Centre for Educational Innovation and Excellence
Xi'an Jiaotong-Liverpool University
XJTLU

AI-migo and XiSPA: Custom-Built AI Tools for Spanish Absolute Beginners —Development, Pilot Implementation, and Initial Evaluation

Raúl Getino-Diez
Raul.GetinoDiez@xjtlu.edu.cn

Cristina García-Sánchez
Cristina.Garcia@xjtlu.edu.cn

Weixuan Dai
Weixuan.Dai@xjtlu.edu.cn

LM-CEIE

CONTENT

Suggested citation:

Getino-Diez, R., García-Sánchez, C., & Dai, W. (2025). *AI-migo & XiSPA: Custom-built AI tools for Spanish absolute beginners—Development, pilot implementation, and initial evaluation.* Centre for Educational Innovation and Excellence, Xi'an Jiaotong-Liverpool University. https://www.learningmall.cn/assets/Files/getino-diez-et-al-2025.pdf

AI-migo and XiSPA: Custom-Built AI Tools for Spanish Absolute Beginners—Development, Pilot Implementation, and Initial Evaluation

Raúl Getino-Diez
Raul.GetinoDiez@xjtlu.edu.cn

Cristina García-Sánchez
Cristina.Garcia@xjtlu.edu.cn

Weixuan Dai
Weixuan.Dai@xjtlu.edu.cn

Xi'an Jiaotong-Liverpool University, China

Abstract

Artificial Intelligence (AI) is rapidly transforming the educational landscape, offering unprecedented opportunities for innovation in second-language learning. At Xi'an Jiaotong-Liverpool University (XJTLU), *SPA001: Spanish Language 1*—an entry-level Spanish module for absolute beginners—has developed two AI-driven tools specifically designed to address students' learning needs: *AI-migo*, a virtual tutor that provides tailored support across multiple aspects of learning, and *XiSPA*, an AI-driven conversation partner that facilitates interactive and level-appropriate communicative practice. Both tools employ a prompt-based approach that incorporates structured knowledge bases to ensure support aligns with curriculum progression.

This article presents the background, development, pilot implementation, and initial evaluation of these tools. A mixed-methods evaluation with 385 students revealed positive perceptions across ease of use, enjoyment, output quality, and perceived learning benefits, with convenience emerging as a particularly salient advantage. Students reported substantial tool adoption, with continued post-assessment use. Qualitative findings identified areas requiring refinement, including technical reliability and accessibility, interaction burden, and adaptation mechanisms, among others. The complementary roles of AI-migo and XiSPA, combined with the prompt-based approach, suggest potential for scalable, resource-efficient AI integration in beginner language education.

Background

SPA001: Spanish Language 1 is the entry-level Spanish module at Xi'an Jiaotong-Liverpool University (XJTLU), designed for absolute beginners aiming to achieve Common European Framework of Reference (CEFR) A1- proficiency within one semester. Serving approximately one thousand students annually—most with no prior Spanish exposure—the module provides foundational support through a gradual, skill-building curriculum.

Despite its current success, the module faces two main challenges: limited out-of-class exposure to Spanish and a shortage of pedagogically appropriate learning materials tailored to students' proficiency levels and curricular progression.

Limited Exposure to Spanish Outside the Classroom

In the Chinese context, Spanish is typically learned as a foreign language in school-based settings, as it is rarely used in learners' immediate environments. There is broad consensus that studying abroad yields greater language learning benefits than classroom-based foreign language instruction (see Tseng et al., 2021, for a multi-level meta-analysis). This aligns with Second Language Acquisition theories emphasizing that target language exposure plays a vital role in acquisition by providing opportunities for comprehensible input (Krashen, 1985), meaningful output (Swain, 1985), and interactional contexts that facilitate the negotiation of meaning (Long, 1996).

In contexts where the target language is not socially dominant, out-of-class exposure remains limited but can still contribute meaningfully to language development. Activities such as digital

gaming, reading, and Internet use have shown potential to enhance learners' vocabulary, oral proficiency, and, in some cases, motivation in learning English as a foreign language (see Sundqvist & Sylvén, 2016, Chapter 5, for a synthesis of empirical studies illustrating these effects). However, the situation of other foreign languages generally differs from that of English, given its unique status as a global language in Kachru's (1985) "Expanding Circle." English learners can readily engage with diverse media and a growing global community of users. This is evident in China, where children's exposure to English from the preschool years has increased markedly in recent years (Nie & Mavrou, 2025; Sun, de Bot & Steinkrauss, 2015), alongside rapid growth in English-medium instruction (EMI) in higher education (Zheng & Choi, 2025).

By contrast, exposure to Spanish and opportunities for contextualized language practice outside the classroom remain considerably more limited. This disparity is reflected, for example, in the relative online presence of the two languages: at the time of this study, English was used by 49.2% of all websites compared to 6% for Spanish (W3Techs, 2025). These limited extramural opportunities are particularly concerning given the central role that out-of-class exposure and practice play in supporting language development.

Lack of Adapted Learning Materials

While target language exposure is crucial for acquisition, input must be carefully adapted for absolute beginners—particularly in contexts where out-of-class opportunities are limited and learners possess minimal lexical and grammatical knowledge. Krashen's *Comprehensible Input Hypothesis* (1985) proposes that optimal learning occurs when linguistic input is slightly beyond learners' current proficiency, a principle known as *i+1.* Empirical studies help clarify what this "+1" entails. Minimal text comprehension requires at least 95% known-word coverage (Laufer, 1989), whereas independent comprehension requires approximately 98% (Hu & Nation, 2000; Schmitt et al., 2011). For listening comprehension, 90-95% coverage appears necessary (Van Zeeland & Schmitt, 2013).

Moreover, texts linguistically beyond learners' knowledge may provoke foreign language reading anxiety (Bahmani & Farvardin, 2017; Liu, 2025; Saito et al., 1999) and, in East Asian contexts, contribute to demotivation. Research has shown that the *experience of failure*—measured through unfamiliar vocabulary hindering understanding, difficulty memorizing words and phrases, limited comprehension of class content or grammar, and lower test performance—is the primary source of demotivation in East Asia, in contrast to Western contexts where teacher-related factors predominate (see Xie et al., 2018, for a discussion).

Accordingly, to enhance language learning and reduce potential anxiety and demotivation, SPA001 learning materials should adhere to the *i+1* principle and progress systematically with the curriculum. Nonetheless, despite recent efforts to develop specialized materials, level-appropriate resources explicitly tailored to the XJTLU Spanish absolute beginner curriculum remain scarce, as creating such materials requires substantial time and resources.

AI Tool Development: AI-migo & XiSPA

Justification

Generative AI—particularly AI chatbots—has shown strong potential to enhance second language learning by providing tailored support, increasing opportunities for practice, and promoting learner autonomy. Research further indicates that chatbots can reduce anxiety by creating supportive, low-pressure environments and can increase motivation through features such as gamification (see Wiboolyasarin et al., 2025, for a comprehensive review of chatbots in language education).

Within this context, SPA001 plays a pivotal role in piloting innovative pedagogical approaches. Its scale, foundational nature, and potential for scalability make it an ideal testing ground for AI-enhanced tools designed to support more engaging, adaptive, and effective language learning.

Goals

To address SPA001's principal challenges—limited Spanish exposure beyond the classroom and scarcity of level-appropriate learning materials—two AI-driven tools were developed: *AI-migo*, a virtual tutor, and *XiSPA*, an interactive conversation partner. Together, these tools expand opportunities for meaningful language exposure and practice, with AI-migo also providing tailored support and feedback.

Beyond meeting these immediate instructional needs, the development of these tools also serves broader pedagogical aims: fostering learner engagement and autonomy, supporting academic success through assessment-oriented practice, and ensuring scalability for wider implementation across similar educational contexts. From a usability perspective, both tools offer 24/7 support for independent learning and provide instructors with flexible means to design targeted tasks for use both inside and outside the classroom.

Acknowledging the affective dimension of language learning, AI-migo and XiSPA integrate human-like attributes—including relatable names, visual representations, and natural conversational styles—to cultivate a sense of personal connection and make the learning experience more approachable and engaging.

Tool Description and Capabilities

AI-migo: The Virtual Spanish Tutor

AI-migo merges "AI" (Artificial Intelligence) with *amigo*, the Spanish word for "friend," evoking the image of a trusted companion who guides, encourages, and supports students throughout their learning journey. Functioning as a virtual tutor, AI-migo provides (1) instant language support by responding to students' questions about grammar, vocabulary, sentence structure, and other module-related topics; (2) personalized feedback on writing tasks, identifying errors, suggesting improvements, and explaining the rationale behind corrections; and (3) custom practice activities—such as vocabulary drills and grammar quizzes—to strengthen areas requiring additional reinforcement.

Explanations, feedback, practice, and examples are aligned with SPA001's content progression. The tool employs an engaging interaction style that emulates a supportive teacher, aiming to create a learning environment where students feel comfortable and confident experimenting with the language and developing their abilities.

XiSPA: The AI-Driven Conversational Partner

XiSPA combines "Xi" from Xi'an Jiaotong-Liverpool University and "SPA" for Spanish, forming a name that evokes *chispa*, the Spanish word for "spark." This association conveys both a sense of belonging and the energy, curiosity, and enthusiasm essential for language learning. Designed as an AI-driven conversational partner, XiSPA engages students in synchronous text-based conversations in Spanish through topic-based exchanges on themes such as personal information, hobbies, weather, and clothing. These interactions provide opportunities for contextualized language use and promote communicative competence development. By reinforcing essential vocabulary and grammatical structures within meaningful communicative contexts, XiSPA aims to help learners build confidence and reduce anxiety during the initial stages of language production in a supportive, non-judgmental environment.

All XiSPA interactions are aligned with the lexical, grammatical, and communicative content covered in SPA001, providing students with exposure to input at an optimal *i+1* level as they progress through the semester.

Tool Design

Linguistic Control Compliance

Generating practice with AI—such as grammar and vocabulary activities, reading comprehension tasks, or conversation practice—requires highly controlled language that adheres to strict lexical and grammatical parameters to align with learners' levels and curricular progression. While AI-migo requires controlled text generation when producing practice activities and examples, XiSPA demands continuous linguistic control throughout every conversation. However, achieving this level of control remains a significant challenge.

Controlled Text Generation (CTG) techniques have proven effective in other forms of linguistic control and can be adapted to the purposes outlined above. However, they generally require substantial expertise and computational resources (see Getino-Diez & García-Madariaga, 2026, for an overview). To mitigate these limitations, we adopted the prompt-based approach proposed in Getino-Diez and García-Madariaga (2026), which offers several advantages: it incorporates explicit mechanisms to enforce adherence to predefined linguistic constraints tailored for absolute beginners, remains resource-efficient, allows for iterative tool refinement, and is easily scalable to other low-level modules.

In this approach, prompts integrate enhanced techniques that enforce adherence to lexical and grammatical constraints specified in structured knowledge base files. Prompts and knowledge bases can be (1) updated as the course progresses without model fine-tuning and (2) integrated seamlessly into AI chatbots and conversational agents. This dynamic approach forms the core of both AI-migo and XiSPA, enabling regular updates to ensure continuous alignment with module progression and chatbot integration in the XIPU AI environment.

System Architecture

Both AI-migo and XiSPA were deployed within XIPU AI Agent 1.0, a component of XJTLU's institutional XIPU AI platform. XIPU AI Agent 1.0 provides an integrated environment for developing and deploying customized intelligent chatbots through a system prompt engine and knowledge base linkage, enabling the creation of domain-specific pedagogical agents without requiring advanced technical expertise.

Each chatbot combines a system prompt with curated knowledge bases aligned with the Spanish curriculum.

System Prompts. The system prompt ensures alignment with each chatbot's intended pedagogical functions. Prompt sections addressing tasks requiring highly controlled text generation—such as lexical and grammatical activities or conversational practice—incorporate strategies essential for maintaining lexical and grammatical control. These include the *persona pattern* to assign the AI model a specific role, *markdown-style formatting* to highlight the structural hierarchy of the prompt, *context priming* to constrain outputs to language characteristic of absolute beginners, *rule reinforcement via redundancy* (pairing affirmative commands with their negative counterparts), *explicit negation of frequent non-compliant linguistic structures* linked to communicative functions studied in class, and a *final reinforcement section* reiterating critical constraints (see Getino-Diez & García-Madariaga, 2026, for a detailed description).

Knowledge Bases. XiSPA's knowledge base follows the structure described in Getino-Diez and García-Madariaga (2026): (1) a *Vocabulary List*—a comprehensive inventory of words organized by category, and (2) a *Functions and Examples* document—an extensive compilation of examples of linguistic structures, organized by communicative function. AI-migo

includes the same files for generating adapted practice activities, plus a *Grammar Rules and Examples* document containing detailed grammar explanations and examples aligned with SPA001. This file enables AI-migo to answer language-related questions and provide feedback tailored to module progression.

System Updates. Knowledge bases contain the linguistic content covered in SPA001 at a given point in the semester. Updating them weekly with newly studied material allows both agents to evolve in step with the module. Prompts also require periodic updates, as non-compliant structures evolve throughout the semester and prompt techniques—particularly the *explicit negation of frequent non-compliant structures*—require corresponding adjustments.

Pilot Implementation

An initial pre-test involving 32 students was conducted during Weeks 10 and 11 of the 2024–25 academic year (Semester 1). This preliminary stage enabled identification of critical issues and verification of functionality before scaling to the entire module.

In Week 12, the final week of teaching, AI-migo and XiSPA were introduced to the 874 students enrolled in SPA001 through a session delivered during their lecture time. This session was designed to familiarize students with the systems and demonstrate how they could be integrated into their regular learning routines. The presentation included an overview of capabilities, followed by guided activities enabling students to engage with the tools and explore their potential uses in preparing for the Speaking Coursework and the Final Examination.

The primary aims of the pilot implementation were to provide hands-on experience using the tools, explore their potential for supporting learning, and identify any immediate usability or technical issues during initial deployment. Although both tools are designed to adapt to the module's pace through incremental knowledge base updates throughout the semester, for this pilot they were optimized specifically for Week 12.

Evaluation Methodology

Design

The evaluation aimed to obtain rapid, actionable feedback on students' first experiences with AI-migo and XiSPA, examining their initial perceptions and self-reported use for assessment and post-assessment learning. Rather than testing hypotheses or theories, this evaluation adopted a descriptive and formative orientation to inform subsequent refinements and guide broader implementation in future semesters.

This evaluation is part of an iterative implementation–evaluation–refinement cycle. Given the compressed nature of the teaching cycle, the novelty of the technology, and the formative purpose of this phase, an exploratory, cross-sectional one-shot case design was adopted to capture early-stage data. Students completed a single post-exposure questionnaire several weeks after their initial interaction with the tools.

A mixed-methods approach was employed, with quantitative survey items as the primary data source and open-ended responses providing complementary qualitative insights.

Participants

All students registered in SPA002 were invited to participate in Week 1 of Semester 2, Academic Year 2024-25, following their exposure to AI-migo and XiSPA at the end of Semester 1 during SPA001. Participation was voluntary and anonymous. Informed consent was

obtained at the beginning of the survey, and ethical approval was granted by the XJTLU Ethics Committee.

Instrument

An online survey was developed to reflect the distinctive characteristics of the AI tools and the pilot implementation context.

The quantitative component measured student perceptions and actual use through self-reported Likert scale, frequency, and yes/no questions. Likert and frequency items used five-point scales with higher values indicating more positive perceptions or greater frequency of use (5 = very positive/frequent, 1 = very negative/infrequent).

Constructs were primarily measured using single-item indicators to minimize respondent burden and maximize completion rates—an approach justified by the exploratory and formative nature of this evaluation, as well as by evidence that single items are sufficient for concrete, singular constructs easily understood by respondents (Bergkvist & Rossiter, 2007).

The survey design was guided by the Technology Acceptance Model (TAM; Davis et al., 1989), incorporating the core constructs of *perceived ease of use* and *perceived usefulness*, the latter operationalized as the *perceived impact of the AI tools on module results* and *post-assessment improvement*. To account for affective influences in language learning, *perceived enjoyment* was included, given its documented association with ease of use (Venkatesh, 2000). Drawing on Wixom and Todd's (2005) framework integrating quality beliefs with technology acceptance, three tool-specific quality dimensions were also examined: *module adaptation* captured perceived ability to generate output aligned with curriculum progression; *response accuracy* reflected perceived precision of AI-migo's output; and *response naturalness* referred to perceived naturalness of XiSPA's interactions. Finally, *actual use* measured whether and how frequently students employed the tools for assessment preparation and post-assessment learning.

The qualitative component included nine optional open-ended questions allowing students to elaborate on their experiences. For each AI tool, questions addressed: (a) self-reported technical issues, (b) perceived advantages, and (c) perceived challenges or limitations. These six questions were designed to provide contextual and explanatory insights into the domains of ease of use, enjoyment, module adaptation, response accuracy and naturalness, and overall usefulness. Three additional questions focused on actual use: (a) how students used AI-migo for the end-of-semester assessment activities, (b) how they used XiSPA for these assessment activities, and (c) how they employed both tools after completing the assessment.

Procedure

The survey was delivered via Learning Mall Core, the university's learning management system, during students' scheduled lecture period.

Data Analysis

For quantitative data, descriptive statistics were employed. Given the ordinal yet approximately interval properties of five-point Likert scales and frequency items, means and standard deviations are reported in text, while medians are provided in the accompanying figures for reference.

For qualitative data, thematic analysis was conducted within the theoretical framework established by the survey constructs, involving systematic coding to identify recurring themes and patterns. The qualitative component provides complementary insights, enriching and contextualizing quantitative findings. However, responses to the optional questions were selective; students who chose to comment may hold particularly strong views—either positive

or negative—compared to the broader population. Accordingly, qualitative statements should be interpreted as illustrative and exploratory rather than as representative confirmation of quantitative patterns. Because comments often express more than one idea, some responses were assigned multiple codes when relevant; consequently, theme frequencies do not necessarily correspond to total number of comments.

Quantitative Results

A total of 385 students completed the post-exposure survey. The sample consisted of first-year undergraduates aged 18–20 years (M = 18.53, SD = 0.60). Nearly all participants (n = 383, 99.5%) were native Chinese speakers; two participants reported other first languages (English, n = 1; Indonesian, n = 1).

All participants rated both AI-migo and XiSPA on perception dimensions regardless of whether they had used the tools during assessment preparation. Usage frequency and impact ratings were collected only from self-reported users.

The following sections provide a descriptive summary of participants' perceptions across key dimensions—ease of use, enjoyment, output quality, and perceived impact—as well as self-reported tool use patterns.

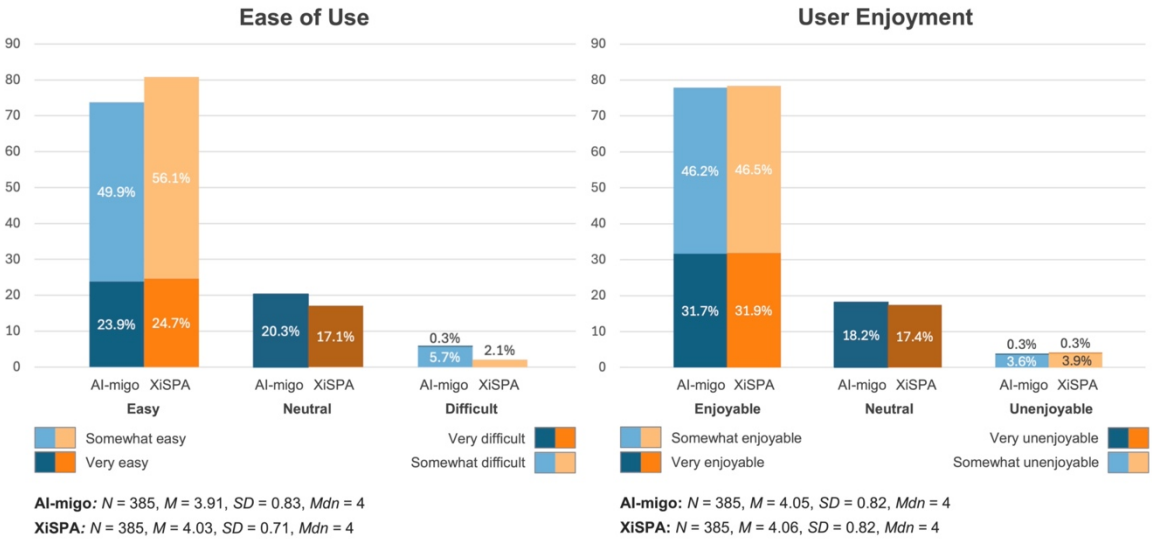
Ease of Use and User Enjoyment

Participants perceived both tools as easy to use, with XiSPA receiving slightly higher ratings (Figure 1). For AI-migo, 73.8% of respondents rated the tool as very or somewhat easy to use (M = 3.91, SD = 0.83), while 80.8% provided comparable ratings for XiSPA (M = 4.03, SD = 0.71). Difficulty ratings were minimal: 6.0% rated AI-migo as difficult or very difficult, compared with 2.1% for XiSPA.

Enjoyment ratings were similarly positive and nearly identical across tools. For AI-migo, 77.9% of participants rated the tool very or somewhat enjoyable (M = 4.05, SD = 0.82), compared with 78.4% for XiSPA (M = 4.06, SD = 0.82). Negative enjoyment ratings were reported by 3.9% of respondents for AI-migo and 4.2% for XiSPA.

Figure 1

Ease of Use and User Enjoyment Ratings for AI-migo and XiSPA



AI-migo: N = 385, M = 3.91, SD = 0.83, Mdn = 4
XiSPA: N = 385, M = 4.03, SD = 0.71, Mdn = 4

AI-migo: N = 385, M = 4.05, SD = 0.82, Mdn = 4
XiSPA: N = 385, M = 4.06, SD = 0.82, Mdn = 4

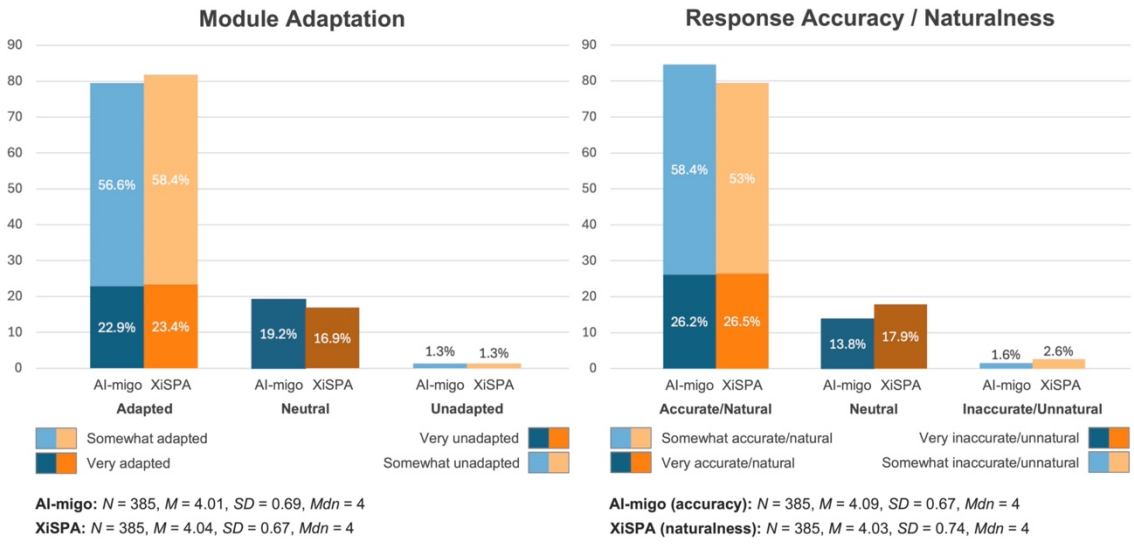
Output Quality: Adaptation, Accuracy, and Naturalness

Perceptions of output quality were positive across all measured dimensions (Figure 2). For curriculum alignment, 79.5% of students rated AI-migo as very or somewhat well adapted to SPA001 content (M = 4.01, SD = 0.69), while 81.8% provided comparable ratings for XiSPA (M = 4.04, SD = 0.67). Ratings indicating poor adaptation were minimal for both tools (1.3% each).

For tool-specific quality dimensions, 84.6% of respondents rated AI-migo's output as very or somewhat accurate (M = 4.09, SD = 0.67), with only 1.6% providing negative accuracy ratings. Similarly, 79.5% described XiSPA's conversational responses as very or somewhat natural (M = 4.03, SD = 0.74), with 2.6% rating naturalness negatively.

Figure 2

Module Adaptation and Response Accuracy/Naturalness Ratings for AI-migo and XiSPA

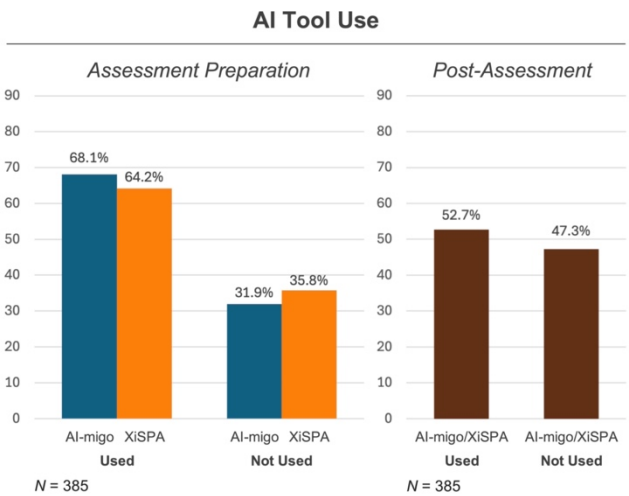


AI Tool Use

A majority of students reported using the AI tools for assessment preparation: 68.1% (n = 262) used AI-migo and 64.2% (n = 247) used XiSPA (Figure 3). Post-assessment engagement remained substantial: 52.7% (n = 203) continued using at least one tool after assessment.

Figure 3

AI Tool Use for AI-migo and XiSPA

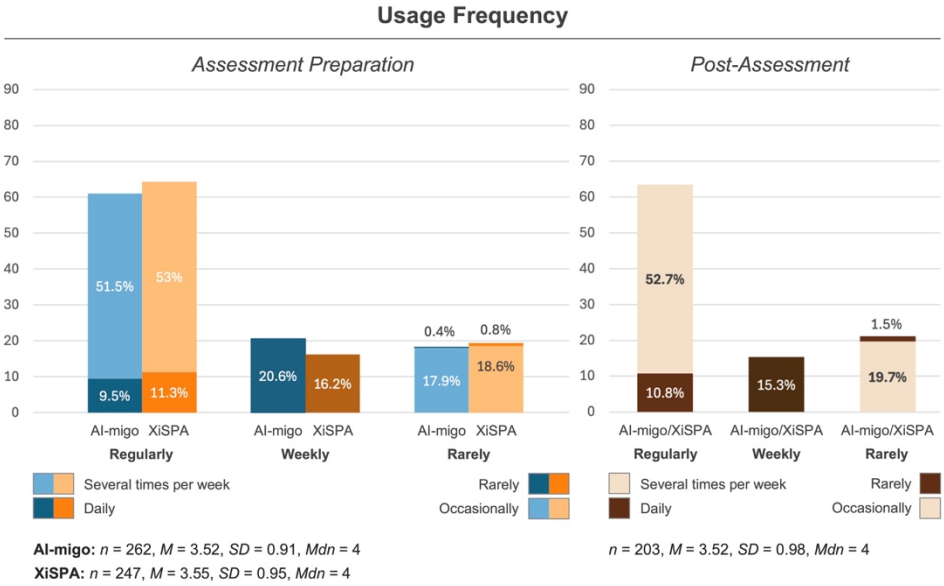


Among students who adopted the tools for assessment preparation, engagement frequency was high (Figure 4). For AI-migo users, 61.0% reported daily or multiple weekly interactions (*M* = 3.52, *SD* = 0.91), while 64.3% of XiSPA users reported comparable frequency (*M* = 3.55, *SD* = 0.95). When including those who engaged at least weekly, proportions increased to 81.6% for AI-migo and 80.5% for XiSPA, representing 55.6% and 51.6% of the total sample, respectively.

Post-assessment usage frequency remained similarly high among continuing users. Of the 203 respondents who used at least one tool, 63.5% engaged daily or multiple times weekly, and 78.8% engaged at least weekly (*M* = 3.52, *SD* = 0.98), representing 41.5% of all participants.

Figure 4

Usage Frequency Ratings for AI-migo and XiSPA



AI-migo: *n* = 262, *M* = 3.52, *SD* = 0.91, *Mdn* = 4
XiSPA: *n* = 247, *M* = 3.55, *SD* = 0.95, *Mdn* = 4

n = 203, *M* = 3.52, *SD* = 0.98, *Mdn* = 4

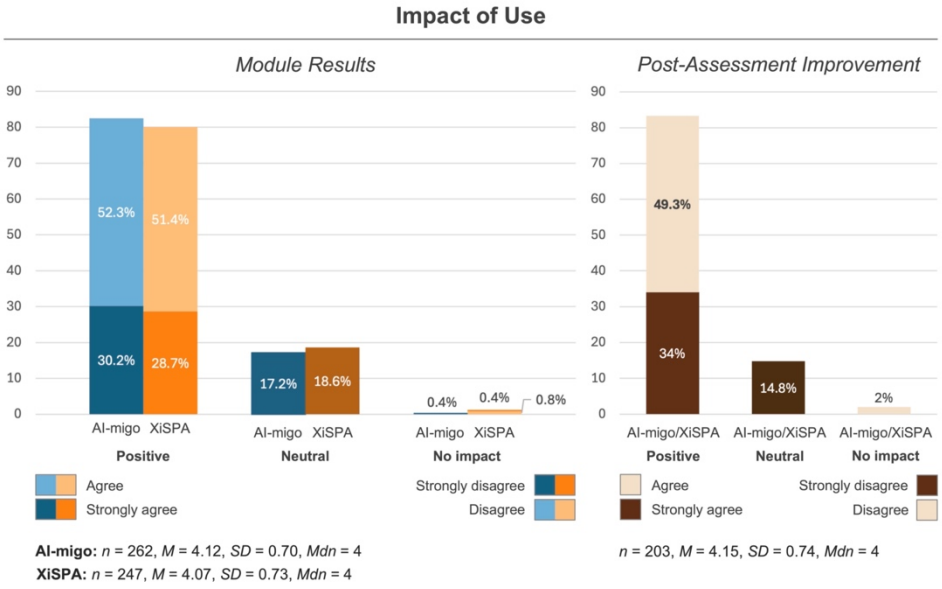
Impact of AI Tool Use

Participants who used the tools for assessment preparation reported strongly positive perceptions of their impact on module results (Figure 5). Among the 262 AI-migo users, 82.5% strongly agreed or agreed that the tool positively impacted their module results (*M* = 4.12, *SD* = 0.70), with only 0.4% reporting negative impact. Similarly, among the 247 XiSPA users, 80.1% strongly agreed or agreed regarding positive impact (*M* = 4.07, *SD* = 0.73), with 1.2% reporting negative impact.

Post-assessment improvement perceptions were similarly positive. Among the 203 students who continued using at least one tool, 83.3% strongly agreed or agreed that tool use contributed to continued language improvement (*M* = 4.15, *SD* = 0.74), with only 2.0% reporting negative impact.

Figure 5

Perceived Impact of AI-migo and XiSPA on Module Results and Post-Assessment Improvement



Qualitative Results

A total of 686 comments were submitted, of which 654 were retained for analysis after excluding those that did not address the question's intent or could not be reasonably interpreted (n = 32). Response rates varied considerably across items due to their voluntary nature (range: 30–115 responses per item).

The response patterns to the six open-ended items on student perceptions provide context for interpreting findings. Among respondents who addressed technical issues, 60 of 92 (65.2%) for AI-migo and 56 of 68 (82.4%) for XiSPA reported no issues, while 29 of 106 (27.4%) for AI-migo and 32 of 73 (43.8%) for XiSPA reported no challenges or limitations. Students reported challenges or limitations more frequently than technical issues for both tools: 77 versus 32 substantive responses for AI-migo and 41 versus 12 for XiSPA. Reported advantages totaled 115 for AI-migo and 74 for XiSPA, exceeding reported technical issues, challenges, and limitations combined (109 and 53, respectively).

The three items on self-reported use yielded 58 responses for AI-migo and 38 for XiSPA regarding use in assessment preparation, and 30 responses describing post-assessment use of both tools. Overall, XiSPA generated fewer qualitative responses than AI-migo (215 vs. 313 for perception items; 38 vs. 58 for use reports).

In the analyses that follow, percentages are calculated from substantive, content-bearing responses only, with "no issue" responses excluded. Of the 654 retained responses, 477 contained substantive content. Tool-specific percentages are based on substantive comments pertaining to that tool: 282 for AI-migo and 165 for XiSPA.

Ease of Use

Convenience as Dominant Theme

Convenience emerged as the most salient theme for AI-migo, appearing in 49 comments (42.6% of AI-migo advantage responses; 17.4% of AI-migo substantive responses) and was also prominent for XiSPA, appearing in 22 comments (29.7% of XiSPA advantage responses; 13.3%

of XiSPA substantive responses). For AI-migo, convenience encompassed multiple dimensions: temporal flexibility ("24h accessible"), spatial flexibility ("we can study wherever we want"), immediacy ("instant reply"), and convenience of accessing academic support ("saving time of going to the professor's office"). For XiSPA, convenience comments focused primarily on immediacy and practice availability ("when I can't find the Spanish speakers").

Technical and Access Barriers

System reliability issues emerged as a concern at comparable rates for both tools, appearing in 21 comments for AI-migo (7.4% of substantive responses) and 12 for XiSPA (7.3%). Students described three problem categories: system crashes ("may break down frequently"), connectivity problems ("frequently displaying loading errors"), and interface glitches ("text typed will disappear").

Accessibility barriers appeared less frequently, with 7 comments for AI-migo and 2 for XiSPA, describing the tools as "hard to find." Four respondents explicitly requested standalone mobile applications.

Effort and Interaction Challenges

Beyond technical issues, interaction challenges differed by tool function.

For AI-migo, *effort* emerged in 17 comments (6% of substantive responses) across three categories:

(a) System-induced effort: Technical instability required repeated input ("to type my question again and again").

(b) Interaction modality effort: Students reported typing issues ("print words slowly") and requested alternatives to text input.

(c) Prompt formulation effort: Users struggled to articulate queries effectively ("think about how to make AI understand my order").

For XiSPA, effort appeared less frequently (4 comments) but reflected different challenges due to its conversational nature:

(a) System-induced effort (similar to AI-migo).

(b) Second-language skill limitation effort ("don't know how to make a conversation with [XiSPA] because my language skill limit[s] me").

Self-attribution patterns—instances where responses acknowledged participants' personal limitations as sources of difficulty—revealed tool-specific differences. For AI-migo, self-attribution appeared in 10 comments: 5 attributing difficulties to prompt formulation effort and tool comprehension issues ("how to ask AI efficiently"), 2 to interaction modality effort (including one admitting being "too lazy to type down my complicated questions"), and 3 to general unfamiliarity. For XiSPA, self-attribution appeared in 5 comments: 2 attributing difficulties to second-language skill limitation effort ("my grammar") and 3 to general unfamiliarity.

Relatedly, 17 comments for AI-migo and 8 for XiSPA described *failures in tool comprehension* without acknowledging potential user-side causes. These instances encompassed comprehension difficulties ("sometimes it cannot understand my question"), misinterpretation ("misunderstanding the purpose"), struggles with complex expressions ("if I ask something difficult"), and perceived insufficient capability ("not smart enough").

Enjoyment

Enjoyment-related comments were sparse, appearing in only 10 comments total: 5 reflected *enjoyment* describing tools as "fun" or "interesting," and 5 reflected *positive motivation* ("helping to improve my interest in learning").

Negative affect emerged less frequently but was more specific. Frustration appeared in 5 comments for AI-migo, which co-occurred with linguistic adaptation mismatches ("too hard to understand") and failures in tool comprehension linked to prompting difficulties or system limitations. For XiSPA, *boredom* appeared in one comment ("boring with talking to [AI]").

Adaptation

Adaptation-related comments were relatively limited compared to ease of use, totaling 28 for AI-migo (9.9% of substantive responses) and 11 for XiSPA (6.7%). For AI-migo, 17 comments reflected perceived adaptation while 9 described inadaptation; XiSPA showed a similar pattern (7 vs. 4). Given the different pedagogical functions of the two tools, findings are presented separately.

AI-migo

Interpreting adaptation comments for AI-migo is complicated by survey design: no open-ended item targeted any specific dimension of "adaptation." Adaptation may concern the nature of the output—linguistic adaptation (difficulty and appropriateness of Spanish vocabulary, grammar, communicative functions) or explanatory adaptation (how the system adjusts explanations and feedback)—and also the target of adaptation—curriculum-level (alignment with lesson content) or student-level (tailored to individual proficiency or knowledge). Although responses consistently identified the target, many comments—such as "adapted to my learning level" or "close to the lesson"—remained ambiguous about whether they referred to linguistic, explanatory, or concurrent adaptation. Therefore, only comments with explicit indicators were coded as linguistic or explanatory; all others were categorized as general adaptation.

General student-level adaptation emerged most frequently, appearing in 10 comments, describing the system as "matches my level." *General curriculum-level adaptation* also appeared in 6 comments, with responses such as "tailored to the lesson."

Reports of inadaptation were more specific. *Student-level linguistic inadaptation* appeared in 4 comments describing AI-migo's Spanish as too difficult due to unfamiliar vocabulary ("hard to understand some words") or overall comprehension problems ("language I didn't understand very well"). *Student-level explanatory inadaptation* emerged in 3 comments, noting difficulties understanding the system's responses ("answers that are not well adapted to my level").

Isolated linguistic cases were also identified: one comment noted *student-level linguistic adaptation* ("answers that match your language skills"), while two described adaptation failures—one finding the Spanish "quite simple" and another describing a recursive vocabulary problem ("uses unfamiliar vocabularies that are not taught in class, and if I say 'No comprendo', it will use another unfamiliar word"), indicating failure to adjust even when confusion was signaled.

XiSPA

Because XiSPA's only function is maintaining written conversations for communicative practice, comments about adaptation refer specifically to linguistic adaptation, targeting either curriculum or student level.

Student-level linguistic adaptation was most prevalent, appearing in 4 comments ("can adapt to your level"), though without specifying linguistic features. *Curriculum-level linguistic*

adaptation appeared in 3 comments (“贴合课程” [fits the course]). Two responses indicated *student-level linguistic inadaptation* (“vocabulary sometimes is too difficult”), and two reflected *curriculum-level linguistic inadaptation* (“sometimes it uses extracurricular vocabulary”).

Response Accuracy

AI-migo

Accuracy perceptions were examined quantitatively and emerged in qualitative data as a contested dimension, with 8 positive and 16 negative responses (24 total, 8.5% of substantive responses).

General accuracy appeared in 8 instances, focusing on the perceived correctness of AI-migo’s answers, described as “accurate,” “precise,” “correct,” or “logical and reasonable,” with two simply affirming accuracy without elaboration. Five of these references co-occurred with convenience.

Among negative responses, *response inaccuracy* was the most frequent theme, emerging in 12 instances describing answers as “not accurate” or “not reliable.” One identified “small grammar questions” as a specific limitation, while another referred to AI hallucinations (“it gives a ridiculous answer”). *Inability to answer questions* also appeared in 6 responses (“can’t answer my questions”). Two accounts attributed these issues to technical constraints, referencing AI-migo’s “knowledge limit” or the broader “limitations of language model.” Two additional mentions simply listed “accuracy” as a concern without further detail.

XiSPA

Accuracy was not an explicit design objective for XiSPA, but it nonetheless appeared in qualitative data. In this case, accuracy referred specifically to linguistic correctness in Spanish, with 4 positive descriptions of XiSPA as “accurate” or as a “grammatically correct friend to practice Spanish.”

Response Naturalness

XiSPA

Naturalness was examined quantitatively and further explored through qualitative data, though responses were very limited, appearing in 9 comments (5.45% of substantive responses). *Conversational discontinuity* emerged in 4 instances, describing difficulty sustaining a “long conversation” and “weak memory.” Perceptions of humanness were mixed: 3 responses reflected *human-like interaction* (“like chatting with your good friend”), while 2 described *lack of human-like interaction* (“cannot act like a human being”).

AI-migo

Although naturalness was not an explicit quantitative survey dimension for AI-migo, 6 relevant qualitative comments emerged. Comments referred to a *lack of human qualities* such as “emotion” and “understanding, empathy, or creativity,” *reduced engagement compared to face-to-face interaction* (“not as concentrating as facial conversation”), and *conversational discontinuity* (“weak memory”).

AI Tool Use Description

A total of 126 responses (26.4% of 477 substantive responses) described how students used the tools for assessment preparation and/or after the assessment.

AI-migo Use

Fifty-eight responses described how students used AI-migo to prepare for final assessments (46% of tool use descriptions). *Question-asking for information seeking* emerged as the predominant usage pattern, appearing in 28 instances and encompassing general inquiries ("asking some questions") and topic-specific queries. Among topic-specific queries, grammar-related questions ("to figure out some grammatical problems") and assessment-related questions ("ask for advice when preparing for the exams") were most common with 5 instances each, followed by vocabulary inquiries and general information searches (3 instances each).

Error correction and feedback constituted the second major pattern, emerging in 11 responses, with grammar correction particularly prominent (4 instances; "correct my grammar mistakes") alongside writing-related feedback (2 instances). *Practice generation* represented another significant cluster, appearing in 9 comments, with requests mainly for grammar-focused exercises (4 mentions; "generate some grammar questions"). *Content generation* also appeared in 7 descriptions, including requests to "do a summary" and "listing wordbank," with some involving assessment-related content generation (3 cases; "presume questions which could appear in my final exam"). *Misuse* was identified in 6 instances, with students attempting conversational practice—a functionality designed for XiSPA—such as "to simulate the speaking test."

XiSPA Use

Students engaged with XiSPA for *conversational practice*, as described in 38 responses (30.2% of tool use descriptions). Question-answer conversation format was most frequently reported, appearing in 9 instances ("ask and answer"). *General conversation* was nearly as common in 8 responses ("have conversations with it"), followed by speaking skill development ("improve my speaking skills"), *topic-specific practice* ("how to describe weather in Spanish"), and *exam-oriented conversation practice* ("based on the scope of the exam"), with 3 mentions each. Some responses also described using XiSPA for vocabulary and grammar practice within conversational exchanges (3 instances each). *Misuse* also appeared in 6 instances, with students attempting to employ XiSPA for feedback, information retrieval, and exercise generation.

AI-migo and XiSPA Post-Assessment Use

Regarding post-assessment use, 30 responses (23.8% of tool use descriptions) described continued engagement with either tool. *Question-asking* remained most frequent, appearing in 15 responses, with vocabulary-related (3 mentions) and grammar-related (2 mentions) inquiries among topic-specific questions. General conversation practice appeared in 12 responses, sometimes oriented toward review and consolidation (2 cases; "literally just chat with it in order not to forget the knowledge I have learnt"). *Practice generation* appeared less frequently (3 responses).

Perceived Usefulness: Practical Benefits of Using AI-migo and XiSPA

Beyond use patterns described in the previous section, qualitative responses framed perceived usefulness in terms of practical, task-oriented benefits. These responses totaled 58 for AI-migo (20.6% of AI-migo's substantive responses) and 28 for XiSPA (17.0% of XiSPA's responses).

General Usefulness

General perceptions of usefulness without specifying particular benefits appeared in 7 comments for AI-migo and 8 for XiSPA, using terms like "useful," "helpful," "efficient," or "professional."

Language Learning and Linguistic Competence Development

For AI-migo, *language learning* emerged in 16 comments, encompassing general learning ("to support learning"), knowledge understanding ("help understand"), and learning efficiency ("making my language learning more efficient"). Isolated instances also pointed to knowledge internalization ("to enhance my knowledge") and the learning of linguistic structures ("grammar and words"). *Linguistic competence development* also appeared in 8 comments, with responses indicating AI-migo helped "improve my foreign language" and "enhance my Spanish skills." For XiSPA, only 1 comment addressed *general language learning*, while 7 focused on *linguistic competence development*, including references to the ability to "communicate with a foreigner" and "apply [knowledge] in conversations."

Practice

Practice appeared minimally for AI-migo in 2 comments, with one describing using it "to practice Spanish" and another emphasizing learning "by practicing instead of reciting." In contrast, *practice* emerged strongly for XiSPA, with 10 mentions (35.7% of XiSPA benefits). Three comments mentioned *general practice* without specification, and the remaining seven explicitly described *conversation-oriented practice*, including human-substitution practice ("when I can't find the Spanish speakers, I can use it to help"), communicative practice ("practice the communication"), and oral skills practice ("practice my oral Spanish"). Two comments noted the absence of a voice feature as a limitation of these benefits ("can't chat with it in audio form").

Feedback (AI-migo)

Feedback appeared in 10 responses for AI-migo (17.2% of AI-migo's benefits). *Functional feedback* appeared in 8 comments, with *grammar-focused feedback* as the most prominent in 6 instances ("check my grammar problems"), followed by writing improvement ("polish my text"), and proofreading ("check passages"). Mentions of punctuation and vocabulary appeared as isolated targets. Additionally, two comments addressed *feedback quality* with opposing assessments: one described it as "detailed," another as "too general."

Additional Benefits

Retention appeared in 5 comments across both tools: 3 comments for AI-migo, referencing retention of class content, post-class review, and exam preparation; and 2 comments for XiSPA, addressing linguistic content like sentence structures. Regarding AI-migo, *assessment preparation* and *information searching* appeared in 3 comments each.

Overreliance

Overreliance was noted in 6 comments for AI-migo and 4 for XiSPA, raising concerns about impact on "critical thinking" and excessive dependence ("depending on it too much" or "may rely on [it] too much").

Discussion

The pilot implementation and initial evaluation of AI-migo and XiSPA provide preliminary evidence that custom-built AI tools can be successfully integrated into beginner-level language instruction. The findings indicate high levels of initial user acceptance, frequent and sustained voluntary use, and perceived learning benefits, while also revealing important areas requiring refinement.

General Reception and Engagement

Both tools demonstrated strong initial acceptance, with positive ratings of 70-85% across dimensions and mean scores clustered around 4.0. This reception, coupled with sustained voluntary use for assessment preparation and post-assessment, and high reported frequencies of usage, suggests successful integration into students' learning routines.

Qualitative data support these patterns while revealing nuanced differences. First, reported advantages exceeded combined technical issues and challenges for both tools. Second, the substantial proportion of "no issue" responses suggests items attracted respondents with varied experiences rather than only those with extreme views. Third, challenges and limitations outnumbered technical issues, suggesting that primary concerns centered on pedagogical effectiveness and usability rather than system stability.

Engagement patterns, however, differed across tools. AI-migo generated substantially more qualitative responses than XiSPA (282 vs. 165 substantive responses) and showed slightly higher use rates for assessment preparation. This disparity likely reflects AI-migo's broader functionality providing more dimensions for comment, higher overall adoption, and possibly survey fatigue due to item ordering. The consistently higher engagement suggests AI-migo occupied a more central role in students' learning experiences, whereas XiSPA served a more focused function as a conversational practice partner.

Usability

Ease of Use and the Primacy of Convenience

Both tools received positive ease-of-use ratings, with XiSPA rated slightly higher (80.8%) than AI-migo (73.8%). This difference likely reflects XiSPA's simpler chat-based interface and students' growing familiarity with conversational AI, compared to AI-migo's multiple functionalities requiring greater user adjustment.

Qualitatively, convenience emerged as the dominant advantage—appearing in 42.6% of AI-migo advantage responses and 29.7% for XiSPA. For AI-migo, students emphasized temporal and spatial flexibility, immediacy, and on-demand access to academic support. For XiSPA, emphasis focused primarily on immediacy and practice availability. This prominence suggests convenience addressed a genuine gap in available Spanish support, practice, and exposure outside the classroom, where target language exposure remains limited and learners have minimal access to level-appropriate materials and practice opportunities.

While convenience is conceptually distinct from perceived ease of use, research indicates that ease of use functions as an antecedent to convenience (Chang et al., 2012; Yoon & Kim, 2007), and this relationship appears consistent with the present findings, where both dimensions received positive evaluations.

Technical and Access Barriers

Two categories of usability challenges emerged despite low frequency. System reliability issues—encompassing system crashes, connectivity problems, and interface glitches—appeared in approximately 7% of substantive responses for both tools. These technical barriers warrant attention as they may undermine the convenience students valued most

highly, and indirectly influence perceived ease of use through system satisfaction (Wixom & Todd, 2005).

Accessibility barriers, though less frequent, are equally critical. Difficulty locating tools and requests for standalone mobile applications indicate that some students faced foundational access challenges. Without reliable access, pedagogical affordances become irrelevant.

Effort and Interaction Challenges

Both tools presented interaction challenges involving three types of effort. *System-induced effort* forced students to retype input after crashes. *Interaction modality effort*, exclusive to AI-migo, reflected the demands of typing extended queries compared to XiSPA's shorter conversation turns. Most significantly, *skill-dependent effort* emerged differently for each tool: *prompt formulation effort* (AI-migo) and *second-language skill limitation effort* (XiSPA).

For AI-migo, students struggled with prompt formulation—the metalinguistic demands of constructing effective queries. This finding highlights a fundamental tension in AI tutor design: while open-ended question-answering provides flexibility, it places substantial demands on students who may lack the knowledge or skills to formulate their questions appropriately. These prompting demands also mobilize the physical effort of typing (interaction modality effort). For XiSPA, the barrier was second-language proficiency itself: limited linguistic repertoire constrained students' ability to sustain conversation, reflecting an inherent challenge for conversational tools targeting absolute beginners—such tools presuppose minimal productive ability, which is precisely what learners at this level are developing.

Self-attribution patterns reinforced these distinctions: Among AI-migo users who acknowledged limitations, half (5 of 10) linked difficulties to prompting issues, while two attributed problems to typing effort. For XiSPA, two of five self-attributing students pointed to language skills limitations. System comprehension difficulties (17 for AI-migo; 8 for XiSPA), reported without recognizing potential user-side causes, could potentially reflect similar prompting or linguistic origins without students' awareness, though other explanations—such as genuine system limitations or reliability issues—are possible and require further investigation.

These findings suggest differentiated interventions. For AI-migo, reducing prompt construction burden through templated interactions or quick-access buttons for common functions could lower both prompt formulation and interaction modality effort barriers. For XiSPA, providing conversational scaffolding—such as suggested responses or vocabulary hints—could help students sustain interactions despite limited linguistic repertoire.

Enjoyment and Frustration

Near-identical enjoyment ratings for both tools (approximately 78% positive, $M \approx 4.05$) indicate that affective engagement was maintained regardless of pedagogical purpose or functional complexity.

The limited negative affect reported differed across tools. For AI-migo, frustration co-occurred with linguistic adaptation mismatches and communication failures, suggesting negative emotions emerged when expectations regarding linguistic appropriateness and communication success were unmet. This pattern aligns with research indicating that texts exceeding learners' comprehension capacity can provoke foreign language reading anxiety (Bahmani & Farvardin, 2017; Liu, 2025; Saito et al., 1999) and, in East Asian contexts, contribute to demotivation (Xie et al., 2018), reinforcing the importance of maintaining outputs within learners' appropriate comprehension level.

For XiSPA, one student characterized AI conversation as "boring." This isolated comment raises questions about a potential design tension for conversational agents targeting absolute beginners: linguistic and thematic constraints ensuring comprehensibility may reduce the

variety and naturalness that make conversation engaging. Whether this tension contributed to the reported boredom cannot be determined from the present data, but maximizing engagement within necessary linguistic boundaries represents an ongoing design challenge requiring further exploration.

Output Quality

Adaptation

Strong positive perceptions of output adaptation (approximately 80% for both tools) suggest students perceived outputs as well-aligned with SPA001 curriculum content at Week 12. For XiSPA, this finding provides preliminary validation of the controlled text generation prompt-based approach described in Getino-Diez & García-Madariaga (2026). For AI-migo, positive perceptions suggest general curricular alignment, though the specific dimensions assessed cannot be determined: the quantitative measure confounds multiple output types—including question responses and explanations that may not employ controlled generation. Despite this ambiguity, these results are encouraging given that providing appropriate-level Spanish exposure, practice, and support represents a primary goal of this implementation. Moreover, the scalability of the prompt-based approach could facilitate adaptation to the course progression through weekly updates to the knowledge bases, as well as extension to other modules.

Qualitatively, more comments reflected perceived adaptation than inadaptation (17 vs. 9 for AI-migo; 7 vs. 4 for XiSPA). However, positive perceptions remained largely underspecified, while negative comments provided concrete descriptions—primarily student-level linguistic mismatches (unfamiliar vocabulary, comprehension difficulty) for both tools and explanatory mismatches (difficulty understanding explanations) for AI-migo. This asymmetry may indicate that adaptation failures are more salient and memorable than successes. The presence of both positive and negative perceptions—including isolated reports of content being "too simple"—points to heterogeneity in learners' proficiency levels and varying degrees of knowledge, creating challenges for a single curriculum-aligned approach to accommodate the full spectrum of learner needs.

Three conclusions emerge from these patterns. First, a general perception exists across both tools that they adapt appropriately to learners or lessons, aligning with quantitative findings, though precise dimensions remain unspecified. Second, linguistic inadaptation for both tools and explanatory inadaptation for AI-migo occur for a subset of students, primarily reflecting outputs more difficult than expected. When AI-generated content proves inaccessible, the tools fail their fundamental purpose of providing accessible learning materials. Third, individual variability in proficiency and knowledge necessitates mechanisms beyond course-level alignment.

Future iterations should strengthen curriculum-level adaptation and incorporate mechanisms for individual-level adaptability. Curriculum-level improvements should ensure that linguistic complexity (both tools) and explanatory complexity (AI-migo) dynamically reflect what students are expected to comprehend at each stage of the course. Individual-level adaptivity might include: (a) user-facing controls such as adjustable difficulty settings or optional "challenge modes," (b) personalized learner profiles tracking mastered structures to introduce new content dynamically or reinforce gaps, (c) real-time adjustment controls for AI-migo to provide simpler or more complex explanations or linguistic output during practice, and (d) on-demand language support during XiSPA interactions, such as inline or pop-up translation or explanations of unfamiliar structures.

Accuracy

Accuracy was measured quantitatively only for AI-migo, assessing perceptions of correctness of answers and explanations. Although quantitative results suggested generally positive

perceptions (84.6%), qualitative data revealed divided views: 12 comments reported response inaccuracy and 6 noted inability to answer questions; conversely, 8 reported accuracy positively—5 of which co-occurred with convenience mentions, suggesting these dimensions may be perceived together.

The asymmetry observed in adaptation comments—where positive perceptions remained general while negative were more specific—appeared here as well, though less pronounced. The data provide limited information on whether differing accuracy perceptions arise from variations in question complexity, topic alignment with the knowledge base, individual prompting strategies, or differing expectations regarding acceptable accuracy levels. Two students demonstrated meta-awareness by attributing limitations to AI-migo's "knowledge" or "language model," recognizing system constraints related to accuracy.

Further investigation of specific inaccuracy instances would inform targeted interventions—including system prompt refinement to improve query interpretation, knowledge base expansion or reorganization in weak content areas, or clearer communication about system limitations.

For XiSPA, although accuracy was not explicitly measured, 4 comments addressed linguistic correctness positively, with no negative accuracy comments emerging. This suggests that stringent lexical and grammatical constraints may not have undermined perceptions of linguistic correctness in conversational output.

Naturalness

Naturalness was measured quantitatively only for XiSPA, assessing perceived naturalness in conversations. XiSPA received high positive ratings (79.5%), suggesting students generally perceived conversations as natural.

Qualitatively, however, naturalness comments were minimal for both tools (9 for XiSPA; 6 for AI-migo), highlighting concerns regarding conversational flow and perceptions of humanness. Naturalness must be interpreted differently across the two tools given their distinct pedagogical functions and design constraints.

For XiSPA, conversational naturalness faces two interrelated constraints. First, it is inherently limited by absolute beginners' restricted linguistic repertoire. Students lack the lexical and grammatical resources for the varied, spontaneous turn-taking, making fully natural dialogue difficult to achieve at this proficiency level. Second, maintaining long conversational memory and, consequently, better coherence across turns may compromise the controlled text generation approach, as increasing contextual information can make it more difficult to enforce strict lexical and grammatical constraints. Comments that reported conversational discontinuity, describing difficulty with longer conversations and reporting "weak memory," potentially reflected this trade-off. Perceptions of humanness were limited and mixed, including lack of human-like interaction and positive human-like experiences, suggesting that the concept of having a conversational partner—even an AI one—may have fostered a sense of social interaction for some learners.

For AI-migo, naturalness comments should be interpreted within its role as support and feedback provider rather than conversational partner. Concerns included a lack of human-like qualities such as emotion, empathy, and creativity, reduced engagement compared to face-to-face interaction, and conversational discontinuity. These responses suggest some students desired more human-like interactional qualities in a tutoring context, perhaps expecting the tool to emulate a real instructor. AI-migo's broader communicative context and less stringent linguistic constraints may provide more flexibility for addressing these concerns in future iterations.

For XiSPA, strengthening conversational continuity without compromising linguistic control remains important, and future iterations should explore mechanisms supporting smoother turn-taking or topic coherence maintenance. For both tools, features such as optional voice-

based interaction or visual embodiments (e.g., animated avatars) could enhance perceptions of human-like interaction. Further research should examine how naturalness relates to other dimensions such as motivation and engagement.

AI Tool Adoption, Usage Patterns, and Perceived Usefulness

Tool Adoption and Sustained Use

A majority of students used the tools for assessment preparation (68.1% for AI-migo; 64.2% for XiSPA), and substantial post-assessment continuation occurred (52.7% using at least one tool). These patterns suggest students found meaningful value in both systems. AI-migo's higher uptake likely reflects its broader functionality compared to XiSPA's primarily conversational format, indicating that higher functionality may encourage adoption even when ease of use is slightly lower.

More important than overall adoption rates is the consistency of engagement among adopters. Approximately 80% of users engaged with the tools at least weekly during assessment preparation, and similar sustained engagement continued post-assessment, suggesting both systems were meaningfully integrated into students' study routines rather than used sporadically. This sustained post-assessment use is particularly encouraging, as it suggests the tools may help address the fundamental challenge of limited out-of-class Spanish exposure and lack of appropriate learning materials. Students voluntarily maintained target language contact and practice even without external performance pressures.

Because engagement was entirely voluntary, quantitative use patterns provide indirect evidence of perceived usefulness—consistent with the TAM relationship between perceived usefulness and actual use (Davis et al., 1989). Qualitative descriptions reveal the specific dimensions of this usefulness.

Use Patterns and Functional Alignment

Qualitative use descriptions confirm that both tools were predominantly employed as intended. AI-migo served diverse functions—question-answering (most frequent, encompassing general inquiries and topic-specific questions focusing particularly on exam preparation, grammar, vocabulary, and information searches), error correction and feedback, practice generation, and content generation. XiSPA primarily supported conversational practice in various formats (e.g., topic-specific practice) and for different goals (e.g., exam preparation). This correspondence between intended and actual use suggests students recognized and leveraged the tools' primary affordances.

However, misuse patterns revealed incomplete understanding of tool-specific boundaries despite explicit introduction sessions. Instances of misuse appeared for AI-migo (attempting conversational practice) and for XiSPA (seeking feedback, information retrieval, and exercise generation). This boundary confusion may reflect: (a) insufficient internalization of functional distinctions during brief introduction sessions, (b) interfaces that did not adequately communicate functional boundaries, (c) convenience-driven behavior, with students using whichever tool was already open, or (d) genuine user preferences for one tool's interaction style adapted beyond design scope.

Additionally, the lower reporting of AI-migo's feedback and practice-generation features suggests these capabilities were less salient than question-answering. This pattern was particularly notable in post-assessment use, where feedback was not mentioned at all. Students may have primarily conceptualized AI-migo as a question-answering resource without fully recognizing its range of capabilities. Future implementations should provide more explicit communication of available functions—potentially through guided tutorials, classroom-integrated use cases and activities, and more prominent interface affordances for accessing different features.

Perceived Usefulness: Impact and Learning Benefits

Students' strongly positive perceptions of learning benefits provide important validation of the tools' value. Over 80% of users agreed that the tools positively impacted module results and post-assessment improvement, indicating that those who engaged with AI-migo and XiSPA generally viewed them as effective aids for both immediate academic performance and ongoing skill development.

From a pedagogical perspective, these perceptions suggest the tools successfully addressed their primary design objectives: expanding opportunities for meaningful language exposure and practice (both tools) while providing tailored support and feedback (AI-migo). However, whether these perceptions correspond to actual learning gains remains an empirical question requiring future investigation through controlled assessment of learning outcomes.

Qualitative benefit descriptions revealed that language learning and linguistic competence development represented the primary reported benefits for AI-migo (41.4% of benefit responses), while XiSPA emphasized skill development over knowledge learning, aligning with its design as a conversational practice partner. This distinction reflects different pedagogical roles: AI-migo serves as a source of explanation, feedback, and structured practice supporting knowledge learning; XiSPA provides an application context where students deploy and develop emerging communicative abilities.

The marked difference in practice-related mentions (2 for AI-migo vs. 10 for XiSPA) likely reflects XiSPA's explicit positioning as a conversation partner, making practice affordances immediately salient. As noted previously, AI-migo's practice generation and feedback capabilities may require more explicit communication or interface prominence.

Although some learners recognized these tools' potential for retention (both tools), assessment preparation (AI-migo), and information-seeking (AI-migo), these benefits were infrequently reported. Future iterations might reinforce these dimensions through intentional pedagogical integration. However, infrequent reporting of assessment preparation does not indicate low importance—as shown in the previous section, many assessment-related uses were reported.

Use descriptions and perceived benefits represent complementary perspectives on perceived usefulness. Use descriptions reveal concrete interaction patterns (question-asking, feedback-seeking, practice generation, and conversation practice) that demonstrate how students found the tools useful in practice, while perceived benefits reflect higher-order learning outcomes students associate with these activities—language learning and competence development. For AI-migo, students may view the tool instrumentally—specific activities collectively contribute to broader learning and development goals. For XiSPA, the pattern was mixed: conversation practice was both the primary activity and most frequently reported benefit, suggesting some students valued practice intrinsically; however, competence development also emerged substantially, indicating others viewed practice instrumentally as contributing to skill development rather than learning. This distinction highlights the different roles these tools play in students' learning processes and how each contributes to their Spanish development.

Overreliance: A Necessary Concern

Ten comments identified overreliance as a concern, raising issues about potential impact on critical thinking and excessive dependence on the tools. Although few in number, these concerns reflect important tensions in AI-supported learning that warrant attention.

On one hand, AI tools like AI-migo and XiSPA are explicitly designed to provide extensive support and practice opportunities; increased use could be interpreted as successful engagement rather than problematic dependence. On the other hand, excessive reliance could potentially undermine the development of autonomous learning strategies, reduce engagement with instructors and peers, or create dependency leaving students struggling

when AI support becomes unavailable. These concerns suggest future implementations should provide guidance to help students develop awareness of potential overreliance and maintain balanced use of AI tools alongside other resources.

Limitations

This evaluation has several limitations. First, the cross-sectional one-shot case design lacks a baseline or comparison group, capturing only initial perceptions following brief exposure rather than longitudinal usage or learning outcomes.

Second, all measures were self-reported, introducing potential bias. For example, self-reported perceptions of usefulness and impact may not correspond to assessment results or actual learning gains. Additionally, self-selection bias may have influenced findings, as participants may have differed from non-participants in motivation or attitudes.

Third, the voluntary nature of qualitative responses may have attracted students with strong views, and the substantial difference in open-ended response rates between AI-migo and XiSPA (313 vs. 215) complicates comparable interpretation.

Fourth, the end-of-semester implementation constrained evaluation scope. Although both tools were designed for semester-long deployment with progressive knowledge base updates, Week 12 introduction prevented evaluation of dynamic curriculum adaptation across a full semester.

Fifth, single-item measures, while justified for exploratory research and response burden minimization, provide less robust measurement than multi-item scales. Relatedly, survey ambiguities emerged, particularly regarding adaptation: the quantitative measure did not distinguish between linguistic and explanatory adaptation, and AI-migo's measure confounded multiple output types. Future research requires instruments that clearly distinguish between adaptation types and assess tool-specific quality dimensions separately.

Conclusions

This article presents the design, pilot implementation, and initial evaluation of AI-migo and XiSPA—two custom-built AI tools developed to address critical challenges in Spanish language instruction for absolute beginners at XJTLU. Initial findings suggest both tools have potential to expand Spanish exposure beyond the classroom, provide convenient on-demand access to level-appropriate materials and practice, and support assessment preparation. Students generally perceived both tools as easy to use, enjoyable, well-adapted to the curriculum, accurate (AI-migo), natural (XiSPA), and beneficial for learning, module performance, and continued improvement. Tool adoption was substantial, with continued post-assessment use indicating voluntary engagement beyond course requirements.

Qualitative findings, however, highlight areas requiring targeted refinement: improving technical reliability and accessibility, reducing prompt and interaction burden, strengthening both curriculum-level and individual-level adaptation, and enhancing conversational continuity and human-like qualities. Addressing these priorities will be essential not only for optimizing AI-migo and XiSPA within SPA001, but also for informing scalable integration of similar AI-driven tools across other modules and languages.

The complementary roles of AI-migo as a virtual tutor and XiSPA as a conversation partner suggest a promising approach for combining structured guidance with interactive practice. The prompt-based linguistic control approach enables efficient development, iterative refinement, and scalability without requiring extensive computational resources or technical expertise—critical factors for sustainable implementation in educational contexts.

Future implementations should deploy these tools from the beginning of the semester with weekly knowledge base updates, allowing full evaluation of their capacity for dynamic

curriculum alignment. Systematic assessment of their impact on actual learning outcomes, rather than perceptions alone, will be essential for understanding their pedagogical value.

The iterative implementation-evaluation-refinement approach adopted here represents a sustainable model for responsible AI integration in education: beginning with clear pedagogical needs, designing purpose-built solutions, piloting with systematic evaluation, refining based on evidence, and scaling progressively while maintaining ongoing assessment. As AI capabilities continue advancing, this evidence-driven approach offers a pathway for harnessing AI to enhance language learning and teaching in contexts where opportunities for authentic interaction and appropriate materials remain limited.

REFERENCES

Bahmani, R., & Farvardin, M. T. (2017). Effects of different text difficulty levels on EFL learners' foreign language reading anxiety and reading comprehension. *Reading in a Foreign Language, 29*(2), 185–202. https://doi.org/10.64152/10125/66912

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, *44*(2), 175–184. https://doi.org/10.1509/jmkr.44.2.175

Chang, C.-C., Yan, C.-F., & Tseng, J.-S. (2012). Perceived convenience in an extended technology acceptance model: Mobile technology and English learning for college students. *Australasian Journal of Educational Technology*, *28*(5), 809–826. https://doi.org/10.14742/ajet.818

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

Getino-Diez, R., & García-Madariaga, M. (2026). Linguistic control in AI text generation: An accessible prompt-based approach targeting L2 Spanish absolute beginners. *Technology in Language Teaching & Learning*, *8*, 103327. https://doi.org/10.29140/tltl.2026.103327

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge University Press.

Krashen, S. (1985). *The input hypothesis: Issues and implications* (Vol. 1). Longman

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.

Liu, Q. (2025). Reading anxiety in first and second language acquisition: A comprehensive literature review. *Journal of Modern Social Sciences, 2*(3), 226–233. https://doi.org/10.71113/JMSS.v2i3.329

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.

Nie, D., & Mavrou, I. (2025). Parents' views on Chinese young learners' foreign language learning attitudes and motivation: A mixed methods study. *Language Teaching*, 1–21. https://doi.org/10.1017/S0261444824000326

Saito, Y., Horwitz, E. K., & Garza, T. J. (1999). Foreign language reading anxiety. *Modern Language Journal, 83*(2), 202–218. https://doi.org/10.1111/0026-7902.00016

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal, 95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Sun, H., de Bot, K., & Steinkrauss, R. (2015). A multiple case study on the effects of temperamental traits in Chinese preschoolers learning English. *International Journal of Bilingualism*, *19*(6), 703–725. https://doi.org/10.1177/1367006914534332 (Original work published 2014)

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.

Tseng, W.-T., Liu, Y.-T., Hsu, Y.-T., & Chu, H.-C. (2024). Revisiting the effectiveness of study abroad language programs: A multi-level meta-analysis. *Language Teaching Research*, *28*(1), 156–200. https://doi.org/10.1177/1362168820988423 (Original work published 2021)

Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457–479. https://doi.org/10.1093/applin/ams074

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research, 11*(4), 342–365. https://doi.org/10.1287/isre.11.4.342.11872

W3Techs. (2025, October). *Usage statistics of content languages for websites*. https://w3techs.com/technologies/overview/content_language

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. *Ampersand, 14*, 100224. https://doi.org/10.1016/j.amper.2025.100224

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research, 16*(1), 85–102. https://doi.org/10.1287/isre.1050.0042

Xie, J., Wei, T., Zeng, Y., Lobsenz, J., & Chen, X. (2018). Learner perceptions of demotivators in the EFL classroom: Experiences of failure on learning outcomes. *The Journal of AsiaTEFL, 15*(2), 491–501. https://doi.org/10.18823/asiatefl.2018.15.2.17.491

Yoon, C. & Kim, S. (2007). Convenience and TAM in a ubiquitous computing environment: The case of wireless LAN. *Electronic Commerce Research and Applications, 6*(1), 102–112. https://doi.org/10.1016/j.elerap.2006.06.009