 224–233. https://doi.org/10.1016/j.system.2018.09.009

Griffith, W. I., & Lim, H.-Y. (2024). Teaching intercultural communicative competence in the foreign language classroom. *MEXTESOL Journal, 48*(1). https://www.mextesol.net/journal/index.php?page=journal&id_issue=194

Hammer, M., Bennett, M., & Wiseman, R. (2003). Measuring intercultural sensitivity: The Intercultural Development Inventory, *International Journal of Intercultural Relations* 27, 421-43, 3.

Hennebry, M. (2014). Cultural awareness: Should it be taught? Can it be taught? In P. Driscoll, E. Macaro, & A. Swerbrick (Eds.), *Debates in Modern Languages Education* (pp. 135–149). Routledge.

Intercultural Development Inventory (IDI). (2025). *Meaningful difference in IDI scores.* Retrieved September 1, 2025, from https://idiinventory.zendesk.com/hc/en-us/articles/360035805253-Meaningful-Difference-in-IDI-Scores

Interagency Language Roundtable (ILR). (n.d.). *ILR scale*. https://www.govtilr.org/Skills/ILRscale2.htm

Joint Chiefs of Staff. (2014). *Joint publication 3-05, Special operations.* U.S. Department of Defense. https://irp.fas.org/doddir/dod/jp3_05.pdf

Kramsch, C. (2013). Culture in foreign language teaching. *Iranian Journal of Language Teaching Research, 1*(1), 57–78. https://files.eric.ed.gov/fulltext/EJ1127430.pdf

Language, Regional Education, and Culture Directorate. (2025). *LREC annual result report for FY 2025* (internal report).

Lemmons, K. (2015). Short-term study abroad: Culture and the path of least resistance. *Journal of Geography in Higher Education, 39*(4), 543–553.

Lemmons, K. (2023). Study abroad academic pre-departure course: Increasing student's intercultural competence pre-sojourns. *Frontiers: the Interdisciplinary Journal of Study Abroad, 35*(3), 128–150.

Lemmons, K., & Schell, J. (2025). Developing global leadership: A case study of U.S. Foreign Area Officers. In J. R. Watson, R. Wolfel, & A. Kalkstein (Eds.), *Language, Regional Expertise, and Culture (LREC) in the Military: State of the Science* (pp. 125-135). West Point Press.

Liddicoat, A. J., & Scarino, A. (2013). *Intercultural language teaching and learning*. Wiley-Blackwell.

López-Rocha, S. (2016). Intercultural communicative competence: Creating awareness and promoting skills in the language classroom. In C. Goria, O. Speicher, & S. Stollhans (Eds.), *Innovative language teaching and learning at university: Enhancing participation and collaboration* (pp. 105-111). Research-publishing.net. https://doi.org/10.14705/rpnet.2016.000411

National Council of State Supervisors for Languages (NCSSFL), & American Council on the Teaching of Foreign Languages (ACTFL). (2021). *NCSSFL-ACTFL can-do statements: Intercultural communication benchmarks and progress indicators*. ACTFL. https://www.actfl.org/resources/ncssfl-actfl-can-do-statements

National Council of State Supervisors for Languages (NCSSFL), & American Council on the Teaching of Foreign Languages (ACTFL). (2026). *NCSSFL-ACTFL can-do statements: Intercultural communication and the intercultural reflection tool*. ACTFL. https://www.actfl.org/uploads/files/general/Can-Do-Intercultural-and-Reflection-Tool.pdf

Petosi, E., & Karras, I. (2020). Intercultural communicative competence: Are Greek EFL teachers ready? *European Journal of Applied Linguistics, 8*(1), 7–22. https://doi.org/10.1515/eujal-2018-0018

Russell, T. L., Crafts, J. L., & Brooks, J. E. (1995). *Intercultural communication requirements for Special Forces teams* (Research Report 1683). U.S. Army Research Institute for the Behavioral and Social Sciences.

Sercu, L. (2006). Foreign language teachers and intercultural competence teaching. Are they in favour or against it? Results from an international investigation. *ITL – International Journal of Applied Linguistics, 151*(1), 15–31. https://doi.org/10.2143/ITL.151.0.2015220

U.S. Army John F. Kennedy Special Warfare Center and School's Office of Strategic Communication (2015). ARSOF Next: A Return to First Principles, Vol 28, Issue 2, accessed at: https://www.swcs.mil/Portals/111/28-2_APR_JUN_2015.pdf

Vears, D. F., & Gillam, L. (2022). Inductive content analysis: A guide for beginning qualitative researchers. *Focus on Health Professional Education: A Multi-Professional Journal*, *23*(1), 111-127.

Walker, G., & Noda, M. (2000). Remembering the future: Compiling knowledge of another culture. In D. Birckbichler (Ed.), *Reflecting on the past to shape the future* (pp. 187–212). National Textbook Company.

Wang, J., & Jia, J. (Eds.). (2022). *Performed culture in action to teach Chinese as a foreign language*. Taylor & Francis. https://doi.org/10.4324/9781003199090

Wang, H., Xu, T., & Zhang, M. (2023). A mixed-method study on Chinese primary school EFL teachers' preparation, affecting factors and support needed to implement intercultural foreign language teaching. *PLOS ONE, 18*(4). https://doi.org/10.1371/journal.pone.0284146

Watson, J. R. (2007). "The Role of Language Proficiency in Cross-cultural Competence (3C): A Fundamental Key to Intercultural Effectiveness in Military Personnel." Center for Languages, Cultures, & Regional Studies, United States Military Academy

Young, T. J., & Sachdev, I. (2011). Intercultural communicative competence: Exploring English language teachers' beliefs and practices. *Language Awareness, 20*(2), 81–98. https://doi.org/10.1080/09658416.2010.540328

Yu, L. (2020). The performed culture approach. In Z. Ye (Ed.), *The Palgrave handbook of Chinese language studies* (pp. 1–34). Palgrave Macmillan. https://doi.org/10.1007/978-981-13-6844-8_7-1

APPENDIX A

Open-Ended Survey Questions

1.	How have the 3C/ICC activities in this course influenced your curiosity, openness, or willingness to engage with people from the target culture (Chinese or Russian)? Can you describe a specific moment or activity that made you more open to learning about or interacting with that culture?
2.	Think about what you have learned from the 3C/ICC activities about the values, beliefs, or everyday practices of the target culture, and list the top three that impacted you the most. How has this knowledge affected the way you think about your own cultural background or assumptions?
3.	How have the 3C/ICC activities helped you interpret culturally unfamiliar behaviors and respond in ways that are appropriate in the target culture? Can you share an example of how your perspective or communication strategy changed as a result?
4.	In what ways did the 3C/ICC activities enhance or hinder your language learning process in this course?
5.	How relevant were the 3C/ICC activities to your future mission or professional role? What aspects of these activities were most helpful, and what suggestions do you have for improving the intercultural component of the basic language program?

Applied Language Learning

Commentary

A Letter from Valley Forge and Words of the Developing World: Knowing the "Why" for Autonomous Language Learning

Daniel R. Seibel
Defense Language Institute Foreign Language Center, Monterey, California

Knowing the "why" has always been critical to our collective drive as an American fighting force. At the Defense Language Institute Foreign Language Center (DLIFLC), instructors are charged with educating a remarkably diverse population of military students from every demographic and professional background. Meanwhile, the DLIFLC language instructor population is just as diverse with civilian faculty from countless nations and military instructors from various branches of service. While having teachers explain the "why" to students in American schools is customary, it is not necessarily a common practice in other countries. This gap in expectations or pedagogical understandings is often invisible to students and teachers alike. In this article, I share my understanding of the "why" for autonomous and well-rounded language learning—an understanding which I developed over time through lessons learned as a DLIFLC student and as an Army Foreign Area Officer—and I reflect on ways to effectively communicate the "why" to our student population.

Keywords: *Autonomous Language Learning, Military Language Education, Military Pedagogy, Learner Motivation, Foreign Area Officer*

> *The genius of this nation is not in the least to be compared with that of the Prussian, Austrians, or French… You say to your soldier 'Do this and he doeth it'; but I am obliged to say [to the American soldier]: 'This is the reason why you ought to do that: and then he does it.'*
>
> ~Baron Friedrich Wilhelm von Steuben, Valley Forge, 1778 (Trickey, 2017)

Baron Friedrich Wilhelm von Steuben, a Prussian military officer charged with training troops of the Continental Army during the American Revolutionary War, observed a key cultural difference

between his American trainees and their European counterparts: communicating the purpose—the "why"—for every command was fundamental to their understanding and ability to perform tasks to standard. By explaining the greater purpose behind menial drills, he transformed an undisciplined and ill-equipped assembly of incompatible militiamen into a professional fighting force capable of challenging a global superpower in a matter of months (Trickey, 2017). Foreign language instructors of American military students will recognize striking parallels between the culturally nuanced didactic challenges that von Steuben faced and those they encounter in their modern classrooms. To cultivate competent military linguists with a strong propensity for autonomous language learning, instructors must consider the roots of the United States military and Friedrich von Steuben's 250-year-old insights about American trainees.

BACKGROUND

How the "Why" Supports Autonomous Learning

Knowing the "why" has always been critical to our collective drive as an American fighting force. Throughout our history, conveying purpose behind our tasks and granting autonomy at the lowest echelons have allowed us to outperform the top-down command structures and aristocratic cultures of foreign counterparts. These concepts enabled the Continental Army to quickly learn the conventional skills necessary to stand against a world superpower during the American Revolution. They empowered scattered, isolated bands of paratroopers from the 82nd and 101st Airborne Divisions to secure key objectives at Normandy in World War II. They allowed a handful of outnumbered U.S. armored cavalry troops to obliterate two enemy brigades at the Battle of 73 Easting in Operation Desert Storm. This approach of empowering subordinate decision making and decentralized execution of tasks is defined in Army command and control doctrine as "mission command" (ADP 6-0, 2019, p. 1-3). It permeates American military culture and is enabled by certain principles including mutual trust, shared understanding, commander's intent, and disciplined initiative (ADP 6-0, 2019).

Language Learning at the Defense Language Institute Foreign Language Center

At the Defense Language Institute Foreign Language Center, instructors are charged with educating a remarkably diverse population of military students from every demographic and professional background: initial entry trainees from every branch of service, careerists and fleet returnees transitioning to new career fields, seasoned non-commissioned officers (NCOs) returning for a second or third language qualification, Foreign Area Officer (FAO) students of various grades, and special operations personnel preparing for unconventional missions. The personal lives and responsibilities of each student are just as diverse. Many are single service members with communal responsibilities in crowded barracks spaces and persistent service unit requirements. Others commute home each day with hours of homework remaining as they navigate their unceasing responsibilities as parents, spouses, siblings, and caregivers.

Meanwhile, the DLIFLC language instructor population is just as diverse with civilian faculty from countless nations and military instructors from various branches of service. Some were raised in the United States and learned the language they teach in a university or growing up in a bilingual household, but most faculty came to this country as adults from a wide range of nations and educational backgrounds. While having teachers explain the "why" to students in American schools is customary, it is not necessarily a common practice in other countries. This gap in expectations or pedagogical understandings is often invisible to students and teachers alike.

Challenges Facing DLIFLC Instructors

The DLIFLC instructor is given the arduous task of getting all their students to attain proficiency in a language at a fraction of the time granted to counterparts in civilian educational institutions, all while fostering a love of the target language and culture. DLIFLC's External Academic Peer Review in 2022 recommended teaching learning strategies and designing online instruction materials to impart autonomous language learning skills on our newest generation of military linguists (pp. 2, 5). Since then, faculty have pushed more low-stakes and non-graded events, self-assessments, flipped classroom practices, and peer tutoring support across the Institute's basic language courses. Additionally, DLIFLC continues to offer online resources and technology through its eLearning webpage to support military linguists in all phases of their studies and careers. Yet the return on investment in these efforts is limited to the extent that faculty convey the purpose and payoff for autonomous language learning to military students. America's service members always perform better when they know the "why."

Instructors must instill a sense of purpose in our military linguists by emphasizing the relevance of all aspects of the curriculum, especially those parts which encourage active participation, independent research, and self-study. The development of these skills is in the best interest of our students' professional growth and our national military capability. Imparting the "why" requires collaboration between civilian and military faculty because of their complementary perspectives on language learning. The civilian faculty at DLIFLC are the most influential component of the students' educational experience. Most are native speakers of the target language and spend more hours with the students than any military faculty or cadre from their respective language schools or military service units. Uniformed faculty are well suited to reinforce the "why" from a military professional lens since most are native-English speakers who learned the target language through formal education and possess experience in relevant military career fields. A theme that commonly appears in End of Course Student Questionnaires is praise for capable Military Language Instructors who convey the relevance of course content by sharing their own operational experiences and successful autonomous language learning practices. Additionally, facilitating opportunities for mature careerist and fleet returnee students to share their professional experiences with junior classmates can also create buy-in organically within the class.

Civilian and military faculty must take special care, though, to monitor and guide these discussions in a positive direction. There can be a tendency for students with some years of military experience to share sophomoric opinions about the curriculum or linguist career fields.

Often these sentiments manifest in periodic sensing sessions with the class or among students during lesson breaks throughout the day. Vigilant instructors can redirect, interject with their own experiences, and transform these counterproductive occurrences into opportunities to discuss the "why" for various aspects of the curriculum and how it relates to developing autonomous language learning skills.

A Personal Autonomous Language Learning Journey

My understanding of the "why" for autonomous and well-rounded language learning was a steady evolution over time through lessons learned as a DLIFLC graduate and Army FAO. I had the opportunity to start my FAO training pipeline in the Spanish Basic Course at DLIFLC in 2013. My class was an unexpectedly diverse mix of Army FAO students, enlisted Marine cryptologists, Navy special operators, Air Force exchange pilots, and a Navy Olmsted Scholar. There were numerous occasions when fellow officer students questioned the utility of certain aspects of the curriculum, including transcription and number drills which they considered to be of minimal relevance to their future jobs, and wished to sacrifice these lessons for additional practice with other skills. I later came to realize these views were incredibly misinformed. Hearing from civilian or military instructors about how these various skills might be applied in our future assignments, why they were important for our autonomous growth in the target language after leaving DLIFLC, or how the lessons could be applied to independently learn other less-spoken languages in our region of assignment, could have set us straight.

Less than a year after graduating from DLIFLC, I found myself using number transcription and gisting skills to record inventory data to uncover a corruption scheme involving a high-ranking partner nation officer selling equipment acquired through U.S. security cooperation programs. I remember standing in front of my Senior Defense Official/Defense Attache's desk as a FAO In-Region Trainee and having to respond with certainty when she asked if the numbers I recorded were correct, because they would have national level implications for our future security cooperation partnership with the host country. The reason why all aspects of the DLIFLC course curriculum are important became instantaneously apparent.

Subsequent assignments revealed the ways in which autonomous language learning skills enhance a military linguist's versatility, and would have helped me understand the "why" during my basic language course if presented by instructors with similar experiences. We are naturally fixated on proficiency in our control languages (CLANGs) for obvious reasons: our careers and Foreign Language Proficiency Bonus depend on it. Department of Defense Instruction (DoDI) 5160.70, Management of the Defense Language, Regional Expertise, and Culture (LREC) Program (2016), provides the policy framework for developing Strategic Language Lists which focus training resources towards widely spoken languages that are most likely to enable our military operations abroad. Yet interactions with military counterparts across many continents taught me that learning to express oneself in local dialects—even at a rudimentary level—often carries far more weight than masterful prose in the country's official language. For these lesser spoken languages, though, there is rarely a basic course available.

While assigned to the U.S. Embassy in Asuncion, Paraguay, I made the mistake of trying to perfect my Spanish at the cost of learning Guarani, Paraguay's indigenous language. Shortly after my arrival and building some initial rapport, my Paraguayan Army partners asked me when I would start learning their treasured native tongue. I would light-heartedly respond that I was still far from knowing Spanish. Sixteen months later, I left the country knowing only a handful of words in Guarani and realized I failed to capitalize on an open invitation to establish deeper ties with the local population.

Years later, when I volunteered for an overseas assignment outside of my FAO Area of Concentration (AOC) in Africa, I committed to learning the language that mattered most to the people. I flew into Lilongwe having already studied basic words and phrases in Chichewa, Malawi's most widely spoken indigenous language. For the next two years, I strove to acquire a new word or phrase every day. A mzungu—meaning "white person" or "wanderer"—expressing himself in the local dialect regularly elicited smiles and chuckles in even the most formal meetings, building rapport in a manner that humbly contrasted with the egocentric actions and communication styles of our adversaries in the region. In this case, autonomous language learning skills born from formal language training secured the most lasting and meaningful ties with the partner nation.

CONCLUSION: THEY NEED THE "WHY" TO THRIVE

Ensuring that military students develop the capacity to become autonomous language learners will enable continued growth in their control languages and benefit them in other unforeseen ways. George Friedman in his 2010 book, *The Next 100 Years: A Forecast for the 21st Century*, reminds us that our ability to anticipate the locations and forms of future conflicts is historically poor from an international security perspective. Few had the strategic foresight to predict America's full scale military operations in Korea, Vietnam, or Afghanistan just 10 to 20 years prior to those conflicts. Arming our military linguists with the will and skill to autonomously learn languages beyond their CLANG increases their survivability, adaptability, and potential to shape outcomes in the future conflicts that will unexpectedly arise in remote corners of the world.

Our American military students are guaranteed to struggle in many aspects of their language learning. Memorizing formidable vocabulary lists, applying complex foreign grammar concepts, and completing repetitious transcription exercises might seem as painstaking and nonsensical to them as the musket drills and rigid military movements of their predecessors at Valley Forge. But if we can remember to explain "why" they need these skills to succeed in their future careers, along with some personal examples drawn from our own life experiences, we will empower them to surpass expectations.

Author Note

The analysis and perspectives offered in this commentary essay are based on the professional experiences of LTC Daniel R. Seibel as a U.S. Army Foreign Area Officer and graduate of the

Defense Language Institute Foreign Language Center. There are no personal or financial conflicts of interest to disclose.

Author

LTC Daniel R. Seibel is the Associate Dean of the Undergraduate Russian Language School, Defense Language Institute Foreign Language Center in Monterey, California. Correspondence concerning this article should be addressed to daniel.seibel@dliflc.edu and daniel.r.seibel.mil@army.mil.

REFERENCES

Army Doctrine Publication 6-0, Mission Command: Command and Control of Army Forces (2019). Headquarters, Department of the Army. Washinton, DC. July 31, 2019.

Defense Foreign Institute Foreign Language Center (2025). *eLearning*. https://www.dliflc.edu/elearning/

Defense Foreign Institute Foreign Language Center (2022). *External Academic Peer Review.* July 16, 2022.

Department of Defense Instruction 5160.70, Management of the Defense Language, Regional Expertise, and Culture (LREC) Program (2016). Office of the Under Secretary of Defense for Personnel and Readiness. December 30, 2016.

Friedman, G. (2010). *The Next 100 Years: A Forecast for the 21st Century*. Doubleday.

Trickey, E. (2017, April 26). The Prussian nobleman who helped save the American revolution. *Smithsonian Magazine*. https://www.smithsonianmag.com/history/baron-von-steuben-180963048/

Applied Language Learning

Research

Language Aptitude as Predictor of Performance

John Pendergast
United States Military Academy, West Point

Thomas Dean
United States Military Academy, West Point

Robert Jannusch
United States Military Academy, West Point

West Point attempts to enroll cadets in their required core language courses based on their top three choices/preferences. Because seats are limited in the most popular languages, it was necessary to devise an objective method for reassigning cadets to courses with available seats, most of which are in Cat III and IV languages, like Arabic, Chinese, Russian, and Persian. West Point has used the Modern Language Aptitude Test (MLAT) and Pre-Defense Language Aptitude Battery (Pre-DLAB) as tools for making this decision. Analyzing the records of 17,000 cadets, this study uses logistic regression to assess the effectiveness of these aptitude tests in predicting expected performance. The analysis finds that aptitude as measured by the MLAT and Pre-DLAB modestly predicts performance.

Keywords: *Language Aptitude, Modern Language Aptitude Test, Defense Language Aptitude Battery, Language Placement, Logistic Regression, L2 Performance*

INTRODUCTION

With each entering class, the Department of English and World Languages (DEWL) at West Point finds more cadets wanting to take certain languages than there are seats available. When this occurs, we try to assign as many of these cadets as possible to one of our Category III or IV "strategic languages": Arabic, Chinese, Persian, and Russian (Savko et al., 2022, p. 96). From 2007 to 2019, DEWL used the Modern Language Aptitude Test (MLAT) to assess overall language aptitude among incoming cadet candidates who needed to be moved into one of their second-

or third-choice strategic languages. An MLAT score of 110 was established as a good predictor that a cadet would achieve a passing grade in the strategic languages (per the Department of Defense's language placement threshold scores on the Defense Language Aptitude Battery, DLAB). In 2020, the MLAT could no longer be used for placement, however, due to restrictions created by the COVID-19 pandemic. As a result, the department switched to using a device DEWL refers to as the DLAB Screening Tool (DST). The DST was created by the Defense Language Institute Foreign Language Center and was based on the DLAB. The department's research and analysis cell determined that a score of 22 on the DST was roughly equivalent to a score of 110 on the MLAT, and for the last three cycles of placement testing, DEWL has used that score for assignment and reassignment purposes.

This paper reports on research comparing the possibility of DST scores and MLAT scores to predict student performance in their language classes.

LITERATURE REVIEW

Language Aptitude Tests

Prior to World War II, the Army used various instruments for recruitment screening, which served as precursors of psychological evaluations, IQ tests, and the Armed Services Vocational Aptitude Battery (ASVAB), but they were not oriented toward linguistic ability (Carroll, 1965, p. 90-91). Interestingly, one of the earliest language aptitude prototypes was administered at West Point in 1954, called the West Point Aptitude Test or WPAT (p. 117-19).

The main shortcoming of the early tests is that they were firmly grounded only in the test taker's mastery of English. Carroll and Sapon revolutionized the testing paradigm by studying decades of previous research on language aptitude and concluding that aptitude depended on factors that indicate an individual's linguistic behavior and observations (Carroll, 1965, p. 95). In the mid 1950s, Carroll and Sapon developed these factors by trial and experiment and released the Modern Language Aptitude Battery in 1959 (later re-designating it as the Modern Language Aptitude Test). It soon became the standard for language screening in the DoD and the Foreign Service, therefore being used at West Point from 2011-2020.

While many studies have determined that language aptitude is a key factor in students' language learning success (Andringa & Dąbrowska, 2019; Carroll, 1959; Curcic et al., 2019; Dahlen & Caldwell-Harris, 2013; Petersen & Al-Haik, 1976), some scholars have noted limitations inherent

[1] In 2024, the Department of Foreign Languages merged with the English division to become the Department of English and World Languages.

[2] The developers call it the Pre-DLAB. Because of the possibility for cadets to confuse this test with the actual DLAB (which has often happened with the MLAT), we chose to call it the DST. ("CASL Research Fact Sheet").

[3] The results on that test were compared to performance in the language classes taught at that time: French, German, Portuguese, Russian, and Spanish. Additionally, these results were compared to the cadets' performance in English and Mathematics. The similarities to our present study are striking.

in these exams. Ehrmann and Oxford (1995), for example, point out that "Some of these measures have been questioned because of their focus on analytic, structural skills, their potential incompatibility with less structured, highly communicative language teaching approaches, and their strong relationship with a general intelligence factor" (p. 68). Nevertheless, aptitude tests continue to provide useful information in many institutions.

Additionally, since student motivation has been identified as an equally important predictor of success in L2 coursework (Ehrman & Oxford, 1995, pp. 68, 72, 76; Gardner & Lambert, 1972; Hardison et al., 2012), DEWL uses both a measure of aptitude as well as a short language survey which allows cadets to rank their top three world language preferences for study at West Point and to indicate whether they have had previous language study experience. Both instruments allow the department to consider multiple datapoints when placing cadets into language classes.

Research Question

This study was guided by the following question: To what degree do the MLAT and DST aptitude tests serve as reliable predictors of student performance in West Point's two-semester core language sequence?

METHOD

Setting

In June 2023, the research team submitted a proposal to West Point's Human Research Protections Program to study the degree to which the MLAT and DST aptitude tests serve as reliable predictors of performance in the two-semester core language sequence. Our proposal appeared before the Academy's Institutional Research Committee in July 2023 and received nearly immediate approval. Cadet records from the class of 2011 to the present, were requested and compiled comprising over 17,000 individual records.

Logistic Regression

Logistic regression was used as a statistical technique to model the probability of success based on predictor variables. These data were used to analyze the relationship between MLAT scores and DST scores on the probability that students will succeed in their world language courses. In this context, success means whether they met/exceeded or failed to achieve expected language course grades. These regression models were standardized across all graduating classes. This reduces the impact of any general inconsistencies across the range of language aptitude test averages.

Data Sources

This study utilized five data sources for analysis: (1) students' scores on the Modern Language Aptitude Test (MLAT) from 2011-2020, (2) students' scores on the DLAB Screening Tool (DST) from 2020-2025, (3) students' grade point averages (called the APSC at West Point), (4) students' language course grades over two semesters, and (5) students' responses on West Point's Language Preference Survey (used to identify students' language study experience and preferences, see Appendix A). Other predictors often used by West Point's registrar's office, such as SAT/ACT and College Entrance Examination Rank (CEER) scores, were also considered, but preliminary analyses indicated that they offered no additional predictive information for this study (See Appendix B for a sample of all placement data considered for this study).

For this study, the operative focus is on performance and not on language proficiency. As important as language proficiency is in second language acquisition, it was not analyzed beyond the grade cadets earned in their language class. The grade in a language class, however, is also not to be construed as an indicator of proficiency, as many graded events were partially assessed based solely on whether they were completed/not completed.

Calculating Student Performance

As mentioned above, student performance was calculated as a binary variable of whether a student either "met/exceeded" a certain performance expectation or not. This was accomplished by comparing cadet grades in their core language courses to their overall GPA. For example, a student with a cumulative GPA of 4.0 and an assigned language course grade of B was coded as "not meeting" expected performance because their GPA suggests their performance should be higher. Comparatively, a student with a cumulative GPA of 2.7 and a language course grade of B was coded as "meeting/exceeding" expected performance because their language course grade aligns more closely with their GPA. This data point was then compared with the cadets' score on the DST and MLAT to determine whether a score of 22 and 110 (respectively) predicts this performance.

Participant Data

The segment of the cadet population of greatest interest (hereafter called "treated group") were those reassigned from their first-choice language to a strategic language due to enrollment limitations (class size, available instructors, etc.). The remaining cadets' records were used as control data. When comparing DST scores to GPA and language core course grades for the classes of 2025 and 2026 across all eight languages, data for roughly 1,250 cadets were collated and analyzed. Because most cadets complete their two-semester core language sequence by the end of the sophomore year, data were available for all 2025 cadets, except those who deferred core language to their junior year (mostly life science majors, often called "pre-med" at other colleges). For the class of 2026, these data were only available for those who completed their core language sequence in their freshman year (approximately 175 cadets).

Because considerably more MLAT data are available for classes 2011-2024, the same comparison was conducted for those class years. This has two primary benefits. It permitted us to validate the initial assessment of the MLAT's reliability as a performance predictor and made possible a consistent comparison between the MLAT and DST using the same criteria across all sets of data. There are important implications to the reassignment variable above. First, reassigned students are those who were not placed in their first-choice language. As a result, in this segment of the study, these students were placed in various strategic languages (Arabic, Chinese, Russian and Persian). Thus, for the following logistic regressions, the impact of reassignment can be seen when the model highlights data for those strategic languages.

Empirical Methodology

Often called Grade Point Average (GPA) at other colleges, at West Point, cadets' GPA is called Academic Program Score Cumulative (APSC). As an aggregate measure of a cadet's academic record, APSC provides a holistic indicator of academic achievement at West Point. For this reason, we incorporated APSC as a benchmark to evaluate whether a student's performance in language classes was consistent with their overall academic performance. Although we compared cadet APSC scores with their DST scores for the target population (2025-2026), we determined APSC to be ill-suited for our control group (2011-2024). Generally speaking, in our model, a good control variable occurs before the analyzed result. In this case, APSC was calculated after an individual's performance in a language class was assessed. Because of this, APSC was not included as a control variable because it interfered with our ability to isolate causal effects.

To create a dynamic and reasonable performance expectation for a cadet with a given APSC, we established a threshold at 1.5 percentage points below the APSC. For example, an individual with a 3.00 GPA equates to a B average. The percentage range for a B in DEWL is 83% to 86.99%. Using the lower bound of the range and subtracting 1.5 percentage points, for example, 83% - 1.5% equates to a performance threshold of 81.5%. We also established this threshold to account for language category variables and minor performance fluctuations. Thereby, academic success is measured against a student's established potential rather than an inflexible standard while also providing a realistic buffer that recognizes individual variation without demanding statistical perfection.

Empirical Analysis

The following logistic regression models were coded using R. For each model, performance was analyzed based on averages of cadets' grades in their two core language courses at West Point. The variable is binary, meaning student performance either met or exceeded the expectation for the model, or not. See Figure 1.

Figure 1

Summary of Balance for Matched Data

Summary of Balance for Matched Data			
	Means Treated	Means Controlled SD	Mean Difference
Distance	0.222	0.222	0.0000
CEER	624.581	623.651	0.9300
MLAT	114.875	114.799	0.0760
SAT_Math	657.684	656.226	1.4580
SAT_Verbal	635.282	631.789	3.4930
Persian_Enrolled	0.205	0.203	0.0020
Russian_Enrolled	0.348	0.341	0.0070
Chinese Enrolled	0.158	0.164	-0.0060
Arabic_Enrolled	0.289	0.292	-0.0030

Figure 1 summarizes the difference in the means of each variable showing minimal difference in means ("matched data") between the control group and the treated group thereby minimizing any confounding variables. As a result of this balancing procedure, variables such as SAT and CEER scores were not included in the final models (Figure 4b). For additional diagnostic data, see the Absolute Standardized Mean Difference figure in Appendix C.

RESULTS

The results have confirmed that there is a modest correlation between aptitude scores and academic performance in language classes. This supports the conclusions of research cited above that scores on language aptitude tests have a direct statistical relationship with academic performance in language courses.

A series of logistic regression models was estimated including aptitude scores (MLAT/DST), language indicators, and reassignment status. See Appendix D for an overview of the underlying formulas used in these models.

[4] "Matched" means that each individual in the treated group has been paired ("matched") with a similar individual in the control group based on key characteristics (e.g., SAT scores, MLAT, CEER). This creates two groups that are as similar as possible, except for the variable being studied.

MLAT Analysis

Figure 2 reflects the impact of MLAT scores on performance. The regression coefficient for MLAT (0.234, $p < .001$), indicates that higher MLAT scores are associated with a higher likelihood of meeting or exceeding the expected performance threshold. In this regression, the intercept is the predicted value when all other independent variables are set to 0. The intercept of 2.285 indicates that a student in the baseline category (with predictors set to zero) has an approximately 91% chance of achieving the expected grade (i.e., a grade at or above their specified threshold).

Figure 2
Effect of MLAT on Performance

MLAT	0.234***
	(0.000)
Intercept	2.285***
	(0.000)

Note. *** represents statistical significance at the 95% confidence level

Figure 3 presents the same variables as Figure 2, with the addition of the "Reassigned" variable, which is negatively associated with performance. This suggests that involuntary reassignment to strategic languages is associated with lower performance compared to those in non-strategic languages. Given the intercept of 1.106, a student in the baseline category (with all predictors set to zero) has a roughly 75% chance of getting a grade at or above their specified threshold.

Figure 3
Effect of MLAT on Performance, including Reassignment

MLAT	0.248***
	(0.000)
Reassigned	-0.539***
	(0.000)
Intercept	1.106***
	(0.000)

Note. *** represents statistical significance at the 95% confidence level

Figure 4a illustrates the correlation of MLAT scores to cadet success in the strategic languages when those strategic languages were chosen by the cadets as their first-choice language. The estimates are in comparison to the average student in a non-strategic language course. This indicates that a cadet choosing to take a strategic language was less likely to meet the expected performance threshold compared to the average cadet not taking a strategic language. Moreover, the effect of the reassignment variable has decreased from the model in Figure 3 and has become statistically insignificant. The reassignment variable captures differences between

students who were involuntarily assigned to a strategic language and those who requested it, holding language enrollment constant.

Figure 4 (a and b)

Effect of MLAT on Performance (Standard vs. Matched)

Effect of MLAT on Performance	
MLAT	0.029 ***
	(0.000)
Russian Enrolled	-0.055 ***
	(0.000)
Persian Enrolled	-0.042 ***
	(0.001)
Arabic Enrolled	-0.084 ***
	(0.000)
Chinese Enrolled	-0.137 ***
	(0.000)
Reassigned	-0.008 ***
	(0.447)
Intercept	0.936 ***
	(0.000)

Effect of MLAT on Performance (Matched)	
MLAT	0.043 ***
	(0.000)
Russian Enrolled	0.039
	(0.059)
Persian Enrolled	0.076 **
	(0.002)
Chinese Enrolled	-0.052 *
	(0.047)
Reassigned	0.0004
	(0.983)
Intercept	0.831 ***
	(0.000)

Note. *** represents statistical significance at the 95% confidence level

Figure 4b depicts a smaller dataset (approximately 1250 students) of students who were involuntarily reassigned into a strategic language. These students had virtually the same ACT, SAT, CEER score and any other pertinent academic credentials with the only difference between the two groups being that one group requested enrollment in the strategic language rather than being involuntarily reassigned into it.[5] In this model, Arabic was used as the baseline variable for each strategic language comparison. The "Reassigned" variable here had a value of 0.0004 and a p-value of 0.983 and the "MLAT" variable had a value of 0.043 with $p < .001$. Further, Russian and Persian language variables were positively associated with cadet performance while Chinese is negatively associated.

DST Analysis

In 2021 after COVID, DEWL switched from the MLAT to the DST. Because of this, only 3,000 cadets have taken the DST and completed their respective core language courses as shown in Figures 5a and 5b.

[5] This model's key component is "ceteris paribus" - a Latin term that translates in English to "all else equal" (Persky, 1990, p. 191-92). Unlike the first three models which use all the students in the dataset, this model finds a set of students who were reassigned with specific academic credentials held constant and compares them to a subset of students with similar academic credentials. The only difference between the subsets of students is that one subset was reassigned to a strategic language, and the other was not.

Figure 5 (a and b)

Effect of DST on Performance (Standard vs. Matched)

Effect of DST on Performance	
DST	0.054 ***
	(0.000)
Russian Enrolled	0.122 ***
	(0.000)
Persian Enrolled	.137 ***
	(0.001)
Arabic Enrolled	.127 ***
	(0.000)
Chinese Enrolled	0.03
	(0.309)
Reassigned	-0.417 ***
	(0.000)
Intercept	0.754 ***
	(0.000)

Effect of DST on Performance (Matched)	
DST	0.065 ***
	(0.000)
Russian Enrolled	0.020
	(0.650)
Persian Enrolled	-0.032
	(0.557)
Chinese Enrolled	-0.103 **
	(0.025)
Reassigned	-0.413 ***
	(0.000)
Intercept	0.877 ***
	(0.000)

Note. *** represents statistical significance at the 95% confidence level

On average students with higher DST scores performed better. However, in both the standard (Fig 5a) and matched model (5b) the "Reassigned" variable was negatively associated with performance and statistically significant ($p < .001$). In other words, students who were involuntarily reassigned into strategic languages had significantly lower performance than those who were not. We found that in the class of 2026, 285 cadets were reassigned to strategic languages compared to an average of 80/year in previous years.[6] Of these 285 cadets, there was a much higher rate of lower-than-expected performance, as shown by the negative reassignment coefficient. This appears to be an outlier.[7] The "Reassigned" variable here had a value of -0.413 with $p < .001$ while the "DST" variable was 0.065 also with $p < .001$. Further, the Russian language variable was positively associated with cadet performance while Persian and Chinese were negatively associated.

[6] For the other class years, an average of 80 cadets were reassigned (with a range of from 32-141). Thus, the class of 2026 had three to four times as many reassigned cadets than other years.

[7] There are at least four factors that played a role in making that class year an outlier. When the class of 2026 were sophomores, a new course called Advanced Beginner Spanish was introduced. The administrative requirements for implementing that course demanded that the planners be a bit more draconian in processing cadets' paperwork. As an example, the first step for cadets who wanted to be considered for Advanced Beginner Spanish was to submit a survey outlining their prior experience with the Spanish language. Cadets who did not submit that survey on time were not given a reminder or second try because the Spanish office already knew they needed to trim down the number of cadets by at least 80 to avoid exceeding their enrollment cap. As a result, more than 120 cadets failed to submit their survey on time. Additionally, two professors were on sabbatical in that academic year, which eliminated another 50-60 seats from European languages that could only be filled in Arabic and Chinese. Finally, 40 life science majors pushed back their language requirement from sophomore to junior year and were almost entirely placed in strategic languages. Therefore, it is plausible that the negative relationship between DST score and its effect on performance is not an accurate result.

DISCUSSION

Recall that we defined performance not as an absolute score or grade but rather how closely the grades in the core language courses match those of the cadets' GPA. Overall, scores on both the MLAT and DST were directly associated with performance, meaning that both tests are good predictors of performance. Scores associated most consistently with expected or better than expected performance were as follows:

MLAT 117 (7 points higher than current cut-score)

DST 24 (2 points higher than current cut-score)

Most notably, for cadets who took the MLAT, the effect of reassignment on performance was statistically insignificant, meaning reassignment did not adversely affect performance. For cadets who took the DST, a negative impact of forced reassignment can be observed, but we believe that this result is an outlier and is attributable more to structural changes in DEWL placement priorities than to the DST itself.

The findings associated with the Class of 2026 should be interpreted as a longitudinal data limitation rather than a reflection of cadet aptitude. Because the data set coincides with significant administrative restructuring within DEWL, the results are likely confounded by institutional transition. In this context, the DST scores do not serve as a clean metric of performance, but rather as a data point captured during a period of high systemic variance.

Additionally, the findings of this study are constrained by several unquantifiable variables that introduced statistical noise into the performance data. A primary limitation lies in the instructor variation across the language programs and within individual languages, as the difference in instructional delivery method means that a cadet's success can be a reflection of an individual instructor's teaching style rather than their own language aptitude. This aligns with Ehrman, who used personality tests to study the compatibility of certain teaching styles with student learning styles but later concluded that such tests are weak predictors of success (Ehrman, 1990; Ehrman & Oxford, 1995). Furthermore, our data do not appear to account for so-called "affective factors," such as cadet motivation, which can cause performance to deviate from predicted benchmarks (Hardison et al., 2012, p. 11). The research is also impacted by environmental and extracurricular stressors, including absences due to intercollegiate sports, club commitments, or unexpected illness. In other words, researchers broadly agree that these factors affect language learning but do not agree about how and to what extent. Because these factors remain unattributed in the current model, they function as confounding variables that obscure the relationship between the MLAT/DST and academic performance.

Implementing the recommended adjustments to enrollment thresholds by increasing the cut-scores, specifically raising the MLAT to 117 and the DST to 24, provides a more rigorous baseline for ensuring cadets possess the necessary aptitude for linguistic success. By establishing these higher benchmarks, the Academy can likely reduce the incidence of training failures and might

lower attrition rates, as cadets will be better equipped to handle the intensive demands of the language curriculum from the outset. Moving forward, policy refinement should focus on an iterative evaluation of these thresholds to ensure they remain calibrated with evolving academic standards.

Future research should prioritize longitudinal tracking of cadet performance to validate these benchmarks, ensuring that West Point continues to optimize its selection process while fostering a linguistically agile officer corps capable of leading in an increasingly complex global environment. Looking beyond the specific goals of this study toward a broader research context, the large corpus available in the study could be used to explore evidence that, in contrast to students with lower attainment in their native (L1) language, those with higher attainment get higher scores on L2 aptitude tests (MLAT) and stronger L2 achievement (similar to the APSC numbers in this study) than students with weaker L1 skills. (Sparks et al., 2011, p. 255).

CONCLUSION

The study establishes a clear association between scores on the MLAT and the DST and performance in core language classes. Those with higher aptitude scores consistently perform at least as well in their language courses as they do in their other academic courses, including those involuntarily reassigned to a strategic language based on their aptitude scores. One change that the results suggest is that the threshold score used to enroll cadets in Cat III and IV languages, especially Chinese, may need to be adjusted several points higher.

Acknowledgement

There are no financial or personal relationships that could influence the manuscript.

Authors

John Pendergast, Associate Professor and Program Director of Russian at West Point, writes about second language acquisition and Russian and German music and letters. Recently Chair of the MLA Opera and Musical Performance Executive Committee, this year he co-authored a chapter on decolonizing Russian study abroad in *Inclusive Strategies and Critical Pedagogies for East European and Eurasian Languages* (Academic Studies Press) and was Editor-in-chief of *Arms, Letters, and the Humanities* (West Point Press).

Thomas Dean, Logistics Officer serving at the 101st Division Sustainment Troops Battalion at Fort Campbell, Kentucky, graduated from the United States Military Academy at West Point in May of 2025. He graduated with a bachelor's degree in science, majoring in Economics with a minor in Applied Statistics with a concentration in systems engineering. He graduated from Logistics Basic Officer Leader Course at Army Sustainment University in Fort Lee, Virginia in November of 2025.

Robert Jannusch, Infantry Officer serving at the 3rd Infantry Division at Fort Stewart, Georgia, graduated from the United States Military Academy at West Point in May of 2025. He graduated with a bachelor's degree in science, majoring in Economics with a concentration in robotics engineering. He graduated from Infantry Basic Officer Leader Course at Fort Benning, Georgia in January of 2025.

REFERENCES

Andringa, S. & Dąbrowska, E. (2019). Individual differences in first and second language ultimate attainment and their causes. *Language Learning*, *69*(S1), 5-12.

Carroll, J. B. (1964). *Language and thought.* Prentice-Hall.

Carroll, J. B. (1959). *Modern language aptitude test: MLAT manual*. The Psychological Corporation.

Carroll, J. B. (1965). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education (pp. 87-136).* John Wiley & Sons.

CASL research fact sheet: Predicting DLAB performance (2009). Center for Advanced Study of Language [CASL], University of Maryland.

Curcic, M., Andringa, S., & Kuiken, F. (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning, 69*(S1), 42-71.

Dahlen, K., & Caldwell-Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal 97*(4), 902-16.

Ehrman, M. E. (1990) The role of personality type in adult language learning: An ongoing investigation. In T. Parry & C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 126-178). Prentice-Hall.

Ehrman, M. E. & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *The Modern Language Journal 79*(1), 67-89.

Gardner, R. C. & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Newberry House.

Hardison, C. M., Miller, L. W., Li, J. J., Schroeder, A. N., Burkhauser, S., Robson, S., and Lai, D. (2012). *Second-language skills for all?* RAND Corporation.

LaValley, M. P. (2008). Logistic regression. *Circulation 117*(18), 2395-2399.

Persky, J. (1990). Retrospectives: Ceteris paribus. *Journal of Economic Perspectives 4*(2), 187-93.

Petersen, C. R. & Al-Haik, A. A. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and Psychological Measurement 36*(2), 369-380.

Savko, P., Savukinas, R., Cutter, N., O'Reilly, E., & Sharif, S. (2022). *DLIFLC general catalog 2021-2022*. Defense Language Institute Foreign Language Training Center. https://www.dliflc.edu.

Sparks, L., Patton, J., Ganschow, L. & Humbach, N. (2011). Subcomponents of second-language aptitude and second-language proficiency. *The Modern Language Journal, 95*(2), 253-273.

APPENDIX A

Language Preference Survey (administered during Cadet Basic Training)

CBT LANGUAGE SURVEY FORM

CO ___ PLT ___ NAME _____

1. **Language Experience:** Fill in the buttons beside each language indicating the number of years' study or experience you have in that language.

Language	Years					
	None	1	2	3	4	5
ARABIC	○	○	○	○	○	○
CHINESE	○	○	○	○	○	○
FRENCH	○	○	○	○	○	○
GERMAN	○	○	○	○	○	○
PERSIAN	○	○	○	○	○	○
PORTUGUESE	○	○	○	○	○	○
RUSSIAN	○	○	○	○	○	○
SPANISH	○	○	○	○	○	○

2. **Other Language Experience.**

Other Language:	

3. **Plebe Year Study:** Select "yes" or "no" to indicate your desire to study a Foreign Language during your Plebe Year.

Plebe Year Study:	Yes ○	No ○

4. **Order of Preference:** Indicate, in order of preference, your three choices for language study by selecting a language for each choice.
Note:
At least one selection must be Arabic, Chinese, Persian, or Russian.

1st CHOICE	2nd CHOICE	3rd CHOICE

ARABIC
CHINESE
FRENCH
GERMAN
PERSIAN
PORTUGUESE
RUSSIAN
SPANISH

5. **Concentration:** Select "yes" or "no" to indicate your desire to concentrate in Foreign Languages.

Concentration:	Yes ○	No ○

APPENDIX B

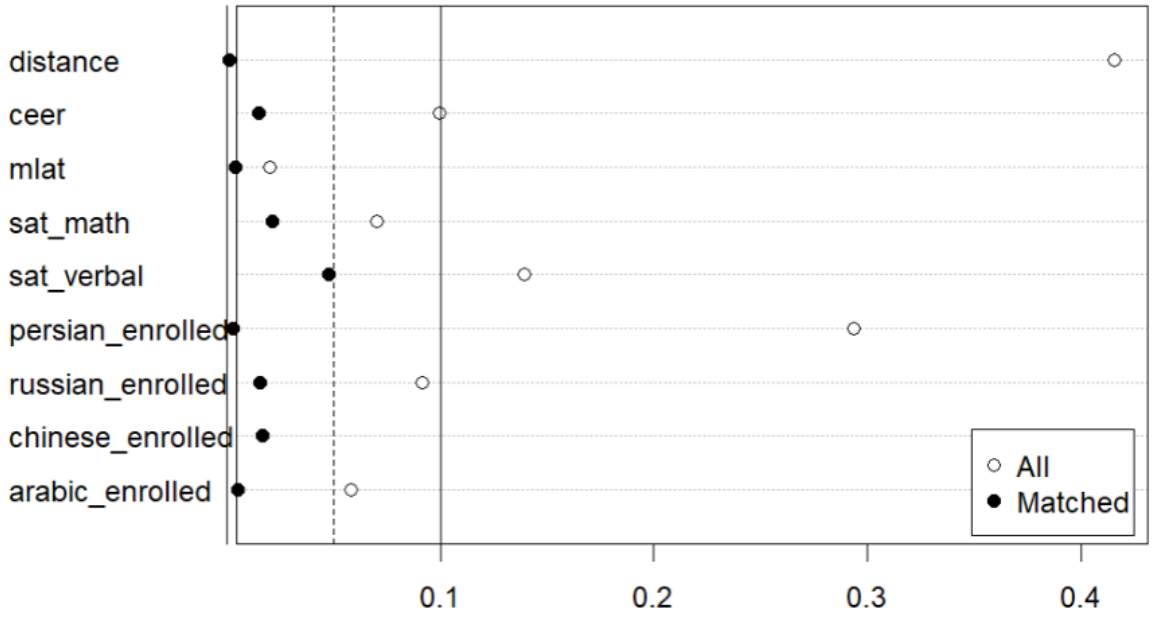
Sample Language Survey Data + Registrar Data (CEER, SAT/ACT, and APSC scores)

Class Year	A	C	F	G	Pe	Po	R	S	Other Exp	1st Choice	2nd Choice	3rd Choice	Placement	Plebe_Lang	Major Minor	MLAT	DST	first sem	first sem grade	first sem pct	second sem	second sem grade	second sem pct	CEER	SAT math	SAT verbal	ACT math	ACT reading comp	APSC
2017	0	0	0	0	0	0	0	2		German	Arabic	Chinese	German	Y	Y	145		LG203	A	94.05	LG204	A	94.55	729	760	750	34	35	3.807
2017	0	0	0	3	0	0	0	0		German	Russian	French	German	N	Y	145		LG371	A+	97.5	LG475	A	94.45	628	650	590	29	22	3.339
2017	0	0	0	3	0	0	0	0		German	Spanish	Arabic	German	N	N	145		LG371	A+	100.25	LG475	A+	97.4	769	720	800	35	34	4.132
2017	0	0	0	1	0	0	0	0		German	French	Persian	German	N	N	143		LG203	A+	97.1	LG204	A	95.15	645	650	720	30	34	3.379
2017	0	0	0	2	0	0	0	0		German	French	Arabic	German	Y	Y	142		LG371	A+	97.75	LG475	A	93.55	738	690	760	35	35	3.743
2017	**0**	**0**	**0**	**2**	**0**	**0**	**0**	**0**		**German**	**Russian**	**Arabic**	**German**	**N**	**N**	**142**		**LA203**	**B+**	**88.13**	**LA204**	**A-**	**90.35**	**653**	**710**	**620**	**34**	**30**	**3.196**
2017	0	0	0	1	0	0	0	0	NONE	German	Russian	Persian	German	N	N	141		LG203	A	93	LG204	B+	87.64	579	550	730	23	36	2.712
2017	0	0	0	0	0	0	0	3		German	French	Russian	German	N	N	139		LG203	A-	91.8	LG204	A	94.45	682	740	700	30	34	3.547
2017	0	0	0	0	0	0	0	0	KOREAN	German	French	Chinese	German	N	N	138		LG203	A	94.5	LG204	A-	90	622	710	540			2.565
2017	0	0	0	0	0	0	0	1		German	Russian	Chinese	German	N	N	137		LG203	A-	90.48	LG204	A-	91.6	724	750	720	30	34	3.507
2017	0	0	0	0	0	0	0	0	LATIN- 1	German	French	Arabic	German	N	Y	137		LG203	B	84.28	LG204	A	93	636	550	670	28	36	2.609
2017	0	0	0	0	0	0	0	1	LATIN 2	German	Russian	French	German	N	N	136		LG203	A-	91.25	LG204	A	93.3	648	630	630			3.4
2017	0	0	1	0	0	0	0	0	NONE	German	French	Russian	German	N	N	136		LG203	A	95.53	LG204	A	95.3	663	680	600	34	29	3.439
2017	0	0	0	2	0	0	0	0	N/A	German	Spanish	Russian	German	N	N	136		LR203	A	94.58	LR204	A	95.58	657			31	24	3.814
2017	0	0	0	0	0	0	0	2		German	Spanish	Arabic	German	N	N	135		LG203	A	94.9	LG204	A-	90.45	704	700	570	35	29	3.742
2017	0	0	0	3	0	0	0	0		German	Persian	Portuguese	German	Y	Y	135		LG371	A	94.33	LG475	B+	88.58	691	750	730	31	30	3.583

APPENDIX C

Absolute Standardized Mean Difference for Sample Subsets

The sample subsets used in this study were very similar. The figure below shows the means for each statistic after matching reassigned students to students that have not been reassigned. This provides a visual demonstration of how similar each subset was as each black mark indicates the mean standardized difference in each category for both subsets.



APPENDIX D

Logistic Regression Formulas for MLAT and DST Analyses

For this study, probability that an outcome y=1, given a set of predictors x was modeled using this general formula:

$$P(y=1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5)}}$$

Here, P(y=1|x) represents the probability of a student receiving a grade equal to or above their given threshold. As y approaches 1, the event is more likely to occur; as y approaches 0, the event is less likely to happen (LaValley, 2008). For this study, the closer y is to 1, the more likely the student is to achieve a grade above the given threshold. As y approaches 0, the student is less likely to achieve this grade. β_0 is the intercept, and the coefficients β_i measure the influence of each predictor x_i on the log-odds of the outcome (where i represents a given number associated with both the coefficient and predictor).

For the MLAT analysis, the predictors were first replaced with meaningful variables such as MLAT (Modern Language Aptitude Test score) and dummy variables representing different languages (x=0 if the student did not enroll in the language and x=1 if they did). MLAT is a continuous variable, while Russian, Persian, Arabic, and Chinese are binary indicators (equal to 1 if the language applies, and 0 otherwise). Each coefficient shows how its predictor changes the log-odds of the outcome, holding everything else constant. After fitting the model with data, the final estimated equation becomes:

$$P(y=1|x) = \frac{1}{1+e^{-(0.936 + 0.029MLAT - 0.0546russian - 0.042persian - 0.084arabic - 0.137chinese - 0.008restrat)}}$$

In this fitted equation, the intercept −.936 represents the baseline log-odds when all predictors are zero. The MLAT coefficient, 0.029, means that each one-point increase in MLAT increases the log-odds of success by 0.029. Conversely, the language variables have negative coefficients, such as −0.546 for Russian and −1.37 for Chinese, indicating a negative association as these languages decrease the log-odds compared to the baseline category (not being in a strategic language). Finally, the "restrat" variable - indicating whether a student was reassigned into a strategic language - has a small negative effect of −0.008 indicating a negative association with student performance compared to students not being reassigned into a strategic language. Thus, logistic regression provides a probability between 0 and 1 for the outcome, with coefficients interpreted in terms of their effect on the log-odds (and, equivalently, the odds) of expected student performance in their foreign language class.

Similarly, the empirical equations for evaluating the DST utilized the same variables and line of reasoning. The completed equation follows:

$$P(y=1|x) = \frac{1}{1+e^{-(0.6502+\ \ 0.0126DST + 0.2307russian + 0.1868persian + 0.2282arabic + 0.1416chinese - 0.6245restrat)}}$$

Applied Language Learning

Research

Beyond News and Documentaries: Developing a Corpus-Based Lexical Resource for Informal North Korean Speech

Mi Hye Lee
Defense Language Institute Foreign Language Center, Monterey, California

Myoyoung Kim
Defense Language Institute Foreign Language Center, Monterey, California

By conceptualizing North Korean (NK) news and NK comedy talk shows as representing distinct Target Language Use domains (Bachman & Palmer, 2010), this study investigates lexical variation across formal and informal registers in contemporary NK discourse and considers whether differences in communicative purpose and discourse conditions are reflected in measurable lexical patterns. Two groups of video clips were compiled, manually transcribed, and normalized: an informal spoken corpus based on 90 minutes of seven NK comedy talk show episodes and a formal spoken corpus based on 90 minutes of NK news broadcasts. The two corpora datasets were analyzed with a focus on part-of-speech (POS) distributions and lexical frequency. The results reveal clear register-based differences in lexical distributions as well as POS categories. In particular, nouns dominate the news corpus, whereas the comedy corpus shows a higher proportion of adverbs, indicating more descriptive and emotionally expressive language use. The comparison was intended not merely to document register contrasts, but to evaluate whether reliance on a single text type, particularly formal discourse, may lead to a skewed lexical representation in instructional materials. In doing so, the analysis provides empirical evidence relevant to questions of domain representativeness in NK language education and highlights the potential limitations of narrowly defined instructional input.

Keywords: *North Korean, Corpus Analysis, Formal and Informal Register, Token and Type, Parts of Speech*

INTRODUCTION AND BACKGROUND

Although North Korean (NK) has often been perceived as largely similar to South Korean (SK), more than 80 years of division between the two have led to significant divergence across multiple linguistic areas, including morphology, syntax, and phonology. Most research on the North Korean language, conducted primarily by South Korean scholars, expanded significantly after North Korea's official announcement of its *Revised Korean Language Education Guidelines* in 2013 (Eum & Seo, 2021; Kim, J., 2015; Kim, N., 2024; Oh, 2015). Prior to 2013, research tended to focus on identifying and comparing linguistic differences between SK and NK. In contrast, after the release of North Korea's national language curriculum in 2013, scholarly attention of SK researchers shifted toward exploring ways to restore linguistic homogeneity between the two languages in preparation for reunification.

Among these efforts, vocabulary-related research has been particularly active, reflecting the growing recognition that vocabulary will play a crucial role in communication in the future. A survey conducted by the National Institute of the Korean Language on the settlement experiences of NK defectors in South Korea further supports the existence of a linguistic gap. Defector respondents identified language, particularly SK vocabulary and pronunciation, as the most challenging aspect of adapting to SK society (Oh, 2015). Testimonies indicating that communication was difficult due to unfamiliar vocabulary clearly demonstrate both the importance of vocabulary learning and the substantial linguistic gap between SK and NK.

At the Defense Language Institute Foreign Language Center (DLIFLC), Korean School students are required to develop proficiency in both SK and NK for their future professional assignments. Accordingly, instructors have consistently sought effective strategies to enhance learners' NK proficiency. Students typically build a foundation in SK vocabulary and grammatical structures before transitioning to NK. In this context, focused vocabulary instruction has been considered one of the most effective means of supporting NK acquisition, as the primary challenge often lies not in syntactic differences but in lexical variation in word choice across contexts, speakers, regions, social groups, and communicative situations.

However, the impact of vocabulary instruction depends critically on the representativeness of the lexical items selected. This concern is particularly relevant in the NK materials used in DLIFLC. Current NK instructional materials rely predominantly on highly formal, government-produced texts such as news broadcasts and official publications. These texts represent a narrowly defined institutional topical domain characterized by scripted discourse, ideological positioning, and conventionalized lexical and grammatical patterns. Even spoken news broadcasts are pre-scripted and informational rather than interactive, reinforcing their classification as formal register. By contrast, informal spoken discourse—featuring spontaneous interaction, colloquial expressions, and pragmatic negotiation—is minimally represented in existing DLIFLC materials. This raises concerns of domain underrepresentation (Bachman & Palmer, 2010). Learners exposed primarily to formal registers may lack exposure to lexical and structural features typical of everyday communication.

Achieving language proficiency requires sustained exposure to the full range of language use situations that learners are likely to encounter in real life. All languages operate across a continuum, extending from highly regulated formal forms to informal forms embedded in everyday life. Although formal language is often associated with written discourse and informal language with spoken communication, this binary distinction is overly simplistic. Register differences cannot be reduced to modality alone. Rather, they emerge from communicative purpose, participant roles, social relationships, situational constraints, and sociocultural norms governing language use. A comprehensive approach to foreign language (FL) education must therefore recognize that proficiency entails not only grammatical accuracy, but also the ability to navigate shifting communicative contexts with appropriate lexical, structural, and pragmatic resources.

This perspective is captured in the concept of the Target Language Use (TLU) domain, proposed by Bachman and Palmer (2010). A TLU domain refers to the set of real-world communicative situations in which language is used and for which language ability is evaluated. It encompasses communicative purposes, language users, settings, tasks, discourse conditions, and the linguistic and pragmatic demands these contexts impose. Within assessment theory, the TLU domain serves as a reference point for ensuring domain representativeness and authenticity. Hence, instructional materials should reflect the communicative realities learners are expected to manage outside the classroom. Although originally developed for language testing, the TLU framework offers a powerful lens for examining instructional materials more broadly. If materials systematically represent only a limited portion of relevant communicative domains, learners may develop competence in that restricted domain while remaining underprepared for others.

Despite the pedagogical importance of such distinctions, research on NK vocabulary remains limited in scope and methodology. Previous studies have examined lexical differences between the two Koreas (Cheong, 2021; Kim, J., 2015; Lee, 2007), proposed the development of NK vocabulary textbooks grounded in intercultural perspectives (Eum & Seo, 2021), explored approaches to NK vocabulary education treating NK as a regional variety (Kim, N., 2024), and suggested the development of instructional materials emphasizing comprehension of NK (Oh, 2015). To date, no study has specifically examined differences between formal and informal registers within NK, which constitutes the primary focus of the present study. This may be because, for native speakers of SK, differences between formal and informal NK registers seldom create substantial comprehension barriers, particularly in listening. By contrast, for foreign learners of NK, such as the students we have at DLIFLC, even minor phonological or grammatical variations may be perceived as distinct lexical and structural features requiring deliberate acquisition.

From a broader research context, corpus-based register studies in Korean and other languages, particularly those adopting multidimensional approaches (e.g., Biber, 2001, 2004; Kim & Biber, 1994; Reppen, 2001), have emphasized grammatical morphology, information density, and global text-structural features, including lexical bundle classification. These studies demonstrate that formal discourse tends toward nominalization and high informational load, whereas informal discourse typically exhibits greater use of modifiers and interactional markers. However,

less attention has been given to lexical distributional patterns at the level of part-of-speech (POS) frequency within a single language variety. To date, no study has systematically examined corpus-based lexical differences between formal and informal registers in NK.

While corpus-based research has primarily advanced our understanding of linguistic patterns, its implications extend beyond theoretical inquiry into language pedagogy. Particularly, corpus-based language data plays an increasingly central role in the development of instructional materials (Curry & Mark, 2024; Moser, 2020), especially in contexts where authenticity and representativeness are essential (Latham, 2025; Li et al., 2025). This empirical foundation supports more principled decisions in vocabulary selection, ensuring that instructional content reflects actual language use within defined communicative domains. Moreover, quantitative comparisons between formal and informal discourse allow instructional material developers to detect imbalances and disparities in lexical and structural representation. In this way, corpus analysis serves not only as a descriptive tool but also as a means of evaluating whether instructional materials adequately represent relevant TLU domains, thereby promoting closer alignment between classroom instruction and real-world language use.

In the present study, NK news broadcasts are classified as formal because of their scripted, institutionalized, and information-dense nature, whereas NK comedy talk shows are classified as informal because they involve spontaneous, interactive exchanges employing everyday colloquial language. Conceptualizing these genres as distinct TLU domains allows register differences to be examined within a principled framework rather than as purely stylistic contrasts.

Building on the theoretical and pedagogical framework outlined above, the following section presents a systematic comparison of formal and informal NK texts through a corpus-based analysis of their lexical features.

Research Questions

The study is guided by two primary research questions:

RQ #1. Are the lexical components used in NK formal and informal texts similar or different in terms of POS distribution and frequency of lexical items?

This question addresses the extent to which register variation manifests at the structural level of lexical organization. If formal discourse is characterized by higher nominal density and informational load, while informal discourse exhibits greater use of modifiers, interactional markers, or other lexical categories, such differences should be observable in corpus-based distributions. By examining POS patterns and lexical frequency profiles across the two corpora, the study aims to provide systematic evidence of intra-varietal register variation within NK.

RQ #2. If differences exist, in what ways are these differences reflected, or not reflected, in current NK language instructional materials?

This question directly connects corpus findings to pedagogical practice within the TLU framework. If instructional materials are predominantly drawn from formal news discourse, they may disproportionately represent lexical features associated with that domain while underrepresenting those typical of informal interaction. Evaluating this alignment allows for a more principled discussion of whether NK instruction adequately prepares learners for a broader range of communicative contexts.

Together, these research questions integrate corpus analysis with TLU-based considerations of domain coverage, positioning the study at the intersection of register research and instructional material development.

METHOD

Data Sources

Approximately 90 minutes of NK news clips and an equal amount of comedy talk show data (seven videos in total) were transcribed and used to build two comparable corpora (see Appendix A for detailed video information). For the analysis of informal speech, comedy talk shows (called 재담, 만담, or 촌극 in North Korea) were selected because they focus on everyday life topics and avoid overt political content, thus making them suitable for analyzing informal language use. While NK news broadcasts are relatively accessible through official state-run broadcasting channels (e.g. Korean Central Television: KCTV; and North Korean News and Media Websites), comedy talk show materials are extremely limited. They were initially obtained from a North Korean defector and are now confirmed to be publicly available on YouTube. The links are provided in Appendix A.

Building Corpora

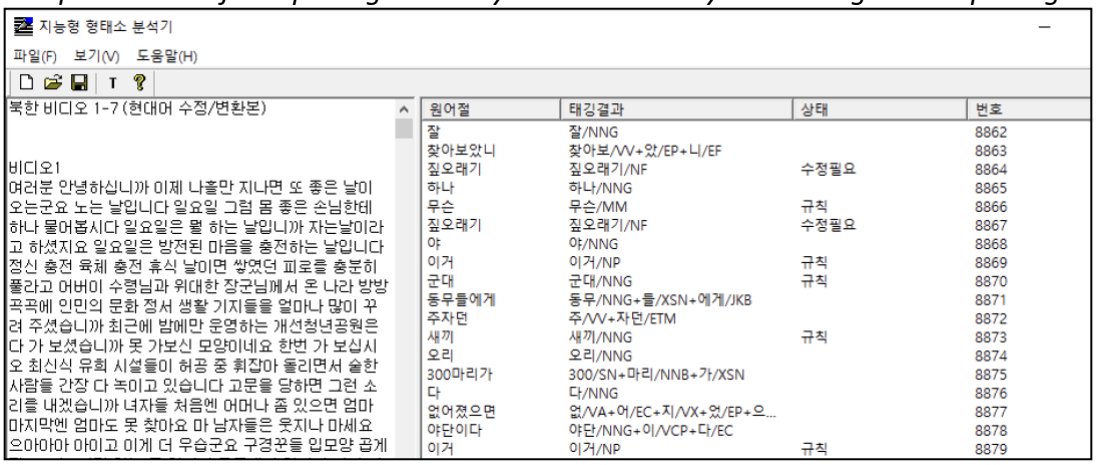
Since both sets of data consisted of spoken materials, a transcription process was required for corpus building. Although existing transcription tools perform reasonably well for standard SK speech, their performance deteriorates markedly when applied to NK speech. The error rate was particularly high in NK comedy talk shows, where diverse intonation patterns, informal expressions, and nonstandard lexical items are common. This reduced accuracy appears to stem from the limited availability of NK speech data for model training. Given these substantial transcription errors, automatic speech recognition was not deemed sufficiently reliable for the purposes of this study. Instead, all recordings were transcribed manually, reviewed in consultation with a North Korean defector, and cross-checked to ensure accuracy and consistency across the dataset. After transcription, colloquial contractions and spoken expressions were normalized into standard South Korean forms, and spacing and punctuation were carefully reviewed and adjusted before finalizing the raw corpus. For this example, refer to Appendix B.

Morphological Analysis and Correction Process

The two raw corpora, news and comedy talk shows, were first automatically annotated using the Intelligent Morphological Analyzer provided by the National Institute of the Korean Language (NIKL) through its Language Information Sharing Service (https://kcorpus.korean.go.kr/). The initial automatic annotation results are as follows.

Figure 1

Sample Results of Morphological Analysis Generated by the Intelligent Morphological Analyzer



지능형 형태소 분석기

파일(F) 보기(V) 도움말(H)

북한 비디오 1-7 (현대어 수정/변환본)

비디오1
여러분 안녕하십니까 이제 나흘만 지나면 또 좋은 날이 오는군요 노는 날입니다 일요일 그럼 몸 좋은 손님한테 하나 물어봅시다 일요일은 뭘 하는 날입니까 자는날이라고 하셨지요 일요일은 방전된 마음을 충전하는 날입니다 정신 충전 육체 충전 휴식 날이면 쌓였던 피로를 충분히 풀라고 어버이 수령님과 위대한 장군님께서 온 나라 방방곡곡에 인민의 문화 정서 생활 기지들을 얼마나 많이 꾸려 주셨습니까 최근에 밤에만 운영하는 개선청년공원은 다 가 보셨습니까 못 가보신 모양이네요 한번 가 보십시오 최신식 유희 시설들이 허공 중 휘잡아 돌리면서 술한 사람들 간장 다 녹이고 있습니다 고문을 당하면 그런 소리를 내겠습니까 녀자들 처음엔 어머나 좀 있으면 엄마 마지막엔 엄마도 못 찾아요 마 남자들은 웃지나 마세요 으아아아 아이고 이게 더 우습군요 구경꾼들 입모양 곱게

원어절	태깅결과	상태	번호
잘	잘/NNG		8862
찾아보았니	찾아보/VV+았/EP+니/EF		8863
짚오래기	짚오래기/NF	수정필요	8864
하나	하나/NNG		8865
무슨	무슨/MM	규칙	8866
짚오래기	짚오래기/NF	수정필요	8867
야	야/NNG		8868
이거	이거/NP	규칙	8869
군대	군대/NNG	규칙	8870
동무들에게	동무/NNG+들/XSN+에게/JKB		8871
주자던	주/VV+자던/ETM		8872
새끼	새끼/NNG	규칙	8873
오리	오리/NNG		8874
300마리가	300/SN+마리/NNB+가/XSN		8875
다	다/NNG		8876
없어졌으면	없/VA+어/EC+지/VX+었/EP+으...		8877
야단이다	야단/NNG+이/VCP+다/EC		8878
이거	이거/NP	규칙	8879

After the initial automatic annotation, manual correction as post-editing was conducted to refine the results and compile the word lists. Figure 2 presents the automatic parsing output for the same transcript shown in Figures 1. Instances of incorrect morphological analysis are highlighted in yellow below. These errors were removed from the dataset prior to analysis.

Figure 2

Example of Automatic Annotation Output and Manual Correction

	A	B	C	D	E	F
1	순번	어절	어절 형태소 분석	비고		
2	1	비디오1	비디오/NNG+1/SN	NULL		
3	2	여러분	여러분/NP	NULL		
4	3	안녕하십니까	안녕/NNG+하/XSA+시/EP+ㅂ니까/EC	NULL		
5	4	이제	이제/MAG	규칙		
6	5	나흘만	나흘/NNG+만/JX	NULL		
7	6	지나면	지나/VV+면/EC	규칙		
8	7	또	또/MAJ	규칙		
9	8	좋은	좋/VA+은/ETM	규칙		
10	9	날이	날/NNG+이/JKS	NULL		
11	10	오는군요	오/VV+는군요/EC	NULL		
12	11	노는	놀/VV+는/ETM	규칙		
13	12	날입니다	날/NNG+이/VCP+ㅂ니다/EC	NULL		
14	13	일요일	일요일/NNG	규칙		
15	14	그럼	그럼/MAG	규칙		
16	15	몸	몸/NNG	규칙		
17	16	좋은	좋/VA+은/ETM	규칙		
18	17	손님한테	손님/NNG+한테/JKB	NULL		
19	18	하나	하나/NR	NULL		
20	19	물어봅시다	묻/VV+어/EC+보/VX+ㅂ시다/EC	NULL	물어보다 과분석	
21	20	일요일은	일요일/NNG+은/JX	NULL		
22	21	뭘	무엇/NP+을/JKO	규칙		
23	22	하는	하/VV+는/ETM	NULL		
24	23	날입니까	날/NNG+이/VCP+ㅂ니까/EC	NULL		
25	24	자는날이라고	자는날이라/NF+이/VCP+고/EC	수정필요	띄어쓰기 분석 오류	
26	25	하셨지요	하/VX+시/EP+었/EP+지요/EF	NULL		
27	26	일요일은	일요일/NNG+은/JX	NULL		

RESULTS AND DISCUSSION

Through the above procedures, the total number of words (token, total word occurrences including duplicates), unique word types (unique lexical items excluding repetitions) were extracted. The results are summarized below followed by a discussion.

Table 1
Token and Type Counts of the NK News and Comedy Talk Show Data

	Token	Type
NK News (Formal Style)	14,415	2,549
NK Comedy Talk Shows (Informal Style)	15,567	2,437
Total	29,982	4,986

In addition to token and type counts, the counts for each POS and function words were also extracted. A total of 31 POS categories were identified in the news corpus and 33 in the comedy talk show corpus (see Appendix C for the complete results of both genres). Although the two corpora show minimal differences in overall type counts and the number of POS categories, the following section examines how the distribution of POS categories differs between the two corpora.

Nouns and Adnominal Items

Table 2 presents the type and token counts for nouns and particles attached to nouns. The percentages are calculated relative to the total number of types and tokens reported in Table 1 above.

Table 2
Token and Type Counts of Nouns and Adnominal Items in NK News and Comedy Talk Shows

Part of Speech	Token (Frequency)		Type	
	News	Comedy Talk Show	News	Comedy Talk Show
Common noun	3,821 (26.51%)	2,790 (17.92%)	1,308 (51.31%)	937 (38.45%)
Proper noun	247 (1.71%)	115 (0.74%)	98 (3.84%)	71 (2.91%)
Adnominal particle	499 (3.46%)	128 (0.82%)	1 (0.04%)	1 (0.04%)
Adnominal ending	1,026 (7.12%)	600 (3.85%)	10 (0.39%)	17 (0.70%)

Common nouns occur in a significantly higher proportion in NK news than in comedy talk shows, in both token and type counts. Among the total 2,549 word types, more than half (1,308, or 51.31%) were common nouns. In terms of frequency, over one-fourth of all tokens (26.51%) were also common nouns. This indicates that a wide variety of nouns are used in news broadcasts,

reflecting the genre's tendency to repeatedly present information about events and to emphasize government policies, particularly in North Korea's case.

For *proper nouns*, while the number of types showed little difference between news and comedy talk shows, their frequency in news was more than twice as high (1.71% vs. 0.74%). This is because news reports frequently include names of people, places, and institutions.

The *adnominal case particle* -의 (*possessive*), which appeared only once as a type in both corpora, showed a frequency of about four times higher in news (3.46%) than in comedy talk shows (0.82%). This can be attributed to the frequent use of noun phrases expanded using -의 (*possessive*) in news texts, whereas in comedy talk shows, short and colloquial expressions often assumed or omitted -의 (*possessive*). Korean examples in this paper are presented in the form of 'a Korean lexical word' followed by its English translation in parentheses. The meanings of grammatical morphemes are italicized to distinguish them from lexical categories. For example, 많이 (*a lot*) vs. -의 (*possessive*).

Similarly, *adnominal endings* appeared more than twice as frequently in news as in comedy talk shows (7.12% > 3.85%), despite similar type counts. This reflects the characteristics of news discourse, which tends to compress information through modifying clauses, whereas comedy talk shows prefer direct description or sequential narration rather than syntactic compression.

In summary, NK news demonstrates a dominant use of common nouns and adnominal particles or modifiers, highlighting its noun-centered and information-focused sentence structures designed for clear and forceful information delivery. This finding aligns with the idea that news discourse is characterized by a predominance of noun-based lexical elements, as reported by Metang and Narathakoon (2025) in their corpus-based study of online news, where they found that nouns, along with noun phrases, serve key informational functions in news texts.

Verbs, Adjectives, and Adverbs

Table 3

Token and Type Counts of Verbs, Adjectives, and Adverbs in NK News and Comedy Talk Shows

Part of Speech	Token (Frequency)		Type (Frequency)	
	News	Comedy Talk Shows	News	Comedy Talk Shows
Verb	1,586 (11.00%)	1,940 (12.46%)	487 (19.11%)	428 (17.56%)
Adjective	520 (3.61%)	575 (3.69%)	165 (6.47%)	136 (5.58%)
General adverb	440 (3.05%)	907 (5.83%)	128 (5.02%)	211 (8.66%)

Verbs

As seen in Table 3, in the case of verbs, the *type* count was higher in the news corpus (487 > 428), but the *token* frequency showed the opposite pattern, comedy talk shows contained more verb tokens (1,940 > 1,586). This means that although news employs a wider range of verb types, comedy talk shows repeat the same verbs more frequently. To examine which verbs were most frequently repeated, the top 30 verbs by frequency were extracted from each genre, revealing several noteworthy distinctions.

The top 30 most frequent verbs in the News and Comedy Talk Show corpora are listed below. Verbs are ordered by frequency within each corpus, and those highlighted in red indicate overlap between the two corpora.

News: 하다 (to do), 경애하다 (to respect), 위하다 (to favor), 심다 (to plant), 보다 (to watch), 만들다 (to make), 가다 (to go), 높이다 (to raise), 찾다 (to find), 받다 (to receive), 품다 (to embrace), 진행하다 (to proceed), 나서다 (to come forth), 생산하다 (to produce), 진행되다 (to be proceeded), 가지다 (to have), 구리다 (to be filthy), 맞다 (to be correct), 수행하다 (to perform), 보내다 (to send), 좋아하다 (to like), 가르치다 (to teach), 들다 (to lift), 일으키다 (to cause), 올리다 (to raise), 나다 (to come out), 이바지하다 (to contribute), 모르다 (to not know), 돌아보다 (to recollect), 간직하다 (to keep)

Comedy Talk Shows: 하다 (to do), 가다 (to go), 되다 (to become), 알다 (to know), 받다 (to receive), 주다 (to give), 그러다 (to be like), 가지다 (to have), 떠나다 (to leave), 죽다 (to die), 먹다 (to eat), 살다 (to live), 맞다 (to be correct), 모르다 (to not know), 나가다 (to go out), 없어지다 (to disappear), 내다 (to submit), 타다 (to ride), 놓다 (to put/drop), 보내다 (to send), 앉다 (to sit), 인정하다 (to acknowledge), 놀다 (to play), 나다 (to be born), 나서다 (to come forth), 막다 (to block), 경애하다 (to respect), 계시다 (to exist), 잘못하다 (to do wrong), 넘겨주다 (to hand over)

Among these, only nine verbs overlapped between the two top-30 lists, representing an overlap rate of about 30%. This reveals a clear lexical divergence that is not visible when looking only at total type or token counts. When the comparison was expanded to the entire verb lists, 157 verbs were shared across both genres, representing 36.7% overlap.

This demonstrates that the sets of verbs used in news and comedy talk shows differ significantly. For example, the verb "경애하다 (to revere, adore a leader)" ranked second in frequency in the news corpus but only 27th in comedy talk shows. This contrast illustrates the political and

ideological nature of NK news discourse compared to the more informal, apolitical tone of comedy talk shows.

Adjectives

A pattern found in verbs is also evident in adjectives. Interestingly, as with verbs, the type count was higher in news (165 > 136), while the token frequency was higher in comedy talk shows (575 > 520). In other words, comedy talk shows used fewer types of adjectives but repeated them more often, whereas news employed a wider variety of adjectives with less repetition. This suggests that the comedy talk show corpus, reflecting the nature of everyday spoken conversation, relies heavily on a limited set of frequently used adjectives. In contrast, news discourse uses a more diverse range of adjectives to describe events and evaluations with precision. Examination of the 30 most frequently used adjectives in both genres further highlights these stylistic and functional contrasts. Among the top 30 adjectives, only six adjectives, 있다 (to be existing), 좋다 (to be good), 크다 (to be big), 고맙다 (to be thankful), 어렵다 (to be difficult), and 훌륭하다 (to be excellent)**,** appeared in both lists, accounting for just about 20% overlap**.** When the entire adjective lists were compared, only 35 adjectives overlapped, resulting in a 25.7% overlap rate**.** This relatively low overlap demonstrates that the two genres draw from largely distinct adjective sets, reflecting their different communicative purposes and stylistic norms.

In the comedy talk show corpus, adjectives such as 고맙다 (to be thankful), 곱다 (to be graceful), 안녕하다 (to be peaceful), 시원하다 (to be refreshing), 아깝다 (to be regretful), 쓸데없다 (to be useless), and 희한하다 (to be strange) frequently appeared, representing emotion-driven or personally evaluative expressions common in casual conversation. These adjectives serve to convey affect, empathy, and speaker attitude, key features of informal, spoken interaction.

By contrast, the news corpus demonstrates a strong ideological and evaluative orientation. Frequently used adjectives such as 위대하다 (to be great), 훌륭하다 (to be excellent), 자애롭다 (to be benevolent), 강력하다 (to be powerful), and 귀중하다 (to be precious) function as politically charged rhetoric, glorifying leadership, national achievement, and socialist values. Such adjectives serve not merely as descriptors but as ideological tools to reinforce loyalty and national pride.

Overall, the usage of adjectives in the two genres illustrates a clear distinction between affective and interpersonal discourse in comedy and ideological and institutional discourse in news.

Adverbs

Among the four major parts of speech (nouns, verbs, adjectives, and adverbs), adverbs showed the most distinctive contrast between the two genres.

Table 4

Token and Type Counts of Adverbs in NK News and Comedy Talk Shows

Part of Speech	Token (Frequency)		Type (Frequency)	
	News	Comedy Talk Shows	News	Comedy Talk Shows
Adverb	440 (3.05%)	907 (5.83%)	128 (5.02%)	211 (8.66%)
Conjunctive adverb	20 (0.14%)	83 (0.53%)	4 (0.16%)	9 (0.37%)

In Table 4, both token and type counts of adverbs were consistently higher in the comedy corpus. The frequency of adverbs in comedy (907, 5.83%) was almost double that in news (440, 3.05%). This can be attributed to the frequent repetition of emotional, attitudinal, and degree adverbs, such as 정말 (really), 잘 (well), 많이 (a lot), and 좀 (a bit), which reflect the speaker's feelings and evaluations. Furthermore, adverbs often co-occur in clusters within the same utterance (e.g., 얼마나 많이, 얼마나 잘), contributing to higher token frequency.

Presented below are the 30 most frequent adverbs in the News and Comedy Talk Show corpora. The adverbs are ranked by frequency within each corpus, and shared adverbs are marked in red.

News: 더 (more), 높이 (highly), 다 (all), 정말 (really), 많이 (a lot), 더욱 (further), 또 (again), 지금 (now), 얼마나 (how much), 잘 (well), 깊이 (deeply), 함께 (together), 아직 (yet), 가장 (the most), 철저히 (thoroughly), 없이 (without), 언제나 (always), 제일 (the first), 바로 (right away), 보다 (than), 특히 (particularly), 이제 (now), 대단히 (greatly), 같이 (with), 계속 (consistently), 정성껏 (carefully), 끝없이 (endlessly), 새로 (newly), 하나하나 (one by one), 또한 (again)

Comedy Talk Show: 다 (all), 또 (again), 안 (not), 좀 (little), 왜 (why), 정말 (really), 못 (not), 빨리 (fast), 얼마나 (how much), 잘 (well), 더 (more), 지금 (now), 이제 (now), 아니 (no), 딱 (perfectly), 절대로 (never), 얼른 (quickly), 그저 (just), 바로 (right away), 없이 (without), 가만(still), 모두(all together), 사실 (in fact), 거저 (for free), 도록 (so that), 몽땅 (all; colloquial), 제발 (please), 다시 (once again), 너무 (too, very), 오직 (only)

Only ten adverbs (33%) overlapped between the two top 30 lists. When expanded to the entire adverb lists, 60 adverbs overlapped, corresponding to a 28.4% overlap rate, meaning that roughly 70% of adverbs used in one genre did not appear in the other. This finding underscores the strong register-based divergence between the adverb usages in formal news and informal comedy talk shows.

A notable pattern was observed in the use of the negative adverbs "안 (do not)" and "못 (cannot)." In the comedy talk show corpus, 안 (do not) ranked third and 못 (cannot) seventh in frequency,

whereas in the news corpus, 안 (do not) appeared only at rank 69, and 못 (cannot) did not appear at all. This absence indicates that negative expressions are rarely used in news discourse, which favors assertive, positive, and ideologically aligned statements.

Conjunctive adverbs, those connecting clauses or sentences while indicating logical relations such as cause and effect, contrast, addition, or transition, also showed significant genre-specific variation. In the comedy talk show corpus, nine types of conjunctive adverbs were identified: for cause-effect 그래서, 그러니까; for contrast 그런데, 근데, 하지만; for addition 그리고; and for transition or condition 그럼, 그러면, 하긴. These appeared with a relatively high frequency of 83 tokens (0.53%), reflecting the genre's conversational and interactive style.

In contrast, the news corpus contained only four types of conjunctive adverbs (그리고, 하지만, 그러나, 그래서) with a total of 20 tokens, roughly one-fourth of the frequency observed in comedy talk shows. The prevalence of colloquial forms such as 그럼 (therefore), 그러니까 (so), and 근데 (but, by the way) in comedy talk shows demonstrates their function in structuring spoken discourse, often used toward the end of an utterance to clarify or wrap up the message.

In summary, adverbs in the comedy talk shows are both more diverse, as reflected in the higher type count, and more frequent, as shown by the higher token count, than in the news. Comedy talk shows employ a wide range of colloquial and affective adverbs repetitively, enhancing rhythm, tone, and emotional engagement. Consistent with the patterns observed in nouns, verbs, and adjectives, the roughly 30% overlap rate between the two genres confirms that they draw on distinct lexical systems, each reflecting its own communicative purpose and stylistic register.

The relatively low overlap in verb, adjective, and adverb usage between the two corpora is noteworthy. The limited shared verbs, adjectives, and adverbs suggest that informal texts introduce many additional words that learners need for broader communicative ability. If instruction is based mostly on formal materials, students may develop vocabulary that works well in formal contexts but is less useful in everyday communication. Including informal materials in instruction can therefore help expand students' vocabulary and strengthen their overall proficiency.

Other POS

Table 5 provides an overview of the distribution of sentence endings, interjections, and vocative particles across the two corpora.

Table 5

Token and Type Counts of Sentence-Final Endings, Interjections, and Vocative Particles in NK News and Comedy Talk Shows

Part of Speech	Token (Frequency)		Type (Frequency)	
	News	Comedy Talk Shows	News	Comedy Talk Shows
Sentence-final ending	317 (2.20%)	1,231 (7.91%)	5 (0.20%)	116 (4.76%)
Interjection	None	746 (4.79%)	None	103 (4.23%)
Vocative particle	None	7 (0.04%)	None	2 (0.08%)

Sentence-Final Endings

Sentence-final endings show a striking contrast between the two genres. In the news corpus, only five formal sentence-final endings were identified, whereas the comedy talk show corpus contained as many as 116 distinct forms, including a wide range of colloquial endings. Their token frequency in the comedy talk show was also substantially higher.

- Endings used in the news (5 types): 습니다, ㅂ니다, 다, 습니까, ㄴ답니다
- Colloquial endings used in comedy talk shows (20 types): 구먼요, 다니요, 는구먼, ㄴ대, 대, ㄴ답니다, ㄴ데요, 던데, 라우, 아야지, 는지, 로군, 외다, 라요, 로구먼, 는군요, 더라, 든가, 너라, 라니

This contrast reflects the stylistic nature of NK news, which maintains a formal, standardized, and monotonous tone aimed at delivering information or official statements. In contrast, the comedy talk show genre exhibits rich variation in colloquial endings, used for declarative, interrogative, imperative, exclamatory, and emphatic expressions, that convey the speaker's emotions, attitudes, intimacy, and social dynamics. The diversity of these sentence-final endings underscores that comedy talk shows are highly interactive and emotionally expressive spoken discourse, while news remains formal and non-interactive.

Interjections

Interjections offer another clear point of contrast between the two genres. No interjections appeared in the news corpus, whereas the comedy talk show corpus included over 100 distinct types and 746 tokens. Interjections typically signal spontaneous emotional responses such as surprise, frustration, admiration, or sympathy and help establish rapport by creating an immediate emotional connection between speakers and listeners. In contrast, news discourse maintains an emotion-neutral, objective style that leaves little room for such expressions. As a result, the presence or absence of interjections provides a clear indication of the emotional stance and communicative distance characteristic of each genre.

- Interjections used in the news (0): None
- Interjections used in comedy talk shows (20 examples): to signal hesitation and uncertainty (뭐, 어, 자, 글쎄, 저); to express sympathy (아이고, 에그, 참, 에고, 에구, 아유, 아이구); for displeasure and frustration (흥, 에이, 아차, 에라); to show a recognition or agreement (그래, 아하)

Vocative Particles

Vocative particles also appeared only in the comedy talk show corpus and were completely absent in news. These particles, such as 야 and 아, do not carry specific lexical meaning; rather they are used to directly address a person or audience to draw attention, serving a conversational and interactive function. The word 동무 (comrade) is a lexical item specific to NK usage and serves as a vocative form of address. Their presence reflects the dialogic and participatory structure of NK comedy talk shows, where emotional exchange and interpersonal connection play a central role. In contrast, NK news delivers information in a one-way, non-interactive format directed at a mass audience, leaving no place for vocative markers. Thus, the appearance of vocative particles exclusively in comedy talk show highlights the oral, interpersonal, and audience-engaging nature of spoken discourse, distinguishing it from the formal, impersonal style of news reporting.

Limitations

While conducting the morphological analysis, some NK endings were manually adjusted to their SK equivalents because the morphological analyzer used in this study was originally designed for the SK language and did not recognize certain NK forms. Therefore, this adjustment may have influenced the distributional results to a limited extent. However, a recently upgraded morphological analyzer (https://kcorpus.korean.go.kr/), specifically designed to accommodate NK linguistic features, has been developed by the National Institute of Korean Language (NIKL) in South Korea. Future research employing improved NK morphological analysis tools is expected to achieve greater accuracy and reduce the need for manual normalization.

CONCLUSION

This study examined the lexical differences between NK formal speech as represented in news broadcasts and informal speech as found in comedy talk shows through a corpus-based POS analysis. A comparison of type and token distributions across major parts of speech revealed clear distinctions between the two genres in NK.

Table 6

Comparison of Token and Type Counts in NK News and Comedy Talk Show Data

Part of Speech	Token	Type
Common Nouns	News > Talk Show	News > Talk Show
Verbs	News < Talk Show	News > Talk Show
Adjectives	News < Talk Show	News > Talk Show
Adverbs	News < Talk Show	News < Talk Show

The findings of this study can be interpreted meaningfully within Bachman and Palmer's (2010) TLU domain framework. As the analysis demonstrates, the lexical profiles of NK news broadcasts and NK comedy talk shows differ systematically in POS distribution, type-token ratios, and lexical overlap. These differences are not merely stylistic; rather, they reflect distinct communicative purposes, discourse conventions, and interactional conditions. In TLU terms, the two genres represent separate domains characterized by different task demands, participant roles, and pragmatic expectations. The higher proportion of common nouns in news discourse corresponds to its expository, information-dense function within a formal setting, while the greater frequency of verbs, adjectives, adverbs, and repetition in comedy talk shows reflects the interactive, affective, and situationally embedded nature of informal spoken communication. Thus, the observed lexical variation aligns with the TLU principle that linguistic features are shaped by the contextual parameters of language use.

The lexical overlap analysis further reinforced the differences between the two genres as shown in Table 7.

Table 7

Lexical Overlap Percentage of the Two Genres by POS

Part of speech	Common Noun	Verb	Adjective	Adverb
Number of Overlapping Items	254	154	35	60
Lexical Overlap Ratio (%)	27.1%	36.7%	25.7%	28.4%

From a domain representativeness perspective, the limited lexical overlap—approximately 30% across major parts of speech—indicates that reliance on a single register provides access to only a partial segment of the broader NK lexicon. If instructional materials are drawn predominantly from formal news discourse, learners are effectively trained within a restricted TLU domain. While this focus may support proficiency in formal or informational contexts, it does not necessarily equip learners to interpret language in informal, interactional settings. Bachman and Palmer (2010) emphasize that meaningful language development depends on alignment between instructional input and the range of real-world communicative situations learners must handle. The present findings suggest that current NK instruction may underrepresent the lexical resources associated with conversational and emotionally expressive discourse, thereby limiting learners' register flexibility. The implication is not that exposure to informal media directly determines performance on assessments such as Defense Language Proficiency Test 5 (DLPT5), but rather that reliance on formal materials alone provides access to only a limited segment of

the broader NK lexicon. When instruction concentrates primarily on formal registers, learners may develop strong comprehension within that domain while remaining less prepared for communicative contexts that depend on different high-frequency vocabulary.

Although this study clearly identified lexical differences between the news and comedy talk show corpora, the disparity in raw corpus length presents a methodological limitation that may have influenced the overlap analysis. Because each corpus varies in size, the absolute number of shared lexical items must be interpreted with caution. While the POS percentage distributions remain stable, as each reflects proportions calculated within its respective corpus, the measures of lexical overlap are based on raw counts and are therefore inherently more sensitive to differences in corpus length. To mitigate this effect, overlap indices were interpreted in relative rather than absolute terms, and corpus size was taken into account when comparing patterns across registers. Accordingly, differences in lexical overlap should be understood as indicative of distributional tendencies rather than direct magnitude comparisons. Future research would benefit from balanced corpora to allow more precise comparisons.

Future research should extend this line of inquiry by examining the relationship between expanded TLU domain coverage and measurable gains in language proficiency. In particular, studies could investigate whether systematic exposure to vocabulary drawn from both formal and informal NK domains correlates with improved comprehension, production, and register flexibility. Such research would help determine the extent to which broader lexical exposure translates into functional communicative competence, thereby providing further empirical grounding for corpus-informed, TLU-aligned curriculum development.

Authors

Mi Hye Lee, Ph.D., is a Korean language instructor at DLIFLC with extensive language teaching experience in both South Korea and the United States. Her professional interests include curriculum development, instructional materials, and assessment practices in language education.

Myoyoung Kim, Ph.D., currently serves at DLIFLC as an educational measurement specialist and institutional evaluator. Her career in language education spans teaching, curriculum development, teacher training, and assessment. With a background in linguistics, she has a strong interest in data management and analysis and enjoys using data-driven approaches in her work.

REFERENCES

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.

Biber, D. (2001). On the complexity of discourse complexity: a multi-dimensional analysis. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 177–202). Longman.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25(3)*, 371–405.

Baron, N.S. (2013). *Words Onscreen: The Fate of Reading in a Digital World*. Oxford University Press.

Cheong, Y. (2021). A study on the concept and classification criteria of North Korean. *Baedalmal,* 69, 213–241. https://doi.org/10.52636/KL.69.7

Curry, N., & Mark, G. (2024). Using corpus linguistics in materials development and teacher education. *Second Language Teacher Education,2(2)*, 187–208. doi:10.1558/slte.25727

Eum, I., Seo, H., & Kwon, S. (2021). Development of North Korean vocabulary textbooks for South Koreans*. Journal of CheongRam Korean Language Education,* 80, 413–459. https://doi.org/10.26589/jockle..80.202103.413

Kim, J. (2015). A critical review on education for understanding of North Korean language in Korean education*. Korean Language Education Research*, 58, 143–170.

Kim, N. (2024). A study on improving Korean language education for unification: Focusing on teaching North Korean as a regional dialect (Thesis). Yonsei University.

Kim, Y. J., & Biber, D. (1994). A Corpus-Based Analysis of Register. *Sociolinguistic perspectives on register*, 157.

Latham, E. (2025). Using corpus-driven TLU domain analysis to increase the authenticity of listening passages. *International Journal of English for Academic Purposes: Research and Practice, vol 5*, n 2. https://doi.org/10.3828/ijeap.2025.8

Li, D., Noordin, N., Ismail, L., & Cao, D. (2025). A Systematic Review of Corpus-Based Instruction in EFL Classroom. *Heliyon 11(2)*, 10.1016/j.heliyon.2025.e42016

Matang, P., & Narathakoon, A. (2025). A Corpus-Based Study of Lexical Bundles of Keywords Found in Online News Articles, *Thai TESOL Journal vol. 38, issue 1*.

Moser, J. (2020) Evaluating Arabic textbooks: A corpus-based lexical frequency study, *International Journal of Applied Linguistics, vol.31,* issue 2. https://doi.org/10.1111/ijal.12321

Oh, J. (2015). *A study on education methods of North Korean vocabularies* (Thesis). Graduate School of Education, Kyung Hee University.

Reppen, R. (2001). Register variation in student and adult speech and writing. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 187–199). Longman.

Tannen, D. (1985). Relative focus on involvement in oral and written discourse. In D. R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning: The nature and consequences of reading and writing* (pp. 124–147). Cambridge University Press.

Yeo, E. (2021). *Several issues on the morphological analysis in North Korean corpus* (Thesis). Jeonbuk University.

APPENDIX A

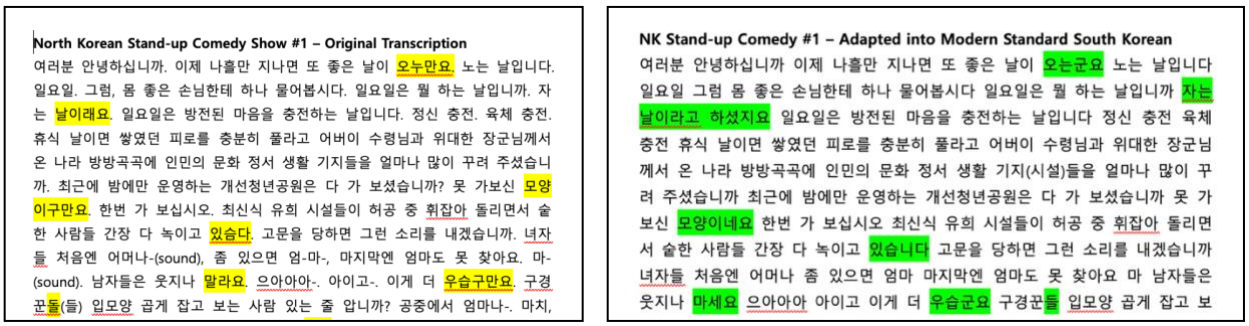
Video Information

	Topic	Genre	Length (min:sec)	Production Year	Link
1	일요일 (Sunday)	Solo comedy Talk show	10:05	Unknown	https://www.youtube.com/watch?v=2-GSHwx6HY0
2	미신을 믿지 말자 (Let's not believe in superstitions)	Two-person comic play	10:40	Unknown	https://www.youtube.com/watch?v=dasJfVO8krE
3	단호한 결심 (A firm resolution)	NK *Jaedam* (comic performance)	15:27	1986	https://www.youtube.com/watch?v=WaY5aCoxCgM&list=PLMHnawEno0Eqk40HcGQjcO2FxLwW_DWut&index=1
4	인심 좋은 작업반장 (A kind-hearted work team leader)	Comedy sketch	10:54	1987	https://www.youtube.com/watch?v=DTTjeCRBWBk
5	정 (Warm Affection)	Comedy sketch	12:00* (20:27)	2016	https://www.youtube.com/watch?v=x02wfE_RQUw&list=PLMHnawEno0Eqk40HcGQjcO2FxLwW_DWut&index=4
6	오가는 정 (Affection Given and Returned)	Comedy sketch	20:00* (29:35)	Unknown	https://www.youtube.com/watch?v=Vy5EpgW2IeI (*currently unavailable)
7	떠나는 마음 보내는 마음 (A Heart Leaving, A Heart Letting Go)	Comedy sketch	20:00* (30:47)	2005	https://www.youtube.com/watch?v=_D4l4ZQIwck

*For videos #5~#7, only the first 12 or 20 minutes (out of the total length shown in parentheses for each video) were transcribed to avoid lexical concentration on a single topic.

APPENDIX B

Example of transcript verification



North Korean Stand-up Comedy Show #1 – Original Transcription

여러분 안녕하십니까. 이제 나흘만 지나면 또 좋은 날이 오누만요. 노는 날입니다. 일요일. 그럼, 몸 좋은 손님한테 하나 물어봅시다. 일요일은 뭘 하는 날입니까. 자는 날이래요. 일요일은 방전된 마음을 충전하는 날입니다. 정신 충전. 육체 충전. 휴식 날이면 쌓였던 피로를 충분히 풀라고 어버이 수령님과 위대한 장군님께서 온 나라 방방곡곡에 인민의 문화 정서 생활 기지들을 얼마나 많이 꾸려 주셨습니까. 최근에 밤에만 운영하는 개선청년공원은 다 가 보셨습니까? 못 가보신 모양이구만요. 한번 가 보십시오. 최신식 유희 시설들이 허공 중 휘잡아 돌리면서 술한 사람들 간장 다 녹이고 있습다. 고문을 당하면 그런 소리를 내겠습니까. 녀자들 처음엔 어머나-(sound), 좀 있으면 엄-마-, 마지막엔 엄마도 못 찾아요. 마-(sound). 남자들은 웃지나 말라요. 으아아아-. 아이고-. 이게 더 우습구만요. 구경꾼들(들) 입모양 곱게 잡고 보는 사람 있는 줄 압니까? 공중에서 엄마나-. 마치,

NK Stand-up Comedy #1 – Adapted into Modern Standard South Korean

여러분 안녕하십니까 이제 나흘만 지나면 또 좋은 날이 오는군요 노는 날입니다 일요일 그럼 몸 좋은 손님한테 하나 물어봅시다 일요일은 뭘 하는 날입니까 자는 날이라고 하셨지요 일요일은 방전된 마음을 충전하는 날입니다 정신 충전 육체 충전 휴식 날이면 쌓였던 피로를 충분히 풀라고 어버이 수령님과 위대한 장군님께서 온 나라 방방곡곡에 인민의 문화 정서 생활 기지(시설)들을 얼마나 많이 꾸려 주셨습니까 최근에 밤에만 운영하는 개선청년공원은 다 가 보셨습니까 못 가보신 모양이네요 한번 가 보십시오 최신식 유희 시설들이 허공 중 휘잡아 돌리면서 술한 사람들 간장 다 녹이고 있습니다 고문을 당하면 그런 소리를 내겠습니까 녀자들 처음엔 어머나 좀 있으면 엄마 마지막엔 엄마도 못 찾아요 마 남자들은 웃지나 마세요 으아아아 아이고 이게 더 우습군요 구경꾼들 입모양 곱게 잡고 보

*The left side shows the output before manual verification, and the right side presents the modified version after verification. Different colors were used to indicate how spellings of the same words were changed. On the right side, green highlight instances in which the same lexical item appears with modified spellings; the corresponding words on the left side were originally transcribed. The colors do not signal differences in meaning, but rather variation in transcription certainty and orthographic representation of the same word. This procedure was applied to NK news broadcasts to minimize the error rate.

APPENDIX C

POS Distribution and Full Results for the News and Comedy Corpora

C-1. Overview

Part of Speech	Tag Set	News	Comedy Talk Shows
Common Noun	NNG	3,821 (26.51%)	2,790 (17.92%)
Verb	VV	1,586 (11%)	1,940 (12.46%)
Adjective	VA	520 (3.61%)	575 (3.69%)
General Adverb	MAG	440 (3.05%)	907 (5.83%)
Others	EF, EC, IC, NNP, NNB, NP, MM, etc.	463 (18.15%)	985 (29.8%)
Total		14,415	15,567

C-2. Comparison of POS count: Tokens

	Part of Speech	Tag Set	News	Comedy Talk Shows
1	Common noun	NNG	3,821 (26.51%)	2,790 (17.92%)
2	Proper noun	NNP	247 (1.71%)	115 (0.74%)
3	Adnominal particle	JKG	499 (3.46%)	128 (0.82%)
4	Adnominal modifier	MM	214 (1.48%)	329 (2.11%)
5	Adnominal ending	ETM	1,026 (7.12%)	600 (3.85%)
6	Positive copula	VCP	144 (1.00%)	352 (2.26%)
7	Pronoun	NP	282 (1.96%)	733 (4.71%)
8	Verb	VV	1,586 (11.00%)	1,940 (12.46%)
9	Verb-derivational suffix	XSV	14 (0.10%)	18 (0.12%)
10	Noun-derivational suffix	XSN	491 (3.41%)	259 (1.66%)
11	Nominalizing ending	ETN	74 (0.51%)	34 (0.22%)
12	Objective particle	JKO	695 (4.82%)	414 (2.66%)
13	Complement particle	JCM	29 (0.20%)	38 (0.24%)
14	Auxiliary verb/adjective	VX	377 (2.62%)	302 (1.94%)
15	Auxiliary particle	JX	394 (2.73%)	599 (3.85%)
16	Adverbial particle	JKB	664 (4.61%)	393 (2.52%)
17	Negative copula	VCN	12 (0.08%)	44 (0.28%)
18	Prefinal ending	EP	360 (2.50%)	490 (3.15%)
19	Numeral	NR	36 (0.25%)	88 (0.50%)
20	Number	SN	93 (0.65%)	8 (0.05%)
21	Connective ending	EC	1,240 (8.60%)	1,383 (8.88%)
22	Dependent noun	NNB	293 (2.03%)	408 (2.62%)
23	Quotative particle	JKQ	3 (0.02%)	5 (0.03%)
24	Adverb	MAG	440 (3.05%)	907 (5.83%)
25	Conjunctive adverb	MAJ	20 (0.14%)	83 (0.53%)
26	Conjunctive particle	JC	159 (1.10%)	55 (0.35%)
27	Final ending	EF	317 (2.20%)	1,231 (7.91%)
28	Subject particle	JKS	356 (2.47%)	492 (3.16%)
29	Nominal prefix	XPN	8 (0.06%)	10 (0.06%)
30	Adjective	VA	520 (3.61%)	575 (3.69%)
31	Adj.-derivational suffix	XSA	1 (0.01%)	1 (0.01%)
32	Interjection	IC	None	746 (4.79%)
33	Vocative particle	JKV	None	7 (0.04%)
	Total		14,415	15,567

C-3. Comparison of POS counts: Types

	Part of Speech	Tag Set	News	Comedy Talk Shows
1	Common noun	NNG	1,308 (51.31%)	937 (38.45%)
2	Proper noun	NNP	98 (3.84%)	71 (2.91%)
3	Adnominal particle	JKG	1 (0.04%)	1 (0.04%)
4	Adnominal modifier	MM	27 (1.06%)	37 (1.52%)
5	Adnominal ending	ETM	10 (0.39%)	17 (0.70%)
6	Positive copula	VCP	1 (0.04%)	1 (0.04%)
7	Pronoun	NP	18 (0.71%)	44 (1.81%)
8	Verb	VV	487 (19.11%)	428 (17.56%)
9	Verb-derivational suffix	XSV	4 (0.16%)	3 (0.12%)
10	Noun-derivational suffix	XSN	31 (1.22%)	20 (0.82%)
11	Nominalizing ending	ETN	3 (0.12%)	2 (0.08%)
12	Objective particle	JKO	3 (0.12%)	3 (0.12%)
13	Complement particle	JCM	2 (0.08%)	2 (0.08%)
14	Auxiliary verb/adjective	VX	20 (0.78%)	22 (0.90%)
15	Auxiliary particle	JX	17 (0.67%)	28 (1.15%)
16	Adverbial particle	JKB	18 (0.71%)	16 (0.66%)
17	Negative copula	VCN	1 (0.04%)	1 (0.04%)
18	Prefinal ending	EP	7 (0.27%)	8 (0.33%)
19	Numeral	NR	15 (0.59%)	22 (0.90%)
20	Number	SN	39 (1.53%)	7 (0.29%)
21	Connective ending	EC	67 (2.63%)	107 (4.39%)
22	Dependent noun	NNB	56 (2.20%)	62 (2.54%)
23	Quotative particle	JKQ	1 (0.04%)	2 (0.08%)
24	Adverb	MAG	128 (5.02%)	211 (8.66%)
25	Conjunctive adverb	MAJ	4 (0.16%)	9 (0.37%)
26	Conjunctive particle	JC	5 (0.20%)	8 (0.33%)
27	Final ending	EF	5 (0.20%)	116 (4.76%)
28	Subject particle	JKS	4 (0.16%)	3 (0.12%)
29	Nominal prefix	XPN	3 (0.12%)	7 (0.29%)
30	Adjective	VA	165 (6.47%)	136 (5.58%)
31	Adj.-derivational suffix	XSA	1 (0.04%)	1 (0.04%)
32	Interjection	IC	None	103 (4.23%)
33	Vocative particle	JKV	None	2 (0.08%)
	Total		2,549	2,437

Applied Language Learning

Research

Motivational Engagement among Persian Farsi Language Learners Across Two Military Institutions

Olivia Holloway
United States Military Academy, West Point

Siyi Gao
Defense Language Institute Foreign Language Center, Monterey, California

Ra'ed Qasem
Defense Language Institute Foreign Language Center, Monterey, California

Zachary Miller
United States Military Academy, West Point

This mixed-methods study investigated motivational engagement among beginner-level Persian Farsi (PF) learners at two U.S. military institutions (N = 59), where foreign language (FL) assignments are mandatory and may not align with learners' interests. Participants included 32 cadets at the United States Military Academy (USMA) and 27 service members enrolled in a 48-week PF program at the Defense Language Institute Foreign Language Center (DLIFLC). The study examined how individual learner characteristics affect students' motivational engagement, operationalized as interaction and interest, in PF courses in the two institutions. The study also explored which factors predict students' motivational engagement in studying PF and which factors sustain motivational engagement over time. Data were collected using three standardized instruments, the Short Grit Scale, the L2 (Second Language) Grit Scale, and the Language-Specific Anxiety Scale, along with guided reflective journals administered over one semester, yielding both quantitative indicators and qualitative narrative data. Results showed that L2 Grit emerged as a significant predictor of motivational engagement and was the strongest predictor of learners' initial and sustained interest in PF and their willingness to participate in the classroom. No significant differences were observed for general Grit or FL anxiety. Both L2 Grit and initial engagement significantly predicted sustained engagement, underscoring the

importance of early involvement and perseverance. Initial emotional states, however, did not forecast learner interest or classroom interaction.

Keywords: *Foreign language (FL) learning, Military Education and Training, Motivation, Grit, L2 (Second Language) Grit*

INTRODUCTION

Foreign language (FL) learning in military environments presents unique challenges that differ markedly from traditional academic settings. Programs operate under compressed timelines and link language proficiency directly to mission readiness, creating high-stakes learning conditions that intensify motivational demands (Jodai et al., 2013; Miller & Crowther, 2024). Within this context, the motivational profiles of learners can vary depending on institutional structure and the degree of learner agency. At the Defense Language Institute Foreign Language Center (DLIFLC), service members typically receive mandatory, full-time language assignments essential to their military roles, whereas cadets at the United States Military Academy (USMA) study foreign languages as part of a broader academic curriculum with comparatively greater choice and wider educational goals. Because these institutions differ in learner autonomy, program intensity, and perceived career relevance, examining motivation across DLIFLC and USMA students provides insight into how distinct military learning environments shape engagement, perseverance, and emotional experience in FL learning.

Although research on motivation in civilian second language acquisition (SLA) is extensive (Dörnyei, 2005; Dörnyei & Ryan, 2015), motivation in military language programs remains comparatively understudied. Because these programs operate under assignment-driven enrollment, accelerated timelines, and mission-critical expectations, understanding motivation in this context requires examining how individual learner characteristics function within such demanding environments. Accordingly, the present study investigates how individual differences, specifically L2 Grit, general grit, and foreign language anxiety, interact with institutional context to shape learners' reported interest and interaction over time. Qualitative indicators of motivational orientations and emotional experiences further illuminate how engagement develops in these high-pressure, fast-paced settings. By situating these factors within two distinct military higher-education institutions, the study offers insights that may help strengthen approaches to sustaining motivation in military foreign language programs.

DLIFLC provides intensive and immersive language instruction aligned with operational military requirements. Learners (enlisted and commissioned officers) are assigned to specific languages based on operational needs, not personal preference. Students attend 6 hours of daily instruction through the Basic Course, which ranges from 36 to 64 weeks. Students studying Persian Farsi complete 48 weeks of instruction and take the Defense Language Proficiency Test (DLPT) for listening and reading, and the Oral Proficiency Interview. Graduation requires at least Interagency Language Roundtable (ILR) levels 2 in reading and listening and 1+ in speaking.

USMA integrates military, physical, and leadership training within a four-year undergraduate degree, aiming to develop commissioned officers equipped for strategic and culturally aware leadership. Cadets must take at least two semesters of an FL regardless of prior coursework, proficiency level, or chosen major. They are assigned to one of eight foreign languages based on the number of instructors, each cadet's language aptitude, and other factors. To graduate as commissioned officers, they must earn passing grades in all their courses across three pillars (military, physical education, and academic courses, which include the two FL courses).

LITERATURE REVIEW

Motivation in SLA is shaped by learners' goals and instructional and sociocultural contexts. Two theories primarily inform this study: Dörnyei's L2 Motivational Self System and Deci and Ryan's Self-Determination Theory. The L2 Motivational Self System conceptualizes motivation around future-oriented self-guides, including the Ideal L2 Self (the learner's envisioned identity as a competent language user), and the Ought-to L2 Self (with duties, obligations, and responsibilities to meet external expectations; Dörnyei, 2005; Dörnyei & Ushioda, 2021). When the Ideal L2 Self is vivid and attainable, learners are likely to invest sustained effort in their language development. In contrast, the Ought-to L2 Self fosters anxiety and cautious learning, leading to lower, less stable motivation (Dörnyei, 2005; Dörnyei & Ushioda, 2021).

Self-Determination Theory (SDT) provides a complementary account of how the motivation categories vary according to fulfillment of needs for autonomy, competence, and relatedness (Deci & Ryan, 1985; Ryan & Deci, 2000, 2017, 2020). SDT positions motivation on a continuum ranging from controlled forms, such as external and introjected regulation, to autonomous forms, including identified, integrated, and intrinsic regulation. As learners internalize the value of language learning and shift toward more autonomous forms of regulation, they tend to exhibit deeper, more persistent engagement (Noels et al., 2000). Considered together, the L2 Motivational Self System and SDT illustrate how future-oriented identity processes and basic psychological needs shape motivated action.

These theoretical insights are particularly relevant in military training environments, which introduce distinctive motivational challenges. Military language learners often study assigned languages under accelerated timelines and stringent proficiency requirements. Such conditions frequently give rise to controlled forms of motivation early on, driven by obligations such as passing assessments or meeting institutional benchmarks. However, as learners develop competence and become more aware of the professional relevance of the language, their motives may shift toward more autonomous forms of regulation (Takahashi & Im, 2020).

Within these environments, individual differences further shape how learners engage with the language. Grit, defined as sustained interest and effort toward long-term goals, has been shown to predict achievement under demanding conditions (Duckworth & Quinn, 2009). L2 Grit reflects perseverance and passion specifically toward learning a foreign language. Research has shown

that L2 Grit predicts persistence, engagement, and achievement in language learning more effectively than general grit (Teimouri et al., 2022; Henry & Liu, 2024). For example, Teimouri and colleagues (2022) found that learners with higher L2 Grit participated more actively in classroom discussions, demonstrated deeper cognitive engagement, sustained interest throughout instruction, and achieved higher FL outcomes—even in rigorous settings. In contrast, foreign language anxiety can undermine confidence and reduce willingness to participate, particularly in evaluative or high-pressure military classrooms (Sudina & Plonsky, 2021).

Motivational types offer additional insight into how learners' goals evolve in military language learning environments where students are assigned to learn FLs. Students often begin with external motives but may internalize these motives as they gain competence and recognize the relevance of the language to their future responsibilities (Ryan & Deci, 2020). Although intrinsic motivation is less common in high-stakes learning contexts characterized by strong external demands, autonomous extrinsic motives can still promote strong engagement (Noels et al., 2000). According to control-value theory, positive activating emotions such as enjoyment enhance interest and participation, whereas negative activating emotions such as anxiety may inhibit engagement unless effectively regulated (Pekrun, 1992; Pekrun & Stephens, 2012).

Finally, institutional context further influences learners' motivational experiences. DLIFLC offers immersive, proficiency-driven instruction, while USMA incorporates language education within a broader undergraduate curriculum. Both institutions assign languages based on operational needs. The structure and intensity of language programs differ in ways that may shape motivation (see Csizér, 2020; Miller & Crowther, 2020).

METHOD

This study employed a mixed-methods design to examine motivational engagement, operationalized as interest and interaction, among 59 U.S. military learners studying PF at USMA and DLIFLC. Quantitative data were collected using validated measures of general grit, L2 Grit, and foreign language anxiety (LAS), along with self-reported indicators of interest and interaction, collected at three points over the first semester. Qualitative data were gathered through three guided reflections to explore learners' motivation categories, grounded in Self-Determination Theory, and emotions. Quantitative data were analyzed using a combination of descriptive statistics, *t*-tests, ANOVAs, regression analyses, and mixed-effects modeling. Thematic coding with established reliability procedures was used for analyzing the qualitative data. This comprehensive approach enabled the study to investigate the interplay between learner characteristics and motivational engagement, identify its predictors, and what sustains it over time in intensive military language learning contexts. The study explored the following three research questions:

1. How do individual learner characteristics (general grit, L2 Grit, and foreign language anxiety) compare between institutions and affect students' motivational engagement (interest and interaction) in PF courses?

2. Which factors predict students' motivational engagement in studying PF?
3. Which factors sustain motivational engagement over time among PF students compared at each institution?

Participants

Fifty-nine students of PF (*N* = 59) from the US military participated in this study. All participants were (a) between the ages of 18 and 28 (*M* = 20.6 years, *SD* = 2.59), (b) native English speakers, (c) not heritage speakers of PF, and (d) enrolled in their 1st semester of PF. Out of the 59 total participants, 43 were male and 16 were female. At USMA, ages ranged from 18 to 23 (*M*=19.31), while at DLIFLC, ages ranged from 18 to 28 (*M*=22.15).

Thirty-two (*n* = 32) were enrolled at USMA and were starting their first semester of PF at the beginner level. Among all students, 21 had not selected PF as their first-choice language to study, and none of them were majoring in PF. All 32 USMA participants needed to pass a year (two semesters) of language study as a requirement to finish their 4-year degrees. USMA students' language class is one academic course out of 5 minimum courses taken per semester. The other twenty-seven (*n* = 27) participants studied at the intensive 48-week course at DLIFLC and were beginning their first semester (18 weeks) of PF education. All DLIFLC students were on track to become military linguists in their respective service and were not granted a choice in which language they study. However, in our initial intake, 24 of the 27 at DLIFLC had preferred to study a language other than PF. DLIFLC students must reach graduation goals in under a year and focus solely on language proficiency.

Materials

Three scales were used in this study: Grit-S Scale, Language-Specific Anxiety Scale, and Second Language Grit Scale. At the semester's start, students completed an intake questionnaire gathering demographic information, language-learning choices, satisfaction with PF program enrollment, and responses for the Grit-S and Language-Specific Anxiety Scales. Additionally, students completed guided reflections at three points in the first semester.

Grit-S Scale

The Short Grit Scale (Grit-S; Duckworth & Quinn, 2009) measures trait-level perseverance and passion to accomplish long-term goals. Several studies have examined Grit-S in FL learning (Keegan, 2017; Khajavy et al., 2021; Oxford & Khajavy, 2021). The Grit-S scale has eight items, with four focused on the subfactor *Consistency of Interest* (measuring passion) and four centered on the subfactor *Perseverance of Effort* (measuring perseverance). The items were rated on a 5-point Likert scale. Individuals with high overall Grit-S scores are considered "gritty" and likely to achieve successful outcomes in challenging environments.

Language-Specific Anxiety Scale

The Language-Specific Anxiety Scale (LAS; Sudina & Plonsky, 2021) gauges FL learners' level of anxiety within the classroom. The scale consists of four items measured on a 5-point Likert scale. Participants with higher scores exhibit greater levels of anxiety while learning an FL. The LAS was used due to its simplicity and brevity.

Second Language Grit Scale

The Second Language Grit Scale (L2 Grit; Teimouri et al., 2022) evaluates individuals' grittiness in L2 settings. The scale measures two related subcomponents using nine items via a 5-point Likert scale: *Consistency of Interest* (L2 learning appeal) and *Perseverance of Effort* (persistence in achieving L2 goals).

Guided Reflections (GRs)

We utilized guided reflections (GRs) combining Likert scale measures to capture levels of interest and interaction (as proxies for motivational engagement) and narrative framing to encourage self-reflective writing and provide insight into learners' emotions, definitions of success, and language-learning goals (Barkhuizen, 2008; Barkhuizen & Wette, 2008; Benson, 2014). There were two Likert scale questions, one for interest and one for interaction, with 6 options ranging from "Very Low" to "Very High." The GR narrative prompts were: (a) *When I think of attending this language class, I feel the following emotion(s)* ___; (b) *For me, success in this language class means* ___; and (c) *My current goals for learning this language are* ___.

Data Collection

This study was conducted with approval from the Institutional Review Boards at both institutions. Students eligible to participate were invited via email, and the informed consent materials emphasized that participation was voluntary. Participants were informed that their responses would remain anonymous, that they could withdraw from the study at any time without penalty, and that their decision to participate or decline participation would not be disclosed to their instructors. The initial intake questionnaire collected demographic information, whether the participant had chosen PF and was satisfied with enrollment in the language, a General Grit score, and a Language-Specific Anxiety score.

GRs were collected at three more points through the participants' 18-week first semester of language study at the DLIFLC and USMA. Written GRs were chosen over interviews to optimize respondents' autonomy and sense of privacy. One month into the participants' semester, we collected GR1, followed by GR2 one month later, and GR3 (along with the L2 Grit survey) one month later, which nearly coincided with the end of the semester. For each GR, we allowed participants one week to complete and return the form back to us. All data were collected using online survey sites (Qualtrics at USMA and Microsoft Forms at DLIFLC). USMA participants were awarded 10 points of extra credit (equivalent to 1%) for full participation in the study; participant

names were reported to instructors in the final week of the semester for this purpose. Of 64 USMA students eligible and invited to participate, 32 of them responded to all surveys. At DLIFLC, 27 of the 62 eligible and invited students volunteered to complete all surveys.

Data Analysis

After data were de-identified by each institution, they were shared securely. One researcher from each institution coded qualitative responses on Academic Emotions and FL Motivation. Inter-rater reliability was 82% for Academic Emotions and FL Motivation.

Academic Emotions in the FL setting drew from the narrative prompt "*When I think of attending this language class, I feel the following emotion(s)* __" from each GR. Utilizing studies on Academic Emotions (Pekrun, 1992; Pekrun et al., 2002; Pekrun & Stephens, 2012), we categorized responses into four categories: Positive, Activating (i.e. enjoyment, pride); Positive, Deactivating (i.e. relief); Negative, Activating (i.e. anxiety); and Negative, Deactivating (i.e. hopelessness). When coding narratives, we focused on the dominant emotion communicated in the responses, and disregarded details about physical state (i.e. "tired"). For responses that were ambivalent or featured multiple meanings (such as both positive and negative or both activating and deactivating), both coders discussed to select a dominant category. Remaining ambivalent responses were resolved with a tie-breaking co-author. To simplify data coding and analysis, we grouped emotions into emotional valences: Positive and Negative.

To categorize motivation, we coded responses to item "*My current goals for learning this language are* ___" from each GR. Drawing from SDT in the learning context (Deci & Ryan, 1985; Noels et al., 2000; Ryan & Deci, 2000, 2017, 2020), we coded along six categories from most autonomous to most controlled. These are Intrinsic Motivation (activities generating inherent satisfaction), Integrated Regulation (finding an activity to be "congruent" with one's "core interests and values"), Identified Regulation ("a relatively high degree of volition" due to personal endorsement of an activity), Introjected Regulation ("partially internalized" extrinsic motivation), External Regulation ("driven by externally imposed rewards and punishments"), and Amotivation ("lacking intentionality;" Ryan & Deci, 2020 n.p.). To simplify data coding and analysis, we grouped the aforementioned spectrum of motivations into two macro-categories for motivation: Autonomous (including Intrinsic, Integrated Regulation, and Identified Regulation) and Controlled (Introjected Regulation, External Regulation, and Amotivation).

The purpose of our analyses was to examine how certain learner characteristics and institutional factors relate to motivational engagement in PF courses and how these factors sustain motivational engagement across the semester. Motivational engagement was operationalized using two self-reported indicators collected throughout the semester: Interest and Interaction. We also chose the term *semester* to represent an academic session, although the timeframes at both institutions are admittedly different. For USMA, a semester of PF is 17 weeks comprising 80 lessons of 55 minutes and approximately 70 minutes of associated homework. At DLIFLC, the full PF program spans 48 weeks and is divided into three semesters: approximately 18 weeks for the first semester, 20 weeks for the second, and 10 weeks for the third. Students attend six class

sessions per day (50 minutes each), five days per week, with approximately two hours of homework assigned on weekdays. At both institutions, students requiring additional academic support may attend extra assistance hours.

RQ1 compared individual learner characteristics (Grit, L2 Grit, and FL Anxiety) between institutions and how they affected students' motivational engagement in PF courses. To establish baseline patterns in the dataset, we calculated descriptive statistics for all variables. Inspection of the distributions indicated that variables were approximately normally distributed within each institutional group, with no extreme skewness or outliers. As such, we used independent samples *t*-tests to compare students from USMA and DLIFLC on their Grit, FL Anxiety, L2 Grit, and average Interest and Interaction scores. To examine how learner characteristics related to engagement across institutions, a series of 2×3 ANOVAs were conducted. Institution (USMA versus DLIFLC) served as one factor, while Grit, FL Anxiety, and L2 Grit were divided into tertiles (low, middle, high) to allow comparison across levels of each trait. These models accounted for Interest and Interaction, allowing us to test the independent effects of learner characteristics and potential interactions with institutional context. Effect sizes for both analyses were measured and reported. To confirm our ANOVA results, we used multiple regression analyses with Grit, FL Anxiety, and L2 Grit predicting Interest and Interaction. Diagnostic checks indicated that our dataset sufficiently met the assumptions for regression testing. Regression allowed us to examine the relative contribution of each learner characteristic to motivational engagement while treating predictors as continuous variables.

RQ2 examined which factors predict students' motivational engagement in studying PF. To identify predictors of motivational engagement across the semester, we conducted multiple regression analyses using early-semester variables as predictors. These included Grit, FL Anxiety, L2 Grit, beginning Interest, beginning Interaction, institutional affiliation, and whether PF was each student's first-choice language assignment. Two models were estimated with end-of-semester Interest and Interaction as outcome variables. To examine motivational development from a theoretical perspective, we examined students' written reflections. Because SDT motivation was measured as a categorical self-classification rather than a continuous scale, categories were collapsed into the theoretically established Autonomous versus Controlled distinction for inferential analyses. We conducted a similar analysis with valence scores, collapsing initial emotional states into either a Positive or Negative category. We then coded students with a positive interest change as "increasers" and those with zero or negative change as "non-increasers." Because the outcome variable represented whether students' Interest increased during the semester, we used binary logistic regression to determine whether initial motivational regulation predicted change trajectories. A parallel analysis examined whether early emotional valence predicted changes in Interest.

RQ3 investigated which factors sustained motivational engagement among students of PF over time. To examine engagement across the semester, we used mixed-effects models with repeated measures of Interest and Interaction collected at three timepoints (beginning, middle, and end of the semester). Mixed-effects modeling was appropriate due to multiple observations of each student. These models allowed us to examine whether motivational engagement changed across

the semester and whether learner characteristics or institutional context influenced those trajectories. To complement these quantitative measures, we also analyzed students' written reflections. Like in RQ2, we collapsed SDT categories into Autonomous or Controlled and emotional responses into Positive or Negative. Because these variables were categorical, we utilized chi-square tests of independence. This analysis allowed us to examine whether the qualitative nature of students' motivation and affect shifted across the semester and whether those patterns differed by institutional context, providing additional insight into how motivational engagement was sustained over time.

RESULTS

Research Question 1

RQ1 compared individual learner characteristics, Grit, L2 Grit, and FL Anxiety, between institutions and how they affected students' motivational engagement in PF courses. We first used descriptive statistics and independent samples *t*-tests to summarize key characteristics and compare our population sets. No significant differences were observed for general Grit or FL Anxiety. However, DLIFLC students scored significantly higher on L2 Grit, Interest, and especially Interaction. The latter effect was large, highlighting a robust institutional difference in self-reported levels of learner interaction. Table 1 highlights our results.

Table 1

Comparison of Students at USMA versus DLIFLC

	Variable	USMA (*N* = 32) *M* (*SD*)	DLIFLC (*N* = 27) *M* (*SD*)	*t*	*p*	*d*
Individual learner characteristics	Grit	3.62 (0.57)	3.49 (0.66)	0.84	.407	0.22
	FL Anxiety	2.62 (0.81)	2.85 (1.01)	-0.97	.335	-0.26
	L2 Grit	3.37 (0.74)	3.90 (0.69)	-2.84	.006**	-0.74
Motivational engagement	Interest (avg)	4.02 (1.16)	4.68 (1.15)	-2.18	.034*	-0.57
	Interaction (avg)	3.69 (1.00)	4.67 (0.78)	-4.23	<.001***	-1.08

* $p < .05$, ** $p < .01$, *** $p < .001$

To investigate whether personal characteristics interacted with institutional context, we used series of 2×3 ANOVAs. For Interest, only L2 Grit emerged as a significant predictor ($F = 5.58$, $p = .006$, $\eta^2p = .174$). Students high in L2 Grit consistently reported stronger motivation regardless of location. A marginal interaction between L2 Grit and Institution ($F = 2.57$, $p = .086$, $\eta^2p = .089$) suggested that the benefits of high L2 Grit may have been somewhat stronger at USMA, though this did not reach significance. General Grit and FL Anxiety were not significant predictors of Interest. Both institutional contexts and personal characteristics contributed to Interaction.

Students at DLIFLC consistently reported higher interaction across models. L2 Grit again showed a robust effect (F = 8.69, p < .001, η^2p = .247), with high-L2-Grit students reporting greater engagement across institutions. To verify the ANOVA results, we conducted multiple regressions using Grit, FL Anxiety, and L2 Grit scores. These analyses yielded the same conclusions: L2 Grit was the strongest predictor of both Interest (β = .46, p < .001) and Interaction (β = .49, p < .001).

Research Question 2

RQ2 explored which factors predicted students' motivational engagement in studying PF. We first ran two multiple regression analyses to determine which early-semester factors (i.e., Grit, FL Anxiety, L2 Grit, Beginning Interest score, Beginning Interaction score, Institution, and PF first-choice selection) best predicted end-of-semester motivational engagement, measured separately as Interest and Interaction. Significant predictors are found in Table 2.

Table 2

Multiple Regressions Predicting End-Of-Semester Motivational Engagement (Interest and Interaction)

Outcome	Predictor	*β (Std)*	*t*	*p*
Interest	L2 Grit	.46	4.69	<.001***
	Beginning Interest	.63	6.79	<.001***
	Beginning Interaction	-.26	-2.40	.020*
Interaction	L2 Grit	.36	3.38	.001**
	Beginning Interaction	.38	3.01	.004**
	Institution	.24	1.98	.053

* p < .05, ** p < .01, *** p < .001

The regression model predicting end-of-semester interest was significant, F (7, 51) = 14.55, p < .001, with an R^2 of .67. The strongest predictors of interest were L2 Grit and Beginning Interest, indicating that students who started the semester with high language-specific grit and strong interest were most likely to remain motivated by the end. That said, beginning interaction negatively predicted end interest, suggesting that students who engaged heavily early on may have seen a slight decline in interest later. No other variables significantly predicted end-of-semester interest. The regression model predicting end-of-semester interaction was also significant, F (7, 51) = 9.57, p < .001, with an R^2 of .57. Both L2 Grit and Beginning Interaction were significant predictors, indicating that persistence in language learning and early engagement set the stage for continued engagement across the semester. Institutional differences approached significance, with students at DLIFLC tending toward higher interaction overall. All other predictors were nonsignificant.

To examine whether initial motivation type was associated with later increases in interest, we analyzed students' opening SDT regulation categories. A binary logistic regression indicated that beginning-of-semester autonomous motivation was a significant predictor of change trajectory, $B = -1.27$, $p = .029$. Interestingly, the odds ratio showed that students starting the semester with autonomous motivation were *less likely* to become "increasers" in interest later in the semester (OR = 0.28, 95% CI [0.09, 0.88]). In fact, the pattern suggested that students who began with more controlled forms of motivation (e.g., obligation or requirement-based) were more likely to increase their interest over time. We conducted a similar analysis with valence scores and results indicated that early semester emotions did not meaningfully forecast which students' interest would rise over the semester.

Research Question 3

RQ3 investigated which factors sustained motivational engagement among students of PF over time at each institution. Here, we utilized mixed-effects models to examine whether institutional and personal factors predicted changes in Interest and Interaction across each institution's initial academic semester (using beginning-, middle-, and end-of-semester self-ratings). Table 3 provides significant results and summarizes the findings.

Table 3
Mixed-effects models for Interest and Interaction

Outcome	Significant Predictors	Findings
Interest	L2 Grit ($\beta = 0.80$, $p < .001$)***	Higher L2 Grit predicted stronger, stable interest across time.
Interaction	L2 Grit ($\beta = 0.65$, $p < .001$)***	Higher L2 Grit predicted more interaction across time.
	Time x Institution (mid-semester, $\beta = 0.51$, $p = .031$)*	DLIFLC students exhibited a mid-semester surge in engagement.

* $p < .05$, ** $p < .01$, *** $p < .001$

Learners with higher L2 Grit reported consistently stronger Interest throughout the semester. However, the overall effect of time was not statistically significant, indicating that interest levels remained relatively stable across the semester. In addition, L2 Grit significantly predicted Interaction. A significant Time × Institution interaction also emerged for the mid-semester timepoint. This interaction indicates that DLIFLC learners exhibited a mid-semester increase in interaction levels; scores surged initially and dropped at end of the semester yet remained higher than the initial interaction levels. The interaction among USMA cadets remained relatively stable throughout the semester. Taken together, these analyses indicate that both institutional context and L2 Grit contributed to differences in motivational engagement. Across all models, L2 Grit emerged as the strongest predictor of motivational engagement.

To examine emotional states and motivations throughout the semester, we coded responses from participants GRs using Pekrun's Academic Emotions Framework and SDT. We used Chi-square tests to determine if differences existed at each institution. Emotion-type distributions did not vary significantly between USMA and DLIFLC at the beginning, middle, or end of the semester, suggesting that students at both locations experienced broadly similar emotions across the term. A different pattern emerged for SDT-type distribution by institution. SDT categorizations for students at USMA and DLIFLC did not significantly differ at the beginning or middle of the courses. By the end, however, USMA cadets reported more autonomous motivations, while those at DLIFLC became more mixed with an increase in controlled motivations. This institutional difference at end-of-semester approached statistical significance, $\chi^2 = 3.43$, $p = .064$. Table 4 provides a sample of reflections from students at both locations and in what SDT category we coded their responses.

Table 4

Motivational commentary from students at USMA and DLIFLC

Institution	Part.	Time* (1, 2, 3)	Reflection	SDT Category
USMA	103	2	*"To become fluent enough that I will be able to get around conversationally if I get selected for [semester abroad]."*	Identified Regulation
	114	2	*"To continue to survive and get through this year."*	External Regulation
	127	2	*"Pass the TEE [final exam] and improve on the last WPR [unit exam]."*	External Regulation
	111	3	*"To be able to basically communicate in Persian and have a base of knowledge for future learning in language."*	Integrated Regulation
DLIFLC	215	1	*"I'd like to learn the language and pass the DLPT."*	External Regulation
	221	1	*"Being able to read faster with less difficulty and have an easier time translating sentences."*	Integrated Regulation
	204	3	*"Being able to journal completely in Persian."*	Intrinsic Motivation
	220	3	*"3, 3, 2 on DLPT, the ability to use Farsi in my daily life confidently."*	Identified Regulation

*Time 1 – beginning of semester; Time 2 – middle of semester; Time 3 – end of semester

DISCUSSION

There were three key findings in this study, which examined how individual learner characteristics and institutional context shape motivational engagement among PF learners in two U.S. military language programs, as well as the factors that predict changes in students' motivational engagement over time.

The first finding was that L2 Grit consistently predicted higher levels of interest and interaction. This aligns with prior research demonstrating that domain-specific grit is a powerful predictor of sustained engagement in SLA (Teimouri et al., 2022; Henry & Liu, 2024). Based on the participants, we see that students from DLIFLC have higher L2 Grit. Students at DLIFLC are recruited to learn languages and become military linguists, whereas USMA students are recruited to become military officers and pursue various majors.

The second finding indicates that students who began the semester with high levels of L2 Grit and strong interest were the most likely to remain highly interested by the end of the course. Nonetheless, analysis of the participants revealed that high initial interaction negatively predicted end-of-semester interest. Although students entered the semester with high levels of self-reported interaction (see Table 2), some experienced a decline in interest over time, possibly due to fatigue with the programs' intensity. Both institutions feature unique instructional environments. As described earlier, DLIFLC students study PF in an accelerated, high-stakes setting that requires approximately six hours of daily instruction, five days per week, and approximately two hours of homework each day. USMA students are responsible for at least four academic classes and other physical education and military duties alongside their language class. At the beginning of the semester, many students reported strong enthusiasm and engagement, likely reflecting both the novelty of the language-learning experience and the perceived professional relevance of PF for future military assignments. Such findings align with Dörnyei's (2005) L2 Motivational Self System, particularly the role of the Ideal L2 Self, in which learners initially envision themselves as competent future language users contributing to military missions.

The third finding illustrated shifts in learners' motivational orientations over the semester. Logistic regression analysis showed that some students who initially demonstrated controlled motivation were more likely to develop increased interest later in the semester, whereas those who began with high levels of autonomous motivation tended to maintain relatively stable interest throughout. This non-linear trajectory echoes findings by Takahashi and Im (2020) and Oxford and Khajavy (2021) that extrinsically motivated learners may later internalize goals as their competence and the perceived relevance of the language grow. It also aligns with the internalization process outlined in SDT, whereby learners gradually adopt more self-endorsed forms of regulation as they begin to find personal value in the activity (Ryan & Deci, 2020).

Although the relationship is not causal, students at DLIFLC reported a marked increase in interaction (see Table 3) during the middle of the semester, followed by a slight decline toward

the end, whereas USMA cadets' remained relatively stable across the semester. These patterns have important implications for intensive military language programs. The observed surge in interaction among DLIFLC participants suggests that immersion-style instruction may initially foster high levels of engagement and participation. However, institutions and instructors should also anticipate potential fatigue as the semester progresses.

These findings carry important implications for intensive military language programs. Maintaining motivational engagement over time may require greater instructional variation and increased opportunities for authentic communicative interaction. In addition, structured support systems that reinforce learners' sense of competence and the perceived L2 relevance may help sustain motivation across intensive training cycles. Incorporating communicative speaking tasks, collaborative activities (Csizér, 2020; Miller & Crowther, 2020), reflective goal-setting, and metacognitive strategy instruction (Allen et al., 2008) may further support learners' long-term motivational engagement. Similarly, early identification and targeted support for students experiencing academic difficulties may help prevent declining confidence and reduced interest from becoming entrenched over time.

Limitations

Despite its contributions, this study presents limitations. First, the sample size (N = 59) was small and focused on two military institutions, potentially limiting the generalizability of the findings. Future research should include a broader range of military learning contexts, languages, and learner profiles to enhance external validity.

Second, although the study adopted a mixed-methods approach, the qualitative data came from guided reflections, which may have been influenced by social desirability bias (e.g., seeking to please the researchers) and followed learners for only one term. Longitudinal designs spanning the full curriculum and the transition to operational assignments would clarify whether autonomous motivation and L2 Grit remain stable beyond classroom settings. Future research could incorporate in-depth, anonymous interviews, longitudinal narrative analyses, students' grade data across assessments, and/or retention data to provide a more comprehensive understanding of instructional effectiveness and learners' evolving motivational patterns.

Finally, while L2 Grit emerged as a strong predictor of learners' reported interest and interaction, the study did not explore the mechanisms through which it develops or interacts with psychological constructs such as resilience, mindset, or identity. Further research could unpack these relationships and explore how to cultivate L2 Grit in military learners.

CONCLUSION

The study explored three key questions: (1) How do individual learner characteristics (general grit, L2 Grit, and foreign language anxiety) compare between institutions and affect students' motivational engagement (interest and interaction) in PF courses? (2) Which factors predict students' motivational engagement in studying PF? (3) Which factors sustain motivational engagement over time among PF students at each institution? By examining these questions across two military foreign language learning contexts, the study contributes to the growing body of research on motivation in military second language acquisition.

Across both quantitative and qualitative analyses, L2 Grit emerged as the most consistent and robust predictor of learners' interest and interaction, surpassing the influence of general grit and foreign language anxiety. Learners with higher levels of L2 grit demonstrated stronger and more sustained motivational engagement throughout the semester, regardless of institutional context. This finding highlights the importance of domain-specific grit in high-stakes language learning environments.

Early-semester engagement also played a critical role in shaping motivational trajectories. Initial levels of interest strongly predicted end-of-semester interest. Similarly, initial levels of interaction predicted end-of-semester interaction. Self-reported interest remained relatively stable throughout the semester at both institutions, suggesting that early interest establishes an enduring foundation. Additionally, institutional context shaped interaction. DLIFLC learners demonstrated higher levels of interaction overall and a mid-semester increase, whereas USMA cadets showed more stable but comparatively lower interaction.

This study underscores the importance of fostering L2 grit and early engagement in language learning programs. For military language education, this suggests a need to design instructional practices and support systems that facilitate meaningful engagement and encourage autonomous motivation. Future longitudinal research spanning full training cycles can further elucidate how motivation evolves across military language learners' trajectories.

Authors

Olivia E. Holloway is an Associate Professor of Portuguese in the Department of English and World Languages at the U.S. Military Academy (USMA). She earned a Ph.D. in Portuguese from Indiana University. Her research interests include embodied learning and motivation in second language acquisition and the warrior ethos as envisioned by Portugal, Brazil, and Japan. At USMA she serves as Director of the Portuguese Program and Officer in Charge of the Portuguese Forum capoeira program.

Siyi Gao is an Associate Professor at the Defense Language Institute Foreign Language Center, where she serves as an Academic Specialist at the Persian Farsi School. She holds a doctoral degree from Johns Hopkins University in Entrepreneurial Leadership in Education. She has nearly

two decades of experience teaching Chinese as a second language across mainland China, Greece, and the United States. Her research interests include teacher professional development, technology integration, motivation, and educational leadership.

Ra'ed Qasem is a Professor at the Defense Language Institute Foreign Language Center, where he currently serves as the Dean of the Persian Farsi School. With over 25 years of experience in the field of Foreign Language Education, he has taught languages and directed language programs across three countries, bringing a broad international perspective to his work.

Zachary F. Miller is an Associate Professor of Portuguese in the Department of English and World Languages at West Point. He earned an M.A. in Portuguese from the University of New Mexico and a Ph.D. in Second Language Studies from Michigan State University. His research interests include second language acquisition in a military context and the effects of emotion and motivation on L2 learning.

REFERENCES

Allen, J. P., Robbins, S. B., Casillas, A., & Oh, I.-S. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education*, *49*(7), 647–664. https://doi.org/10.1007/s11162-008-9098-3

Barkhuizen, G. (2008). A narrative approach to exploring context in language teaching. *ELT Journal, 62*(3), 231–239. https://doi.org/10.1093/elt/ccm043

Barkhuizen, G., & Wette, R. (2008). Narrative frames for investigating the experiences of language teachers. *System, 36*(3), 372–387. https://doi.org/10.1016/j.system.2008.02.002

Benson, P. (2014). Narrative inquiry in applied linguistics research. *Annual Review of Applied Linguistics, 34*, 154–170. https://doi.org/10.1017/S0267190514000099

Csizér, K. (2020). The L2 motivational self system. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning* (pp. 71–93). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3_4

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* Springer. https://doi.org/10.1007/978-1-4899-2271-7

Dörnyei, Z. (2002). The motivational basis of language learning tasks. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 137–158). John Benjamins. https://doi.org/10.1075/lllt.2.10dor

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Routledge. https://doi.org/10.4324/9781410613349

Dörnyei, Z. (2019). Towards a better understanding of the L2 learning experience, the Cinderella of the L2 motivational self system. *Studies in Second Language Learning and Teaching, 9*(1), 19–30. https://doi.org/10.14746/ssllt.2019.9.1.2

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited.* Routledge. https://doi.org/10.4324/9781315779553

Dörnyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation* (3rd ed.). Routledge. https://doi.org/10.4324/9781351006743

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit–S). *Journal of Personality Assessment, 91*(2), 166–174. https://doi.org/10.1080/00223890802634290

Henry, A., & Liu, M. (2024). L2 motivation and self-regulated learning: An integrated model. *System, 123,* 103301. https://doi.org/10.1016/j.system.2024.103301

Jodai, H., Zafarghandi, A. M., & Tous, M. D. (2013). Motivation, integrativeness, organizational influence, anxiety, and English achievement: Evidence from a military university. *ERIC.* https://files.eric.ed.gov/fulltext/ED586751.pdf

Keegan, K. (2017). Identifying and Building Grit in Language Learners. *English Teaching Forum*, *55*(3), 2-9. https://eric.ed.gov/?id=EJ1156469

Khajavy, G. H., MacIntyre, P. D., & Hariri, J. (2021). A closer look at grit and language mindset as predictors of foreign language achievement. *Studies in Second Language Acquisition*, *43*(2), 379–402. https://doi.org/10.1017/S0272263120000480

Miller, Z. F. (2016). Military and civilian L2 instructors: Decoding perceptions of U.S. service academy cadets. *Applied Language Learning, 26*(2), 25–52. https://www.dliflc.edu/wp-content/uploads/2021/12/Applied-Language-Learning_Vol26_Pt2.pdf

Miller, Z. F., & Crowther, D. (2020). Foreign language learning motivation in a U.S. military academy: A comparative case study on the effects from the learning environment. *Applied Language Learning, 30*(1–2), 41–59. https://www.researchgate.net/publication/339887540

Miller, Z. F., & Crowther, D. (2024). Foreign language learning in a specialized institution: Opportunities and challenges. *Foreign Language Annals, 57*(1), 48–66. https://doi.org/10.1111/flan.12730

Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning, 50*(1), 57–85.

Oxford, R. & Khajavy, G. H. (2021). Exploring Grit: "Grit Linguistics" and Research on Domain-General Grit and L2 Grit. *Journal for the Psychology of Language Learning*, *3*(2), 7–36. https://doi.org/10.52598/jpll/3/2/2

Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology, 41*(4), 359–376. https://doi.org/10.1111/j.1464-0597.1992.tb00712.x

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Vol. 2. Individual differences and cultural and contextual factors* (pp. 3–31). American Psychological Association. https://doi.org/10.1037/13274-001

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* Guilford Press. https://doi.org/10.1521/978.14625/28806

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://selfdeterminationtheory.org/SDT/documents/2000_RyanDeci_SDT.pdf

Sudina, E., & Plonsky, L. (2021). Language learning grit, achievement, and anxiety among L2 and L3 learners in Russia. *ITL-International Journal of Applied Linguistics*, *172*(2), 161-198. https://doi.org/10.1075/itl.20001.sud

Teimouri, Y., Plonsky, L., & Tabandeh, F. (2022). L2 Grit: Passion and perseverance for second-language learning. *Language Teaching Research, 26*(5), 893–918. https://doi.org/10.1177/1362168820921895

Takahashi, C., & Im, S. (2020). Comparing self-determination theory and the L2 motivational self system and their relationships to L2 proficiency. *Studies in Second Language Learning and Teaching, 10*(4), 673–696. https://doi.org/10.14746/ssllt.2020.10.4.2

Applied Language Learning

Commentary

Framing Intercultural Reflection in the Language Classroom

Anjel Tozcu
Defense Language Institute Foreign Language Center, Monterey, California

Clarissa Eagle
Defense Language Institute Foreign Language Center, Monterey, California

*Intercultural (communicative) competence (ICC), defined as the ability to communicate and behave appropriately and effectively across cultural contexts (Deardorff, 2006), is widely recognized as an essential component of language learning. However, because ICC resists standardized, objective measurement (Deardorff, 2011, 2021; Byram, 1997), educational institutions have historically struggled to incorporate it meaningfully into instructional programs (2015, 2023). This paper demonstrates how ICC can nevertheless be incorporated into curricula in a structured and transparent way through the use of SMART—**s**pecific, **m**easurable, **a**chievable, **r**elevant, and **t**ime-bound—learning objectives, helping to make intercultural engagement visible without reducing it to simplistic metrics. To illustrate this approach, we draw on a classroom implementation of the cross-cultural simulation game BARNGA, in which learners experience and analyze the consequences of differing unspoken rules. By aligning the activity with SMART objectives, we show how instructors can help students engage in key ICC processes, such as articulating underlying assumptions and practicing perspective-taking, in and through the target language. This structured approach offers a clear, pedagogically grounded way to embed ICC into language learning while meeting curricular demands for clarity and accountability.*

Keywords: *Intercultural Competence, Intercultural Communicative Competence, SMART Learning Objectives, Language Education, Cross-Cultural Simulation, BARNGA, Reflective Practice, Communicative Language Teaching, Unspoken Assumptions*

INTRODUCTION

Intercultural (communicative) competence (ICC), conceptualized by Deardorff (2006) as a combination of internal shifts in perspective and the ability to communicate appropriately and effectively across cultures, develops through context-dependent experiences that shape attitudes, knowledge, and skills. Because of this complexity, ICC remains notoriously difficult to assess through standardized, objective measures (Deardorff, 2011, 2021; Byram, 1997). Yet in an increasingly interconnected world, the ability to navigate cultural difference is more essential than ever—particularly for language learners. This tension poses a challenge for institutions like the Defense Language Institute Foreign Language Center (DLIFLC), which prioritize cultural learning while also depending on objective, measurable outcomes to ensure grading consistency and accountability. How can such institutions meaningfully incorporate ICC activities into their curricula if the learning they foster cannot be reliably measured?

One practical solution is to shift the focus from *assessing* ICC to *documenting and supporting student engagement* with foundational ICC concepts, such as perspective-taking, flexibility, and awareness of unspoken assumptions (Dervin, 2014; Deardorff, 2006). Indeed, this paper demonstrates how ICC can be effectively embedded into language curricula through the use of SMART—**s**pecific, **m**easurable, **a**chievable, **r**elevant, and **t**ime-bound—learning objectives. Rather than serving as assessments, SMART objectives function as a pedagogical tool that transforms intercultural engagement into observable classroom practices, helping instructors guide reflection and learners recognize their own growth. To illustrate this approach, we first review key frameworks of ICC, then show how SMART objectives can be applied to intercultural learning, and finally present a classroom implementation of the simulation game BARNGA. Together, these elements highlight how ICC can be integrated into curricula in a structured and transparent way, even when direct measurement remains challenging.

LITERATURE REVIEW

Scholars have developed several influential frameworks to conceptualize ICC. As mentioned above, Deardorff (2006, 2021, 2023) emphasizes that ICC develops through a dynamic process, beginning with *attitudes* such as respect, openness, curiosity, and tolerance for ambiguity. These attitudes shape the learner's ability to acquire *knowledge* (including cultural self-awareness, sociolinguistic awareness, and deep understanding of cultures) and to practice *skills* such as listening, observing, interpreting, analyzing, evaluating, and relating. From this foundation, learners may achieve *internal outcomes*—flexibility, empathy, and shifts in perspective—that in turn enable *external outcomes* such as behaving and communicating appropriately across cultural contexts. Deardorff underscores that ICC is never "finished"; it is an iterative process of continuous growth.

Byram (1997, 2008) offers a complementary framework that identifies five overlapping areas of competence, the *savoirs*. These include the *attitudes (savoir être)* of curiosity, openness, and readiness to suspend assumptions about self and other; *knowledge (savoirs)* of social groups,

practices, and interactional processes; *skills of interpreting and relating (savoir comprendre),* such as explaining and contextualizing cultural phenomena; *skills of discovery and interaction (savoir apprendre/faire)* that allow one to acquire new cultural knowledge in real-time interactions; and *critical cultural awareness (savoir s'engager)*, or the ability to evaluate different perspectives and practices on the basis of explicit criteria. Importantly, Byram stresses that evidence of intercultural communicative competence is difficult to isolate in learner interactions, as it often cuts across multiple *savoirs*. This has led to greater emphasis on reflective practice and subjectively graded, performance-based assessment, where learners demonstrate intercultural engagement through application, analysis, and self-awareness (Byram et al., 2017).

Similarly, Fosher and Mackenzie (2023), in the *Culture General Guidebook: Globally Applicable Concepts and Skills for Military Professionals*, argue that culture general education—that is, education centered on transferable intercultural concepts and analytic skills rather than country-specific facts—should move beyond static cultural information toward frameworks that can be applied across operational contexts (pp. 5–12). Their guide includes discussion of teaching tools and classroom activities designed to foster perspective-taking, critical inquiry, and reflection—processes closely aligned with broader ICC frameworks. This emphasis on developing culture general capability within military education aligns with broader analyses of U.S. military cross-cultural training programs over the past two decades (see Abbe, 2021).

Taken together, these frameworks—from foundational ICC models to recent military scholarship on culture general competence—highlight both the richness and the elusiveness of intercultural development. They show that intercultural growth involves well-defined attitudes, knowledge, and skills, but also reveal why standardized testing struggles to capture it: ICC is developmental, context-dependent, and expressed indirectly through reflection and interaction. This creates an opening for pedagogical approaches that do not attempt to measure ICC reductively, but instead focus on documenting and supporting student engagement with ICC processes.

Simulation-based tasks, such as BARNGA, provide a powerful context for this kind of engagement. By intentionally creating moments of confusion, negotiation, and perspective-shifting, BARNGA elicits the very dispositions and skills described by Deardorff and Byram—curiosity, openness, interpreting, relating, and critical awareness—in an experiential, language-rich environment. Our paper builds on this insight by showing how SMART learning objectives can be used to structure and track student engagement with these intercultural processes, making them more visible for both learners and instructors while maintaining the complexity and authenticity of ICC.

SMART Learning Objectives

SMART learning objectives offer a structured and transparent way to guide instruction and encourage purposeful reflection. While they are not designed as formal assessments of ICC, they serve as a valuable pedagogical tool by making student thinking visible and providing a framework for facilitating, observing, and discussing intercultural learning in the classroom. The SMART framework originated in the field of management, when Doran (1981) proposed it as a practical method for setting goals that were clear, actionable, and attainable. Since then, the

concept has been widely adopted across disciplines, including education, where it has proven useful in designing precise learning objectives that link instructional activities to observable outcomes. In language education, SMART objectives help educators articulate expectations in ways that are both transparent to learners and aligned with curricular goals.

When applied to intercultural learning, SMART objectives enable instructors to foreground specific aspects of ICC—such as perspective-taking or recognition of unspoken assumptions—in ways that can be meaningfully practiced and reflected upon within classroom activities. In this sense, they help to transform complex and often abstract dimensions of intercultural communicative competence into concrete opportunities for engagement, providing not only clarity for teachers but also scaffolding for students.

A Classroom Activity Built Around BARNGA

To illustrate this approach, we draw on a classroom activity centered on BARNGA, a cross-cultural simulation game created by Sivasailam Thiagarajan (for more information, see here). This game provides a powerful context for learners to confront and reflect on hidden normative assumptions. In BARNGA, small groups of learners learn to play a simple card game in silence, unaware that each group has been taught slightly different rules. As learners rotate tables and begin playing with new group members, confusion and mild conflict arise. The disorienting experience mirrors real-world intercultural encounters, where unspoken assumptions and differing norms often produce tension and misunderstanding.

In our implementation, the activity unfolds in three stages, spanning two fifty-minute instructional periods:

1. **Silent gameplay** under conflicting rule sets (35 minutes);
2. A **structured written reflection in the target language** immediately following the game, in which learners describe their reactions and begin to analyze what led to their confusion (20 minutes);
3. A **guided oral discussion in the target language**, where students explore themes such as responding to rule violations, recognizing assumptions, and understanding how communication may break down when shared norms are lacking (45 minutes).

Students use the target language not only to recount and reflect on their experience, but also to engage with key intercultural themes such as navigating misunderstanding, responding to difference, and interpreting others' behavior—all while using communicative functions appropriate to their proficiency level. These may include describing cause and effect, making comparisons, listing reasons for something, expressing uncertainty, or hypothesizing about others' intentions. To support this work, the activity should be introduced to the students at the culmination of a unit where relevant communicative functions (e.g., comparing and contrasting) and accompanying linguistic structures ("is different from", "the same as yours/mine") have been taught and practiced. It is important to remember that even the most basic communicative

functions can support a conceptually rich engagement with ideas at the heart of ICC—for instance, articulating different perspectives and exploring one's own assumptions—without the necessary use of complex linguistic structures. Lower-proficiency students may say, for example, "I was confused because the rules were different," "My rules were not the same," or "Maybe I was wrong." Mid-proficiency students might say, "I thought they made a mistake, but actually, their rules were different," or "It seemed like they broke the rules on purpose." Higher-proficiency students might say, "I assumed we all had the same understanding of the rules, which caused some tension when we started playing together," or "This experience reminded me how easily misunderstandings can arise when assumptions aren't shared." At all levels, students are using the target language to analyze a shared experience from different perspectives—a core goal of intercultural communication.

To guide this instructional sequence, we use SMART objectives, grounded in Deardorff's (2006) Process Model of Intercultural Competence:

- *After completing the simulation, students will be able to describe in writing one assumption they made during the game that was not shared by others* (Skills: interpreting and relating; Attitudes: openness).
- *At the end of the class period, students will be able to explain in simple spoken language how the confusion in the game reflects real-world intercultural challenges* (Knowledge: of cultural self-awareness and communication styles).

Finally, to evaluate how well students are engaging with intercultural concepts during the activity, teachers may use a simple rubric (Figure 1). The rubric is not intended to measure ICC as a fixed trait but rather to provide instructors with observable indicators of growth in areas such as perspective-taking, recognition of assumptions, use of specific examples, and application of target-language functions.

Figure 1
Rubric for Tracking Student Reflections on BARNGA

Criteria	Beginning	Developing	Proficient	Advanced
Perspective-taking	Shows little awareness of others' perspectives; responses are self-focused	Acknowledges others' perspectives but without detail	Identifies and contrasts perspectives with some explanation	Consistently interprets situations from multiple perspectives with some nuanced analysis
Recognition of assumptions	Does not identify assumptions or confuses them with rules	Identifies at least one assumption but without linking to outcomes	Clearly identifies unspoken assumptions and connects them to misunderstandings	Insightfully analyzes how assumptions shape interactions and outcomes
Use of examples	Provides vague or general statements (e.g., 'It was confusing')	Gives one simple example with limited explanation	Provides specific examples that clarify sources of confusion or conflict	Uses multiple, detailed examples to illustrate points and deepen analysis
Language functions	Limited use of target language; relies on single words or memorized phrases	Uses simple structures (e.g., 'I was confused,' 'The rules were different')	Employs a range of functions such as cause-effect, comparison, or hypothesis	Uses varied and complex functions fluently to analyze and reflect (e.g., hypothesizing, contrasting perspectives)

When filling out this rubric, it is important that linguistic competence and ICC not be conflated. For example, even while their language remains simple, learners at lower proficiency levels can still demonstrate advanced intercultural awareness, as shown in the sample student responses discussed previously ("my rules were not the same").

CONCLUSION

By embedding SMART objectives into a carefully structured, language- and reflection- rich activity, BARNGA becomes more than a classroom game: it functions as a vehicle for both communicative practice and intercultural inquiry, creating conditions for key ICC concepts—such as perspective-taking, recognition of assumptions, and flexibility in face of unexpected differences—to surface in authentic classroom interactions. This activity also demonstrates that intercultural learning does not need to be separated from linguistic development. The same classroom task can simultaneously support communicative growth and intercultural insight. For institutions like DLIFLC, this offers a concrete model for embedding ICC as a developmental process alongside existing curricular goals without sacrificing clarity or accountability, with SMART objectives providing a practical bridge between the complexities of intercultural growth and the accountability requirements of outcomes-based education.

Authors

Anjel Tozcu has a Ph.D. in Second Language Acquisition and Teaching from the University of Arizona. Her research interests include CALL, vocabulary acquisition, reading comprehension, intercultural competence, and bilingualism. She has presented at national and international conferences. Prior to retiring from DLIFLC in 2025, Dr. Tozcu served in many roles across the Institute.

Clarissa Eagle is an Associate Professor and Student Learning Specialist at the Defense Language Institute Foreign Language Center (DLIFLC) in Monterey, CA. Before joining DLIFLC, she taught French language and intercultural competence classes at the Middlebury Institute, where she also coordinated the Middlebury French Summer Intensive Language Program. She has presented research in language and intercultural studies at conferences ranging from the International Association for Languages and Intercultural Communication (IALIC) forum in Helsinki, Finland, to the Action Research Network of the Americas (ARNA) conference in Knoxville, TN. Her publications focus on second language acquisition, language and intercultural studies, and language teaching methodology and materials.

REFERENCES

Abbe, A. (2021). Evaluating military cross-cultural training programs. *Expeditions with MCUP*. https://doi.org/10.36304/ExpwMCUP.2021.06

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Byram, M. (2008). *From foreign language education to education for intercultural citizenship: Essays and reflections*. Multilingual Matters.

Byram, M., Golubeva, I., Han, H., & Wagner, M. (2017). *From principles to practice in education for intercultural citizenship*. Multilingual Matters.

Clarke, S., & Warr Pedersen, L. (2023). *Embedding intercultural learning in higher education: Moving from theory to practice. International Journal of Intercultural Relations, 97*, 102805. https://doi.org/10.1016/j.ijintrel.2023.102805

Deardorff, D. K. (2006,). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, *10*(3). 241-266.

Deardorff, D. K. (2011). *Assessing intercultural competence. New Directions for Institutional Research, 2011*(149), 65–79. https://doi.org/10.1002/ir.381

Deardorff, D. [IACCP 2023: European Regional Conference]. (2023). Keynote: Developing intercultural competence for all [Video]. YouTube. https://www.youtube.com/watch?v=GYTaVcnMQ1I

Doran, G. T. (November 1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review, 70*(11), 35-36.

Fosher, K., & Mackenzie, L. (2023). *Culture general guidebook: Globally applicable concepts and skills for military professionals*. Marine Corps University Press. https://www.usmcu.edu/Portals/218/Culture%20General%20Guidebook_web.pdf

Jones, D. K. (2010). *Intercultural competence: An emerging focus in international higher education. Observatorio Iberoamericano de la Educación.* Retrieved from https://observatorioieb.com.br/docs/docs873529691.pdf

Thiagarajan, S. & Thigarajan, R. (2006). *BARNGA: A simulation game on cultural clashes* (25th Anniversary ed.). John Murray Business Academic.

Intercultural-Learning.eu. (n.d.). *BARNGA: Five Tricks (10 Versions).* Retrieved from https://intercultural-learning.eu/wp-content/uploads/2017/08/BARNGA-Five-tricks-10-Versions.pdf.

Call for Papers

Applied Language Learning, a refereed journal published semiannually by the Defense Language Institute Foreign Language Center and Presidio of Monterey, is soliciting articles for publication.

The Journal (US ISSN 1041-679X and ISSN 2164-0912 for the online version) is to provide a forum for the exchange of ideas and information on

- Instructional methods and techniques
- Curriculum and materials development
- Assessment of needs within the profession
- Testing and evaluation
- Implications and applications of research from related fields such as linguistics, education, communications, psychology, and the social sciences

The journal seeks to serve the professional interest of language teachers, administrators, and researchers concerned with the teaching of foreign languages to adult learners. We welcome articles that describe innovative and successful practice and methods and/or report educational research or experimentation.

Deadline: Submissions are welcome at any point. Send your manuscript electronically to the Editor: journals@dliflc.edu.

View the **Guide for Authors** and read the recent and past issues of ***Applied Language Learning*** at **https://www.dliflc.edu/ojs/ALL**

Send all questions, comments, and/or submissions to the Editor: journals@dliflc.edu.

Thank You Reviewers

Allison Abbe, Ph.D.
Jamil Al-Thawahrih, Ed.D.
Netta Avineri, Ph.D.
Gyseon Bae, Ph.D.
Natalia Barley, Ed.D.
Christine Campbell, Ph.D.
Elliott Casal, Ph.D.
Sonia Estima, Ed.D.
Laura Hook
Donald Harootian
Dan Henk, Ph.D.
Matthew Kessler, Ph.D.
Myoyoung Kim, Ph.D.
Young Ae Kim, Ph.D.
LTC Shoshannah B.J. Lane
Saekyun Lee, Ed.D.
Edna Lima, Ph.D.
Lauren Mackenzie, Ph.D.
Tatiana McCaw, Ed.D.
Sun Young Park, Ed.D.
Jon Phillips
Kathleen Phillips
Viktoriya Shevchenko, Ph.D.
Pete Swanson, Ph.D.
LTC Scott Womack, Ph.D.
Rong Yuan, Ph.D.
Shen S. Zhu, Ph.D.