

卓越教创中心
西交利物浦大学

XJTLU

**AI-migo 与 XiSPA:
面向西班牙语绝对初学者的定制化 AI 驱动工具
——开发、试点实施与初步评估**

翻译：戴维萱、韩巍

Raúl Getino-Diez

Raul.GetinoDiez@xjtlu.edu.cn

Cristina García-Sánchez

Cristina.Garcia@xjtlu.edu.cn

Weixuan Dai (戴维萱)

Weixuan.Dai@xjtlu.edu.cn

LM-CEIE

目录

背景	1
AI 工具开发: AI-migo 与 XiSPA	2
试点实施	4
评估方法	5
定量结果	6
定性结果	9
讨论	15
局限性	20
结论	20
参考文献	21

推荐引用格式 (APA):

Getino-Diez, R., García-Sánchez, C., & Dai, W. (2025). *AI-migo 与 XiSPA: 面向西班牙语绝对初学者的定制化 AI 驱动工具——开发、试点实施与初步评估* (W. Dai & W. Han, Trans.). 卓越创新中心, 西交利物浦大学. <https://www.learningmall.cn/assets/Files/getino-diez-et-al-2025-CHI.pdf>



© 2025 Raúl Getino-Diez, Cristina García-Sánchez & 戴维萱。翻译: 戴维萱、韩巍。
本作品采用知识共享署名 4.0 国际许可协议进行许可:
<https://creativecommons.org/licenses/by/4.0/deed.zh-hans>

AI-migo 与 XiSPA: 面向西班牙语绝对初学者的定制化 AI 驱动工具 ——开发、试点实施与初步评估

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

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

Weixuan Dai (戴维萱) 
Weixuan.Dai@xjtlu.edu.cn

翻译: 戴维萱、韩巍

西交利物浦大学, 中国

摘要

人工智能 (Artificial Intelligence, AI) 正迅速重塑教育格局, 为二语习得中的创新提供了前所未有的机遇。在西交利物浦大学 (Xi'an Jiaotong-Liverpool University, XJTLU), 面向绝对初学者的入门级课程 SPA001: 西班牙语 1, 开发了两款专门用于满足学生学习需求的 AI 驱动工具: AI-migo——一款在多个学习维度提供定制化支持的虚拟导师, 以及 XiSPA——一款 AI 驱动的对话伙伴, 用于促进互动性强且水平适配的交际练习。两款工具均采用基于提示词的方法, 并结合结构化知识库, 以确保所提供的支持与课程进度保持一致。

本文系统介绍了这两款工具的研发背景、开发过程、试点实施以及初步评估结果。基于 385 名学生的混合方法评估显示, 学生在感知易用性、感知愉悦、输出质量以及感知学习效益等方面均持积极看法, 其中, 便利性尤为突出。评估显示, 学生有较高的工具使用率, 并在考试后仍持续使用。定性结果进一步揭示了若干亟需改进的方面, 包括系统可靠性与访问便利性、交互负担以及适配机制等。AI-migo 与 XiSPA 在功能上的互补性, 结合基于提示词的方法, 表明其在初级语言教育中实现可扩展、资源高效的 AI 融合具有良好潜力。

背景

SPA001: 西班牙语 1 是西交利物浦大学 (Xi'an Jiaotong-Liverpool University, XJTLU) 开设的入门级西班牙语课程, 面向绝对初学者, 课程目标是在一个学期内达到欧洲语言共同参考框架 (Common European Framework of Reference, CEFR) A1- 水平。该课程每年服务约一千名学生, 其中大多数此前从未接触过西班牙语。课程通过循序渐进、以技能建构为导向的课程设计, 为学生提供基础性的语言学习支持。

尽管该课程在实践中取得了一定成效, 但仍面临两个主要挑战: 一是学生在课堂之外接触西班牙语的机会有限, 二是缺乏在教学法上适切、且能够与学生水平等级和课程进度相匹配的学习材料。

课堂之外西班牙语接触的不足

在中国, 西班牙语通常作为外语在学校情境中学习, 学习者的日常生活环境中几乎不使用该语言。学界普遍认为, 相较于课堂内的外语教学, 海外留学能够带来更显著的语言学习成效 (参见 Tseng et al., 2021 的多层次元分析)。这一观点与二语习得理论相一致, 这些理论强调目标语接触在语言习得中的关键作用: 通过提供可理解输入 (Krashen, 1985)、有意义输出 (Swain, 1985), 以及促进意义协商的互动情境 (Long, 1996), 支持语言发展。

在目标语并非社会主导语言的情境中, 课堂之外的语言接触总体上仍然有限, 但依然可能对语言发展产生实质性贡献。研究表明, 数字游戏、阅读以及互联网使用等活动, 在英语作为外语的学习中, 具有提升学习者词汇量、口语能力, 并在某些情况下增强学习动力的潜力 (参见 Sundqvist & Sylvén, 2016, 第五章, 对相关效应的综合实证研究)。然而, 由于英语在 Kachru (1985) “扩展圈”框架中所具有的全球语言地位, 其他外语的学习情境通常与英语存在显著差

异。英语学习者能够较为容易地接触多样化的媒体资源以及不断扩大的全球使用者群体。在中国，这一点尤为明显：近年来，儿童从学前阶段开始接触英语的程度显著提高（Nie & Mavrou, 2025; Sun, de Bot & Steinkrauss, 2015），同时，高等教育中英语授课（EMI）也迅速发展（Zheng & Choi, 2025）。

相比之下，课堂之外接触西班牙语以及进行情境化语言练习的机会仍然相当有限。这种差异也体现在两种语言的网站使用比例上：在本研究开展时，英语用于全球 49.2% 的网站内容，而西班牙语仅占 6%（W3Techs, 2025）。鉴于课外接触与练习在支持语言发展中的核心作用，这种有限的课外机会尤为令人担忧。

缺乏适合的学习材料

尽管目标语接触对语言习得至关重要，但对于绝对初学者而言，输入内容必须经过谨慎的水平适配——尤其是在缺乏课外练习机会、且学习者词汇和语法知识极为有限的情况下。Krashen 的可理解输入假说（1985）指出，当语言输入略高于学习者当前水平时，学习效果最佳，这一原则即所谓的 $i+1$ 原则。相关实证研究进一步阐明了 $i+1$ 原则的具体内涵：至少达到 95% 的已知词汇覆盖率（Laufer, 1989），才能实现对文本的最低理解，而实现独立理解则需约 98%（Hu & Nation, 2000; Schmitt et al., 2011）；在听力理解中，所需的词汇覆盖率约为 90%–95%（Van Zeeland & Schmitt, 2013）。

此外，语言难度超出学习者已有知识范围的文本，可能引发外语阅读焦虑（Bahmani & Farvardin, 2017; Liu, 2025; Saito et al., 1999），而在东亚环境中，这将进一步导致学习动力不足。研究表明，在西方环境中，教师相关因素导致的学生学习动力不足占主导，而与此不同的是，在东亚地区，学习失败的体验——具体表现为陌生词汇阻碍理解、词汇和短语记忆困难、对课堂内容或语法理解不足，以及考试成绩偏低——是学习动力不足的主要来源（参见 Xie et al., 2018 的相关讨论）。

因此，为了促进语言学习、降低潜在焦虑并解决缺乏学习动力的问题，SPA001 的学习材料应当遵循 $i+1$ 原则，并与课程进度保持系统性的同步推进。然而，尽管近年来已开展了一些专门材料的开发工作，明确针对 XJTLU 西班牙语绝对初学者课程体系、且在水平上高度适配的教学资源仍然相对匮乏，其主要原因在于此类材料的开发需要投入大量时间与资源。

AI 工具开发：AI-migo 与 XiSPA

理由

生成式人工智能——尤其是 AI 聊天机器人——在二语习得领域展现出显著潜力，能够通过提供定制化支持、增加练习机会以及促进学习者自主性来增强语言学习效果。相关研究进一步表明，聊天机器人可通过营造支持性、低压力的学习环境来降低学习焦虑，并通过游戏化等功能提升学习动力（参见 Wiboolyasarin et al., 2025，有关语言教育中聊天机器人的系统性综述）。

在此背景下，SPA001 在探索创新教学路径方面具有关键的试点意义。该课程规模大、处于语言学习的起始阶段，且具备良好的可扩展性，使其成为测试 AI 增强型工具的理想场域，以支持更加投入、自适应且高效的语言学习。

目标

为应对 SPA001 面临的两大核心挑战——课堂之外西班牙语接触有限以及水平适配材料的匮乏——本研究开发了两款 AI 驱动工具：AI-migo（虚拟导师）和 XiSPA（互动式对话伙伴）。两者共同拓展了有意义的目标语接触与练习机会，其中 AI-migo 还提供定制化支持与个性化反馈。

除满足上述直接教学需求外，这两款工具的开发亦服务于更广泛的教学目标：促进学习者参与度与学习者自主性，通过提供针对考试的练习来帮助学生取得更好的成绩，并确保在相似教育情境中的可扩展性。从可用性角度看，两款工具均提供全天候（24/7）的自主学习支持，同时也为教师在课堂内外设计针对性练习提供了灵活手段。

鉴于语言学习的情感维度，AI-migo 与 XiSPA 融合了类人特征——包括易于产生亲近感的名称、视觉呈现以及自然的对话风格——以培养学习者的个人联结感，使学习体验更加亲和且富有吸引力。

工具描述与功能

AI-migo: 虚拟西班牙语导师

AI-migo 将“AI”（Artificial Intelligence，人工智能）与西班牙语中意为“朋友”的 *amigo* 相结合，唤起一种值得信赖的陪伴者形象，在学习过程中为学生提供引导、鼓励与支持。作为一名虚拟导师，AI-migo 主要提供以下三类功能：（1）即时语言支持，回应学生在语法、词汇、句法结构及其他课程相关主题方面的提问；（2）针对写作任务的个性化反馈，识别错误、提出改进建议，并解释修改依据；（3）定制化练习活动——如词汇和语法练习——以强化需要进一步巩固的薄弱环节。

AI-migo 所提供的讲解、反馈、练习与示例均与 SPA001 的课程进度保持一致。该工具采用富有吸引力的互动风格，模拟支持性的教师角色，旨在营造一种让学生能够安心、积极地进行语言尝试并发展自身能力的学习环境。

XiSPA: AI 驱动的对话伙伴

XiSPA 由“Xi”（取自西交利物浦大学）与代表西班牙语的“SPA”组合而成，其名称在语音与语义上联想到西班牙语中的 *chispa*（“火花”）。这一联想既传达了归属感，也体现了语言学习所需的活力、好奇心与热情。作为一款 AI 驱动的对话伙伴，XiSPA 通过围绕个人信息、兴趣爱好、天气、服饰等主题的交流，引导学生开展同步文本对话。这些互动为情境化语言使用提供了机会，并促进交际能力的发展。通过在有意义的交际语境中强化核心词汇与语法结构，XiSPA 旨在帮助学习者在语言产出的初始阶段建立信心，并在支持性、非评判性的环境中降低焦虑。

XiSPA 的所有互动内容均与 SPA001 所涵盖的词汇、语法及交际内容保持一致，使学生在整个学期的学习过程中持续接触符合 *i+1* 原则的最优输入水平。

工具设计

语言控制合规性

利用人工智能生成练习内容——如语法与词汇活动、阅读理解任务或会话练习——要求对语言进行高度控制，使其严格遵循词汇与语法参数，以与学习者的水平等级和课程进度保持一致。AI-migo 在生成练习活动和示例时需要受控文本生成，而 XiSPA 则要求在每一次对话中持续保持语言控制。然而，实现如此程度的语言控制仍然是一项重大挑战。

受控文本生成（Controlled Text Generation, CTG）技术已在其他形式的语言控制中被证明行之有效，并可被调整以服务于上述目的。然而，这类技术通常需要较高的专业技术能力和大量计算资源（概述性讨论可参见 Getino-Diez & García-Madariaga, 2026）。为缓解这些限制，本文采用了 Getino-Diez 与 García-Madariaga（2026）提出的基于提示词的方法。该方法具有多方面优势：其一，内嵌明确机制以强制遵循为绝对初学者量身定制的预设语言约束；其二，资源高效，支持工具的迭代优化；其三，易于扩展至其它初级水平的课程。

在该方法中，提示词整合了增强型技术，用以确保生成内容严格遵循结构化知识库文件中所规定的词汇约束和语法约束。提示词与知识库具有两项关键特性：（1）可在课程推进过程中随时更新，而无需进行模型微调；（2）可无缝集成至 AI 聊天机器人和对话代理中。这种动态方法构成了 AI-migo 与 XiSPA 的核心设计基础，使二者能够通过定期更新持续与课程进度保持对齐，并顺利集成于 XIPU AI 环境中的聊天机器人系统。

系统架构

AI-migo 与 XiSPA 均部署于 XIPU AI Agent 1.0 中，该系统是西交利物浦大学（XJTLU）校级 XIPU AI 平台的组成部分。XIPU AI Agent 1.0 通过系统提示词引擎与知识库联动，提供一个用于开发与部署定制化智能聊天机器人的集成环境，使教师无需具备高级技术专长即可创建面向特定领域的教学型智能体。

每个聊天机器人均由系统提示词和与西班牙语课程内容同步的精选知识库共同构成。

系统提示词。 系统提示词用于确保各聊天机器人与其既定教学功能保持一致。针对需要高度受控文本生成的任务（如词汇与语法活动或会话练习），提示词相应部分整合了多种维持词汇与语法控制的关键策略，包括：通过人设模式为 AI 模型设定特定角色；采用类 Markdown 的格式以突出提示词的结构层级；通过语境预设将输出限定在绝对初学者所特有的语言范围内；冗余强化规则（将肯定性指令与其对应的否定性指令成对呈现）；对高频出现的不符合课堂所学交际功能的语言结构进行明确否定；以及在提示词末尾设置强化段落，重申关键约束条件（详见 Getino-Diez & García-Madariaga, 2026 中的具体描述）。

知识库。 XiSPA 的知识库结构遵循 Getino-Diez 与 García-Madariaga（2026）所描述的框架，包括：（1）**词汇表**——按类别整理的完整词汇清单；（2）**语用功能与例句文档**——按交际功能系统整理的大量语言结构示例。AI-migo 在生成水平适配练习活动时同样使用上述两类文件，此外还包含一份**语法规则与例句文档**，其中汇集了与 SPA001 课程同步的详细语法讲解与示例。该文件使 AI-migo 能够回答语言相关问题，并提供与课程进度相匹配的反馈。

系统更新。 知识库涵盖了 SPA001 在学期中每个阶段所教授的内容。通过每周将新学习的内容纳入知识库，可使两个智能体与课程进度同步演进。系统提示词同样需要定期更新，因为不合规范语言结构会随着学期推进而变化，而提示词技术——尤其是对高频不合规结构的明确否定——也需随之进行相应调整。

试点实施

在 2024–25 学年（第一学期）第 10 与第 11 周，研究团队首先开展了一项包含 32 名学生的前测。该初步阶段旨在全面推广至整个课程之前，识别关键问题并验证系统功能的可行性。

在第 12 周（即教学的最后一周），AI-migo 与 XiSPA 通过课堂时间内的一次专门介绍环节，正式向修读 SPA001 的 874 名学生推出。该环节旨在帮助学生熟悉系统，并展示如何将其融入日常学习流程。介绍内容包括对工具功能的整体概述，随后通过引导式活动让学生亲自使用工具，探索其在准备口语测验和期末考试中的潜在用途。

本次试点实施的主要目标在于：为学生提供使用工具的实践体验，探索其对学习支持的潜在价值，并在初始部署阶段识别任何即时的可用性技术问题。尽管两款工具在设计上均通过学期内逐步更新知识库以适应课程节奏，但在本次试点中，其配置专门针对第 12 周的课程内容进行了优化。

评估方法

研究设计

本次评估旨在快速获取具有可操作性的反馈，了解学生对 AI-migo 与 XiSPA 的首次使用体验，考察他们的初步感知并自述在备考和考试后对工具的使用情况。本评估并非用于检验假设或理论，而是采取描述性与形成性导向，为后续优化提供依据，并为未来学期的更大范围实施提供指导。

本次评估是一个迭代式实施-评估-改进循环的一部分。鉴于教学周期较为紧凑、技术本身具有新颖性，且本阶段以形成性目标为主，研究采用了探索性的横断面单次案例设计，以捕捉工具部署初期的数据。学生在首次接触工具数周后完成了一次事后问卷。

研究采用混合方法评估，以定量调查问题作为主要数据来源，并通过开放式问题补充定性洞见。

参与者

所有注册 SPA002 的学生均在 2024–25 学年第二学期第 1 周受邀参与本研究。此前，这些学生已在第一学期末修读 SPA001 期间接触过 AI-migo 与 XiSPA。问卷参与为自愿且匿名。调查开始前已获取参与者的知情同意，研究亦获得西交利物浦大学（XJTLU）伦理委员会的伦理审批。

测量工具

为反映 AI 工具的独特特征及试点实施情境，本研究开发了一份在线问卷。

定量部分通过自我报告的李克特量表题、频率题以及是/否题，测量学生的感知与实际使用情况。李克特量表与频率题均采用五点量表，数值越高表示感知越积极或使用频率越高（5 = 非常积极/非常频繁，1 = 非常消极/非常不频繁）。

各构念主要采用单项指标进行测量，以降低受访者负担并最大化问卷完成率。这一做法既符合本次评估的探索性与形成性特征，也有研究证据表明，对于受访者易于理解的具体、单一构念，单项指标已足以进行有效测量（Bergkvist & Rossiter, 2007）。

问卷设计以技术接受模型（Technology Acceptance Model, TAM; Davis et al., 1989）为理论框架，将感知易用性与感知有用性纳入其核心构念。其中，感知有用性具体化为学生对 AI 工具在课程成绩与考试后提升方面影响的感知。为考虑语言学习中的情感因素，鉴于其与感知易用性之间已有明确关联（Venkatesh, 2000），本研究还纳入了感知愉悦。基于 Wixom 与 Todd（2005）将质量信念与技术接受相整合的框架，研究进一步考察了三个工具特定的质量维度：课程适配，用以反映对生成内容与课程同步性的感知；回答准确性，用以衡量对 AI-migo 输出精确性的感知；以及回复自然度，用以表示学生对 XiSPA 交互自然性的感知。最后，实际使用用于测量学生是否使用以及在备考及考试后的学习中对工具的使用频率。

定性部分包含九个可选的开放式问题，允许学生进一步阐述其使用体验。针对每一款 AI 工具，问题分别涉及：（a）自我报告的技术问题，（b）感知到的优势，以及（c）感知到的挑战或局限性。这六个问题旨在为感知易用性、感知愉悦、课程适配、回答准确性与回复自然度，以及整体有用性等维度提供情境化与解释性的补充洞见。此外，另设三道问题聚焦实际使用情况：

（a）学生如何在学期末考试活动中使用 AI-migo，（b）学生如何在这些考试活动中使用 XiSPA，以及（c）学生在完成考试后如何使用这两款工具。

实施过程

问卷通过学校的学习管理系统 Learning Mall Core 发放，并在学生既定的课堂授课时段内完成填写。

数据分析

在定量数据分析方面，采用描述性统计方法。鉴于五点李克特量表与频率题在测量层级上具有序数但近似等距的特性，正文中提供了均值与标准差，相关图表中则提供中位数以供参考。

在定性数据分析方面，在问卷构念所确立的理论框架内开展主题分析，通过系统化编码识别反复出现的主题与模式。定性分析为定量结果提供了补充性洞见，有助于对其进行丰富与情境化解读。然而，由于开放式问题为自愿作答，回答具有选择性，选择发表评论的学生相较于整体样本，可能持有更为强烈的正向或负向观点。因此，定性陈述应被视为说明性与探索性的材料，而非对定量结果的代表性验证。此外，由于单条评论往往包含多个观点，部分回答在分析中被赋予多个编码；因此，各主题的出现频率不必然对应评论总数。

定量结果

共有 385 名学生完成了接触工具后的问卷。样本为 18–20 岁的大学一年级本科生 ($M = 18.53$, $SD = 0.60$)。几乎所有参与者 ($n = 383$, 99.5%) 的母语为中文；另有两名参与者报告其第一语言为其他语言（英语， $n = 1$ ；印尼语， $n = 1$ ）。

无论学生在备考阶段是否实际使用过工具，所有参与者均对 AI-migo 和 XiSPA 的各项感知维度进行了评分。使用频率与感知影响仅向自我报告为使用者的学生收集。

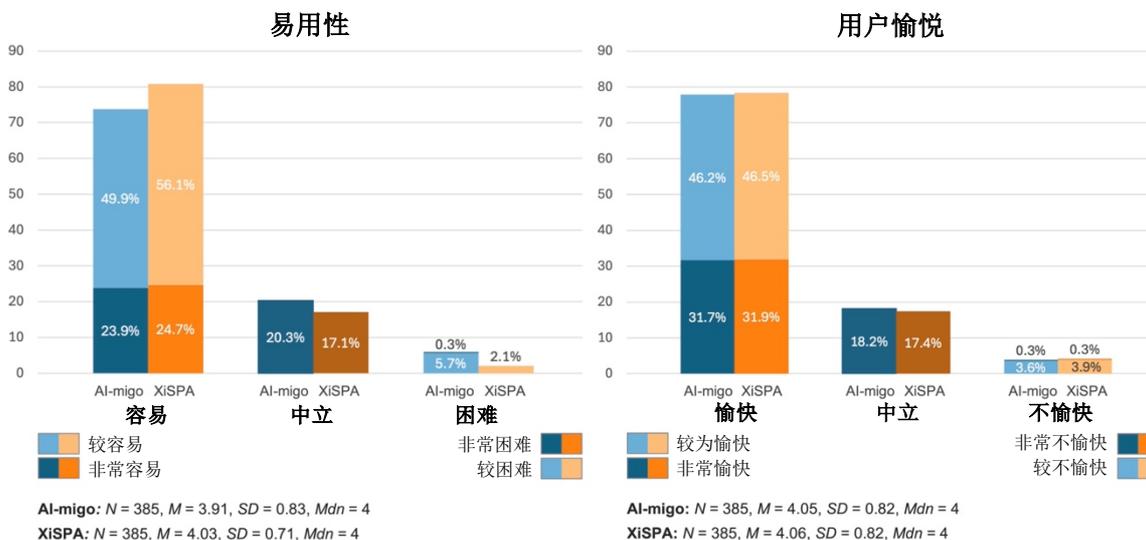
下文从描述性角度概述参与者在关键维度上的感知结果——包括感知易用性、感知愉悦、输出质量与感知影响——以及自我报告的工具使用模式。

易用性与用户愉悦

参与者普遍认为两款工具易于使用，其中 XiSPA 的评分略高（见图 1）。在 AI-migo 方面，73.8% 的受访者将该工具评为“非常容易”或“较为容易” ($M = 3.91$, $SD = 0.83$)；XiSPA 的相应比例为 80.8% ($M = 4.03$, $SD = 0.71$)。对“困难”或“非常困难”的评价比例较低：AI-migo 为 6.0%，XiSPA 为 2.1%。

两款工具在感知愉悦方面的评价同样积极，且结果几乎一致。就 AI-migo 而言，77.9% 的参与者将其评为“非常愉悦”或“较为愉悦” ($M = 4.05$, $SD = 0.82$)；XiSPA 的相应比例为 78.4% ($M = 4.06$, $SD = 0.82$)。对愉悦感的负面评价比例较低，AI-migo 为 3.9%，XiSPA 为 4.2%。

图 1
AI-migo 与 XiSPA 的感知易用性与用户愉悦评分

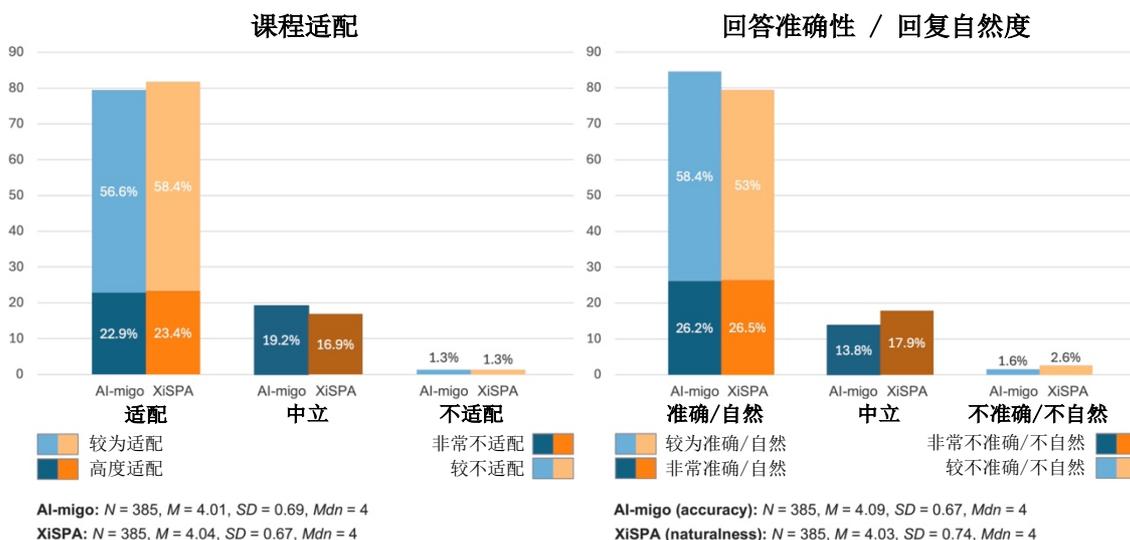


输出质量：适配、准确性与自然度

在所有测量的维度上，参与者对输出质量的感知总体呈现积极结果（见图 2）。在课程适配方面，79.5%的学生将 AI-migo 评为“高度适配”或“较为适配”于 SPA001 的课程内容（ $M = 4.01, SD = 0.69$ ），而对 XiSPA 给出相同评价的比例为 81.8%（ $M = 4.04, SD = 0.67$ ）。两款工具被评为“适配较差”的比例均较低（均为 1.3%）。

在工具特定的质量维度上，84.6%的受访者认为 AI-migo 的输出“非常准确”或“较为准确”（ $M = 4.09, SD = 0.67$ ），仅有 1.6%的参与者给出了准确性的负面评价。类似地，79.5%的参与者将 XiSPA 的对话回复评价为“非常自然”或“较为自然”（ $M = 4.03, SD = 0.74$ ），对回复自然度给出负面评价的比例为 2.6%。

图 2
AI-migo 与 XiSPA 的课程适配及回答准确性 / 回复自然度评分

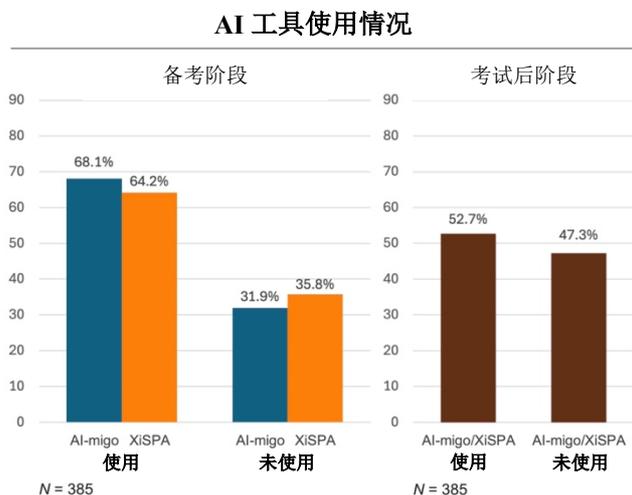


AI 工具使用情况

多数学生报告在备考阶段使用了 AI 工具：68.1% (n = 262) 的学生使用了 AI-migo, 64.2% (n = 247) 的学生使用了 XiSPA (见图 3)。在完成考试之后，学生的使用参与度仍然较高：52.7% (n = 203) 的学生在考试结束后仍持续使用至少一款工具。

图 3

AI-migo 与 XiSPA 的 AI 工具使用情况

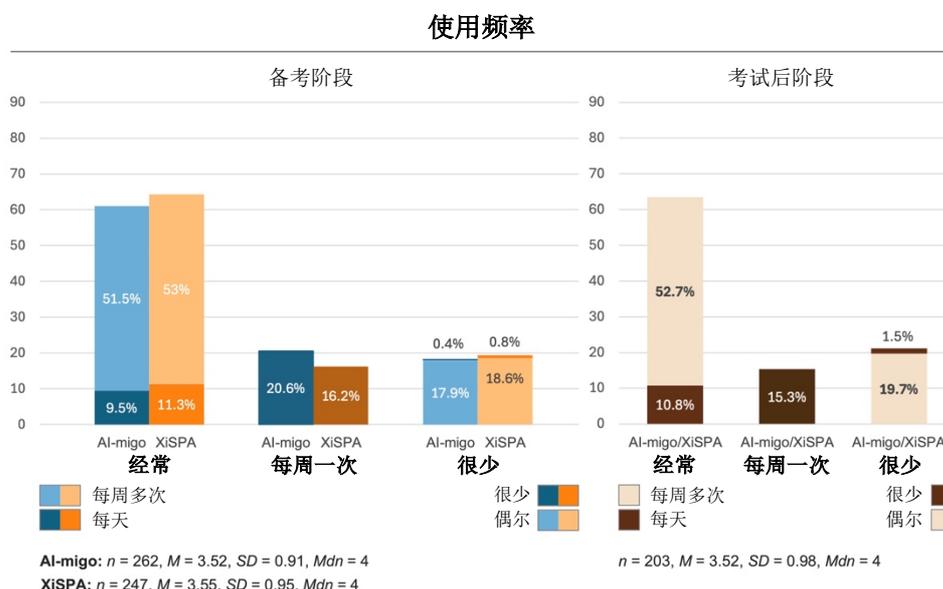


在采用这些工具进行备考的学生中，使用参与频率整体较高（见图 4）。在 AI-migo 使用者中，61.0% 的学生报告其使用频率为“每天”或“每周多次” (M = 3.52, SD = 0.91)；XiSPA 使用者中，有 64.3% 的学生报告了类似的使用频率 (M = 3.55, SD = 0.95)。若将“至少每周使用一次”的学生纳入统计，比例进一步上升：AI-migo 为 81.6%，XiSPA 为 80.5%，分别占总体样本的 55.6% 和 51.6%。

在考试结束后继续使用工具的学生中，使用频率同样保持在较高水平。在 203 名至少使用过一款工具的受访者中，63.5% 的学生表示其使用频率为“每天”或“每周多次”，78.8% 的学生表示“至少每周使用一次” (M = 3.52, SD = 0.98)，相当于全部参与者的 41.5%。

图 4

AI-migo 与 XiSPA 的使用频率评分

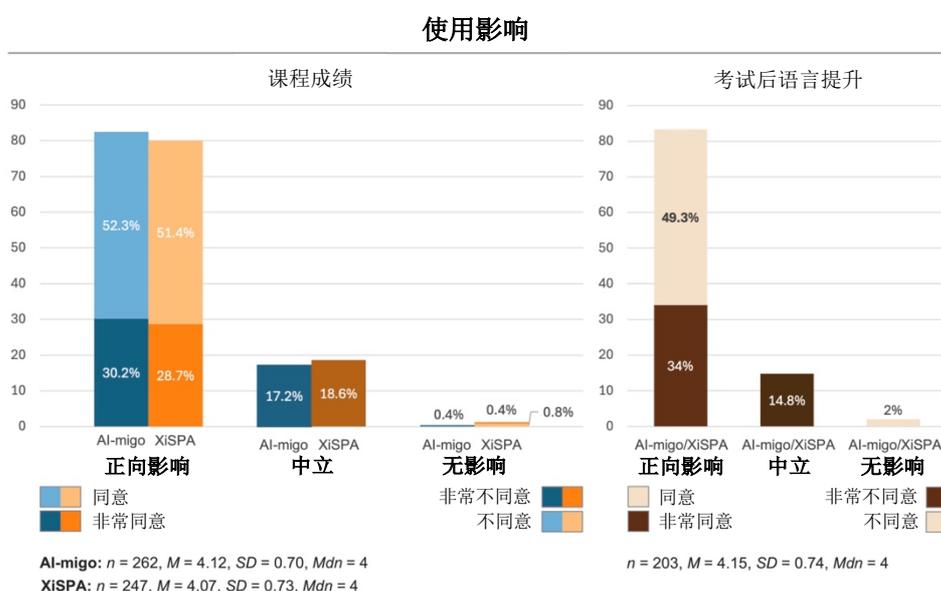


AI 工具使用的影响

在备考阶段使用过工具的参与者，对其对课程成绩的影响给出了高度积极的评价（见图 5）。在 262 名 AI-migo 使用者中，82.5% 的学生“非常同意”或“同意”该工具对其课程成绩产生了积极影响（ $M = 4.12$, $SD = 0.70$ ），仅有 0.4% 的学生报告了负面影响。类似地，在 247 名 XiSPA 使用者中，80.1% 的学生对其积极影响表示“非常同意”或“同意”（ $M = 4.07$, $SD = 0.73$ ），报告负面影响的比例为 1.2%。

对于考试后语言提升的感知同样呈现出积极结果。在考试结束后仍持续使用至少一款工具的 203 名学生中，83.3% 的学生“非常同意”或“同意”工具使用有助于其语言能力的持续提升（ $M = 4.15$, $SD = 0.74$ ），仅有 2.0% 的学生报告了负面影响。

图 5
AI-migo 与 XiSPA 对课程成绩及考试后提升的感知影响



定性结果

共收集到 686 条评论，其中 32 条因不符合问题意图或无法合理解读而被剔除，最终保留 654 条用于分析。由于所有开放式题目均为自愿作答，各题目的回答数量差异较大（每题 30–115 条不等）。

六个关于学生感知的开放式题目的回答情况为理解研究结果提供了重要背景。在涉及技术问题的回答中，AI-migo 有 92 名作答者，其中 60 人（65.2%）表示未遇到技术问题；XiSPA 有 68 名作答者，其中 56 人（82.4%）表示未遇到技术问题。在涉及挑战或局限的回答中，AI-migo 有 106 名作答者，其中 29 人（27.4%）表示未感知到任何挑战或局限；XiSPA 有 73 名作答者，其中 32 人（43.8%）表示未感知到挑战或局限。总体而言，两款工具中，学生报告挑战或局限的频率均高于报告技术问题的频率：AI-migo 为 77 条对 32 条，XiSPA 为 41 条对 12 条。就优势而言，AI-migo 共报告 115 条，XiSPA 为 74 条，均高于技术问题、挑战与局限三类合计的数量（分别为 109 条和 53 条）。

关于自我报告使用情况的三个题目中，涉及备考用途的回答为：AI-migo 58 条、XiSPA 38 条；另有 30 条回答描述了考试后对两款工具的使用情况。总体来看，针对 XiSPA 定性问题的回答少于 AI-migo（感知类题目分别为 215 条与 313 条；使用描述分别为 38 条与 58 条）。

在后续分析中，百分比仅基于具有实质内容的回答计算，“无问题”类回答不计入在内。在保留的 654 条回答中，共有 477 条包含实质性内容。工具层面的百分比基于与该工具相关的实质性评论计算：AI-migo 为 282 条，XiSPA 为 165 条。

易用性

便利性作为主要话题

便利性是 AI-migo 最为突出的主题，共出现于 49 条评论中（占 AI-migo 优势类回答的 42.6%；占 AI-migo 实质性回答的 17.4%），同时在 XiSPA 中也较为显著，出现于 22 条评论中（占 XiSPA 优势类回答的 29.7%；占 XiSPA 实质性回答的 13.3%）。就 AI-migo 而言，便利性涵盖多个维度：时间灵活性（“全天可使用” [“24h accessible”]）、空间灵活性（“我们可以在任何地方学习” [“we can study wherever we want”]）、即时性（“即时回复” [“instant reply”]），以及获取学术支持的便利性（“节省去老师办公室的时间” [“saving time of going to the professor’s office”]）。相比之下，XiSPA 中有关便利性的评论主要集中在即时性和练习可获得性（“当我找不到西班牙语使用者时” [“when I can’t find the Spanish speakers”]）。

技术与访问壁垒

系统可靠性问题在两种工具中的出现频率相近：AI-migo 有 21 条评论提及（占实质性回答的 7.4%），XiSPA 有 12 条（占 7.3%）。学生描述的问题主要分为三类：系统崩溃（“可能会经常崩溃” [“may break down frequently”]）、连接问题（“经常显示加载错误” [“frequently displaying loading errors”]），以及界面故障（“输入的文字会消失” [“text typed will disappear”]）。

访问壁垒出现频率较低，AI-migo 有 7 条评论、XiSPA 有 2 条，主要将工具描述为“很难找到” [“hard to find”]。此外，有 4 名受访者明确提出希望提供独立的移动端应用程序。

使用负担与交互挑战

除技术问题外，交互层面的挑战因工具功能不同而呈现出差异。

对于 AI-migo，使用负担相关内容共出现于 17 条评论中（占实质性回答的 6%），主要可归纳为三类：

- (a) 系统导致的负担：技术不稳定导致需要反复输入（“不得不一遍又一遍地输入我的问题” [“to type my question again and again”]）。
- (b) 交互模态负担：学生提到打字本身带来的困难（“打字很慢” [“print words slowly”]），并希望提供文本输入以外的替代方式。
- (c) 提示词构造负担：用户在有效表达问题时感到困难（“需要思考怎样才能让 AI 理解我的指令” [“think about how to make AI understand my order”]）。

相比之下，XiSPA 中与使用负担相关的评论出现频率较低（4 条），但由于其对话式特性，所反映的挑战有所不同：

- (a) 系统导致的负担（与 AI-migo 类似）；
- (b) 二语能力受限所致的负担（“不知道该如何和 [XiSPA] 进行对话，因为我的语言能力限制了我” [“don’t know how to make a conversation with [XiSPA] because my language skill limit[s] me”]）。

自我归因的情况——即回答中明确将困难归因于自身限制的情况——显示出明显的工具差异。对于 AI-migo，自我归因出现在 10 条评论中：其中 5 条将困难归因于提示词构造负担和对工具理解不足（“如何高效地向 AI 提问” [“how to ask AI efficiently”]），2 条归因于交互模态负担（其中一名学生承认自己“懒得把复杂的问题打出来” [“too lazy to type down my complicated

questions”])，另有 3 条归因于总体上的不熟悉。对于 XiSPA，自我归因出现在 5 条评论中：其中 2 条将困难归因于二语能力受限（“我的语法” [“my grammar”]），3 条归因于对工具的不熟悉。

与此相关，AI-migo 有 17 条、XiSPA 有 8 条评论描述了工具理解失败的情况，但不能排除潜在的用户端原因。这些表述包括理解困难（“有时它无法理解我的问题” [“sometimes it cannot understand my question”]）、理解错误（“误解了目的” [“misunderstanding the purpose”]）、在复杂表达上的困难（“如果我问一些比较复杂的问题” [“if I ask something difficult”]），以及对工具能力不足的感知（“不够聪明” [“not smart enough”]）。

愉悦感

与愉悦感相关的评论数量较少，总计仅 10 条：其中 5 条体现了愉悦体验，将工具描述为“有趣”或“很好玩” [“fun / “interesting”]；另有 5 条反映了积极的学习动机，认为工具“有助于提升我对学习的兴趣” [“helping to improve my interest in learning”]。

负向情感出现频率更低，但表现得更为具体。在 AI-migo 中，有 5 条评论提及挫败感，其出现往往伴随语言适配不匹配（“太难理解” [“too hard to understand”]），以及与提示词构造困难或系统限制相关的工具理解失败。相比之下，XiSPA 仅有 1 条评论提到无聊感（“和 [AI] 聊天很无聊” [“boring with talking to [AI]”]）。

适配性

与易用性相比，涉及适配的评论相对较少：AI-migo 共 28 条（占 AI-migo 实质性回答的 9.9%），XiSPA 共 11 条（占 XiSPA 实质性回答的 6.7%）。在 AI-migo 的评论中，17 条反映了感知适配性，9 条描述了不适配；XiSPA 呈现出相似的情况（7 条 vs. 4 条）。鉴于两种工具在教学功能上的差异，以下将分别阐述相关研究结果。

AI-migo

对 AI-migo 适配类评论的解读受到问卷设计的影响而较为复杂：开放式问题并未针对“适配”的任何具体维度进行提问。适配既可能涉及输出的性质——包括语言适配（西班牙语词汇、语法与交际功能的难度与恰当性）或解释性适配（系统如何调整解释与反馈）——也可能涉及适配的对象——课程层面（与课堂内容的对齐）或学生层面（针对个体水平或知识的定制）。尽管受访者在评论中通常能够明确指出适配对象，但许多表述——如“适配我的学习水平” [“adapted to my learning level”] 或“贴近课程” [“close to the lesson”]——并未明确其所指的是语言适配、解释性适配，还是二者并存的适配。因此，只有包含明确指示的评论才被编码为语言适配或解释性适配，其余均归类为一般性适配。

一般性的学生层面适配最为常见，共出现于 10 条评论中，描述系统“符合我的水平” [“matches my level”]。一般性的课程层面适配也出现于 6 条评论中，例如“为课程量身定制” [“tailored to the lesson”]。

关于不适配的报告则更为具体。学生层面的语言不适配出现在 4 条评论中，受访者认为 AI-migo 的西班牙语因不熟悉的词汇而过难（“有些词很难理解” [“hard to understand some words”]），或存在整体理解困难（“我不太能理解的语言” [“language I didn't understand very well”]）。学生层面的解释性不适配出现在 3 条评论中，指出难以理解系统的回答（“并不太符合我水平的回答” [“answers that are not well adapted to my level”]）。

此外，还识别出个别语言相关案例：一条评论提到学生层面的语言适配（“回答符合你的语言能力” [“answers that match your language skills”]），而另有两条描述了适配失败——一条认为西班牙语“相当简单” [“quite simple”]，另一条描述了词汇循环释义困境（“使用课堂未教授的

生词，如果我说‘我不理解’，它又会使用另一个不熟悉的词” [“uses unfamiliar vocabularies that are not taught in class, and if I say ‘No comprendo’, it will use another unfamiliar word”]), 表明即便在已明确表达困惑的情况下，系统仍未能进行调整。

XiSPA

由于 XiSPA 的唯一功能是维持用于交际练习的书面会话，关于适配的评论均明确指向语言适配，其对象为课程层面或学生层面。

学生层面的语言适配最为常见，共出现于 4 条评论中（“可以适应你的水平” [“can adapt to your level”]), 但未具体说明语言特征。课程层面的语言适配出现在 3 条评论中（“贴合课程”）。另有两条回答指出学生层面的语言不适配（“有时词汇太难” [“vocabulary sometimes is too difficult”]), 以及两条反映课程层面的语言不适配（“有时会使用课外词汇” [“sometimes it uses extracurricular vocabulary”])。

回答准确性

AI-migo

对准确性的感知既通过定量方式进行考察，也在定性数据中作为一个存在争议的维度浮现出来：共出现 24 条相关回答（占实质性答的 8.5%），其中正向评价 8 条，负向评价 16 条。

一般性的准确性在 8 条评论中出现，主要聚焦于对 AI-migo 回答正确性的感知，描述其为“准确” [“accurate”]、“精确” [“precise”]、“正确” [“correct”] 或“合乎逻辑且合理” [“logical and reasonable”]; 其中有两条仅简单肯定其准确性而未作进一步说明。这 8 条评论中，有 5 条与便利性评价同时出现。

在负向回答中，回答不准确是最为突出的主题，共出现于 12 条评论中，受访者将回答描述为“不准确” [“not accurate”] 或“不可靠” [“not reliable”]。其中一条将“小的语法问题” [“small grammar questions”] 指出为具体局限，另一条则提及 AI 幻觉（“它会给出荒谬的回答” [“it gives a ridiculous answer”])。无法回答问题亦出现在 6 条回答中（“无法回答我的问题” [“can’t answer my questions”])。另有两条将这些问题归因于技术层面的限制，提及 AI-migo 的“知识限制” [“knowledge limit”] 或更广义的“语言模型的局限性” [“limitations of language model”]。此外，还有两条仅将“准确性”列为关注点，而未提供进一步说明。

XiSPA

准确性并非 XiSPA 的明确设计目标，但该维度仍在定性数据中有所体现。在此情境下，准确性特指西班牙语的语言正确性，共有 4 条正向描述，将 XiSPA 评价为“准确” [“accurate”]，或称其为“用于练习西班牙语、语法正确的朋友” [“a grammatically correct friend to practice Spanish”]。

回应自然度

XiSPA

自然度通过定量方式进行测量，并在定性数据中进一步加以探讨，但相关回答数量极为有限，仅出现于 9 条评论中（占实质性回答的 5.45%）。其中，对话不连贯在 4 条评论中被提及，受访者描述难以维持“较长的对话” [“long conversation”]，并指出系统“记忆较弱” [“weak memory”]。对类人性的感知呈现出分化态势：3 条回答体现了类人交互体验（“就像在和好朋友聊天” [“like chatting with your good friend”]），而另有 2 条则认为缺乏类人交互（“不能像人一样行动” [“cannot act like a human being”]）。

AI-migo

尽管自然度并未作为 AI-migo 的明确量化调查维度，但仍出现了 6 条相关的定性评论。这些评论涉及缺乏人类特质，如“情感” [“emotion”] 以及“理解、共情或创造力” [“understanding, empathy, or creativity”]；与面对面交流相比参与度较低（“不如面对面交流那样专注” [“not as concentrating as facial conversation”]）；以及对话不连贯问题（“记忆较弱” [“weak memory”]）。

AI 工具使用描述

共有 126 条回答（占 477 条实质性回答的 26.4%）描述了学生在备考阶段和/或考试完成后如何使用这些工具。

AI-migo 的使用

共有 58 条回答描述了学生如何使用 AI-migo 为期末考试做准备（占工具使用描述的 46%）。以信息获取为目的的提问是最主要的使用模式，共出现 28 次，既包括一般性提问（“问一些问题” [“asking some questions”]），也包括特定主题的提问。在特定主题提问中，语法相关问题（“弄清一些语法问题” [“to figure out some grammatical problems”]）和考试相关问题（“在准备考试时寻求建议” [“ask for advice when preparing for the exams”]）最为常见，各出现 5 次；其次是词汇查询和一般信息搜索（各 3 次）。

纠错与反馈构成第二大使用模式，共出现于 11 条回答中，其中语法纠错尤为突出（4 次；“纠正我的语法错误” [“correct my grammar mistakes”]），并辅以与写作相关的反馈（2 次）。生成练习构成另一重要类别，共出现于 9 条评论中，主要是要求生成以语法为重点的练习（4 次；“生成一些语法题” [“generate some grammar questions”]）。内容生成也出现在 7 条描述中，包括请求“做一个总结” [“do a summary”] 和“列出词汇表” [“listing wordbank”]，其中部分涉及考试相关内容生成（3 例；“预测可能出现在期末考试中的题目” [“presume questions which could appear in my final exam”]）。此外，还识别出 6 次误用情况，学生尝试进行本应由 XiSPA 承担的会话练习，例如“模拟口语考试” [“to simulate the speaking test”]。

XiSPA 的使用

学生使用 XiSPA 进行会话练习的情况出现在 38 条回答中（占工具使用描述的 30.2%）。最常见的是问答式对话形式，共出现 9 次（“问答” [“ask and answer”]）。一般性对话紧随其后，出现在 8 条回答中（“和它进行对话” [“have conversations with it”]）。其后依次为口语技能提升（“提高我的口语能力” [“improve my speaking skills”]）、特定主题练习（“如何用西班牙语描述天气” [“how to describe weather in Spanish”]）以及以考试为导向的会话练习（“基于考试范围” [“based on the scope of the exam”]），各出现 3 次。部分回答还描述了在对话过程中使用 XiSPA 进行词汇和语法练习（各 3 次）。同样也出现了 6 次误用情况，学生尝试将 XiSPA 用于反馈获取、信息检索和练习生成。

AI-migo 与 XiSPA 的考试后使用

在考试后的使用方面，有 30 条回答（占工具使用描述的 23.8%）描述了学生在考试完成后继续使用其中一种或两种工具。提问仍是最常见的使用方式，共出现 15 条回答，其中在特定主题提问中，词汇相关问题（3 次）和语法相关问题（2 次）较为突出。一般性会话练习出现在 12 条回答中，有时以复习和巩固为导向（2 例；“真的就是和它聊天，以免忘记我学过的知识” [“literally just chat with it in order not to forget the knowledge I have learnt”]）。练习生成的提及频率较低，仅出现于 3 条回答中。

感知有用性：使用 AI-migo 与 XiSPA 的实际成效

除前一节所述的使用情况外，定性回答还从实践性、任务导向的角度界定了学生对工具有用性的感知。这类回答中，AI-migo 共计 58 条（占 AI-migo 实质性回答的 20.6%），XiSPA 共计 28 条（占 XiSPA 回答的 17.0%）。

总体有用性

在未具体说明成效类型的情况下，对工具体有用性的概括性评价在 AI-migo 中出现 7 条，在 XiSPA 中出现 8 条，常使用“有用”“有帮助”“高效”或“专业”等表述。

语言学习与语言能力发展

就 AI-migo 而言，语言学习在 16 条评论中被提及，涵盖对总体学习的支持（“支持学习” [“to support learning”]）、对知识理解的帮助（“帮助理解” [“help understand”]）以及学习效率的提升（“让我的语言学习更高效” [“making my language learning more efficient”]）。个别回答还指出其有助于知识内化（“提升我的知识” [“to enhance my knowledge”]）以及语言结构的学习（“语法和词汇” [“grammar and words”]）。此外，语言能力发展也在 8 条评论中出现，回答表明 AI-migo 有助于“提高我的外语水平” [“improve my foreign language”] 和“提升我的西班牙语能力” [“enhance my Spanish skills”]。相比之下，XiSPA 中仅有 1 条评论涉及一般性的语言学习，而有 7 条集中于语言能力发展，包括能够“与外国人交流” [“communicate with a foreigner”] 以及“在对话中运用 [所学知识]” [“apply [knowledge] in conversations”]。

练习

在 AI-migo 中，练习相关表述仅在 2 条评论中出现，其中一条描述将其用于“练习西班牙语” [“to practice Spanish”]，另一条强调通过“练习而非背诵”来学习 [“by practicing instead of reciting”]。相比之下，练习在 XiSPA 中显著突出，共出现 10 次（占 XiSPA 成效评价的 35.7%）。其中 3 条仅笼统提及练习而未作具体说明，其余 7 条则明确描述了以对话为导向的练习，包括替代真人的练习（“当我找不到西班牙语使用者时，可以用它来帮忙” [“when I can't find the Spanish speakers, I can use it to help”]）、交际练习（“练习交流” [“practice the communication”]）以及口语技能练习（“练习我的西班牙语口语” [“practice my oral Spanish”]）。另有 2 条评论指出语音功能缺失限制了这些成效（“不能用音频形式和它聊天” [“can't chat with it in audio form”]）。

反馈 (AI-migo)

反馈在 AI-migo 的 10 条回答中被提及（占 AI-migo 成效评价的 17.2%）。其中，功能性反馈出现在 8 条评论中，以语法导向的反馈最为突出（6 次；“检查我的语法问题” [“check my grammar problems”]），其次是写作改进（“润色我的文本” [“polish my text”]）和校对（“检查段落” [“check passages”]）。标点和词汇仅作为个别目标被提及。此外，还有 2 条评论对反馈质量作出了相反评价：一条认为反馈“很详细” [“detailed”]，另一条则认为“过于笼统” [“too general”]。

其他成效

记忆保持在两种工具中共出现 5 条评论：其中 3 条来自 AI-migo，涉及课堂内容的保持、课后复习以及备考；另有 2 条来自 XiSPA，主要指向句式等语言内容。就 AI-migo 而言，备考和信息检索也各在 3 条评论中被提及。

过度依赖

过度依赖在 AI-migo 的 6 条评论和 XiSPA 的 4 条评论中被指出，主要引发了对“批判性思维”受影响以及过度依赖的担忧（“过于依赖它” [“depending on it too much”] 或“可能会过度依赖 [它]” [“may rely on [it] too much”]）。

讨论

AI-migo 与 XiSPA 的试点实施及初步评估为定制化 AI 工具成功融入绝对初学者层级的语言教学提供了初步证据。研究结果显示，学生在初始阶段表现出较高的用户接受度、频繁且持续的使用意愿，并感知到学习成效；与此同时，结果亦揭示了若干亟需进一步完善的关键方面。

整体接受度与参与情况

两款工具均表现出强劲的初始接受度，在各评价维度上的正向评分达到 70%–85%，均值集中在 4.0 左右。结合学生在备考与考试后续阶段的持续自愿使用，以及较高的使用频率，这一结果表明工具已成功融入学生的日常学习常规。

定性数据在支持上述趋势的同时，也揭示了更为细致的差异。首先，两款工具中，学生关于工具优势的评价数量均超过技术问题与挑战的合计。其次，“无问题”回答所占比例较高，表明开放题吸引了具有不同使用体验的受访者，而非仅汇集持极端观点者。第三，挑战与局限的数量多于技术问题，说明学生关注的核心主要集中在教学有效性与可用性，而非系统稳定性。

然而，不同工具之间的参与模式存在差异。针对 AI-migo 的定性回答显著多于 XiSPA（282 条 vs. 165 条实质性回答），且在备考阶段 AI-migo 的使用率略高。这一差异很可能源于 AI-migo 更为广泛的功能为学生提供了更多可评论的维度、更高的整体采用率，以及问卷题目排序可能引发的调查疲劳。总体而言，更高且更稳定的参与度表明 AI-migo 在学生在学习体验中占据了更为核心的地位，而 XiSPA 的功能则聚焦于作为学生的对话练习伙伴。

可用性

易用性与便利性的首要地位

两款工具在感知易用性方面均获得了积极评价，其中 XiSPA 的评分略高（80.8%），高于 AI-migo（73.8%）。这一差异很可能反映了 XiSPA 更为简洁的基于聊天的界面，以及学生对对话式人工智能日益增长的熟悉度；相比之下，AI-migo 具备多项功能，因而需要用户进行更多的适应。

在定性数据中，便利性成为最突出的优势——在关于 AI-migo 的优势回答中占 42.6%，在 XiSPA 中占 29.7%。对于 AI-migo，学生主要强调时间与空间的灵活性、即时性以及按需获取学术支持的便利；而在 XiSPA 中，评论重点则主要集中在即时性与练习可获得性上。这表明，在目标语接触有限、学习者几乎无法获得与水平适配的材料和练习机会的情况下，便利性切实弥补了课堂之外西班牙语支持、练习和接触机会的真实缺口。

尽管在概念上，便利性不同于感知易用性，但相关研究表明，感知易用性是便利性的前因变量（Chang et al., 2012; Yoon & Kim, 2007）。这一关系也与本研究的发现一致，即这两个维度均获得了积极评价。

技术与访问壁垒

尽管出现频率较低，但仍出现了两类可用性挑战。系统可靠性问题——包括系统崩溃、连接故障以及界面异常——在两款工具的实质性回答中均约占 7%。这些技术障碍值得关注，因为它们可能削弱学生最为看重的便利性，并通过影响系统满意度而间接影响感知易用性（Wixom & Todd, 2005）。

访问虽然出现得更少，但同样至关重要。学生反映难以找到工具入口，以及希望提供独立的移动端应用，表明部分学生在基础访问层面面临困难。若无法提供可靠的访问途径，教学层面的可供性便失去意义。

使用负担与交互挑战

两款工具在交互过程中均呈现出涉及三种使用负担类型的挑战。系统诱发的负担使学生在系统崩溃后不得不重复输入内容。交互模态负担则仅出现在 AI-migo 中，反映了与 XiSPA 较短的对话轮次相比，输入较长查询语句所带来的打字负担。最为关键的是，技能依赖型负担在两款工具中呈现出不同形态：AI-migo 体现为提示词构造负担，而 XiSPA 则体现为二语能力受限所致的负担。

就 AI-migo 而言，学生在提示词构造方面面临困难，即在构建有效提问时所涉及的元语言层面的要求。这一发现凸显了 AI 虚拟导师设计中的一种根本矛盾：开放式问答虽然提供了高度灵活性，但同时也对学生提出了较高要求，而这些学生可能尚不具备恰当构建问题所需的知识或技能。这种提示词构造上的要求还会进一步调动身体层面的付出，即较大的打字负担（交互模态负担）。相比之下，XiSPA 的主要障碍在于二语水平本身：有限的语言储备限制了学生持续展开对话的能力，这反映了面向绝对初学者的对话式工具所固有的挑战——此类工具预设学习者至少具备最低限度的语言产出能力，而这正是该水平学习者正在发展的能力。

自我归因的情况进一步强化了上述区分：在承认自身限制原因的 AI-migo 使用者中，一半（10 人中的 5 人）将困难归因于提示词构造问题，另有两人将问题归因于打字负担；而在 XiSPA 中，进行自我归因的五名学生中有两人明确指出语言技能受限。与此同时，还发现了系统理解困难的情况，但未能排除潜在的用户原因（AI-migo 为 17 条，XiSPA 为 8 条）。这些情况可能在学生未加察觉的情况下，同样源于提示词构造或语言能力限制，但也不排除其他解释，例如系统本身的限制或可靠性问题，仍有待进一步探究。

上述发现提示有必要采取差异化干预策略。对于 AI-migo，可通过模板化交互或为常用功能设置快捷入口，降低提示词构造与交互模态两方面的负担；而对于 XiSPA，提供支架式会话——例如回复建议或词汇提示——有助于学生在语言储备有限的情况下维持对话的持续进行。

愉悦感与挫败感

两款工具在感知愉悦方面的评分几乎一致（约 78% 为正向评价， $M \approx 4.05$ ），表明无论教学目的或功能复杂程度如何，学习者的情感参与度均得以维持。

所报告的负面情绪虽较为有限，但在两款工具中呈现出不同特征。对于 AI-migo，挫败感往往与语言适配失调及沟通失败同时出现，这表明当学习者对语言適切性和沟通成功的期待未能得到满足时，负面情绪便会产生。这一模式与既有研究相一致：当文本难度超出学习者理解能力时，可能引发外语阅读焦虑（Bahmani & Farvardin, 2017; Liu, 2025; Saito et al., 1999），并在东亚环境中进一步导致学习动力不足（Xie et al., 2018）。这一结果再次凸显了将输出控制在适合学习者理解水平的范围之内的重要性。

对于 XiSPA，有一名学生将 AI 对话描述为“无聊”。这一极个别的评论引出了一个值得关注的设计矛盾：针对绝对初学者的对话代理，为确保可理解性而施加的语言与主题限制，可能会削弱对话本身的多样性与自然性，而这些恰恰是促成对话吸引力的关键因素。现有数据尚无法确定

这一矛盾是否直接导致了评论中所说无聊感，但在必要的语言约束条件下如何最大化学习者参与度，仍是一项有待进一步探究的持续性设计挑战。

输出质量

适配性

对输出适配性的高度正向评价（两种工具均约为 80%）表明，在第 12 周时，学生普遍认为两种工具的输出与 SPA001 的课程内容具有良好对齐性。就 XiSPA 而言，这一结果为 Getino-Diez & García-Madariaga（2026）所提出的基于提示词的受控文本生成方法提供了初步验证。就 AI-migo 而言，正向评价表明其在总体上实现了课程对齐，但具体所评估的适配维度尚无法明确：该量化指标混合了多种输出类型——包括问题回答和解释性说明，而这些输出未必均采用受控生成方式。尽管存在这一不确定性，鉴于为绝对初学者提供水平适配的西班牙语接触、练习与支持是本次实施的核心目标之一，这些结果仍具有积极意义。此外，基于提示词的方法具有良好的可扩展性，可通过对知识库的每周更新实现与课程进度的持续适配，并有潜力推广至其他课程。

在定性数据中，反映适配性的评论数量多于不适配性评论（AI-migo 为 17 条对 9 条；XiSPA 为 7 条对 4 条）。然而，正向评价大多缺乏具体说明，而负向评论则提供了较为明确的描述——主要包括两种工具中均出现的学生层面的语言不匹配（如不熟悉的词汇、理解困难），以及 AI-migo 中出现的解释性不匹配（难以理解系统给出的解释）。这种不对称性可能表明，相较于适配成功，适配失败对学习者的影响更为显著、也更容易被记住。与此同时，正负并存的评价——包括个别认为内容“过于简单”的反馈——反映出学习者在语言能力和知识水平上的差异性，使得单一的、以课程内容同步为导向的设计在满足全体学习者的不同需求方面面临挑战。

基于上述模式，可以得出三点结论。第一，两种工具均存在一种普遍认知，即其输出能够较为恰当地适配学习者或课程内容，这一点与定量调查结果一致，但具体适配维度仍不清晰。第二，两种工具均在部分学生中出现语言层面的不匹配，而 AI-migo 还存在解释性不匹配，其主要表现为输出难度高于学习者预期。当 AI 生成内容对学习者的理解而言不可理解时，工具便无法实现其提供适合学习者水平的学习材料的基本功能。第三，学习者在语言能力与知识水平上的个体差异表明，仅在课程内容同步层面的适配并不足以满足需求，还需要引入更精细的适配机制。

未来的迭代应在强化课程层面适配的同时，引入个体层面的适配机制。在课程层面，应确保语言复杂度（两种工具）与解释复杂度（AI-migo）能够动态反映学生在课程各阶段应当具备的理解水平。在个体层面，适配性设计可以包括：（a）面向用户的控制选项，如可调节的难度设置或可选的“挑战模式”；（b）个性化学习者档案，用于追踪已掌握的语言结构，从而动态引入新内容或强化薄弱环节；（c）为 AI-migo 提供实时调节功能，使其在练习过程中能够根据需要提供更简单或更复杂的解释或语言输出；以及（d）在 XiSPA 交互过程中提供按需语言支持，如文中翻译或弹出式翻译，或对不熟悉语法结构的即时解释。

准确性

准确性仅针对 AI-migo 进行了定量测量，评估的是学生对其回答与解释正确性的感知。尽管定量结果显示总体感知较为积极（84.6%），定性数据却呈现出分化的观点：12 条评论指出回答不准确，6 条提及无法回答问题；相对而言，8 条评论对准确性作出积极评价，其中 5 条与便利性同时出现，表明这两个维度在学生感知中可能存在联结。

关于适配性评论中的不对称性——即正面感知较为笼统，而负面感知更为具体——在准确性维度中同样存在，但程度较轻。现有数据对于不同准确性感知的来源仅提供了有限信息，尚无法判断其是否源于提问复杂度差异、问题主题与知识库的契合程度、个体提示词构造策略的不同，或对可接受准确性水平的期望差异。有两名学生表现出元认知意识，将相关局限归因于 AI-migo 的“知识”或“语言模型”，从而认识到与准确性相关的系统约束。

对具体不确实实例的进一步调查，将有助于制定有针对性的干预措施——包括通过优化系统提示词以改进查询理解、在薄弱内容领域扩展或重组知识库，或更清晰地向用户传达系统的局限性。

对于 XiSPA，尽管未对准确性进行明确测量，但有 4 条评论对其语言正确性给予了积极评价，且未出现任何负面的准确性评论。这表明，严格的词汇与语法约束并未削弱学生对其对话输出语言正确性的感知。

自然度

自然度仅针对 XiSPA 进行了定量测量，评估的是学生对对话自然性的感知。XiSPA 获得了较高的正向评价（79.5%），表明学生整体上认为其对话具有较高的自然度。

然而，从定性角度看，两种工具中与自然度相关的评论数量均较少（XiSPA 为 9 条；AI-migo 为 6 条），主要集中于对话流畅性及感知类人性方面的担忧。鉴于两种工具在教学功能与设计约束上的差异，自然度在此需作差异化解读。

对于 XiSPA 而言，其对话自然性面临两个相互关联的限制。首先，绝对初学者的语言储备本身十分有限，学生缺乏进行多样化、自发性轮流对话所需的词汇与语法资源，使得在该水平上实现完全自然的对话本就较为困难。其次，维持较长的对话记忆、从而实现更好的跨轮次连贯性，可能会削弱受控文本生成方法的效果，因为语境信息的增加会提高严格执行词汇与语法约束的难度。部分评论提及对话不连贯，描述了维持较长对话的困难，并指出“记忆较弱”，这可能正是权衡上述利弊的结果。关于类人性的感知则较为有限且呈现多样化，既包括对缺乏类人交互的评价，也包括积极的类人体验，这表明“拥有一个对话伙伴”这一概念——即便是 AI 对话伙伴——对部分学习者而言仍可能激发一定的社会互动感。

对于 AI-migo，自然度相关评论应结合其作为支持与反馈提供者、而非对话伙伴的角色来理解。相关担忧包括缺乏情感、共情与创造性等类人特质，与面对面互动相比参与度较低，以及对话不连贯等问题。这些回答表明，部分学生在辅导情境中期望工具呈现更具类人特征的互动方式，甚至希望其在一定程度上模拟真实教师。鉴于 AI-migo 所处的交际情境更为宽泛，且语言约束相对不那么严格，未来版本在满足这些期待方面可能具有更大的灵活性。

对于 XiSPA 而言，在不削弱语言控制的前提下增强对话连贯性仍是重要方向，未来迭代可探索支持更顺畅的对话轮换或维持话题连贯性的机制。对于两种工具而言，引入可选的语音交互或视觉化呈现（如动画形象）等功能，或有助于提升类人交互的感知。后续研究亦有必要进一步考察自然度与学习动机、学习者参与度等其他维度之间的关系。

AI 工具采用、使用模式与感知有用性

工具采用与持续使用

多数学生在备考阶段使用了这些工具（AI-migo 为 68.1%；XiSPA 为 64.2%），且在考试结束后仍有相当比例继续使用（52.7% 至少使用过一种工具）。这些使用情况表明，学生在两套系统中均感知到了切实的价值。与 XiSPA 以对话为主的功能定位相比，AI-migo 具有更为广泛的功能，其更高的采用率很可能反映了这一差异，说明即便在感知易用性略低的情况下，更高的功能丰富度仍可能促进工具采用。

相较于总体采用率，更为关键的是采用者之间的参与一致性。在备考期间，约 80% 的用户至少每周使用一次工具；考试结束后也呈现出类似的持续参与水平。这表明，两套系统均被有意义地融入学生的学习常规之中，而非零散、偶发地使用。考试后的持续使用尤为令人鼓舞，因为这意味着这些工具可能有助于应对课外西班牙语接触有限以及缺乏适当学习材料这一根本性挑战。即便在缺少外在成绩压力的情况下，学生仍自愿保持与目标语的接触与练习。

由于所有使用行为均为自愿，关于工具使用的定量研究数据可作为感知有用性的间接证据——这与技术接受模型（TAM）中“感知有用性”与“实际使用”之间的关系一致（Davis et al., 1989）。而定性描述则进一步揭示了这种有用性的具体维度。

使用情况与功能契合度

定性使用描述表明，两种工具在很大程度上均按照其设计初衷被使用。AI-migo 承担了多样化功能——以问答为主（最为高频，涵盖一般性提问以及以备考、语法、词汇和信息检索为重点的专题性问题），同时还包括纠错与反馈、练习生成以及内容生成。XiSPA 则主要用于不同形式的会话练习（如专题化练习），并服务于不同目标（如备考）。预期用途与实际使用之间的高度一致性表明，学生能够识别并有效利用这些工具的核心功能优势。

然而，误用情况显示，尽管在引入阶段进行了明确说明，学生对工具特定功能边界的理解仍不充分。误用现象在两种工具中均有出现：在 AI-migo 中，部分学生尝试进行会话练习；在 XiSPA 中，则出现了寻求反馈、信息检索以及练习生成的情况。这种功能边界混淆可能源于以下因素：（a）在简短的介绍环节中，学生未能充分内化不同工具之间的功能区分；（b）界面设计未能充分传达功能边界；（c）以便利性为导向的使用行为，即学生倾向于使用当时已打开的工具；或（d）学生对某一工具的交互方式存在真实偏好，并将其使用方式扩展至超出原有设计范围。

此外，关于 AI-migo 的反馈与练习生成功能的报告相对较少，表明这些能力相较于问答功能并不突出。这一情况在考试后使用中尤为明显，其中完全未提及反馈功能。学生可能主要将 AI-migo 概念化为一种问答资源，而未充分意识到其功能范围的多样性。未来的实施应更为明确地向学生介绍可用功能，例如通过引导式教程、融入课堂的使用案例与活动，以及在界面中更为突出的功能入口设计，以促进学生对不同功能的理解与使用。

感知有用性：影响与学习成效

学生对学习成效的高度积极感知为这些工具的价值提供了重要验证。超过 80% 的使用者认同这些工具对课程成绩和考试后提升产生了积极影响，这表明参与使用 AI-migo 和 XiSPA 的学生普遍将其视为既有助于即时提升学习成绩、又能支持持续技能发展的有效工具。

从教学角度来看，这些感知结果表明，两款工具成功地实现了其核心设计目标：为有意义的目标语接触和练习提供更多机会（两款工具），并提供定制化的支持与反馈（AI-migo）。然而，这些主观感知是否能反映真实的学习成效，仍有待通过对学习结果的受控评估在未来研究加以实证检验。

对学习成效的定性描述显示，语言学习与语言能力发展构成了 AI-migo 最主要的感知成效（占成效类回答的 41.4%），而 XiSPA 则更强调技能发展而非知识学习，这与其作为对话练习伙伴的设计定位一致。这一区分反映了两种不同的教学角色：AI-migo 作为解释、反馈与结构化练习的来源，主要支持知识学习；XiSPA 则提供一个应用性语境，使学生能够运用并发展其正在形成的交际能力。

练习相关提及数量的显著差异（AI-migo 为 2 次，XiSPA 为 10 次）很可能源于 XiSPA 明确将自身定位为对话伙伴，使其练习可供性更加直观、突出。如前所述，AI-migo 的练习生成与反馈功能可能需要通过更明确的功能说明或更突出的界面呈现来强化。

尽管部分学习者意识到这些工具在巩固记忆（两款工具）、备考（AI-migo）以及信息检索（AI-migo）方面的潜在价值，但相关提及频率较低。未来的迭代可通过有意的教学整合来强化这些维度。然而，备考相关评论较少并不意味着其重要性不足——如前一节所示，报告显示有大量与考试相关的用途。

使用描述与感知成效构成了对感知有用性的互补视角。使用描述揭示了具体的互动模式（提问、寻求反馈、练习生成以及对话练习），展示了学生在实践中如何认为这些工具“有用”；而感知

成效则反映了学生将这些活动与高阶学习结果相关联——语言学习与能力发展。对 AI-migo 而言，学生可能以工具性方式看待其作用：具体活动共同促进实现更广泛的学习与发展目标。对 XiSPA 而言，这一模式则较为混合：对话练习既是最主要的活动，也是最常被报告的成效，表明部分学生内在认同练习的价值；与此同时，语言能力发展也成效显著，说明另一些学生将练习视为促进技能发展的工具性手段，而非直接的知识学习。这一区分凸显了两款工具在学生学习过程中的不同角色，以及它们各自对西班牙语水平发展的贡献。

过度依赖：一个不容忽视的担忧

共有 10 条评论表达了对过度依赖的担忧，提出了其可能对批判性思维发展造成影响以及对工具产生过度依赖的问题。尽管数量不多，这些担忧反映了 AI 支持型学习中存在的重要矛盾，值得加以重视。

一方面，诸如 AI-migo 和 XiSPA 等 AI 工具本就被明确设计为提供广泛的支持与练习机会，因此使用频率的提升可以被视作成功参与的体现，而非必然的问题性依赖。另一方面，过度依赖也可能削弱学习者自主学习策略的发展，减少其与教师和同伴的互动，或形成一种依赖状态，使学生在 AI 支持不可用时面临学习困难。这些担忧表明，未来的实施应提供相应指导，帮助学生潜在的过度依赖倾向提高警惕，并在使用 AI 工具的同时，保持与其他学习资源之间的平衡。

局限性

本次评估存在若干局限。首先，所采用的横断面单次个案研究设计缺乏基线或对照组，仅捕捉了学生在短期接触工具后的初始感知，而无法反映长期使用情况或真实的学习成效。

其次，所有评估均基于自我报告，不可避免地存在潜在偏差。例如，学生自我报告的工具感知有用性及感知影响，未必与实际考试结果或真实学习收获相一致。此外，自我选择偏差亦可能影响研究结果，因为参与者在学习动机或态度方面可能不同于未参与者。

第三，定性问题的自愿作答特性可能更容易吸引观点较为强烈的学生参与，同时，AI-migo 与 XiSPA 在开放式问题回答数量上的显著差异（313 vs. 215）也增加了跨工具类比解读的难度。

第四，工具在学期末才进行实施，这限制了评估范围。尽管两款工具均设计为可在整个学期内使用，并通过逐步更新知识库实现持续适配，但第 12 周才引入使得无法评估其在完整学期内进行动态课程适配的能力。

第五，尽管在探索性研究和降低受访者负担的背景下采用单项测量具有合理性，但其测量稳健性仍不及多项量表。与此相关，问卷设计中亦暴露出一定的歧义，尤其体现在适配维度上：定量测量未区分语言适配与解释性适配，而 AI-migo 的测量题目则混合了多种输出类型。未来研究亟需开发能够清晰区分不同适配类型、并分别评估工具特定质量维度的测量工具。

结论

本文系统呈现了 AI-migo 与 XiSPA——两款为西交利物浦大学（XJTLU）绝对初学者西班牙语教学而定制开发的 AI 驱动工具——的设计理念、试点实施过程及初步评估结果。研究初步表明，这两款工具具有在课堂之外扩展西班牙语目标语接触、提供按需可得水平适配材料与练习，并支持备考的潜力。学生整体上将两款工具评价为易于使用、具有愉悦性、与课程高度对齐，并认为其在准确性（AI-migo）、自然度（XiSPA）以及对学习、课程成绩与持续提升方面均具有积极作用。工具采用率较高，且在考试结束后仍保持使用，表明学生在课程要求之外自愿持续参与。

然而，定性结果也揭示了若干亟需有针对性改进的方面，包括：提升系统技术可靠性与可及性，降低提示词构造负担与整体交互负担，同时加强课程层面与个体层面的适配机制，并进一步改善对话连贯性与类人交互特征。针对这些重点进行改进，不仅对在 SPA001 中优化 AI-migo 与 XiSPA 至关重要，也将为类似 AI 驱动工具在其他课程与语言中的可扩展整合提供重要参考。

AI-migo 作为虚拟导师、XiSPA 作为对话伙伴所形成的互补角色配置，展示了一种将结构化指导与互动式练习相结合的有前景的方案。所采用的基于提示词的方法在语言控制方面兼具开发效率、可迭代优化能力与可扩展性，且无需大量计算资源或高水平技术专长，这些特征对于教育情境中的可持续实施尤为关键。

未来实施应在学期伊始即部署上述工具，并通过每周知识库更新实现与课程进度的动态对齐，从而全面评估其在整个学期内实现课程动态适配的能力。同时，有必要通过系统性的研究，评估其对实际学习结果的影响，而非仅停留在感知层面，以更准确地理解其教学价值。

本文所采用的迭代式实施-评估-改进循环为教育领域中可靠的 AI 融合提供了一种可持续范式：从明确的教学需求出发，设计目的导向的解决方案，通过系统性评估进行试点，根据证据持续优化，并在保持持续评估的前提下逐步推广。随着 AI 能力的不断发展，这种证据驱动的方法为在真实交际机会与適切材料仍然有限的情境中，利用 AI 促进语言学习与教学提供了切实可行的路径。

参考文献

- 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/tl.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> (原作发表于 2014 年)
- 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> (原作发表于 2021 年)
- 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
- Wiboolyasarini, W., Wiboolyasarini, 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>