

Beyond the interface: Supporting beginner-level CEFR mediation and learner autonomy with ChatGPT

Language Teaching Research

I-35

© The Author(s) 2025

Article reuse guidelines:
sagepub.com/journals-permissionsDOI: [10.1177/13621688251393481](https://doi.org/10.1177/13621688251393481)
journals.sagepub.com/home/ltr**Alice Lukešová** 

Czech Technical University in Prague, Czech Republic

Petra Juna Jennings 

Czech Technical University in Prague, Czech Republic

Abstract

This study explores how ChatGPT, used as a dialogic mediator, supports CEFR-based mediation and learner autonomy among multilingual university students at A1 to A2 proficiency levels in Spanish. A key finding is the contrast between learners' observable autonomy during AI-mediated tasks and the deeper strategic awareness revealed in post-task reflections. Grounded in sociocultural theory and the CEFR (Common European Framework of Reference for Languages) Companion Volume (2020), the study examines how learners used ChatGPT to interpret texts, clarify vocabulary, and engage in structured argumentation. All tasks were conducted in Spanish, which was neither the students' native language nor the primary language in which ChatGPT was trained. Twenty-four students completed structured Spanish-language tasks with ChatGPT and submitted written reflections in their native languages. Qualitative content analysis, guided by CEFR mediation descriptors, revealed frequent use of strategies related to Mediating Texts and Facilitating Communication. During the tasks, students often relied on ChatGPT for linguistic support. However, their reflections demonstrated higher levels of metacognitive autonomy, including goal-setting, critical evaluation of AI feedback, and comparison of digital tools. These findings highlight the importance of combining AI-supported interaction with structured reflection to promote communicative competence and digital agency. The study also points to the need for adapting CEFR mediation frameworks to better capture the dynamics of AI-mediated language learning.

Corresponding author:

Alice Lukešová, Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, Prague 166 27, Czech Republic.

Email: lukesali@fel.cvut.cz

Keywords

AI-assisted language learning, CEFR, ChatGPT, language mediation, learner autonomy, multilingual learners

I Introduction

Since its public release in 2022, ChatGPT has become one of the most widely used large language models (LLMs) in education, supporting learners across diverse disciplines (Crompton & Burke, 2023; Wang et al., 2024; Wollny et al., 2021). Among the available LLMs, such as Google Gemini, Claude, and Microsoft Copilot, ChatGPT remains especially popular due to its accessibility and ease of use. For this study, we intentionally used ChatGPT 3.5, the freely available version, to ensure equal access to learning tools, regardless of students' financial means. This approach allowed all participants to engage fully in the activity without being limited by access to premium AI features.

While prior research has explored how ChatGPT supports discrete language skills such as grammar, vocabulary, and writing (Imran & Almusharraf, 2023), less attention has been paid to its potential in fostering mediation and learner autonomy, two core competencies in the CEFR Companion Volume (Council of Europe, 2020), which defines mediation as the act of relaying or adapting meaning for others in a communicative context. These competencies are central to authentic communication but remain under-examined in AI-mediated environments.

Godwin-Jones (2023) highlights the potential of AI to scaffold learner interaction and promote exploratory dialogue, yet also warns that LLMs may encourage passive consumption if learners do not critically engage with the tool. This study builds on that tension by revealing a key discrepancy: while beginner learners often relied on ChatGPT during tasks, their post-task reflections demonstrated advanced metacognitive strategies such as goal-setting, error evaluation, and tool comparison. This contrast underscores the importance of structured reflection in surfacing otherwise invisible dimensions of learner autonomy.

Similarly, Jeon and Lee (2023) emphasize the importance of pedagogical frameworks that align AI interaction with language learning goals, particularly in low-proficiency settings. However, their work does not explicitly engage with CEFR mediation. By applying CEFR mediation descriptors to student-AI interactions, this study extends Lee's insights and provides a concrete framework for evaluating communicative development in AI-supported contexts.

This study addresses this gap by examining how multilingual learners at A1–A2 proficiency levels utilize ChatGPT to perform CEFR-based mediation tasks and cultivate autonomous learning behaviours. Drawing on interaction data and learner reflections, the research explores ChatGPT not only as a linguistic scaffold but also as a dialogic mediator that facilitates the development of communicative and cognitive competencies.

II Research questions

- Research question 1: How do A1–A2 learners use ChatGPT to perform CEFR-defined mediation activities?

- Research question 2: What learner autonomy behaviours are observable in AI-mediated language tasks?
- Research question 3: How do students engage critically with ChatGPT during and after interaction?

III Literature review

1 CEFR mediation and the companion volume

The CEFR Companion Volume (Council of Europe, 2020) defines mediation as the process of constructing meaning by relaying, simplifying, or adapting content for others. Mediation is organized into three categories: mediating texts, mediating concepts, and facilitating communication. These categories reflect integrated communicative skills, such as summarizing, clarifying, paraphrasing, and adapting messages for specific audiences. While originally developed for human-to-human communication, CEFR descriptors provide a robust framework for analysing human–AI interactions.

2 Sociocultural theory and AI as a mediating tool

Rooted in Vygotsky's (1978) concept of the Zone of Proximal Development, sociocultural theory emphasizes that learning occurs through mediated interaction. While ChatGPT is technically a tool, in the context of this study it functions as a dialogic mediator, an agent-like interface that interacts responsively, shaping learner output and co-constructing meaning. Following Lantolf (2000), we conceptualize ChatGPT as a semi-agentive scaffold: a system that offers dialogic affordances similar to peer interaction, even though it lacks human intentionality or consciousness.

Although dialogic mediation and AI-assisted scaffolding both involve supportive processes during learning, they differ significantly in their dynamics and implications for learner agency. AI-assisted scaffolding typically refers to instructional support that enables learners to complete tasks beyond their current capabilities. This support often includes vocabulary assistance, grammar correction, or content explanations and tends to be system-initiated and unidirectional in nature. In contrast, dialogic mediation involves reciprocal interaction in which meaning is actively co-constructed through negotiation, clarification, and adaptation.

In the context of this study, ChatGPT is understood not merely as a scaffold that delivers linguistic assistance but as a dialogic mediator that participates in responsive exchanges. These exchanges enable learners to formulate questions, refine their output, and critically reflect on the interaction. While ChatGPT does not possess intentionality or consciousness, its dialogic functionality creates opportunities for learners to engage in meaning-making processes and develop metacognitive strategies. This interpretation aligns with the CEFR's conceptualization of mediation as a collaborative, audience-aware activity that integrates comprehension, production, and reflection.

3 Learner autonomy and self-regulated learning

Learner autonomy, defined as the ability to take charge of one's own learning (Holec, 1981), is closely linked to self-regulated learning processes such as planning, monitoring, and evaluating (Little, 1991; Zimmerman, 2002). In AI-mediated environments, this autonomy includes initiating interactions, managing digital tools, and critically evaluating feedback. Previous research indicates that autonomous learners tend to perform better in computer-assisted language learning (CALL) when they are given opportunities for reflection and control (Reinders & Benson, 2017).

In this study, learner autonomy refers broadly to the learner's capacity to direct their own language learning, both during tasks and in independent study. It involves behaviours such as initiating interactions, selecting appropriate strategies, and responding adaptively to feedback. Within this broader framework, metacognitive autonomy is defined as the ability to reflect on, evaluate, and regulate one's strategic use of language tools and resources. It includes actions such as setting learning goals, critically assessing AI-generated feedback, comparing digital tools, and engaging in post-task reflection. Whereas learner autonomy is often observable during task performance, metacognitive autonomy tends to emerge more clearly through reflective analysis after the learning activity.

4 Critical digital and AI literacy

As learners increasingly engage with AI systems, developing critical digital literacy is essential, not only for navigating tools effectively but also for evaluating the reliability, accuracy, and socio-technical implications of their outputs. Pangrazio and Selwyn (2019) argue that digital literacy must extend beyond functional skills to encompass reflective and ethical dimensions of technology use. Similarly, Jones and Hafner (2012) emphasize that language learners must critically examine how digital tools shape communication, identity, and knowledge production. Buckingham (2010) further highlights that critical digital literacy includes awareness of the power dynamics embedded in digital environments and algorithmic outputs.

AI literacy frameworks (Luckin et al., 2016) build on these perspectives, urging learners and educators to understand not only how AI systems work but also how to question their assumptions, biases, and pedagogical roles. In the context of AI-mediated language learning, this form of literacy intersects with learner autonomy and reflective practice, fostering a capacity to use AI tools like ChatGPT not just as linguistic aids, but as objects of inquiry and critique.

5 Mediation in multilingual contexts

This study also addresses concerns about linguistic bias in LLMs, which are often trained primarily on English data (Brown et al., 2020; Seghier, 2023). To explore mediation in non-English environments, this research was conducted in Spanish, a major global language that is underrepresented in LLM training data. Participants were multilingual students whose native languages were neither English nor Spanish. This multilingual setting

allowed for the analysis of cross-linguistic mediation, language alternation, and meaning-making using learners' complete linguistic repertoires.

6 AI in language learning: Efficacy, autonomy, and multilingual affordances

A wave of 2025 research has expanded our understanding of how conversational and generative AI can be harnessed in language education, offering robust empirical evidence and nuanced insights into learner experience. A meta-analysis by Gu (2025) synthesising 55 studies reported consistent positive effects of conversational AI on language proficiency, motivation, and learner engagement, underscoring its pedagogical value across diverse contexts. These benefits are often linked to adaptive learning pathways, with Nwanakwaugwu et al. (2025) demonstrating that AI-driven personalization promotes learner autonomy by tailoring instruction to individual needs and proficiencies. In the affective domain, Kittridge et al. (2025) found that generative AI integrated into mobile language applications significantly enhanced learners' self-efficacy and communicative confidence, while Kruk and Kałužna (2024) identified its capacity to foster motivation, positive emotional engagement, and translation skills development. Beyond discrete skills, Lin et al. (2025) highlighted that AI-supported multimodal composition tasks can strengthen critical thinking, digital literacy, and multilingual competence, particularly in multicultural classrooms. Furthermore, Akhter et al. (2025) showed that embedding AI activities within a CEFR-informed framework measurably improves EFL learners' communication skills, reinforcing the relevance of CEFR descriptors in structuring AI-mediated pedagogy. Collectively, these 2025 studies establish a contemporary evidence base for examining how AI tools such as ChatGPT can facilitate CEFR-based mediation and learner autonomy in multilingual beginner-level contexts.

IV Methodology

This study employed a qualitative content analysis framework guided by CEFR mediation descriptors (Council of Europe, 2020). The aims were threefold: (1) to investigate how ChatGPT supports CEFR-based mediation in multilingual learners, (2) to identify the types of mediation strategies students employed, and (3) to explore the development of learner autonomy and critical engagement with AI.

I Participants

The study involved 24 university students from the Czech Technical University, all enrolled in the Computer Science and Open Informatics program. Participants were purposefully selected based on their prior exposure to AI tools, ensuring that the study's focus remained on language learning and mediation, rather than on digital literacy or tool familiarization. Given their routine engagement with advanced technologies, their digital competencies were presumed to be high, and no formal assessment of digital skills was conducted. Although no pre-task questionnaire was administered, all participants

informally confirmed prior experience with ChatGPT or comparable large language models and reported confidence using conversational AI interfaces. This was further supported by the program's academic curriculum, which includes a detailed study of large language models (LLMs) within artificial intelligence and natural language processing courses.

A CEFR-aligned diagnostic test was administered to assess Spanish language proficiency. Based on the results, students were divided into two groups: Group 1 (A1, $n=12$) and Group 2 (A2, $n=12$). Participants ranged in age from 19 to 26 years and included both bachelor's and master's students (16 males and 8 females). Their linguistic backgrounds were diverse. In addition to Spanish, participants used Czech, Slovak, Ukrainian, French, German, English, Italian, and Serbian during the tasks, languages that supported contrastive mediation and metalinguistic reflection.

2 Task design and procedure

The activity was divided into two parts to address the CEFR categories of mediation: Mediating Texts, Mediating Concepts, and Facilitating Communication. The tasks also promoted learner autonomy and digital critical engagement.

The study was conducted over a one-month period, from March 1 to March 31, 2024. During this time, students completed a series of structured language learning tasks in individual sessions with ChatGPT. Each session was designed to be completed within a flexible timeframe, allowing students to engage with the tasks at their own pace outside of classroom settings. All tasks were completed independently to ensure that learners interacted with the AI autonomously and without influence from peers or instructors. Students received activities with model prompts, described in detail in Appendix A, which they could use as given or modify to suit their preferences. They were encouraged to adapt the tasks to their needs, seek clarification on relevant points, and explore topics of personal interest in greater depth. The aim was to maximize opportunities for individualized task adaptation.

a Text interpretation and vocabulary clarification (Part 1). Each group received an original text tailored to their proficiency level. One text discussed lifestyle differences between rural and urban areas, while the other focused on Spanish holidays and cultural practices. Texts were authored by the researchers to avoid copyright issues (Lucchi, 2023). The following is a description of the texts:

- A1 group: 'Fiestas españolas': 288 words. Topic: selection of well-known and less familiar Spanish festivals. Vocabulary focused on celebration-related nouns, adjectives, and present-tense verbs. Syntax was predominantly simple main clauses with frequent use of *hay*, *es/son*, and *se celebra*, and limited coordination (*y*, *también*). Verb forms were almost exclusively present indicative, with occasional periphrastic future (*va a haber*). All vocabulary was selected to match the A1 level, with some culturally relevant terms retained in L2 for clarification during mediation (e.g. *cofradía*, *encierro*).

- A2 group: ‘¿Pueblo o ciudad?’: 285 words. Topic: comparison of the advantages and disadvantages of rural and urban living. Vocabulary included terms related to the environment, transport, and lifestyle. Syntax featured mostly simple sentences with occasional coordination (*y, pero*) and limited subordination (*porque, aunque*). Verb tenses included the present indicative, using *hay, puede*, and modal constructions to describe possibilities. Lexical choice was controlled to align with A2 descriptors, with some higher-level words retained for mediation practice (e.g. *inconveniente, desplazarse*).

Students worked independently with ChatGPT to interact with the texts. Common prompts included ‘What does this word mean?’ and ‘Can you explain this expression to me?’ to support lexical mediation. Learners also requested summaries and comprehension questions to rephrase, simplify, and clarify text content.

b Argumentation and meaning negotiation (Part 2). Students were assigned one side of a structured debate, with ChatGPT taking the opposing role. Topics included urban versus rural living and the cultural significance of Spanish holidays. Learners received a prompt bank with expressions for agreeing, disagreeing, and requesting clarification. These exchanges promoted the use of mediation strategies and encouraged student initiative. Students also asked ChatGPT to correct and explain their output, supporting metacognitive reflection.

c Duration and scope. The activity was designed for short, focused blocks to keep cognitive load manageable for beginner learners working alone. The recommended schedule is presented in Table 1, which summarizes the timing and sequencing of the activity steps for both parts of the task.

This structure was presented as a guideline, but actual completion times varied considerably among students. Some reported *esta conversación ha durado aproximadamente entre 20 y 30 minutos* (‘this conversation lasted approximately between 20 and 30 minutes’), while others indicated *aproximadamente entre 60 y 90 minutos* (‘approximately between 60 and 90 minutes’). Given that ChatGPT does not have an internal timer to record interaction length, these self-reported durations cannot be verified with complete accuracy. The average self-reported completion time across all participants was 63 minutes. There was no statistically significant difference between the A1 and A2 groups: while A2 students tended to produce longer responses, A1 students reported needing more time to think through their replies. This variation in timing does not undermine comparability, as both groups completed the full set of steps in the prescribed order, with differences reflecting individual pacing preferences rather than unequal task exposure.

3 Data collection

Primary data included anonymized transcripts from student–ChatGPT interactions. Students also submitted written reflections in their native language to elicit more detailed insights into their experiences, perceived benefits, challenges, and suggestions.

Table 1. Recommended time allocation per activity step.

Step	Activity description	Estimated time
<i>Comprehension:</i>		
1	Summary read/compare	10–15 min
2	Vocabulary translate/explain + quick re-read	15–25 min (A1 ≈ 15–20 min; A2 ≈ 20–25 min)
3	Answer five comprehension questions	10–15 min
4	Review corrections + revise answers	10–15 min
5	Read explanation, find evidence, adjust	10–15 min
<i>Discussion:</i>		
1	Expressions list + quick rehearsal	8–12 min
2	Role-play discussion with corrections	20–30 min (A1 ≈ 20–25 min; A2 ≈ 25–30 min)

4 Data analysis

a Analysis of student conversations. Student utterances were coded using CEFR mediation categories:

- Mediating texts (e.g. summarizing, simplifying);
- Mediating concepts (e.g. paraphrasing, clarifying);
- Facilitating communication (e.g. adapting language, negotiating meaning).

While originally intended for human-to-human communication, CEFR descriptors were used here as a flexible analytical lens to examine how learners mediated meaning in their interactions with ChatGPT. Particular attention was paid to how mediation behaviours emerged within the unique dynamics of human–AI dialogue.

Additional dimensions included:

- Learner autonomy (initiative, strategic behaviour, self-regulation);
- Digital literacy and critical thinking (questioning AI output, identifying errors);
- Linguistic complexity (syntactic range, lexical variety, pragmatic use).

Each student's interaction with ChatGPT was analysed for the presence of autonomy-, digital-thinking-, and critical-thinking-related behaviours. Behaviours were coded as 'present' if they occurred at least once during the interaction, regardless of how many times. Sessions in which no examples of these behaviours appeared are labelled 'No coded autonomy/critical thinking behaviours observed'.

Rationale for presence/absence coding

Counting exact frequencies could lead to misleading interpretations because:

1. Opportunity bias – A student may not display certain behaviours simply because they had no need (e.g. knowing all vocabulary, so no requests for meaning).

2. Error dependence – The number of corrections depends on how many errors ChatGPT produced; fewer AI errors reduce opportunities to correct.
3. Contextual variability – Some behaviours arise only under certain conversational or task conditions.

This study, therefore, records the range of behaviours demonstrated in a given session, not their raw frequency, treating the data as a record of capabilities observable in context rather than a complete inventory of a learner's abilities. This aligns with qualitative interaction analysis in CALL (Chanier & Lamy, 2017) and learner autonomy research (Benson, 2011), which emphasize the types and variety of strategies employed over their counts.

To ensure the reliability and validity of the coding process, the data were independently analysed by two researchers. Each researcher applied the predefined coding scheme to the full dataset, and their annotations were subsequently compared. Inter-rater agreement was assessed using both percent agreement and Cohen's kappa. The average percent agreement across all coded segments was 90.1%, reflecting a high level of consistency in the application of the categories. The average Cohen's kappa was 0.81, which, according to the interpretative scale proposed by Landis and Koch (1977), indicates substantial to almost perfect agreement. Discrepancies in coding were reviewed collaboratively, and consensus was reached through discussion, further enhancing the rigor and trustworthiness of the analysis. Comparative analyses were conducted within and between the two learner groups to identify key patterns, contrasts, and common strategies in mediation behaviour and language use across proficiency levels.

b Qualitative analysis of student feedback. Reflections were analysed using a hybrid approach combining thematic analysis with grounded theory techniques. Drawing on autonomy and digital literacy frameworks (Holec, 1981; Little, 1991; Luckin et al., 2016; Pangrazio & Selwyn, 2019; Zimmerman, 2002), the analysis proceeded in two stages:

Stage 1: Thematic Analysis: Five overarching themes were identified:

1. Planning and strategy selection
2. Monitoring and reflection
3. Autonomous interaction
4. Critical evaluation of AI
5. Tool comparison and digital navigation

Stage 2: Grounded Analysis: Open coding was employed to label specific learner behaviours (e.g. self-correction, prompting AI, rejecting suggestions). These were grouped into axial categories such as metacognitive engagement, digital tool use, and language control. A collaborative coding process ensured reliability, with recurring themes compared among participants to ensure consistency and thematic saturation.

Table 2. Learner autonomy coding scheme.

Code	Category	Sub-category	Definition	Example
LA-1	Planning and strategy	Explicit goal-setting	Student defines learning goals during interaction.	My goal is to practice vocabulary related to festivals.
LA-2	Monitoring and reflection	Self-correction	Student identifies errors or evaluates their progress.	Oops, I used the wrong tense here.
LA-3	Independent action	Autonomous initiation	Student initiates tasks or prompts without instruction.	Student spontaneously asks ChatGPT for examples.

Table 3. Critical engagement coding scheme.

Code	Category	Sub-category	Definition	Example
CE-1	Evaluating AI responses	Explicit scepticism	Student questions accuracy or usefulness of AI output.	Is that really correct? It doesn't match what I remember.
CE-2	Identifying errors	Error identification	Student spots factual or linguistic errors in AI responses.	I think ChatGPT got the tense wrong.
CE-3	Tool comparison	Cross-referencing sources	Student compares AI output with other tools or resources.	I checked this sentence in DeepL and got something different.

Tables 2 and 3 present the coding schemes used to analyse learner autonomy and critical engagement, respectively.

V Results

This section presents the analysis of student–AI interactions, organized by CEFR mediation categories, learner autonomy behaviours, and critical digital engagement, corresponding to research question 1, research question 2 and research question 3, respectively.

I Learner use of ChatGPT for CEFR-based mediation tasks (research question 1)

AI group

- Mediating texts: Students frequently requested summaries, translations, and clarifications of difficult words. Several (e.g. Students 1, 3, 5, 6) translated between

Spanish and Czech, demonstrating cross-linguistic mediation. Some also initiated comprehension questions to monitor understanding.

- Mediating concepts: Paraphrasing was rare, as most relied directly on ChatGPT's formulations. Nonetheless, cultural engagement was high, particularly in festival-related discussions (e.g. Students 6, 8, 10, 11).
- Facilitating communication: Clarification and correction requests were common, often framed in a conversational tone. Several used metacognitive expressions (e.g. 'correct me', 'ask me questions') to manage the dialogue.

A2 group

- Mediating texts: All students regularly asked ChatGPT to summarize or clarify vocabulary and complex phrases (e.g. 'to move around', 'nevertheless', 'congestion'), supporting meaning simplification and clarification.
- Mediating concepts: Most reflected on and interpreted key ideas, especially in debates. Some (e.g. Students 6, 8, 10) deepened arguments in response to ChatGPT's feedback, indicating reciprocal conceptual engagement.
- Facilitating communication: Clarifying questions (e.g. 'What does this word mean?') and adaptive rephrasing were common. A few engaged in discourse management, using connectors such as 'on the other hand' and 'moreover'.

Spontaneous prompts across groups. Structured prompts (Appendix A) were excluded from the quantitative analysis, as they were identical for all participants. Only four students from the A1 group and four from the A2 group adapted the model prompts, while the remaining students used them in their original form. Instead, the analysis focused on spontaneous prompts, that is, learner-generated requests and questions that emerged naturally during AI interaction. A visual representation of these proportions is shown in Figure 1.

- A1 learners: 92.6% of spontaneous prompts fell into Mediating Concepts, with only 3.7% each for Mediating Texts and Facilitating Communication.
- A2 learners: 77.8% were Mediating Concepts, 18.9% Mediating Texts, and 3.3% Facilitating Communication.

This distribution suggests both groups relied primarily on conceptual mediation beyond the given tasks. However, A2 learners more often initiated additional text-based mediation (e.g. clarifications, summaries, vocabulary checks), while A1 learners focused on idea clarification and maintaining basic understanding. In both groups, spontaneous Facilitating Communication prompts were rare, reflecting a focus on comprehension over discourse management.

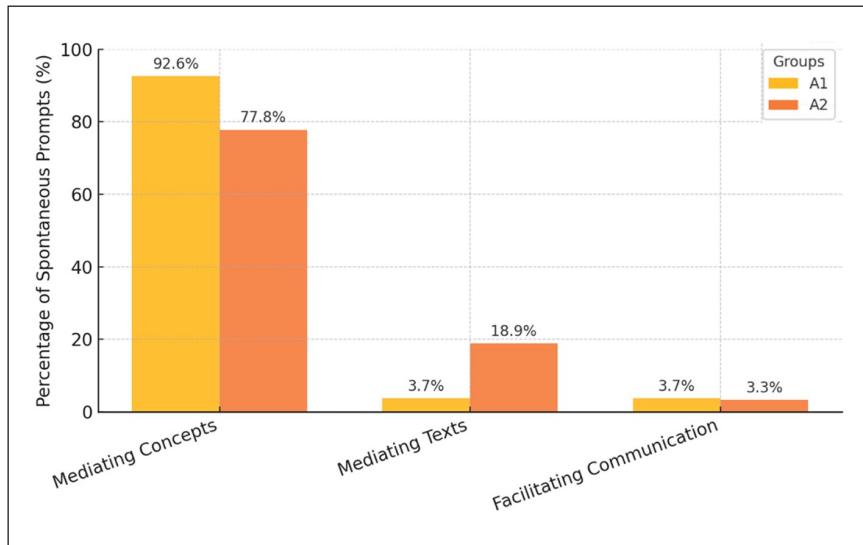


Figure 1. Distribution of spontaneous prompt types by proficiency level.

Notes. This figure shows the percentage of learner-generated prompts (excluding model prompts) classified under the CEFR categories *Mediating Concepts*, *Mediating Texts*, and *Facilitating Communication*. Both groups relied predominantly on *Mediating Concepts* (A1: 92.6%, A2: 77.8%), but A2 learners initiated more *Mediating Texts* prompts (18.9% vs. 3.7%). Spontaneous *Facilitating Communication* prompts were rare in both groups (A1: 3.7%, A2: 3.3%).

Examples of spontaneous prompts

- Mediating concepts:
 - A1: *¿Por qué esta fiesta es peligrosa?* ('Why is this festival dangerous?')
 - A1: *¿Cuál fiesta es más famosa que las otras?* ('Which festival is more famous than the others?')
 - A2: *¿Puedes explicar mejor por qué vivir en un pueblo es más seguro?* ('Can you explain better why living in a village is safer?')
- Mediating texts:
 - A1: *Tradúcame 'hermandad' al checo.* ('Translate "hermandad" into Czech.')
 - A2: *Explícame qué significa 'desplazarse' en este contexto.* ('Explain what "desplazarse" means in this context.')
- Facilitating communication:
 - A1: *Hazme otra pregunta sobre las fiestas.* ('Ask me another question about the festivals.')
 - A2: *Puedes darme un ejemplo para responder mejor.* ('Can you give me an example so I can answer better?')

2 Observable behaviours of learner autonomy in AI-mediated tasks (research question 2)

A1 group

- Overall: Autonomy ranged from moderate to moderately high. Many students set explicit goals (e.g. ‘practice superlatives’) and guided ChatGPT accordingly.
- Initiative: 83% (10/12) initiated prompts or directed the conversation.
- Self-regulation: 42% (5/12) monitored their own output or adjusted strategies mid-task.
- Strategic behaviour: 25% (3/12) used targeted approaches to elicit desired responses.

A2 group

- Overall: Autonomy was moderate. Students expressed opinions and posed grammar-related questions, but often deferred to ChatGPT for validation or correction, limiting independent problem-solving.
- Initiative: 75% (9/12) initiated or expanded discussion topics.
- Self-regulation: 17% (2/12) showed active monitoring or adjustment.
- Strategic behaviour: 17% (2/12) demonstrated targeted prompting.

Interpretive contrast. To be considered as having achieved full autonomy, students were required to demonstrate all three aspects of learner autonomy: initiative, strategic behaviour, and self-regulation. Only one student in the A1 group met this criterion. Although A1 learners had lower proficiency, they exhibited slightly higher levels of initiative and self-regulation, which may suggest a greater willingness to manage the flow of interaction. In contrast, A2 learners produced more advanced linguistic structures but relied more heavily on AI confirmation.

Figure 2 illustrates students’ autonomy behaviours across the A1 and A2 groups. Initiative was the most frequently observed behaviour in both groups (83% in A1 and 75% in A2). However, A1 students demonstrated higher levels of self-regulation (42%) and strategic behaviour (25%) than A2 students (17% each), suggesting a stronger tendency toward self-directed learning among A1 participants.

3 Critical engagement with ChatGPT (research question 3)

A1 group

- Questioning AI output: 33% (4/12), technical clarifications (e.g. rewording prompts, confirming instructions) rather than content challenges.
- Identifying errors: 17% (2/12), all technical (e.g. correcting ChatGPT’s task interpretation).
- Verification: No instances of real-time cross-checking with external tools were observed. In one case, a student noted, ‘. . . I had to use Google Translate’;

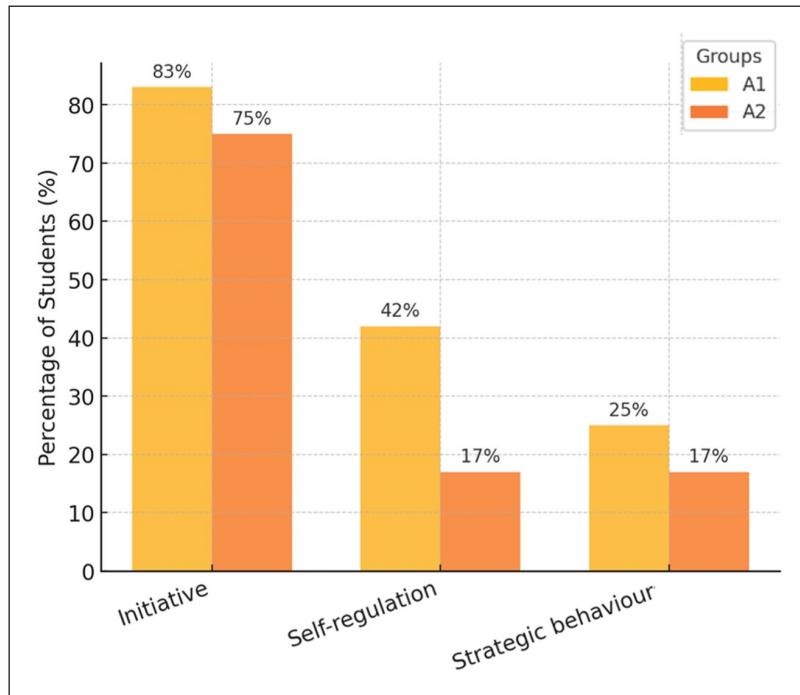


Figure 2. Autonomy behaviours by proficiency level.

Notes. The figure shows the percentage of students demonstrating autonomy behaviours: *initiative*, *self-regulation*, and *strategic behaviour*. A1 learners reported higher initiative (83% vs. 75%) and self-regulation (42% vs. 17%) than A2 learners, while strategic behaviour was low in both groups.

however, this reflected an attempt to overcome a communication breakdown with the chat rather than an effort to verify information.

A2 group

- Questioning AI output: 42% (5/12), with one case of content-level challenge; the remainder were technical clarifications.
- Identifying errors: 50% (6/12), primarily technical corrections, one case involving the accuracy of informational content.
- Verification: No examples of triangulating AI output during the task.

Interpretive contrast. Critical engagement behaviours were more frequent in A2 (especially error detection), but still heavily skewed toward technical rather than substantive challenges. Across both groups, the absence of real-time verification suggests over-reliance on ChatGPT output during interaction.

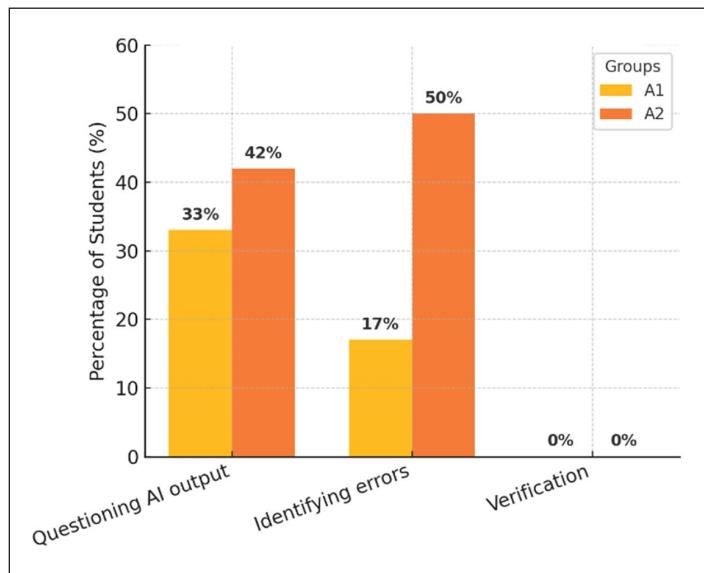


Figure 3. Critical thinking behaviours by proficiency level.

Notes. The figure shows critical thinking behaviours, with A2 learners more frequently *questioning AI output* (42% vs. 33%) and *identifying errors* (50% vs. 17%). No instances of real-time verification were observed in either group, indicating a limited tendency to cross-check ChatGPT's responses using external tools.

Figure 3 presents students' critical engagement with ChatGPT during the task. A2 students showed slightly higher levels of questioning AI output (42%) compared to A1 (33%) and identified errors more frequently (50% vs. 17%). No instances of real-time verification were observed in either group, indicating a limited tendency to cross-check ChatGPT's responses using external tools.

4 Linguistic complexity (cross-cutting observation)

This section synthesizes linguistic complexity findings across both learner groups (A1 and A2), highlighting syntactic variety, lexical richness, discourse management, and common developmental patterns.

A1 learners

Syntactic variety:

- Predominantly simple declaratives linked with basic coordination (*y, pero, porque*).
- Occasional causal subordination (*porque*), rare concessives (*aunque*).
- Reliance on memorized opinion frames (*Me parece que . . ., Creo que . . .*).
- Verb use limited to present indicative, occasional *gustar + infinitive*, and *hay*.

- Minimal embedding; one idea per sentence. *Example: Me parece que Las Fallas es la más divertida porque hay fuegos artificiales y música.*

Lexical richness:

- Thematic lexis from the fiestas españolas task (e.g. *Las Fallas, encierro, muñecos*).
- Target adjectives integrated (*divertida, peligrosa, tradicional*), occasional beyond-prompt evaluatives (*hermosas, únicas*).
- Few synonyms; high reliance on task-provided vocabulary. *Example: Es una fiesta muy original y bonita.*

Pragmatic markers and discourse management:

- Consistent opinion openers (*Para mí . . ., Me parece que . . .*).
- Frequent justification with *porque*; minimal hedging (*quizás*).
- Limited acknowledgement of opposing views. *Example: Creo que San Fermín es peligroso porque los toros corren por la calle.*

Accuracy patterns:

- Frequent gender/number mismatches (*el más peligrosa → la más peligrosa*).
- Article omission (*casi todas fiestas → casi todas las fiestas*).
- Subject–verb agreement errors (*Las Fallas es → Las Fallas son*).
- L1 transfer in adverbs/pronouns (*muy quiero → me gustaría mucho*).

In-task development:

- Gains in grammatical accuracy after feedback.
- Expansion into comparative/superlative forms post-modelling.
- Longer turns with cause–and–effect justification toward the task end. *Before: Es peligrosa. → After: Es la fiesta más peligrosa de España porque los toros pueden hacer daño.*

A2 learners

Syntactic variety:

- Frequent coordination and subordination, including concessives (*aunque*), contrastives (*sin embargo*), and relative clauses (*que tiene más transporte*).
- Multi-part arguments with introductions, justifications, and conclusions.
- Present indicative dominant, with some future/conditional for hypotheticals (*podría, sería mejor*). *Example: Aunque la ciudad tiene más oportunidades, el pueblo es más tranquilo y seguro para los niños.*

Lexical richness:

- Topic-specific lexis from the *pueblo* vs. *ciudad* task (*infraestructura, coste de vida, congestión*).
- More varied evaluatives (*ventajoso, económico*), intensifiers (*mucho más, bastante*), exemplifiers (*por ejemplo*).
- Broader synonym range than A1. Example: *Vivir en un pueblo es más económico y ofrece una mejor calidad de vida, por ejemplo, menos estrés y menos contaminación.*

Pragmatic markers and discourse management:

- Regular connectors for addition (*además*), contrast (*sin embargo*), concession (*aunque*), exemplification (*por ejemplo*).
- Balanced argumentation: acknowledgment of counterarguments before preference.
- Clear paragraph-style structuring in oral turns. Example: *Entiendo que en la ciudad hay más transporte, sin embargo, en el pueblo hay más tranquilidad y seguridad.*

Accuracy patterns:

- Occasional *ser/estar/hay* confusion, resolved after modelling.
- Sporadic article omission (less frequent than A1).
- Agreement errors mostly in early turns, corrected after feedback.

In-task development:

- Increase in syntactic complexity during exchanges (from single-clause to multi-clause).
- Adoption of new vocabulary and connectors mid-task.
- Greater control of agreement and complex connectors over time. *Before: En ciudad es mejor porque hay trabajo. → After: En la ciudad es mejor vivir porque hay más oportunidades de trabajo y estudio, y además el transporte público es más rápido.*

In summary, the A1 group produced functional, topic-relevant output but relied heavily on memorized frames, simple coordination, and high-frequency vocabulary, with persistent morphosyntactic errors. Development during the task was visible mainly in accuracy and slightly longer turns. The A2 group showed greater syntactic flexibility, richer and more precise vocabulary, and more coherent discourse organization. Development was evident in form and content, with learners integrating advanced connectors, concessives, and topic-specific vocabulary into increasingly complex, multi-clause arguments. A comparative overview is presented in Table 4.

Table 4. Comparative overview: A1 vs. A2 linguistic complexity.

Feature	A1 group (<i>festas españolas</i>)	A2 group (<i>pueblo o ciudad</i>)
Syntactic variety	Simple declaratives with basic coordination; rare concessives; reliance on fixed opinion frames.	Coordination + subordination, including concessives, contrastive markers, and relative clauses; multi-part arguments.
Verb use	Present indicative; occasional <i>gustar</i> + infinitive; <i>hay</i> . Few tense variations.	Present indicative with some future/conditional in hypotheticals; better control of <i>hay/estar/ser</i> .
Lexical range	Topic vocabulary and target adjectives; limited synonyms.	Broader topic-specific lexis; varied evaluatives, intensifiers, exemplifiers, and synonyms.
Discourse organization	Short turns, one idea per sentence; limited counterargument use.	Longer turns with structured arguments, counterarguments, and conclusions.
Pragmatic markers	Opinion openers, minimal hedging; few contrastive signals.	Variety of markers for addition, contrast, concession, exemplification.
Accuracy patterns	Frequent agreement and article errors; L1 transfer in adverbs/pronouns.	Fewer agreement errors; occasional <i>ser/estar/hay</i> confusion; less frequent article errors.
In-task development	Gains in accuracy and turn length after feedback; minor structural expansion.	Gains in both form and discourse structure; adoption of advanced connectors and topic-specific vocabulary.

VI Analysis

Across all three research questions, clear patterns emerged in learners' engagement with ChatGPT. Both proficiency groups relied heavily on conceptual mediation, though A2 learners diversified more into text-based mediation and produced linguistically richer output. A1 learners showed unexpectedly high initiative and self-regulation, while A2 learners identified more errors but tended to seek AI confirmation rather than manage challenges independently. Critical engagement remained limited in both groups, with most questioning focused on technical rather than substantive aspects of AI output. Linguistic complexity findings reinforced these contrasts, revealing broader syntactic variety, lexical range, and discourse organization in A2, alongside more frequent mor-phosyntactic errors and simpler structures in A1. Together, these patterns provide a comprehensive learner–AI interaction profile to inform the thematic analysis presented in the following section.

I Qualitative analysis of student feedback

Building on the observable interaction patterns reported in Section V.1, this section examines learners' post-task written reflections to explore how they perceived and managed their learning processes. These qualitative data reveal dimensions of strategic thinking, affective response, and critical engagement with AI that were not always visible in real-time transcripts. The thematic and grounded analyses link directly to the three research questions:

- Research question 1 through evidence of language mediation strategies;
- Research question 2 through accounts of learner autonomy and self-regulation; and
- Research question 3 through critical evaluation of ChatGPT, comparison with other tools, and awareness of its limitations.

a Thematic analysis. To provide an overview of recurring patterns, Table 5 summarizes five themes emerging from student reflections, illustrating how each theme connects to the relevant research question and theoretical constructs. As shown in Table 5, learners demonstrated a combination of planning, monitoring, and evaluative behaviours that reflect increasing levels of autonomy and critical engagement.

b Grounded analysis: Open and axial coding. The grounded analysis complemented the thematic coding by identifying specific learner behaviours and grouping them into broader conceptual categories. These categories (Table 6) reflected dimensions of autonomy, critical thinking, and digital literacy, grounded in the theoretical framework. This analysis highlights how learners combined reflective thinking, technical skills, and strategic behaviours to enhance their interaction with ChatGPT. Although not always visible in real-time transcripts, these behaviours indicate a developing autonomy and critical

Table 5. Thematic analysis results.

Theme	Research question link	Representative evidence (student quotes)	Interpretation / conceptual link
1. Planning and strategy selection	Research question 2: Learner autonomy	<p>'I had as much time as I needed . . . I could go through the text in detail.'</p> <p>'I like making written notes . . . I filled about two-thirds of an A4 with phrases and vocabulary.'</p> <p>'I use DeepL . . . switch the languages to get the correct form.'</p> <p>'I asked it why it used that verb form, and it replied it was a mistake.'</p> <p>'Sometimes I wasn't sure if ChatGPT's corrections were actually accurate.'</p> <p>'I think I benefited from trying to form sentences in a more complex way and checking words I didn't know.'</p> <p>'It was useful that I could ask for a translation directly in the chat.'</p> <p>'It forced me to think in Spanish again instead of defaulting to English.'</p> <p>'It helped me create sentences, which I think is the most important part of language.'</p> <p>'Sometimes it just repeated arguments from the original text instead of coming up with new ones.'</p> <p>'I had to instruct it how to argue, which was annoying.'</p> <p>'It corrected words that were actually fine, just not copied from the original text.'</p>	Strategic planning and personalized resource selection; aligns with Holc's learner autonomy and Zimmerman's planning phase of self-regulated learning.
2. Monitoring, reflection, and evaluation	Research question 2: Learner autonomy	<p>'I asked it why it used that verb form, and it replied it was a mistake.'</p> <p>'Sometimes I wasn't sure if ChatGPT's corrections were actually accurate.'</p> <p>'I think I benefited from trying to form sentences in a more complex way and checking words I didn't know.'</p> <p>'It was useful that I could ask for a translation directly in the chat.'</p> <p>'It forced me to think in Spanish again instead of defaulting to English.'</p> <p>'It helped me create sentences, which I think is the most important part of language.'</p> <p>'Sometimes it just repeated arguments from the original text instead of coming up with new ones.'</p> <p>'I had to instruct it how to argue, which was annoying.'</p> <p>'It corrected words that were actually fine, just not copied from the original text.'</p>	Reflective monitoring and evaluation of learning processes; evidence of metacognitive autonomy.
3. Tool use and autonomous interaction	Research question 1, research question 2	<p>'Phind . . . lets you set a profile, use search mode, and shows where it found information.'</p> <p>'I use DeepL for rewriting sentences, more of a habit, but it works.'</p>	Independent and intentional use of ChatGPT for learning management; digital learner autonomy.
4. Critical evaluation of AI	Research question 3: Critical engagement	<p>'It's not bad, but in school, face-to-face is better.'</p>	Awareness of AI limitations; aligns with critical digital literacy frameworks.
5. Technical navigation and tool comparison	Research question 3: Critical engagement	<p>'Phind . . . lets you set a profile, use search mode, and shows where it found information.'</p> <p>'I use DeepL for rewriting sentences, more of a habit, but it works.'</p>	Evidence of AI literacy and multimodal tool integration; critical evaluation of technological affordances.

Table 6. Grounded analysis results.

Open code	Axial category	Theoretical connection
Asked ChatGPT to explain grammar choices	Metacognitive monitoring	Self-regulated learning (Zimmerman); reflection on input
Took notes during ChatGPT interaction	Strategy use and Personalization	Learner autonomy (Holec); strategic note-taking as planning
Attempted more complex sentence structures	Productive risk-taking	Self-regulated strategy; active cognitive engagement
Critiqued ChatGPT's repetition	Evaluation of AI output	Critical digital literacy (Panagrazio & Selwyn); AI limitations awareness
Compared ChatGPT to other tools	Tool selection and triangulation	AI literacy (Luckin et al.); multimodal critical use
Controlled interaction pace	Temporal autonomy	Learner regulation of task engagement (Little; Zimmerman)
Questioned ChatGPT corrections	Critical stance	Metacognitive reflection and digital agency
Combined tools strategically	Multimodal integration	Technical fluency and strategic learner autonomy
Used ChatGPT for contextual translation	Functional mediation strategy	Language problem-solving and initiative
Expressed positive emotion about task	Affective engagement	Motivation and sustained autonomy (Little)

engagement with AI that is consistent with the CEFR and digital competence frameworks.

2 Divergence between observed behaviour and self-perceived autonomy

The transcript analysis captures what students did in the moment, showing autonomy and critical thinking as enacted behaviours. The feedback analysis reveals what students thought, planned, and felt, including strategies and evaluations that were invisible in real-time interactions. Combining the two provides a fuller picture. A comparison of observed and reported behaviours is presented in Table 7.

- Some autonomy and critical engagement were observable during interaction.
- Other dimensions (planning, tool comparison, metacognitive reflection, affective factors) emerged only in post-task self-reports.
- Many aspects of critical thinking or digital competence may not be evident in the conversation transcript. For example, when ChatGPT made an error, some students simply redirected it back to the task but subsequently ceased to trust it as a reliable source of information. As a result, they did not verify further information with ChatGPT during the task; instead, they reported in their feedback that they had checked the accuracy of ChatGPT's responses afterwards using sources they considered more trustworthy.
- Another behavioural pattern observed was that students sometimes ignored ChatGPT's mistakes during the interaction, which, in the transcript, appeared as an absence of critical thinking at that moment. However, their feedback clarified that they had, in fact, noticed these errors and reflected on them after the task. Similarly, if students verified certain information through other sources, they typically did not mention this in the ChatGPT conversation but reported it in their feedback.
- Feedback also revealed students' reasoning for modifying certain prompts in real time in response to dissatisfaction with ChatGPT's output. For instance, one student found ChatGPT's responses excessively long and noted in their feedback that, in future tasks, they would include a formula within the prompt to limit the length of the AI's replies.

A central finding of this study is the gap between students' observed behaviour during ChatGPT interactions and their self-reported perceptions of autonomy and critical engagement. While transcript data revealed moderate autonomy, with student-initiated interactions and requests for clarification, students also relied heavily on ChatGPT for linguistic decisions. Few questioned or challenged AI outputs.

By contrast, student reflections conveyed more advanced strategic thinking. Learners described goal-setting, tool comparison, metacognitive insight, and awareness of AI limitations. These reflective accounts suggest developing digital agency and self-regulated learning beyond what is evident in real-time behaviour.

This discrepancy suggests that performance-based autonomy (visible in task execution) may differ from metacognitive autonomy (revealed in reflection). ChatGPT's

Table 7. Comparison of observed vs. reported behaviours.

Dimension	Observed in conversations	Reported in feedback
Initiative	Frequent in both groups, seen in prompt adaptation and topic extension.	Often linked to purposeful use of AI for targeted learning aims.
Self-regulation	More common in AI; visible in vocabulary requests, clarification-seeking.	Expanded to include strategic planning, note-taking, productive risk-taking, and pacing control.
Strategic behaviour	Relatively infrequent in transcripts, mostly in AI.	Stronger presence in reflections, with detailed strategies and tool integration.
Questioning AI output	Mostly focused on language form; limited factual checking.	Included awareness of AI's argumentative limits, unnecessary corrections, and repetition.
Identifying errors	More common in A2; mostly technical corrections, rare content checks.	Broader critical stance including evaluation of logic and alternative tool use.
Digital literacy	Seen only indirectly in transcripts (e.g. directing AI format/style).	Strongly present: students explicitly compared tools and integrated multiple resources.
Affective engagement	Not observable. Although certain student responses exhibit indications of frustration or irony. For instance, following an unsatisfactory reply from ChatGPT to a prompt, one student commented: 'Well, you really did a great job there.'	Frequently reported: motivation, satisfaction, and increased willingness to engage in Spanish.

dialogic and semi-agentive interaction may scaffold learners so effectively that outward signs of autonomy are masked, even as cognitive engagement occurs internally. Moreover, task design itself, particularly when scaffolded, can shape how autonomy is expressed externally.

These findings suggest that autonomy is not always directly observable during interactions, especially in AI-mediated settings where the tool's responsiveness can obscure learners' internal strategies. Post-task reflections are essential for capturing the full scope of learner development.

VII Limitations

This study has several limitations. First, the CEFR mediation descriptors were designed for human-to-human interaction and may not fully capture the unique dynamics of human–AI communication. Although the descriptors offered a helpful analytical framework, their application in AI-mediated contexts requires cautious interpretation and further theoretical development.

Second, participants were purposefully selected for their high level of digital literacy. While this ensured a focus on language learning rather than technical navigation, it limits the generalizability of the findings to learners with less experience using AI tools. The sample was also institutionally homogeneous, consisting entirely of students from a single technical university, which may not reflect broader educational contexts. However, this homogeneity helped enhance the internal validity of the study by reducing variability in participants' technical competence. For the purposes of our research, it was important that all participants had a comparable level of technical education, minimizing factors that could affect the outcomes. Nevertheless, students did not encounter any technically demanding aspects during the process, suggesting that the activity may also be suitable for learners with technical knowledge at a typical secondary-school level. This hypothesis could be further examined in future research.

Third, the tasks were guided by structured prompts and designed scenarios. While necessary for consistency, this structure may have constrained learners' spontaneous language use and autonomy. Consequently, observable behaviours might not fully reflect learners' capacity for independent mediation in more open-ended tasks.

Fourth, focusing exclusively on lower CEFR proficiency levels (A1–A2) was intentional, as mediation activities and AI interactions are typically the most challenging for these learners. Nevertheless, observations revealed that some A2 participants were able to summarize, explain, and reformulate their ideas when ChatGPT did not respond adequately, behaviours commonly associated with B1-level mediation competences. Although these attempts were generally comprehensible, they contained numerous grammatical inaccuracies, suggesting that students were reaching the upper limits of their current linguistic control. The successful execution of such tasks at basic proficiency levels supports the hypothesis that learners with greater linguistic resources could engage more effectively in similar or more complex mediation scenarios. This indicates that mediation skills may begin to emerge earlier than expected, even when linguistic accuracy remains limited. Future research could therefore test this empirically

by including B1–B2 participants in tasks designed to foster more advanced forms of mediation, which would help to trace the developmental continuum of these abilities and refine pedagogical strategies for different proficiency levels.

Fifth, the reliance on self-reported reflections introduces the possibility of social desirability bias. While reflections in a native language likely increased honesty and depth, some students may have overstated or understated their strategic thinking.

Finally, the short duration of the study limits the ability to conduct a longitudinal analysis of learner development. Long-term tracking would provide stronger evidence of how mediation and autonomy evolve over time and across various AI-supported tasks.

1 Discussion

This study investigated how beginner-level multilingual learners engaged with ChatGPT to perform CEFR-based mediation tasks and develop learner autonomy. Regarding research question 1, the findings indicate that learners at both A1 and A2 proficiency levels participated in meaningful communicative activities, including summarizing texts, clarifying vocabulary, and negotiating meaning. These behaviours align with the CEFR Companion Volume, particularly within the categories of Mediating Texts and Facilitating Communication, and suggest that mediation skills can be developed even at lower proficiency levels when learners receive appropriate digital and pedagogical support.

More complex strategies associated with Mediating Concepts, such as paraphrasing and synthesizing information, were less frequently observed. Students predominantly used prompts to clarify or compare. This tendency may be attributed both to linguistic limitations and to the nature of ChatGPT's output. Because the tool produces fluent and grammatically accurate responses, learners may be less inclined to reformulate ideas in their own words. This finding is particularly relevant to research question 1, as it underscores the importance of designing tasks that explicitly require learners to adapt or explain content rather than passively accepting AI-generated formulations.

With respect to research question 2, a central finding is the contrast between learner behaviours during interactions and the strategic thinking revealed in post-task reflections. While students initiated prompts and responded to feedback, they often relied on ChatGPT to validate their choices or resolve uncertainties. However, their written reflections revealed deeper engagement, including goal-setting, tool comparison, and critical evaluation of AI-generated responses. This supports the conclusion that meta-cognitive autonomy is expressed more clearly in reflective writing than during real-time interaction.

These observations imply that outward signs of autonomy may be less visible in AI-mediated settings, particularly when the tool responds efficiently to learner input. ChatGPT's supportive output can obscure the cognitive effort involved in decision-making, making post-task reflection essential for capturing the full range of learner strategies and mental processes.

Another notable finding is that some A1 learners exhibited greater interactional autonomy than their A2 counterparts. This observation is grounded in the higher

percentage of A1 participants demonstrating initiative (83% vs. 75%) and self-regulation (42% vs. 17%) during the tasks, rather than a subjective judgment of overall quality. While higher proficiency generally provides greater linguistic flexibility, these results suggest that autonomy may be influenced more by learner disposition than by language level alone. Factors such as motivation, self-confidence, and willingness to experiment appear to play a significant role. Some lower-proficiency learners took more linguistic risks and persisted in exploring meaning, whereas some more advanced learners occasionally simplified their output to avoid errors.

In relation to research question 3, the study identified early signs of critical digital and AI literacy. Although learners rarely challenged ChatGPT during the tasks, many demonstrated reflective thinking in their feedback. Several participants noted inaccuracies or repetitive reasoning in the AI's responses, while others compared ChatGPT's output with that of alternative tools. These behaviours suggest that while critical engagement may be minimal in real time, it emerges more clearly in post-task reflection, supporting an expanded definition of critical digital literacy. Furthermore, when students recognized factual errors in ChatGPT's output, they often lost trust in the tool as a reliable source of information and saw little value in pursuing further clarification during the task.

The data also indirectly address research question 3 by raising questions about the emotional and psychological dimensions of AI-mediated learning. For some learners, interacting with a non-human partner reduced performance anxiety and encouraged experimentation. For others, however, the absence of human feedback appeared to limit motivation. Further research is needed to investigate how affective factors influence learner behaviour in AI-supported contexts.

Finally, although the CEFR mediation descriptors provided a useful analytical framework, they were not designed with AI interaction in mind. This has implications for both research question 1 and research question 3. Their application to AI-mediated tasks requires adaptation to account for emerging dynamics such as dialogic co-construction, intrapersonal meaning-making, and digital agency. These issues are examined in greater detail in the following section.

2 Rethinking CEFR mediation and plurilingualism in AI-enhanced contexts

While the CEFR offers a solid foundation for describing mediation competence, its descriptors were developed for interpersonal communication and do not fully account for the dynamics of learner–AI interaction. In this study, learners used ChatGPT to develop mediation strategies aligned with CEFR categories, including Mediating Texts, Mediating Concepts, and Facilitating Communication. However, AI tools such as ChatGPT pose theoretical and pedagogical challenges that warrant closer examination.

a Plurilingual and pluricultural competence: Receptive vs. experiential knowledge. Plurilingual and pluricultural competence in the CEFR emphasizes the dynamic, context-sensitive use of multiple languages and cultural knowledge in real-world social interaction. Human mediators draw on lived experience, emotional awareness, and social positioning

to navigate communication. ChatGPT, by contrast, has no experiential or affective grounding. Although it can access and generate content across many languages, it lacks receptiveness to cultural diversity and cannot authentically engage in the negotiation of cultural meaning.

Learners interacting with AI thus encounter a tension: they benefit from the tool's expansive linguistic knowledge but must take full responsibility for interpreting and applying cultural information. This suggests that AI-mediated mediation should be seen as an opportunity to foster learner-led meaning-making, provided it is supported with tasks that promote reflection and cultural awareness.

b Limitations in paralinguistics and situated communication. Another limitation of ChatGPT is its inability to process or respond to paralinguistic features such as facial expression, gesture, prosody, or shared physical context. Human mediation often relies on these cues to interpret affect, gauge understanding, and adjust communicative strategies. AI-mediated communication, in contrast, is limited to written or spoken language. While this constraint may promote clearer linguistic expression, it removes a core dimension of natural communication.

At the same time, this limitation can be pedagogically useful. Because learners must compensate for the absence of contextual cues, they are encouraged to be more explicit, structured, and reflective in their language use skills that align with CEFR goals for organized and audience-aware mediation.

c Linear dialogue and artificial conversational dynamics. AI conversations are linear and explicit, lacking the inferential flow of human dialogue. To co-construct meaning, human speakers draw on shared knowledge, incomplete utterances, and emotional states. ChatGPT, in contrast, requires full verbalization of assumptions, repetition of context, and high degrees of precision to maintain coherence.

This changes the communicative conditions under which mediation occurs. Learners must take increased responsibility for structuring discourse and managing turn-taking. These demands may develop helpful metacognitive strategies, but they also highlight the artificiality of interaction. Unlike a human interlocutor, ChatGPT does not evolve during the conversation or incorporate affective context. To account for these differences, CEFR descriptors should be contextualized and expanded for AI-mediated settings. Key areas for adaptation are presented in Table 8.

In addition, CEFR-based mediation could recognize AI use as a transitional space between reception and production. When learners ask ChatGPT to clarify unknown words or rephrase texts, they engage in intrapersonal mediation – constructing understanding with digital support. These processes can prepare learners for interpersonal mediation, where they apply that understanding in social or academic contexts.

In summary, CEFR remains highly relevant in AI-supported learning, but must be rethought in terms of the nonlinear, text-dependent, and cognitively distinct features of AI conversation. Recognizing the difference between intrapersonal and interpersonal mediation, and designing tasks that foster both, will be essential to future pedagogy.

Table 8. Adapting CEFR mediation descriptors for AI contexts.

Standard CEFR element	Proposed AI contextualization
Mediating a text	Retain with scaffolding; AI supports summarizing, rephrasing, and simplifying with learner prompts.
Mediating concepts	Include intrapersonal meaning construction via AI before interpersonal use.
Facilitating communication in plurilingual settings	Differentiate between simulated cultural explanation with AI and experiential negotiation with people.
Managing interaction	Emphasize explicit turn-taking and clarification strategies specific to AI interaction.
Pluricultural awareness	Encourage learners to critically reflect on AI's cultural neutrality, gaps, or inaccuracies.

VIII Conclusions and pedagogical recommendations

This study shows that beginner-level multilingual learners can engage in CEFR-based mediation tasks using ChatGPT as a supportive tool. Learners at A1 and A2 levels successfully performed activities such as summarizing information, clarifying language, and negotiating meaning. These findings indicate that mediation can be effectively developed even at the early stages of language learning when learners are supported by structured tasks and responsive digital tools.

Learners' written reflections revealed deeper levels of strategic thinking than were visible during real-time interactions. While students often relied on ChatGPT throughout the tasks, their reflections demonstrated goal-setting, evaluation of feedback, and emerging critical awareness of the tool's limitations. These metacognitive behaviours suggest that reflection plays a vital role in making learner autonomy more visible and in fostering critical engagement with AI.

The CEFR framework provided a useful lens for analysing mediation in this context, even though it was originally developed for human-to-human communication. As AI becomes more integrated into language learning, there is a growing need to adapt CEFR descriptors to better capture the dynamics of human–AI interaction, including dialogic patterns and digital agency.

Based on the findings of this study, the following recommendations are proposed to support both mediation competence and learner autonomy in AI-assisted language learning environments:

1. Acknowledge AI as a unique, nonhuman partner in language learning. Tasks should frame AI not as a replacement for human interaction, but as a tool that supports learner-driven meaning-making, planning, and reflection.
2. Encourage linguistic clarity and structured output by guiding learners to formulate effective prompts, ask for clarification, and rephrase content. These skills align with CEFR mediation descriptors and promote intentional language use.
3. Integrate critical digital literacy by prompting learners to reflect on the reliability, neutrality, and limitations of AI-generated responses. Learners should be encouraged to compare outputs from multiple digital tools and question inconsistencies.

4. Incorporate structured reflection tasks that help learners set goals, monitor their progress, and evaluate the quality of AI feedback. These reflections make cognitive and metacognitive processes more visible and support the development of learner autonomy.
5. Design open-ended activities that require learners to adapt language for different audiences, purposes, or levels of formality. This supports the development of adaptive mediation strategies and audience awareness.
6. Promote the use of learners' full linguistic repertoires to support cross-linguistic mediation. Learners should be encouraged to draw on their plurilingual competencies when clarifying meaning or rephrasing content.
7. Guide learners in identifying and reflecting on errors or limitations in AI-generated language. This includes noticing repetitive phrasing, unnatural formulations, or culturally inappropriate references.
8. Include follow-up interpersonal tasks where learners apply knowledge or language structures developed through AI interaction in human communication, such as peer discussions, role-plays, or presentations.
9. Create opportunities for independent task completion, allowing learners to take responsibility for managing their own learning process, making decisions, and setting their own pace.
10. For A1 especially, it might be helpful to explicitly instruct ChatGPT to limit summaries and explanations to X words or use only sentences of ≤ 10 words to ensure that cognitive load stays low.

Future research should explore how mediation, learner autonomy, and AI literacy develop over time and across different educational contexts. Longitudinal studies could offer further insight into the evolution of these competencies and how the CEFR can be adapted to foster pedagogical innovation in AI-mediated environments.

Acknowledgement

not applicable.

Credit author statement

Alice Lukešová: conceptualization, methodology, formal analysis, investigation, resources, writing – original draft; Petra Juna Jennings: conceptualization, validation, investigation, resources, data curation, writing – review and editing.

Data availability statement

The datasets generated and analysed during the current study, including the raw data supporting the findings, are available from the corresponding authors upon reasonable request and within the article; however, some data cannot be provided to protect the students' anonymity.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Ethical approval

The research was conducted, and the data were handled in accordance with the Ethical Code of the Czech Technical University in Prague, approved by the Ministry of Education, Youth, and Sports under § 36, paragraph 2 of Act No. 111/1998 Coll., on Higher Education Institutions and on the Amendment and Supplementation of Other Acts (the Higher Education Act), on April 22, 2022, under file no. MSMT-9911/2022-2. The research also followed the ethical guidelines outlined in the Declaration of Helsinki.

Informed consent

All participants were thoroughly informed about the research and gave consent to the use of the data collected.

GenAI use disclosure statement

During the preparation of this work, the authors used ChatGPT for language editing and formatting references. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

ORCID iDs

Alice Lukešová  <https://orcid.org/0009-0009-5470-6969>

Petra Juna Jennings  <https://orcid.org/0009-0002-6852-9325>

References

Akhter, S., Ahmed, K., Tambi, F., et al. (2025). Enhancing EFL learners' communication skills: A CEFR-inspired approach for high school students. *Frontiers in Language Studies*, 7, Article 9182.

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

Brown, T.B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *ArXiv preprint, 2020*, Article 14165. Available at: <https://arxiv.org/pdf/2005.14165.pdf> (accessed November 2025).

Buckingham, D. (2010). Defining digital literacy. In B. Bachmair (Ed.), *Medienbildung in neuen Kulturräumen* (pp. 59–71). VS Verlag für Sozialwissenschaften.

Chanier, T., & Lamy, M.-N. (2017). Researching technology-mediated multimodal interaction. In C.A. Chapelle & S. Sauro (Eds.), *The Handbook of technology and second language teaching and learning* (pp. 428–443). Wiley-Blackwell.

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – companion volume*, council of Europe publishing, Strasbourg. Available at: www.coe.int/lang-cefr

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, Article 22.

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning and Technology*, 27, 6–27.

Gu, W. (2025). A meta-analysis of the efficacy of conversational AI use in EFL learning contexts. In G. Lampropoulos & S. Papadakis (Eds.), *Social robots in education* (Studies in Computational Intelligence, Vol. 1194) (pp. 335–374). Springer.

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

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15, Article ep464.

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28, 15873–15892.

Jones, R.H., & Hafner, C.A. (2012). *Understanding digital literacies: A practical introduction*. Routledge.

Kittredge, A.K., Hopman, E.W.M., Reuveni, B., et al. (2025). Mobile language app learners' self-efficacy increases after using generative AI. *Frontiers in Education*, 10, Article 1499497.

Kruk, M., & Kałużna, A. (2024). Investigating the role of AI tools in enhancing translation skills, emotional experiences, and motivation in L2 learning. *European Journal of Education*, 60, Article e12859.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.

Lantolf, J.P. (2000). Second language learning as a mediated process. *Language Teaching*, 33, 79–96.

Lin, C.-H., Zhou, K., Li, L., & Sun, L. (2025). Integrating generative AI into digital multimodal composition: A study of multicultural second-language classrooms. *Computers and Composition*, 75, Article 102895.

Little, D. (1991). *Learner autonomy: Definitions, issues and problems*. Authentik Language Learning Resources.

Lucchi, N. (2023). ChatGPT: A case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation*. Advance online publication. <https://doi.org/10.1017/err.2023.59>

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L.B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

Nwanakwaugwu, A.C., Obi-Akwarri, O.E., Ojaide, E., et al. (2025). AI-driven personalization in English as a foreign language (EFL). In Y.B. Khouya & A.I. Alaoui (Eds.), *Application of AI in the teaching and learning of English as a foreign language (EFL)* (pp. 73–106). IGI Global.

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media and Society*, 21, 419–437.

Reinders, H., & Benson, P. (2017). Research agenda: Language learning beyond the classroom. *Language Teaching*, 50, 561–578.

Seghier, M.L. (2023). ChatGPT: Not all languages are equal. *Nature*, 615, 216.

Vygotsky, L.S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Wang, S., Wang, F., Zhu, Z., et al. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252, Article 124167.

Wollny, S., Schneider, J., Di Mitri, D., et al. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, Article 654924.

Zimmerman, B.J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41, 64–70.

Appendix A

Activity outlines for A1 and A2 groups with model prompts.

Part 1: Comprehension stage

Step 1: Prompt to ChatGPT

Spanish: *Hazme un resumen del texto.*

English: *Make me a summary of the text.*

- Type: Mediating Texts
- ChatGPT does: Produces a short, simple summary of the source text.
- Student does: Reads the summary, compares it to the original, and checks whether it matches their understanding.
- CEFR A1 (learning outcome): After using the summary, the student can recognize the gist and 1–2 key ideas of a short, simple text and confirm whether these ideas are present.
- CEFR A2 (learning outcome): After using the summary, the student can verify completeness of main points, notice missing details, and adjust their understanding accordingly.

Step 2: Prompt to ChatGPT

Spanish: *Tradúcelme al . . . / explícame lo significado de las palabras que son más difíciles que el nivel de español A1/A2 CEFR.*

English: *Translate into . . . / explain to me the meaning of the words that are more difficult than the Spanish A1/A2 CEFR level.*

- Type: Mediating Texts
- ChatGPT does: Provides L1 translations and/or simple L2 explanations of difficult words from the text.
- Student does: Re-reads the text with the support list and checks they now understand those items; optionally uses the words in very simple examples.
- CEFR A1 (learning outcome): After the explanations, the student can identify unknown words and grasp their basic meaning in context.
- CEFR A2 (learning outcome): After the explanations, the student can integrate clarified vocabulary to make sense of short passages and paraphrase them simply if needed.

Step 3: Prompt to ChatGPT

Spanish: *Escríbeme cinco preguntas para que pueda comprobar que he entendido bien el texto.*

English: *Write me five questions so that I can check if I have understood the text well.*

- Type: Mediating Concepts
- ChatGPT does: Generates five comprehension questions (predictable, text-based).
- Student does: Answers the questions based on the text.
- CEFR A1 (learning outcome): After answering, the student can respond to simple, concrete questions about main facts from a familiar text.
- CEFR A2 (learning outcome): After answering, the student can handle a small range of factual and very simple inference questions and keep answers clear and relevant to the text.

Step 4: Prompt to ChatGPT

Spanish: *Corrígeme las respuestas.*

English: *Correct my answers.*

- Type: Mediating Concepts
- ChatGPT does: Corrects the student's answers and aligns them with the source text.
- Student does: Compares corrections with their own answers and makes changes.
- CEFR A1 (learning outcome): After feedback, the student can spot where their answer diverged from the text and adjust it to match basic facts.
- CEFR A2 (learning outcome): After feedback, the student can revise answers for accuracy and completeness, incorporating corrected information from the text.

Step 5: Prompt to ChatGPT

Spanish: *Explícame por qué mi respuesta es incorrecta.*

English: *Explain to me why my answer is incorrect.*

- Type: Mediating Concepts
- ChatGPT does: Gives a simple explanation with text evidence (quotes/line references) and, where relevant, highlights target grammar (e.g. superlatives in the A1 'fiestas' task).
- Student does: Locates the evidence in the text and adjusts their understanding/ answer.
- CEFR A1 (learning outcome): After the explanation, the student can point to the text's relevant part supporting the correct answer.
- CEFR A2 (learning outcome): After the explanation, the student can justify why an answer is correct/incorrect by referring to the text in simple terms.

Part 2: Discussion stage

Step 1: Prompt to ChatGPT (expressions list)

A1 Spanish: *Escríbeme una lista corta (5–6) de expresiones muy simples que puedo usar para dar mi opinión, estar de acuerdo o no, y pedir más información. Usa solo el nivel de español A1.*

A1 English: *Write me a short list (5–6) of very simple expressions I can use to give my opinion, agree or disagree, and ask for more information. Use only Spanish at the A1 level.*

A2 Spanish: *Escríbeme una lista de 8–10 expresiones que puedo usar para dar mi opinión, comparar ideas, expresar acuerdo o desacuerdo, y pedir aclaraciones, usando solo el nivel de español A2.*

A2 English: *Write me a list of 8–10 expressions I can use to give my opinion, compare ideas, express agreement or disagreement, and ask for clarification, using only Spanish at the A2 level.*

- Type: Facilitating Communication
- ChatGPT does: Provides an expression bank appropriate to the level; for A1 ‘fiestas’, includes superlative frames like *La más divertida es . . . / La menos peligrosa es . . .*
- Student does: Uses selected expressions during the discussion.
- CEFR A1 (learning outcome): After using the list, the student can produce memorized expressions to give a simple opinion, agree/disagree, and ask for more information.
- CEFR A2 (learning outcome): After using the list, the student can vary expressions to compare ideas, qualify agreement/disagreement, and ask for clarification.

Step 2: Prompt to ChatGPT (discussion setup)

A1 Spanish (fiestas & superlatives): *Soy estudiante de español y necesito practicar la mediación y el superlativo relativo de los adjetivos. Vamos a hablar sobre las fiestas españolas del texto. Yo voy a decir cuál creo que es la más divertida, la más aburrida, la más peligrosa, la más interesante, la más original, la más rara, la más religiosa, la más tradicional, la más alegre, la más conocida, la más increíble, la más famosa y la más tonta. Tú vas a decir una fiesta diferente y explicar por qué. Después vamos a comparar nuestras opiniones y decidir juntos cuál es la mejor, la peor, la más peligrosa, etc. Usa solo el nivel de español A1 según el CEFR. Si cometí algún error gramatical, primero corrige mi respuesta y después continúa la discusión.*

A1 English: *I am a Spanish student, and I need to practice mediation and the relative superlative of adjectives. We are going to talk about the Spanish festivals from the text. I am going to say which one I think is the most fun, the most boring, the most dangerous, the most interesting, the most original, the strangest, the most religious, the most traditional, the happiest, the best known, the most incredible, the most*

famous, and the silliest. You are going to say a different festival and explain why (to talk about a different festival . . .). Afterwards, we are going to compare our opinions and decide together which one is the best, the worst, the most dangerous, etc. Use only the A1 level of Spanish according to the CEFR. If I make any grammatical mistake, first correct my answer and then continue the discussion.

A2 Spanish (city vs. village): *Soy estudiante de español y necesito practicar la mediación. Voy a defender la opinión de que vivir en una ciudad es mejor que en un pueblo. Tú, en cambio, vas a decir que es mejor vivir en un pueblo. Puedes usar solo el nivel de español A2. Después de mi respuesta, corrige todos los errores gramaticales que cometa y luego respóndeme.*

A2 English: *(I am a Spanish student, and I need to practice mediation. I am going to defend the opinion that living in a city is better than living in a village. You, on the other hand, are going to say that it is better to live in a village. You can only use A2-level Spanish. After my answer, correct all the grammatical mistakes I make and then reply to me.)*

- Type: Mediating Concepts
- ChatGPT does: Takes the opposing position, keeps the discussion going, and corrects the student's language before continuing.
- Student does: Defends a position with simple reasons from the text; negotiates a shared conclusion.
- CEFR A1 (learning outcome): After interaction, the student can state a basic opinion, use superlatives in set frames (fiestas task), respond to a different view, and co-decide using very simple language.
- CEFR A2 (learning outcome): After interaction, the student can sustain a short, structured discussion, respond to counter-arguments, and incorporate corrections into subsequent turns while keeping to familiar language.