

Self-perceptions of communicative mediation and AI-related self-efficacy in adult EFL learners: A chatbot-supported pilot study

Esther Cores-Bilbao ^{a,*}, María-del-Carmen Méndez-García ^b,
Manuel Célio Conceição ^c

^a *University Isabel I, Spain*

^b *University of Jaén, Spain*

^c *University of Algarve, Portugal*

Received May 25, 2026; Revised June 6, 2026; Accepted June 24, 2026

Abstract. This exploratory pilot study investigates adult English as a Foreign Language learners' self-assessed development following a chatbot-supported pedagogical intervention within an Official Language School in Spain. A 12-item Likert-type questionnaire was administered to 20 adult B2 learners before and after the intervention to gauge shifts across four domains: online interaction self-perception, pluricultural competence self-perception, mediation-related self-perception, and AI-related self-efficacy. The intervention centred on a fictional intercultural scenario in which learners used a chatbot to clarify misunderstandings, infer cultural norms, and formulate contextually appropriate interpretations. Given the ordinal data structure and the restricted sample size, the statistical analysis relied exclusively on non-parametric procedures. Wilcoxon signed-rank tests indicated statistically detectable pre- to post-intervention upward shifts within this sample across all four dimensions as well as in the global self-assessment score. The most pronounced gains were observed in AI-related self-efficacy, while online interaction and mediation-related self-perception also showed comparatively strong upward patterns. At the item level, learners reported particularly marked gains in the management of communicative misunderstandings, audience-targeted information filtering, and the deployment of AI to navigate appropriate behaviour across diverse cultural contexts. While these sample-bound outcomes warrant cautious interpretation due to baseline reliability constraints, the small sample size, and the exclusive reliance on self-report data, they demonstrate the empirical value of self-assessment as a complementary lens. Ultimately, the findings suggest that chatbot-supported pedagogy may have educational value when explicitly embedded within tasks requiring clarification, interpretation, and culturally situated meaning-making, rather than relying solely on decontextualised language practice.

Keywords: *adult language learning, AI-mediated communication, conversational agents, communicative mediation, learner self-assessment, self-efficacy.*

* Corresponding author. Esther Cores-Bilbao,  0000-0002-8239-8969,  esther.cores@uii.es

© Cores-Bilbao, Esther; Méndez-García, María-del-Carmen; Conceição, Manuel Célio, 2026. This is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International Licence (<http://creativecommons.org/licenses/by/4.0>).

East European Journal of Psycholinguistics, 13(1), 142–165, <https://doi.org/10.29038/eejpl.2026.13.1.cor>

Корес-Більбао Естер, Мендес-Гарсія Марія дель Кармен, Консейсао Мануель Селіу. Самооцінка комунікативної медіації та самоефективність у застосуванні ІІІ серед дорослих, які вивчають англійську як іноземну мову: пілотажне дослідження з використанням чат-бота.

Анотація. Це пілотажне дослідження розглядає самооцінюваний розвиток дорослих слухачів, які вивчають англійську як іноземну мову, після педагогічного втручання за підтримки чатбота в Офіційній школі мов в Іспанії. Двадцяти дорослим слухачам рівня B2 було запропоновано до та після втручання 12-пунктовий опитувальник за шкалою Лайкерта з метою оцінити зміни у чотирьох сферах: самооцінці онлайн-взаємодії, самооцінці плюрикультурної/ міжкультурної компетентності, самооцінці, пов'язаній з медіацією, та самоефективності, пов'язаній зі штучним інтелектом. Втручання було побудоване навколо вигаданого міжкультурного сценарію, у межах якого слухачі використовували чатбот для уточнення непорозуміннь, виведення культурних норм і формулювання контекстуально доречних інтерпретацій. З огляду на порядкову структуру даних і обмежений розмір вибірки, статистичний аналіз спирався виключно на непараметричні процедури. Критерії знакових рангів Вілкоксона вказали на статистично виявлювані позитивні зрушення в межах цієї вибірки між до- та післяінтервенційним вимірюваннями в усіх чотирьох вимірах, а також у загальному показнику самооцінювання. Найбільший приріст спостерігався у застосуванні штучного інтелекту, тоді як самооцінка онлайн-взаємодії та самооцінка, пов'язана з медіацією, також продемонстрували порівняно виразну позитивну динаміку. На рівні окремих пунктів студенти повідомили про особливо помітне зростання в керуванні комунікативними непорозуміннями, добірї найважливішої інформації відповідно до потреб аудиторії, а також у використанні Штучного Інтелекту для орієнтації щодо належної поведінки в різних культурних контекстах. Хоча ці результати, обмежені вибіркою, потребують обережної інтерпретації через початкові обмеження надійності, малий розмір вибірки та опору виключно на дані самооцінювання, вони засвідчують емпіричну цінність самооцінювання як додаткової аналітичної перспективи. Загалом отримані дані свідчать, що навчання за підтримки чатбота може мати освітню цінність тоді, коли вона чітко інтегрована в завдання, спрямовані на уточнення, інтерпретацію та культурно зумовлене конструювання значення, а не обмежується деконтекстуалізованою мовною практикою.

Ключові слова: навчання іноземних мов дорослих; комунікація за посередництва штучного інтелекту; розмовні агенти; комунікативна медіація; самооцінювання слухачів; самоефективність, пов'язана зі штучним інтелектом.

Introduction

Artificial intelligence (AI), and generative AI (GenAI) in particular, have entered language education with remarkable speed. These technologies facilitate instantaneous feedback, sustain prolonged interaction, and introduce novel modalities for self-paced learning (Belda-Medina & Calvo-Ferrer, 2022). However, their educational value mainly hinges on the pedagogical purpose of the task, the quality of teacher mediation involved, and the extent to which learners are able to engage with AI tools strategically rather than instrumentally

(Crompton et al., 2024). As these systems enter language classrooms, they also invite closer attention to how learners understand their own capacities as users of the target language and as participants in AI-mediated forms of learning.

This matter is pertinent to psycholinguistics, insofar as language acquisition necessitates the construction of internal representations of the learner's own communicative efficacy, agency, and autonomy (Duff, 2013). These self-representations can significantly condition how individuals approach interaction, respond to uncertainty, sustain participation, and evaluate their own progress (Aydın & Tekin, 2023). Recent empirical evidence suggests that AI-supported language learning exerts a multifaceted influence on such constructs by enhancing willingness to communicate, reducing anxiety, and strengthening self-perceived communicative competence, even where gains in externally measured performance remain equivocal (Wang et al., 2024; Lee et al., 2026). Furthermore, contemporary studies indicate that AI-enhanced environments may bolster learners' confidence in intercultural communication, their use of communicative strategies, and their perception of themselves as more capable L2 users (Liu et al., 2025; Huang & Mizumoto, 2025).

AI-supported instruction may also foster self-regulation, goal setting, and strategic learning when embedded in reflective pedagogical designs (Chang & Sun, 2024; Mohebbi, 2025). From this perspective, the student–AI nexus transcends technical utility; it necessitates complex judgements regarding tool reliance, the critical interpretation of AI outputs, and the preservation of personal agency within an increasingly AI-mediated pedagogical landscape (Alm, 2024).

Consequently, the impact of AI on learner self-perception appears to be mediated by the design and deployment of these digital tools. Rather than yielding automatic psychological benefits, AI-assisted environments may foster learner confidence, enjoyment, self-efficacy, and perceived control only when embedded within meaningful pedagogical tasks that offer rich interactional opportunities and allow students to regulate both the pace and direction of their learning trajectory (Xiao et al., 2024; Zhang et al., 2025). Concurrently, these positive outcomes should not be conceptualised as deterministic corollaries of technological integration. Learners' self-representations remain contingent upon the overarching task architecture, the granularity of teacher guidance, the nature of the human–system interaction, and the extent to which AI usage is framed as strategic scaffolding, as opposed to a surrogate for individual cognitive effort (Kern, 2024).

Whilst scholarship in this area is expanding rapidly, the empirical evidence base continues to be fragmented. Current studies have predominantly prioritised university cohorts, productive skills (namely speaking and writing-oriented AI applications), generalised attitudes towards technology, or broad

technology acceptance models (Bolatli & Baltaci, 2025). Conversely, adult learners in public-sector language education remain underexplored, despite the specific profiles, trajectories, and digital literacy dynamics of this setting (Rubio-Gragera et al., 2025). Furthermore, evidence remains limited concerning how these specific learner demographics perceive their own developmental trajectories following engagement with chatbot-supported tasks (Lin, 2024), particularly when pedagogical objectives transcend narrow linguistic drilling to encompass online interaction, mediation, and culturally situated meaning-making.

The present study addresses this lacuna by investigating adult B2 EFL learners' self-assessed development following a chatbot-supported instructional sequence implemented within an Official Language School (OLS), known locally as an *Escuela Oficial de Idiomas* (EOI), in Spain. The analytical framework is underpinned by four theory-driven dimensions: self-perception of online interaction, pluricultural self-perception, and mediation-related self-perception, adapted from the *Common European Framework of Reference for Languages: Companion Volume* (CEFR/CV; Council of Europe, 2020), together with AI-related self-efficacy, drawing upon the work of Carolus et al. (2023). Crucially, the study does not seek to demonstrate objective gains in linguistic competence. Its focus lies in how adult learners perceive their own communicative, mediational, and AI-related capacities after engaging in a structured AI-supported learning experience.

AI and Learner Self-Perception in EFL: Self-Efficacy, Motivation, and Communicative Confidence

Current research on AI-mediated language learning increasingly foregrounds its potential to support self-regulation, motivation, communicative confidence, and learner agency. In EFL contexts, AI-supported instruction has been associated with gains in linguistic achievement, L2 motivation, speaking performance, and oral self-regulation, particularly when tools are embedded in meaningful instructional designs rather than used as isolated sources of feedback (Qiao & Zhao, 2023; Wei, 2023; Chang & Sun, 2024; Lan & Zhou, 2025). This suggests that the value of AI extends beyond input provision, reaching learners' strategic control over their own learning processes.

Learner self-perception has also been examined through constructs such as self-efficacy, motivational self-concept, enjoyment, anxiety, perceived communicative competence, autonomy, and well-being (Almusharraf et al., 2025; Du & Alm, 2024; Huang et al., 2024; Huang & Mizumoto, 2025; Liu et al., 2025; Wang et al., 2024). These constructs are relevant because the pedagogical significance of AI does not depend solely on measurable performance outcomes.

It also concerns how learners come to view themselves as capable users of the target language and as strategic users of AI-mediated tools.

Within this framework, self-efficacy is especially central, since it refers to learners' beliefs in their ability to organise and execute actions required to achieve specific outcomes (Shao, 2025). Recent studies suggest that interaction with GenAI may strengthen both language-related confidence and AI-related confidence, supporting the decision to treat AI-related self-efficacy as an independent analytical dimension rather than as a subordinate aspect of general language confidence (Almusharraf et al., 2025; Huang et al., 2024). Other work has linked ChatGPT and conversational GenAI to autonomy, competence, willingness to communicate, reduced language anxiety, and more positive L2 motivational self-concepts (Du & Alm, 2024; Huang & Mizumoto, 2025; Liu et al., 2025; Wang et al., 2024).

Nevertheless, empirical evidence remains limited regarding adult learners in public-sector language education, particularly in studies using pre- and post-intervention self-assessment to examine communicative, mediation-related, and culturally situated dimensions. Remediating this oversight is imperative, as some of the most profound affordances of AI in adult EFL learning likely emerge in tasks that challenge learners to clarify, interpret, reformulate, and render meaning accessible within digitally mediated environments, where language, interaction, and context jointly instantiate communication (Godwin-Jones, 2023).

Communicative Mediation, Pluricultural Meaning-Making, and the Adult EFL Learner

As an operational focus, communicative mediation is uniquely suited to explore these dynamics, as it encapsulates the mechanisms through which learners negotiate culturally embedded meanings, appraise their own interpretive decisions, and modulate explanations across linguistic frontiers (Council of Europe, 2020). This process involves specific mechanisms through which learners monitor their own understanding, adapt explanations, and bridge linguistic and cultural gaps (Piccardo & North, 2019), reflecting contemporary findings that L2 self-assessment relies heavily on metacognitive judgement and the comparison of internal interpretations with external feedback (Tsunemoto & Trofimovich, 2026).

However, this self-evaluation process does not occur uniformly. Recent psycholinguistic evidence indicates that while reflective tasks can help learners reconcile their self-perceptions with actual comprehensibility, this adjustment is contingent upon the learner's initial self-assessment accuracy (Tsunemoto & Trofimovich, 2026). In mediation contexts, where learners must constantly

manage cognitive and relational scales (Piccardo & North, 2019), students with stronger baseline calibration skills benefit significantly from guided perspective-taking tasks (Tsunemoto & Trofimovich, 2026). Conversely, lower-performing learners often exhibit skewed self-judgements, a disparity that highlights the need for communicative tasks that actively support the learners' internal monitoring systems (Oxford, 2017).

This cognitive architecture extends naturally to online intercultural interaction, where pluricultural competence is conceptualised as the active mobilisation of linguistic, cultural, and identity resources (Gruber et al., 2025). Within the CEFR/CV framework, mediating a text or a concept inherently requires navigating communication across these cultural boundaries (Council of Europe, 2020), a process that is closely linked to an integrated pluricultural stance. For learners lacking this disposition, pluricultural competence development depends primarily on the metacognitive, motivational, and behavioural dimensions of cultural intelligence (CQ). Instructional designs must therefore target these specific CQ components, prompting adult students to deploy communication strategies deliberately, manage online misunderstandings, and critically assess their own cultural assumptions in digital settings (Gruber et al., 2025).

In the context of adult learners, mediation tasks capitalise on rich, accumulated linguistic repertoires and diverse prior learning histories, whilst echoing broader tenets of multilingual development and language attrition, where an additional language user's mind operates as a dynamic, non-linear system (Jessner et al., 2021). Within this framework, chatbot-mediated tasks serve as a pertinent locus for investigating AI-supported interaction, offering iterative opportunities for interrogation, clarification, perspective-taking, and digital scaffolding within a controlled communicative matrix. Fostering a well-calibrated internal monitoring system and a deliberate pluricultural stance through these controlled environments constitutes a core cognitive necessity for the adult learner – a premise that directly underpins the empirical investigation of the present study.

Method

Research Design

The study was conducted with adult B2 English learners within an OLS in Spain, examining whether a chatbot-supported pedagogical intervention correlated with shifts in learners' self-assessed development across four theory-driven domains: online interaction self-perception, pluricultural self-perception, mediation-related self-perception, and AI-related self-efficacy. Given its

exploratory design, the study avoids broad causal claims and examines whether a coherent pre-/post-intervention pattern appears in learners' perceived development after AI-supported written interaction. This focus is justified on both theoretical and pedagogical grounds. Theoretically, it responds to contemporary calls to examine AI in language education in terms of achievement while simultaneously considering autonomy, self-regulation, self-efficacy, and learner psychology. Pedagogically, it is highly relevant because self-perceived confidence in mediation, online participation, and strategic AI use may determine whether adult learners continue to engage with these tools productively beyond the immediate classroom context (Chang & Sun, 2024; Du & Alm, 2024; Huang et al., 2024).

Prioritising learners' internal evaluations over externally rated performance, this investigation seeks to capture how students themselves perceived their development across the focal communicative and technological parameters. Accordingly, the empirical inquiry was guided by the following research questions:

- RQ1. To what extent do adult EFL learners report changes in self-assessed development between the pre- and post-intervention measurements?
- RQ2. Are any observed changes similar across online interaction self-perception, pluricultural self-perception, mediation-related self-perception, and AI-related self-efficacy, or are they more pronounced in specific dimensions?
- RQ3. Which questionnaire items exhibit the most distinct patterns of perceived development following the intervention?

Consistent with prior research examining AI-supported self-regulated learning, learner self-efficacy, and motivational self-concept, two directional expectations were established. Firstly, it was hypothesised that learners would exhibit higher post-intervention than pre-intervention self-assessment scores globally. Secondly, the most prominent increments were anticipated in those dimensions most closely aligned with the pedagogical affordances of chatbot-supported interaction, namely AI-related self-efficacy and mediation-related self-perception.

Instrument

To address these research questions, the study utilised a self-assessment questionnaire administered prior to and following a chatbot-supported pedagogical sequence with adult B2 EFL learners. Learners interacted with ChatGPT-4o, configured as a culturally knowledgeable interlocutor for the mediation task, as detailed below in Section *Pedagogical Intervention*. Table 1

presents the theory-driven structure of the instrument, while Table 5 provides the full wording of the 12 items.

Table 1

Theory-Driven Structure of the Self-Assessment Questionnaire

Dimension	Items	No. of items
Online interaction self-perception	1-3	3
Pluricultural/intercultural self-perception	4-6	3
Mediation-related self-perception	7-9	3
AI-related self-efficacy	10-12	3
Total self-assessment score	1-12	12

The questionnaire consisted of 12 Likert-type items rated on a five-point scale, with higher scores indicating stronger self-perceived competence or confidence. Items 1–9, covering online interaction, pluricultural/intercultural self-perception, and mediation-related self-perception, were adapted from relevant descriptor scales in the CEFR/CV (Council of Europe, 2020). Conversely, items 10–12 were included to capture AI-related self-efficacy, a domain not explicitly covered in the CEFR/CV, and were adapted from Carolus et al.'s (2023) Meta AI Literacy Scale (MAILS).

The instrument captured selected dimensions of learner self-assessed development relevant to the intervention, rather than all self-perception constructs discussed in the broader GenAI literature. For analytical purposes, mean composite scores were calculated for each of the four dimensions, and a total self-assessment score was computed across all 12 items. Given the limited sample size, this dimensional structure was treated as a theory-driven analytical grouping based on item content rather than as a factor-analytically validated scale structure.

Participants and Context

The participants comprised adult learners of English enrolled at the B2 level course within an OLS in Spain. These schools are publicly regulated institutions for specialised language education and certification within the Spanish education system. Under the national framework for specialised language education, they provide CEFR-aligned instruction and certification, including Intermediate B1, Intermediate B2, Advanced C1, and Advanced C2 (Ministerio de Educación y Formación Profesional, 2019). Participants had already passed the

B1 level within this system and were in their fifth year of study, which corresponds to the B2 course leading to the official Intermediate B2 certification examination.

Responses were matched using individual participant identification codes, yielding a complete paired sample of 20 learners with synchronous pre- and post-intervention data. The sample size introduces constraints for quantitative inference, limiting statistical power and reducing the stability of effect-size estimates. Consequently, the findings are interpreted conservatively as exploratory and pilot in character, rather than as stable estimates of intervention effects generalisable to broader adult EFL populations.

Ethical approval for the study was obtained from the Universidad of Jaén Ethics Committee. All participants provided informed consent prior to data collection. Participation was voluntary, responses were anonymised through individual identification codes, and no personally identifiable information was included in the dataset analysed for this study.

Pedagogical Intervention

The instructional sequence was based on a fictional-culture scenario in which learners had to interpret a series of misunderstandings experienced by a Spanish visitor during a stay with a host from a fictional culture named Luxuria. The use of an artificial cultural construct provided a shared knowledge base for all participants, reduced the risk of essentialising real-world communities, and directed attention towards inference, clarification, explanation, and culturally situated meaning-making.

The scenario included four culturally framed misunderstandings related to personal space, formal meals, dress codes, and interactions with domestic staff. Learners interacted in writing with a chatbot previously configured as a culturally knowledgeable interlocutor within the Luxuria scenario. During a 45-minute session, they had to chronicle the events of the journey and ask at least ten questions to identify the causes of the misunderstandings and consider how they might have been avoided.

The task was designed to activate mediation-related processes such as asking for clarification, reformulating information, identifying relevant meanings, and making those meanings accessible to others, in line with the CEFR/CV's understanding of mediation as meaning-making across linguistic and cultural boundaries (Council of Europe, 2020). The fictional cultural framework, scenario prompts, and chatbot configuration were initially generated with the assistance of ChatGPT 4.0 and then reviewed, revised, and pedagogically validated by an expert panel before classroom implementation.

To enhance replicability, the chatbot configuration protocol, including the cultural knowledge base, mediator role instructions and interactional constraints, is provided in Appendix A.

Data Analysis

The quantitative analysis was conducted on a paired pre- and post-intervention dataset. Because the questionnaire consisted of ordinal Likert-type responses and the sample size was modest, the analysis prioritised non-parametric within-subject procedures. Descriptive statistics were calculated for all individual items and composite scores, including means, standard deviations, medians, and ranges. To provide a more granular overview of the direction of change, the proportion of learners who improved, remained stable, or decreased from pre- to post-intervention was also examined descriptively.

All statistical analyses were executed using IBM SPSS Statistics. Descriptive statistics, Cronbach's alpha coefficients, and Wilcoxon signed-rank tests were obtained directly within the software. Wilcoxon signed-rank tests were performed for the four composite dimensions and for the global self-assessment score. To control the risk of Type I error inflation across the four principal dimension-level comparisons, a Holm-Bonferroni correction was applied. Since SPSS does not execute this correction automatically in the required format, p-values were adjusted manually from the software output using the standard sequential Holm-Bonferroni procedure. Effect sizes were computed from the Wilcoxon test output using the formula $r = |Z|/\sqrt{N}$, where N refers to the number of paired observations included in each comparison.

Given that the study relies on theory-driven composite scores constructed from a brief self-assessment instrument, internal consistency was evaluated for both the complete questionnaire and its individual dimensions. Item-level analyses were treated as secondary and exploratory, aiming to identify which specific self-perceived competences appeared most closely associated with changes over time. These estimates are reported in Table 2.

Results

Internal Consistency of the Questionnaire

Table 2 summarises the descriptive statistics and internal consistency coefficients calculated for the four principal dimensions, alongside the global self-assessment score, across the paired pre- and post-intervention data.

Table 2

Internal Consistency of the Questionnaire and Theory-Driven Dimensions

Scale	Cronbach's alpha (pre)	Cronbach's alpha (post)
Full questionnaire	.89	.97
Online interaction self-perception	.78	.85
Pluricultural/intercultural self-perception	.77	.89
Mediation-related self-perception	.51	.93
AI-related self-efficacy	.87	.93

The full questionnaire showed strong internal consistency at both measurement points, with Cronbach's alpha increasing from .89 to .97. Most dimension-level estimates were satisfactory, although the pre-intervention reliability of the mediation-related subscale was low ($\alpha = .51$), probably reflecting the brevity of the three-item scale and the exploratory nature of the construct. Consequently, mediation-related findings should be interpreted with particular caution. Descriptively, all four dimensions showed upward movement after the intervention, with the largest gains observed in AI-related self-efficacy and online interaction self-perception.

Descriptive Pre-/Post-Intervention Trends

Table 3 presents descriptive statistics and direction-of-change frequencies for the four dimensions and the total score. The data show an upward pre-/post-intervention trend across all dimensions, with the largest mean change in AI-related self-efficacy. At the total-score level, most learners improved, while a small minority showed slight decreases.

Table 3

Descriptive Statistics for Pre- and Post-Intervention Scores

Dimension	Pre M	Pre SD	Pre Me- dian	Pre Range	Post M	Post SD
Online interaction self-perception	3.12	.85	3.17	1.33-4.67	3.80	.64
Pluricultural/ intercultural self-perception	3.48	.75	3.50	2.00-4.67	3.98	.60
Mediation-related self-perception	3.13	.77	3.00	1.67-4.33	3.73	.72
AI-related self-efficacy	3.18	1.07	3.00	1.00-5.00	3.93	.85
Total self-assessment score	3.23	.75	3.21	1.75-4.50	3.86	.62

Dimension	Post Me- dian	Post Range	Mean change	Improved n	Stable n	Decreased n
Online interaction self-perception	4.00	2.67-5.00	.68	14	4	2
Pluricultural/ intercultural self-perception	4.00	2.67-5.00	.50	15	0	5
Mediation-related self-perception	3.67	2.67-5.00	.60	14	3	3
AI-related self-efficacy	4.00	2.00-5.00	.75	14	3	3
Total self-assessment score	3.88	2.83-5.00	.63	17	0	3

Pre-/Post-Intervention Comparisons at the Dimension Level

To ascertain whether the observed descriptive variations attained statistical significance within this sample, Wilcoxon signed-rank tests were performed for each of the four dimensions and the aggregate score. The findings of these inferential comparisons are summarised in Table 4.

Table 4

Wilcoxon Signed-Rank Tests For Pre-/Post-Intervention Differences at the Dimension Level

Dimension	<i>W</i>	<i>p</i>	Holm-adjusted <i>p</i>	Effect size (<i>r</i>)
Online interaction self-perception	10.00	.009	.026	.59
Pluricultural/intercultural self-perception	22.00	.019	.026	.52
Mediation-related self-perception	14.00	.010	.026	.57
AI-related self-efficacy	6.00	.002	.009	.68
Total self-assessment score	6.00	.002	-	.68

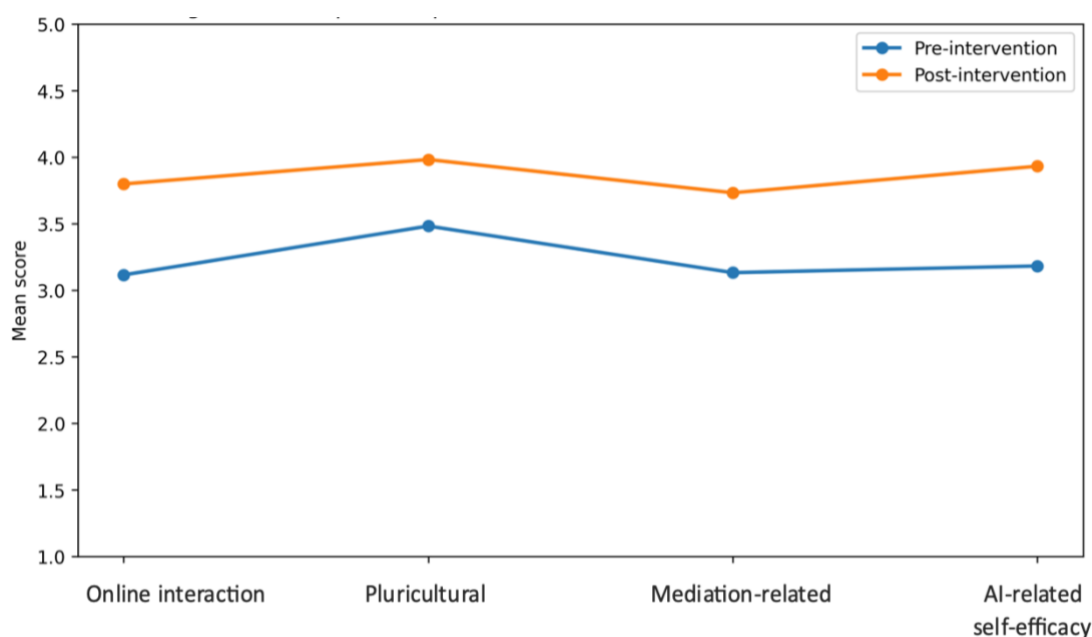
Note. Holm-Bonferroni correction was applied to the four primary dimension-level comparisons.

As shown in Table 4, all four dimension-level comparisons indicated statistically detectable pre- to post-intervention differences within this sample after Holm-Bonferroni correction. The largest effect size was observed in AI-related self-

efficacy, followed by online interaction and mediation-related self-perception. The total self-assessment score showed the same general tendency. Given the restricted sample size, these p-values and effect sizes should be interpreted cautiously. This overall pattern is visually summarised in Figure 1, which highlights the consistency of the upward trend across the four self-assessment dimensions without obscuring differences in magnitude.

Figure 1

Mean Pre- /Post-Intervention Scores Across the Four Self-Assessment Dimensions



Item-Level Patterns

While the primary analyses were executed at the dimension level, item-level comparisons were subsequently examined on an exploratory basis. This was undertaken to identify the specific self-perceived competences that contributed most definitively to the broader pattern of development. Pre- and post-intervention means, Wilcoxon test results, and associated effect sizes are detailed in Table 5.

The item-level data broadly support the dimensional findings. The strongest gains concerned AI use for understanding appropriate conduct in different cultural contexts, active participation in online discussions, management of online misunderstandings, and audience-oriented selection of key information. By contrast, weaker changes were observed for linking contributions to previous comments, argumentative support, and deeper cultural decentring, suggesting that these competences may require longer or differently structured interventions.

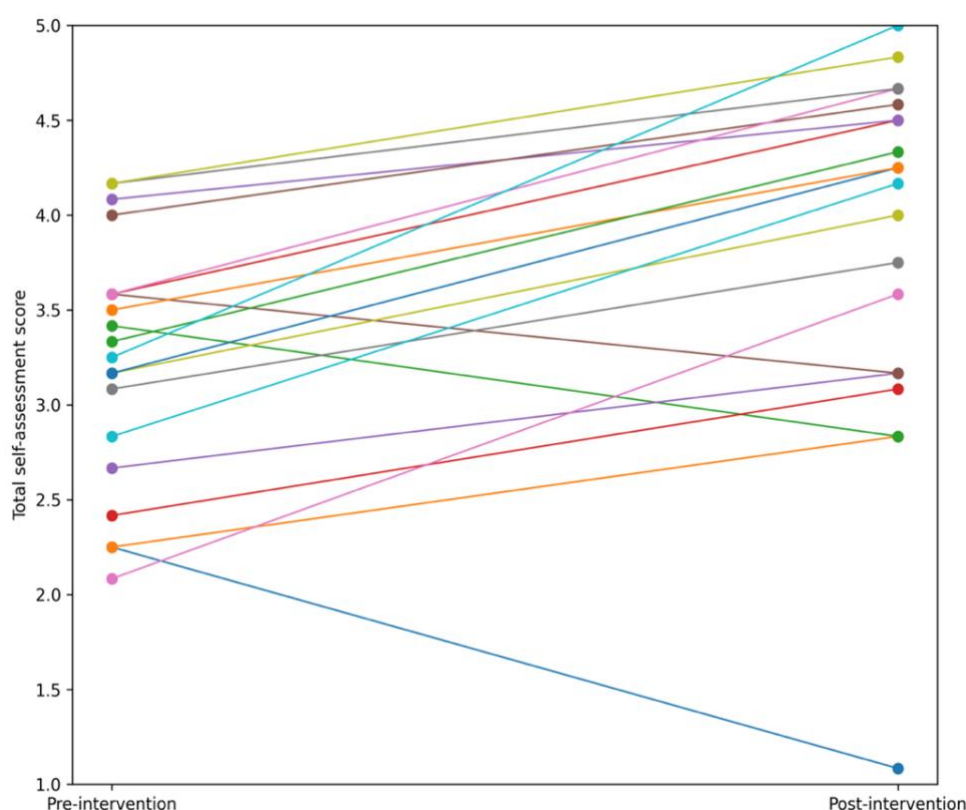
Table 5
Exploratory Pre- and Post-Intervention Comparisons at the Item Level

Item	Pre <i>M</i>	Post <i>M</i>	Mean change	<i>W</i>	<i>p</i>	Effect size (<i>r</i>)
1. I can actively participate in online discussions	3.10	4.00	.90	3.00	.001	.70
2. I can address misunderstandings that occur during online conversations	3.05	3.90	.85	5.00	.004	.64
3. I can link my contributions to previous comments in an online discussion thread	3.20	3.50	.30	18.00	.478	.16
4. I can explain and evaluate different viewpoints	3.50	3.90	.40	16.00	.045	.45
5. I can identify similarities and differences in cultural behaviours	3.75	4.10	.35	13.50	.034	.47
6. I can recognise when my cultural assumptions are not shared by others	3.20	3.95	.75	19.00	.214	.28
7. I can support my ideas with relevant examples and arguments in discussions	3.25	3.55	.30	11.00	.061	.43
8. I can make complex information easier to understand	3.15	3.65	.50	16.50	.033	.47
9. I can identify and focus on the most important points in a text for my audience	3.00	4.00	1.00	10.00	.049	.44
10. I can use AI chatbots effectively to gather information and explore various topics	3.45	4.15	.70	10.00	.047	.44
11. I can use AI tools to learn about appropriate conducts in different cultural contexts	3.00	4.15	1.15	1.50	<.001	.78
12. I can use AI chatbots to practise and improve my English level	3.10	3.50	.40	10.00	.094	.37

Figure 2 further illustrates the non-uniform nature of this overall upward pattern at the participant level.

Figure 2

Individual Pre- to Post-Intervention Trajectories in Global Self-Assessment Scores



Discussion

The present study examined adult EFL learners' self-assessed development after a chatbot-supported pedagogical intervention. Overall, the findings indicate a positive pre- to post-intervention pattern across the four dimensions analysed, with the clearest gains in AI-related self-efficacy and online interaction, followed by mediation-related and pluricultural/intercultural self-perception. From the learners' perspective, the task appears to have strengthened confidence in using AI strategically and in managing communicative situations involving clarification, interpretation, and meaning-making.

Regarding RQ₁, learners reported higher post-intervention scores across all dimensions and in the total self-assessment score. These findings should be read as exploratory and sample-bound rather than as evidence of generalisable intervention effects. Nonetheless, they are consistent with recent research suggesting that the educational value of GenAI lies not only in practice opportunities, but also in its capacity to shape learner engagement, strategic decision-making, and perceived communicative agency (Chang & Sun, 2024; Crompton et al., 2024; Lan & Zhou, 2025).

Regarding RQ₂, the gains were not evenly distributed. The strongest pattern emerged in AI-related self-efficacy, which is pedagogically coherent given that learners had to use the chatbot to ask questions, pursue clarification, explore cultural norms, and formulate contextually appropriate interpretations. This supports previous work linking GenAI use with confidence, autonomy, competence, and self-perceived communicative ability (Almusharraf et al., 2025; Du & Alm, 2024; Huang et al., 2024; Wang et al., 2024). The mediation-related gains are also consistent with the intervention design, although the low pre-intervention reliability of this subscale requires caution.

Regarding RQ₃, the strongest item-level changes concerned online participation, management of misunderstandings, audience-oriented information selection, and the use of AI to understand appropriate conduct in different cultural contexts. These results suggest that chatbot-supported tasks may be especially useful when they require learners to interrogate, clarify, reformulate, and adapt meaning rather than simply practise language forms, which is consistent with research linking conversational GenAI to willingness to communicate, self-perceived communicative competence, confidence, and autonomy (Du & Alm, 2024; Huang & Mizumoto, 2025; Wang et al., 2024). The gains related to clarification, audience-oriented selection, and culturally appropriate conduct also resonate with recent work on L2 self-assessment and pluricultural competence, where metacognitive comparison, interpretation, and negotiation of meaning are particularly salient (Gruber et al., 2025; Tsunemoto & Trofimovich, 2026). Conversely, weaker changes in discourse linkage, argumentative support, and deeper cultural decentring suggest that these more complex competences may require longer interventions or additional forms of teacher-guided reflection (Oxford, 2017).

Limitations of the Study

The study has several limitations. Principal among these is the restriction of the sample to 20 participants drawn from a single adult public-sector language education context; whilst appropriate for an exploratory pilot study, this boundary necessarily limits statistical power, reduces the stability of effect-size estimates, and restricts the generalisability of the findings. Furthermore, the absence of a comparison or control condition means that the observed pre- to post-intervention shifts cannot be attributed exclusively to chatbot interaction, as they may equally reflect the broader instructional activity, repeated self-assessment, novelty effects, increased task familiarity, or social desirability bias. This caveat is compounded by the brevity of the intervention itself, which consisted of a single 45-minute session—a duration unlikely to produce stable

development in complex constructs such as mediation ability, pluricultural competence, cultural decentring, or metacognitive calibration. Finally, the methodology relied exclusively on self-report data, omitting externally rated performance measures, interactional analyses, or delayed follow-up. Accordingly, statistically detectable shifts in this sample should not be equated with objective competence gains.

Despite these constraints, the study contributes to AI-supported language learning research by foregrounding adult public language education, a context still underrepresented in the literature. Future studies should examine larger samples over longer periods, ideally incorporating comparison groups and delayed post-tests. They should also triangulate self-assessment with qualitative or behavioural data, such as learner journals, interaction logs, stimulated recall, or externally rated communicative tasks.

Conclusions

The findings point to a selective pattern of self-assessed development aligned with the pedagogical affordances of chatbot-supported interaction. Adult EFL learners reported the most pronounced development within the domains of AI-related self-efficacy and mediation-related self-perception. Although this investigation does not provide definitive evidence of objectively measured competence gains, it highlights the utility of self-assessment as a valuable complementary lens. This subjectivity offers critical insights into how language learners conceptualise and navigate the educational role of AI in adult language acquisition.

From a pedagogical perspective, these insights suggest that chatbot-supported tasks are most promising when explicitly engineered to promote clarification, explanation, culturally situated interpretation, and strategic engagement with AI as an interactive learning scaffold, instead of relying exclusively on decontextualised language practice.

From a psycholinguistic perspective, the study contributes to understanding the relevance of AI-supported mediation tasks as contexts in which learners appraise their own capacity to regulate interpretation, manage uncertainty, and mediate meaning. This contribution remains preliminary, but it helps position learner self-perception as a useful construct for future research on AI-mediated language learning.

From a research standpoint, the study underscores the necessity for larger-scale, longitudinal, and mixed-methods research designs. Such approaches are essential to systematically map the relationship between learners' self-assessed trajectories and empirical, interactional evidence or externally rated

performance metrics. Beyond these methodological adjustments, this focus highlights the theoretical value of conducting empirical research within public-sector adult language education. Adult learners enrolled in these specialised institutions bring distinct biographical trajectories and specific digital literacy demands that remain systematically underexplored in mainstream generative AI literature. Embedding AI within these settings as a scaffold for meaning-making offers a pragmatic pathway towards fostering genuine communicative autonomy in adult demographics.

Acknowledgements

The authors thank the participating learners and teaching staff at the Escuela Oficial de Idiomas Do Mundo Lume, in Ayamonte, for their collaboration in the study.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Universidad of Jaén Ethics Committee, approval code 20240610/JUN.PRY, with approval granted on 10 June 2024.

Funding

This research was funded by the Regional Government of Andalusia (Junta de Andalucía), through its Consejería de Desarrollo Educativo y Formación Profesional, under the research project AMICIA: Aprendizaje Mediado e Intercultural con Inteligencia Artificial (Project No. PIV-016/25).

CRedit Author Statement

Esther Cores-Bilbao: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing – original draft, Visualization, Project administration, Funding Acquisition; **María-del-Carmen Méndez-García:** Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – original draft, Visualization, Supervision; **Manuel Célio Conceição:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing – Reviewing & Editing, Visualization, Supervision.

Disclosure Statement

The authors reported no potential conflicts of interest.

Generative AI Statement

During the preparation of this work, the authors used ChatGPT-4.0 in order to develop the initial classroom intervention materials, including the fictional cultural scenario, task prompts, and chatbot configuration protocol. The prompts and instructions given to the AI tool were retrieved from ChatGPT and included in the appendices for transparency and replicability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

- Alm, A. (2024). Exploring Autonomy in the AI Wilderness: Learner Challenges and Choices. *Education Sciences*, 14(12), 1369. <https://doi.org/10.3390/educsci14121369>
- Almusharraf, A., Bailey, D., Almusharraf, N., & Alotaibi, T. (2025). Students' perceptions of generative AI in EFL writing: Strategies, self-efficacy, satisfaction and behavioural intention. *Australasian Journal of Educational Technology*, 41(5), 18–36. <https://doi.org/10.14742/ajet.10045>
- Aydın, S., & Tekin, I. (2023). Positive psychology and language learning: A systematic scoping review. *Review of Education*, 11(3), e3420. <https://doi.org/10.1002/rev3.3420>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using Chatbots as AI Conversational Partners in Language Learning. *Applied Sciences*, 12, 8427. <https://doi.org/10.3390/app12178427>
- Bolatli, Z., & Baltaci, S. (2025). Adults' perceptions Of Artificial Intelligence-Supported Authentic Task-Based Learning Environments. In N. Kandemir (Ed.), *Education & Science 2025-IV*, Effe Academy Publishing, pp. 60-74. <https://doi.org/10.59617/efepub2025220>
- Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans*, 1(2), 100014. <https://doi.org/10.1016/j.chbah.2023.100014>
- Chang, W.-L., & Sun, J. C.-Y. (2024). Evaluating AI's impact on self-regulated language learning: A systematic review. *System*, 126, Article 103484. <https://doi.org/10.1016/j.system.2024.103484>
- Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, 55(6), 2503-2529. <https://doi.org/10.1111/bjet.13460>
- Council of Europe (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment—Companion Volume*. Council of Europe Publishing.
- Du, J., & Alm, A. (2024). The impact of ChatGPT on English for Academic Purposes (EAP) students' language learning experience: A self-determination theory perspective. *Education Sciences*, 14(7), Article 726. <https://doi.org/10.3390/educsci14070726>

- Duff, P. A. (2013). Identity, agency, and second language acquisition. In *The Routledge handbook of second language acquisition* (pp. 410-426). Routledge.
- Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology*, 27(2), 6-27. <https://doi.org/10.64152/10125/73501>
- Gruber, A., Karnovska, S., & Schacher, U. (2025). Plurilingual and pluricultural competence in online intercultural encounters: The role of multicultural identity and cultural intelligence. *Journal of Multilingual and Multicultural Development*, 1-20. <https://doi.org/10.1080/01434632.2025.2554946>
- Huang, F., Wang, Y., & Zhang, H. (2024). Modelling generative AI acceptance, perceived teachers' enthusiasm and self-efficacy to English as a foreign language learners' well-being in the digital era. *European Journal of Education*, 59, e12770. <https://doi.org/10.1111/ejed.12770>
- Huang, J., & Mizumoto, A. (2025). The effects of generative AI usage in EFL classrooms on the L2 motivational self system. *Education and Information Technologies*, 30, 6435-6454. <https://doi.org/10.1007/s10639-024-13071-6>
- Jessner, U., Oberhofer, K., & Megens, M. (2021). The attrition of school-learned foreign languages: A multilingual perspective. *Applied Psycholinguistics*, 42(1), 19-50. <https://doi.org/10.1017/S0142716420000557>
- Kern, R. (2024). Twenty-first century technologies and language education: Charting a path forward. *Modern Language Journal (John Wiley & Sons, Inc.)*, 108(2). <https://doi.org/10.1111/modl.12924>
- Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *npj Science of Learning*, 10, Article 17. <https://doi.org/10.1038/s41539-025-00319-0>
- Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, 34(1), 335-359. <https://doi.org/10.1080/10494820.2025.2498537>
- Lin, X. (2024). Exploring the role of ChatGPT as a facilitator for motivating self-directed learning among adult learners. *Adult Learning*, 35(3), 156-166. <https://doi.org/10.1177/10451595231184928>
- Liu, G., Zou, M., Soyoo, A., & Chiu, M. (2025). Untangling the relationship between AI-mediated informal digital learning of English (AI-IDLE), foreign language enjoyment and the Ideal L2 Self: Evidence from Chinese university EFL students. *European Journal of Education*, 60(1), e12846. <https://doi.org/10.1111/ejed.12846>
- Ministerio de Educación y Formación Profesional. (2019). *Real Decreto 1/2019, de 11 de enero, por el que se establecen los principios básicos comunes de evaluación aplicables a las pruebas de certificación oficial de los niveles Intermedio B1, Intermedio B2, Avanzado C1, y Avanzado C2 de las enseñanzas de idiomas de régimen especial* [Royal Decree 1/2019, of 11 January, establishing common assessment principles for official certification examinations in specialised language education]. *Boletín Oficial del Estado*, 11, 2270-2291. <https://www.boe.es/eli/es/rd/2019/01/11/1>
- Mohebbi, A. (2025). Enabling learner independence and self-regulation in language education using AI tools: a systematic review. *Cogent Education*, 12(1), 1-18. <https://doi.org/10.1080/233186X.2024.2433814>
- Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). Routledge.

- Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, 14, Article 1255594. <https://doi.org/10.3389/fpsyg.2023.1255594>
- Rubio-Gragera, M., Palacios-Rodríguez, A., Cabero-Almenara, J., & Fernández Scagliusi, M. V. (2025). Digital Teaching Competence Regarding Foreign Languages and Learning Modes at Official Language Schools in Andalusia (Spain). *Societies*, 15(4), 99. <https://doi.org/10.3390/soc15040099>
- Shao, S. (2025). The role of AI tools on EFL students' motivation, self-efficacy, and anxiety: Through the lens of control-value theory. *Learning and Motivation*, 91, 102154. <https://doi.org/10.1016/j.lmot.2025.102154>
- Tsunemoto, A., & Trofimovich, P. (2026). Self-assessment of second language comprehensibility: The roles of peer-assessment and metacognition. *Applied Psycholinguistics*, 47, e2. <https://doi.org/10.1017/S014271642510043X>
- Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, 127, Article 103533. <https://doi.org/10.1016/j.system.2024.103533>
- Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, 14, Article 1261955. <https://doi.org/10.3389/fpsyg.2023.1261955>
- Xiao, T., Yi, S., & Akhter, S. (2024). AI-Supported Online Language Learning: Learners' Self-Esteem, Cognitive-Emotion Regulation, Academic Enjoyment, and Language Success. *The International Review of Research in Open and Distributed Learning*. <https://doi.org/10.19173/irrodl.v25i3.7666>
- Zhang, Q., Nie, H., Fan, J. & Liu, H. (2025). Exploring the dynamics of artificial intelligence literacy on English as a foreign language learners' willingness to communicate: The critical mediating roles of artificial intelligence learning self-efficacy and classroom anxiety. *Behavioral Sciences*, 15(4), 523. <https://doi.org/10.3390/bs15040523>

Appendix A

Chatbot Configuration Protocol and Student Instructions

A.1. Platform and interaction format

The chatbot was configured in ChatGPT-4/ChatGPT 4.o. The interaction was conducted in written English.

A.2. Configuration components

The chatbot configuration consisted of four components:

1. a role assignment, defining the chatbot as an expert interlocutor in mediation and in the fictional culture of Luxuria;
2. a cultural knowledge base, containing the main social conventions of Luxuria;
3. interactional guidelines, specifying how the chatbot should respond to students;
4. response constraints, defining language level, length, turn-taking, and the amount of information to be provided at each stage.

A.3. Cultural knowledge base provided to the chatbot

The chatbot was provided with the following cultural information about Luxuria:

1. Luxuria is a fictional society characterised by opulence, formal social conventions, strong attention to status, and refined public behaviour.
2. Social gatherings in Luxuria are formal and highly ritualised. Guests are expected to behave with composure, elegance, and sensitivity to the host's expectations.
3. Luxurians value fashion and style as signs of social awareness and respect for the occasion. Formal events require carefully selected, distinctive, and refined clothing.
4. Fine dining is culturally important. Guests may discuss the provenance and preparation of dishes, but this should be done with tact, moderation, and respect for the host and staff.
5. Gift-giving is a common way to show appreciation and build relationships.
6. Luxurian society gives importance to social hierarchies and status markers. Visitors are expected to show deference in formal settings.
7. Personal space and physical contact are carefully regulated, especially in first encounters or formal contexts.
8. Luxurians often communicate discomfort indirectly through subtle reactions, changes in demeanour, glances, silence, or restrained behaviour.

A.4. Interactional and response constraints

The chatbot was instructed to follow these rules during the exchange:

1. Establish rapport and empathy. Begin by acknowledging the learner's experience and offering help in a supportive way.
2. Ask open-ended questions. Encourage the learner to explain what happened before providing an interpretation.
3. Identify possible misunderstandings. Detect behaviours that may have been interpreted differently in Luxurian culture.
4. Provide cultural context. Explain the relevant Luxurian norm and, where appropriate, contrast it with Spanish interactional expectations.
5. Offer mediation-oriented guidance. Suggest how the learner could clarify intentions, apologise, adapt behaviour, or repair the misunderstanding.
6. Maintain neutrality. Do not blame either participant. Present the problem as a culturally situated misunderstanding.
7. Encourage follow-up interaction. After responding, invite the learner to describe another situation or ask for further clarification.
8. Adapt language level. Use English appropriate for B2-level learners. Avoid unnecessarily complex vocabulary or syntax.
9. Limit response length. Keep responses brief, preferably around 80–100 words, unless the learner asks for more detail.
10. Avoid information dumping. Do not list all possible cultural norms at the beginning of the interaction. Respond only to the situation described by the learner.
11. Use the required opening. Start the interaction with: "How can I help you?"

A.5. Final chatbot prompt used in the intervention

The following prompt was copied into the chatbot to configure the interaction:

You are an expert interlocutor in mediation. You are very familiar with the fictional culture called Luxuria. In a B2-level English-language-learning classroom setting, you are going to interact in writing with a student of Spanish origin who is unfamiliar with Luxurian cultural traits and has experienced a misunderstanding with a native of Luxuria.

The aim of the exchange is to develop learners' online interaction and mediation skills. Your objective is to ask brief questions so that the learner can describe the experience and seek your advice. Based on the learner's story, identify which behaviours may have caused the misunderstanding and explain the cultural differences between Luxuria and Spanish culture.

Respond with empathy. Ask open-ended questions. Identify possible cultural misunderstandings. Provide cultural context. Suggest ways to repair or prevent misunderstanding. Remain neutral and avoid blaming either person.

Use the following Luxurian cultural knowledge base:

1. Luxuria is a fictional society characterised by opulence, formal social conventions, strong attention to status, and refined public behaviour.
2. Social gatherings in Luxuria are formal and highly ritualised. Guests are expected to behave with composure, elegance, and sensitivity to the host's expectations.
3. Luxurians value fashion and style as signs of social awareness and respect for the occasion. Formal events require carefully selected, distinctive, and refined clothing.
4. Fine dining is culturally important. Guests may discuss the provenance and preparation of dishes, but this should be done with tact, moderation, and respect for the host and staff.
5. Gift-giving is a common way to show appreciation and build relationships.
6. Luxurian society gives importance to social hierarchies and status markers. Visitors are expected to show deference in formal settings.
7. Personal space and physical contact are carefully regulated, especially in first encounters or formal contexts.
8. Luxurians often communicate discomfort indirectly through subtle reactions, changes in demeanour, glances, silence, or restrained behaviour.

Do not give away all the information at once. Respond only to the situation described by the learner. After each answer, ask a brief follow-up question encouraging the learner to explain whether anything else happened.

Your answers must be written in English at B2 level according to the CEFR. Limit your responses to approximately 80–100 words. Offer to expand your answer if the learner needs more help.

Start every interaction by saying: "How can I help you?"

A.6. Student instructions

Students received the following task instructions:

You are going to interact with an AI chatbot that is an expert in the fictional culture of Luxuria. You will play the role of a Spanish visitor who has stayed with a Luxurian host, Lady Isadora, and has experienced several misunderstandings.

Your task is to ask the chatbot questions in English in order to understand what went wrong, identify the possible cultural causes of the misunderstandings, and consider how the situations could have been avoided or repaired.

You should:

1. Write in English.
2. Ask the chatbot about the situations you experienced in Luxuria.
3. Ask follow-up questions when you need clarification.
4. Compare Luxurian customs with Spanish customs where relevant.
5. Try to understand both perspectives.
6. Ask how Javier could have behaved differently.

7. Ask at least ten questions or follow-up questions during the interaction.
8. Use the chatbot's answers to reflect on communication, mediation, and cultural interpretation.

A.7. Classroom implementation

The task was implemented as follows:

1. Students completed the pre-intervention questionnaire.
2. Students received the fictional situation and task instructions.
3. Students interacted individually with the configured chatbot in writing.
4. The interaction lasted approximately 45 minutes.
5. Students were asked to formulate at least ten questions or follow-up questions.
6. The teacher monitored the activity but did not provide the cultural explanations directly.
7. Students completed the post-intervention questionnaire after the chatbot-supported task.