Impact of AI Integration in Bilingual Learning Environments at Private Institutions in Colombia: Implications for Cultural and Cognitive Dimensions

Laura Carolina Bernal Flórez, Luis Fernando López Metaute and José David Azuero

Facultad de Educación, Universidad Santo Tomás

Maestría en Ambientes Bilingües de aprendizaje

Director: Carlos David Díaz Díaz

Author Note

José David Azuero Sánchez

https://orcid.org/0009-0005-8415-1600

jose.sanchezazu@usantotomas.edu.co

https://scienti.minciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0002373247

Luis Fernando López Metaute

https://orcid.org/0009-0007-9198-9588

luis.metautelop@usantotomas.edu.co

https://scienti.minciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0002418615

Laura Carolina Bernal Flórez

https://orcid.org/0009-0004-4631-1637

laura.florezber@usantotomas.edu.co

https://scienti.minciencias.gov.co/cvlac/visualizador/generarCurriculoCv.do?cod_rh=0002373285

Table of Contents

Abstract

This study examines the impact of Artificial Intelligence (AI) integration on the cognitive and cultural dimensions of students in bilingual learning environments at private institutions in Colombia. As AI tools such as ChatGPT and Google Gemini become increasingly prevalent in classrooms, concerns arise regarding their superficial use, potential erosion of critical thinking, and cultural homogenization in language learning. Grounded in a qualitative comparative case study design, data were collected through semi-structured interviews, non-participant classroom observations, and think-aloud protocols. Findings reveal that AI can support language acquisition, personalization, and intercultural awareness when mediated by skilled educators; however, uncritical reliance on these tools risks undermining cognitive engagement and cultural identity. The study concludes that effective bilingual education requires purposeful AI integration alongside teacher expertise and intercultural task design, positioning AI as a pedagogical complement rather than a replacement for meaningful human instruction.

Este estudio analiza el impacto de la integración de la inteligencia artificial (IA) en las dimensiones cognitivas y culturales de los estudiantes en entornos de aprendizaje bilingües de instituciones privadas en Colombia. A medida que herramientas de IA como ChatGPT y Google Gemini se vuelven cada vez más comunes en las aulas, surgen preocupaciones respecto a su uso superficial, la posible erosión del pensamiento crítico y la homogeneización cultural en el aprendizaje de idiomas. Basándose en un diseño de estudio de caso comparativo cualitativo, los datos se recopilaron a través de entrevistas semiestructuradas, observaciones

no participantes en el aula y protocolos de pensamiento en voz alta. Los resultados revelan que la IA puede apoyar la adquisición del idioma, la personalización y la conciencia intercultural cuando está mediada por educadores capacitados; sin embargo, la dependencia acrítica de estas herramientas corre el riesgo de socavar el compromiso cognitivo y la identidad cultural. El estudio concluye que una educación bilingüe eficaz requiere una integración intencionada de la IA junto con la experiencia de los docentes y el diseño de tareas interculturales, posicionando a la IA como un complemento pedagógico en lugar de un sustituto de la instrucción humana significativa.

Key words: Artificial Intelligence, bilingual education, cognitive development, cultural identity, language acquisition, intercultural competence.

INTRODUCTION

The educational system has undergone drastic transformations due to the rapid evolution of technology. Among these innovations, Artificial Intelligence (AI) stands out as a breakthrough in language teaching by enabling personalization, automating processes, and offering dynamic learning resources. Its growing presence in bilingual classrooms makes it essential to examine not only its potential benefits but also its broader cultural and cognitive implications.

Bilingual education goes beyond mastering two languages; it fosters communication, cognition, and intercultural understanding. Within this process, language is not just a tool for expression but also a way to engage with culture and identity. AI tools, when thoughtfully implemented, can support these goals by offering adaptive platforms, real-time feedback, and resources that accommodate diverse learner needs.

Despite these advantages, AI in education also presents challenges. Many tools prioritize efficiency and standardization over cultural depth, often overlooking the sociolinguistic richness of language. AI-generated English, for example, may be grammatically correct but culturally sterile, limiting opportunities for intercultural awareness and authentic communication. These risks are amplified by the global imbalance in digital content, as less than 2% of online resources exist in African languages despite the continent's linguistic diversity (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2023). Such limitations risk reducing language learning to a mechanical exercise rather than a culturally meaningful experience (García, Johnson, & Seltzer, 2021; Hovy & Spruit, 2016).

In the Colombian context, particularly in private institutions, recent practices reveal a growing dependence on AI platforms like ChatGPT and Google Gemini. While these tools are widely used to complete reading, writing, and translation tasks, students often engage with them superficially, without critically verifying accuracy or reflecting on meaning. This reliance undermines problem-solving, independent thinking, and deeper cognitive engagement with the target language (Crompton, Edmett, Ichaporia, & Burke, 2024).

These issues highlight the urgent need for a balanced approach. AI should not replace traditional teaching but rather complement it in ways that encourage inclusion, cultural awareness, and meaningful communication. Educators play a key role in guiding students toward critical use of these tools, ensuring that technology enhances rather than weakens bilingual learning (Cummins, 2021).

This study therefore seeks to explore how AI integration in bilingual learning environments shapes both cognitive development and cultural identity. By focusing on the

lived experiences of students in private Colombian classrooms, the project aims to reveal not only the opportunities and risks of AI but also its potential to strengthen bilingual education. Ultimately, it seeks to answer how AI can reinforce bilingual environments from a cultural and cognitive perspective.

Research Question

How does the integration of AI in bilingual learning environments within private Colombian institutions impact students' cultural and cognitive development?

General Objective

To explore the impact of AI integration on the cognitive and cultural dimensions of students in bilingual learning environments at private institutions in Colombia.

Specific Objectives

1. To examine the ways in which AI tools can be used by students in bilingual classrooms.
2. To examine the influence of AI integration on students' cognitive development, particularly in relation to language acquisition and critical thinking.
3. To compare the impact of students exposed to AI tools versus those in traditional bilingual environments.

THEORETICAL FRAMEWORK

Despite the rapid adoption of AI in education, its impact on bilingual learning environments remains underexplored, leaving important gaps in understanding how AI connects with cognitive and cultural dimensions of individuals in bilingual settings.

To delve into the use of the AI in bilingual education, it is necessary to consider previous studies that underline its use and some of the challenges that this technology poses. For instance, Edinoh Kingsley (2025) in his study “Exploring the Challenges of Artificial Intelligence in Basic Education in Nigeria” examines the overall impact of AI in education; however, it lacks specific analysis of bilingual education, including the potential challenges and opportunities presented by AI-powered teacher's aides in multilingual contexts. It is very interesting to see how AI has become one of the main focuses for different education scenarios, due to its employability and versatility. New studies emerge and new insight is given by scholars and the education community. It connects with the view of AI integration that this study intends to carry out and paves the way for the future of education and the different variables from its use.

Another study relevant to this research is titled “Incorporating Artificial Intelligence and Gamification in EFL Education: Strategies for Enhancing Teaching and Learning Outcomes of EFL Teachers.” from the authors Claro Narváez, M. F., Díaz Díaz, C. D., & Román Pérez, K. J. (2024) While it is important to discuss how AI can improve a learning

environment, it is also important to mention how it can be of great help in the acquisition of a new language and in bilingual education. Likewise, it is important to understand the role of the student when they encounter new technologies. However, the role of the teacher cannot be overshadowed in this new change, as it is their responsibility to facilitate the acquisition of a second language through new technologies. Therefore, it is pertinent to take the teachers' role into account. Consequently, teacher training and professional development are important when using AI, leading to acknowledging the importance of considering the ethical issues that bilingual education faces, since the constant use of AI may harm human development not only in social but in cultural fields as well.

While AI's potential is being explored for academic use, its consequences for bilingual learners' cognition and cultural identity remain poorly examined, therefore, It is important to take into consideration studies such as "Understanding Bilingualism, Bilinguality, and Bilingual Education in an Era of Globalization" from Wang, A.-L. (2023). where Dr Wang provides the cultural and cognitive foundation that AI-enhanced bilingual education must be built upon. Her research offers a theoretical compass pointing toward: culturally responsive AI design ethically, aware integration of technology cognitive alignment in bilingual contexts even if she doesn't focus directly on AI, her ideas are crucial for guiding how AI should be used in bilingual education making her work highly relevant to this project.

With the purpose of understanding more the concepts that surround this study, it is critical to investigate the pivoting points of the research to give a more comprehensive conceptualization of AI integration, bilingual learning environments, and cultural and cognitive dimensions in language learning.

Therefore, this study explores how the integration of AI in bilingual learning environments influences students' cultural identity and cognitive development. By grounding the analysis in theoretical perspectives on language, culture, and cognition, it acknowledges both the opportunities and risks that AI poses in educational settings. These foundations justify the use of a qualitative comparative case study, as this methodological approach allows for an in-depth examination of students lived experiences and the nuanced ways AI shapes their learning processes.

CONCEPTUALIZATION

1. Artificial Intelligence

AI refers to the development of systems capable of performing tasks that typically require human-like cognitive functions such as learning, problem-solving, and decision-making. Barhoi, Gupta, Kumar, and Kaushik (2025) define AI as algorithms and models that interpret data, adapt to new information, and improve performance over time. Similarly, Russell and Norvig (2020) describe AI as the study of agents that perceive their environment and act to achieve goals. Taken together, these perspectives highlight AI as both a scientific discipline and a set of technological applications that increasingly influence education, language learning, and society at large.

Given its rapid evolution, AI is no longer limited to basic automation but now encompasses complex tools that can personalize learning, analyze student progress, and facilitate

intercultural communication. This makes it essential to evaluate its implications for bilingual environments, particularly regarding cognitive development and cultural identity.

1.1 Artificial Intelligence in Language Learning

Yina Xia, Seong-Yoon Shin, and Jong-Chan Kim (2024) show in their study that "… Cross-Cultural Intelligent Language Learning System (CILS), are a novel approach to integrating Artificial Intelligence (AI) into language education to enhance cross-cultural communication." Specifically, their CILS system applies artificial intelligence to provide adaptive, real-time feedback and personalized learning pathways based on learners' linguistic backgrounds and progress. This technological innovation helps optimize bilingual instruction on a practical level. While the authors provide a technologically robust and culturally aware system, demonstrating how AI can operationalize aspects of personalized bilingual education, they don't deeply address cognitive or ethical implications.

In this vein, Xia, Shin, and Kim offer a highly pragmatic and technological contribution. For instance, their Cross-Cultural Intelligent Language Learning System (CILS) demonstrates how AI can be effectively leveraged to deliver personalized, culturally sensitive bilingual instruction. Additionally, their work shows that when thoughtfully designed, AI tools can adapt to learners' cultural contexts and linguistic needs, making education more accessible and tailored. Despite these strengths, their system excels in operationalizing key aspects of AI-enhanced bilingual learning but leaves open the need for a richer, more humanistic philosophical framework to guide its application.

1.1.2 AI Integration in Bilingual Learning Environments

Ma Baoyi (2024) explores the impact of AI on the education of bilingual students with a focus on academic language learning experience, cultural sensitivity, and inclusiveness. As AI technologies like ChatGPT and Grammarly become more prevalent in education, they can be used to enrich the learning experience of bilingual students. By focusing on this goal, this research provides insight into how AI can be harnessed to enrich bilingual skills, as its main purpose is to find out whether the use of this technology can impact bilingual education. At the same time, it shows an example of challenges that both educators and students face when this technology takes part in the learning process.

Some key aspects that make this study like ours include the authors' demonstration of how AI can take part in the cognitive development of those who use it. Thereby, one of the main constructs of this study "cognitive dimension," and like what the researchers aimed for in this study, we want to investigate the extent to which AI can help and impact the cognitive process of learners. Furthermore, the study illustrates how this technology is being integrated into U.S. education, suggesting its potential applicability within a Colombian context and prompting further exploration of its various implications and effects. Taking the before mentioned, in this technological age, technology plays a crucial role in education, necessitating an examination of its transformative potential. To address this, the study explores

AI's capacity as a powerful educational tool and its potential to enhance language acquisition. Finally, it investigates the role of culture in this process, acknowledging the inherent relationship between language and culture and the importance of integrating both in effective language teaching and learning.

2. Bilingual learning environments

In the present study, bilingual learning environments are defined as any educational setting where two languages are used for teaching and learning. In these environments, students are exposed to and engage with content in both languages, often with the goal of promoting proficiency in each language. (Garcia 2009) In that order, it can be inferred that the true strength of bilingual learning environments lies in their ability to integrate languages organically, without one prevailing over the other. Students should feel equally comfortable using both languages, not only as communication tools, but also as fundamental elements for their academic and personal development.

Conversely, Araque (2021), in her publication "Understanding Learning Environments," expands the concept beyond a simplistic view of physical space, defining it as a natural environment fostering relationships that provide meaning to human existence. This broader definition facilitates the integration of bilingualism in diverse settings.

3. Bilingual education

Likewise, bilingual education is any system of schooling in which instruction takes place through two languages, and in which the use of those languages is sustained over time as part of the students' learning process. This approach not only involves teaching two languages, but also using both as vehicles of instruction, with goals that may include the maintenance or development of the mother tongue, the acquisition of a second language, the cognitive and academic development in both languages, and the appreciation of cultural and linguistic diversity (Baker 2017) In that sense, we can state that when we understand bilingual education as a dynamic, flexible and student- centered process, we can see that this could be a paradigm shift in the way we conceive language and learning. Since, it is not about teaching both languages separately, but to recognize that every bilingual student comes to the classroom with their own linguistic repertory, which is full of experiences, identities, meanings and cultural background.

Building on this understanding, it is evident that Johannessen's perspective (2019) goes hand in hand with the definition of bilingual education proposed by Garcia (2014), since both coincide in a critical and complex vision about bilingualism. Far from understanding bilingual education as a simple instruction in a second language, mostly known as a second foreign language, these authors underline the importance of conceiving it as a dynamic and inclusive process that recognizes, values and promotes the speaker's mother tongue.

In this sense, García (2014) highlights that bilingual education must start from a translanguaging pedagogy, that is, one that allows not only students but also teachers, to mobilize all their entire linguistic repertoire without fragmenting it or hierarchizing it. Johannessen (2019), for his part, reinforces the idea by warning about the risks of a bilingual education that, by exclusively prioritizing the acquisition of a dominant language, this could

lead not only to the displacement of the native of minority language but also to a form of cultural invisibility.

Therefore, understanding bilingual education from these perspectives implies recognizing that the mother tongue is not only a means of communication, but also an essential component of both individual and collective identity. The abandonment or devaluation of the mother tongue can lead not only to the loss of fundamental cultural and symbolic references, but also to a depletion of the sociolinguistic diversity that enriches communities. Consequently, a genuine bilingual education cannot be built on the exclusion of the mother tongue but must foster a space in which the multiple languages and cultures coexist and mutually strengthen each other.

4. Cultural dimension

On the other hand the concept of Cultural dimension can be seen as a sensory, practical and bodily experience that is lived in everyday life and goes beyond language.(Pink 2015) In that way, one can say that, culture is understanding the ways of life, of expression, of walking, of feeling and behaving, of intoning a word when speaking and of physical expressions or gestures that go far beyond language and that can only be experienced through human contact. It is important to keep this in mind, since one of the intentions of the research project is to examine how AI may or may not interfere with the linguistic-cultural processes in the acquisition of a second language through artificial intelligence.

This study, however, finds Pink's (2015) perspective to be somewhat subjective. While acknowledging the sensory, practical, and physical aspects of culture, this research refutes the assertion that language connects to culture merely from a physical standpoint. This study posits language not only as a component of culture, but, as Kim (2003) suggests, as one of its most essential and structuring elements, given that knowledge, values, beliefs, and social norms are transmitted through language. To suggest that culture transcends language minimizes its crucial role as the primary vehicle of cultural expression. Furthermore, language is key to understanding how individuals construct their reality, communicate their experiences, and negotiate their identity within a cultural group. Therefore, cultural experience and language are deeply intertwined and cannot be easily separated.

5. Cognitive dimension

According to Ai-Ling Wang (2023) "cognitive dimensions in bilingual education can be defined as the mental processes activated and developed through culturally situated language learning, particularly those that involve metalinguistic awareness, cognitive flexibility, and intercultural reasoning" by framing bilingual education as a process that develops higher-order thinking skills such as metalinguistic awareness, cognitive flexibility, and reflective learning. Her emphasis on the interplay between language and thought supports the idea that AI tools should not only deliver content but also foster learners' ability to think critically across languages and cultures. In the Colombian context, her insights encourage the design of AI-integrated environments that actively promote cognitive development through

culturally relevant and intellectually engaging bilingual tasks, ensuring that technology enhances (not replaces) the deep mental processes essential to meaningful language learning.

Learning a second language is not just about acquiring vocabulary or grammar—it's about engaging the mind to compare, contrast, and navigate between linguistic and cultural systems. This process strengthens the learner's ability to reflect on language use, adapt to diverse contexts, and think critically about meaning, identity, and communication. Thus, cognitive dimensions in Wang's framework are deeply intertwined with culture, identity, and emotional awareness, forming part of a holistic, humanistic educational process rather than isolated intellectual skills.

In contrast Astghik Chubaryan and Mariam Vardanyan (2024) pose an interesting perspective on language acquisition as their study focuses on cognitive linguistics which offer a comprehensive model that considers various theories of how languages are learnt and store in the brain within a holistic framework, allowing researchers and scholars to move from traditional linguistics that underscore syntactical patterns, word structures, grammar rules, phonology, semantics, and lexical meanings, to a broader, more complex and complete view of language itself as a mental phenomenon and its use as merely being a cognitive activity. However, it overlooks the significant role culture plays in language acquisition. It's crucial to view culture as an integral part of language learning, as it shapes individual understanding and fosters a more profound comprehension.

This study suggests that understanding language acquisition requires investigating the intricate cognitive processes involved, moving beyond traditional modular views that emphasize grammar rules and coherence, and toward a perspective that integrates language

with broader cognitive abilities and the learner's individual characteristics. Therefore, it is critically important for this study to explore these ideas, as they encompass aspects that contribute to a holistic understanding of language in a modern era where the rapid growth of AI presents challenges for researchers and scholars in integrating this technology while considering the broader context of language learning and acquisition.

METHDOLOGICAL FRAMEWORK

Building on these theoretical foundations, the study adopts a qualitative comparative case study design to examine how AI integration shapes cultural and cognitive dimensions in bilingual classrooms. This approach is particularly suited to exploring students lived experiences, as it allows for a rich, contextualized understanding of how AI tools are used in practice and how they influence learning processes. As Yin (2018) emphasizes, case studies provide an in-depth investigation of contemporary phenomena within real-world contexts, making this method ideal for analyzing classroom dynamics where traditional and AI-supported practices coexist.

In order for a qualitative-comparative research to be successful it has to follow an order that we considered encompasses the objectives of our study, which are: to account for the ways AI tools can be used in the classroom, to look into the influence of these tools in relation to language acquisition and development of critical thinking and finally create a point of comparison between those students exposed to AI versus the students in a traditional bilingual classroom. Therefore, it was critical for the study to follow certain stages in the methodology implemented; these stages are:

1. Develop or use a detailed theory of change
2. Identify cases of interest
3. Develop a set of factors
4. Score the factors
5. Analyze the dataset
6. Interpret the finding and revise the Theory of Change

This structured methodology allows for an in-depth examination of complex causal relationships within a set of cases. It systematically compares different combinations of conditions to identify patterns leading to specific results. The method's focuses on identifying necessary and sufficient conditions that provide valuable insights. Therefore, utilizing this framework is the most effective way to examine the possible outcomes. Taking into consideration the before mentioned, it was imperative to ground these steps and adjust them to our implementation.

In the first step (develop or use of a theory of change), we reflected on the way that AI as a disruptive technology in the classroom could generate changes in the students' cultural and cognitive development in our study, therefore, connecting with Yin's (2018) idea of the first stage which states that "The theory of change should be design to identify two things: the change the qualitative comparative analysis (QCA) is interested in and the factors that (in theory) help bring about those changes.". Accordingly, the second step (identify case of interest), refers to the fashion in which the cases were selected to evaluate how the "outcome" happened and those cases in which it did not, that in concordance with our study, the results were accounted for in both scenarios. The third step (develop a set of factors) refers to how the results depend on certain key factors, which in our case is the presence or absence of AI tools

in two scenarios within the same context. The fourth step (score the factors) helped us create an intial qualitative coding for the variables that were developed (e.g., “frustration with AI,” “using translation for speed” “cultural misunderstanding”) which permitted to go through the data line by line and attach short descriptive labels. Finaly, in the two last steps (analyze the dataset and interpret the findings), the information was collected, analyzed and the theory of change of our study was revised to be able to conclude whether there were any results regarding the use of artificial intelligence tools for cultural and cognitive development.

By defining two scenarios (an AI integrated classroom and a traditional one) we can directly address our third specific objective while exploring the nuances of the other two. Therefore, it is highly important to delimit the population and sampling that were chosen to carry out this study. Consequently, two bilingual classrooms were involved in the process. Case **A** which is the AI- integrated one in which 5 fifth grade students participate by using tools like chatGPT, google Gemini etc. and case **B** which represented the comparable other 5 fifth-grade students that does not formally integrate these AI tools. The population selected for this study is made up by two bilingual classrooms within the same private school located on the outskirts of a town called Facatativá in Colombia. Both groups have students ranging between 10-11 years old with an English proficiency from A1 to B1 who at the same time have access to an “English lab” where different activities take place using pedagogical and e-learning materials. These students were selected thanks to an assent form signed by their parents and the institution. In this school, English is taught every day with a 10-hour intensity per week.

Fifth grade students were chosen as the focus of this study because they are at a crucial stage of both cognitive and linguistic development. At ages 10 to 11, learners are transitioning

from concrete operational to more abstract thinking, which allows them to begin engaging critically with language tasks and reflecting on meaning. In terms of second language acquisition, they are typically moving from basic communicative competence (A1–A2) toward greater fluency and accuracy (B1), making them an ideal group to observe how AI tools influence the development of higher-order skills such as problem-solving, critical thinking, and intercultural awareness. Studying this age group provides valuable insight into how AI integration can support or hinder students as they consolidate foundational skills while preparing for more advanced bilingual learning.

To understand how the information was collected, it was necessary to examine different methods for data collection that were aligned to this qualitative comparative study. Such instruments were useful to create an environment in which both scenarios were thoroughly addressed to pinpoint the outcomes as accurately as possible.

Data Collection Instruments

Within the framework of a **qualitative comparative case study**, the selection of data collection instruments is essential to ensure the richness, validity, and reliability of the information obtained. As Yin (2018) and Stake (1995) highlight, case studies require the use of multiple sources of evidence to understand phenomena in their complexity and from different perspectives.

Based on these methodological considerations, this research will employ three complementary data collection instruments: semi-structured interviews, non-participant classroom observations, and think-aloud protocols. The combination of these methods will allow for exploration of the impact of artificial intelligence on our specific participant group (bilingual 10–11-year-olds with A1-B1 English proficiency) addressing their perceptions, observed practices, and the cognitive processes that emerge in real time within their unique learning context.

Instrument 1: Semi-Structured Interview

A **semi-structured interview** is a qualitative data collection method that uses a set of predetermined questions while allowing flexibility for the interviewer to explore emerging topics in more depth. As Kallio et al. (2016) point out, this method strikes a balance between structure and openness, making it particularly suitable for exploring participants lived experiences and perceptions in a systematic yet flexible way.

Semi-structured interviews will be employed as a data collection tool, since they represent one of the most effective methods in qualitative case study research, allowing for an in-depth exploration of participants' perceptions and experiences. This type of interview combines a pre-designed guide of questions with enough flexibility for the interviewer to formulate additional questions depending on the responses and the context of the conversation (Wilson, 2014).

According to Fontana and Frey (1994), semi-structured interviews are situated in an intermediate position between the rigidity of structured questionnaires and the openness of unstructured interviews, offering a balance between data comparability and the richness of

individual narratives. In the same vein, Kvale (1996) emphasizes that this format facilitates understanding the meanings that interviewees attribute to their experiences, which is essential in research that aims to interpret educational phenomena from the perspective of the actors involved.

For the design of the interview guide, the steps proposed by Naz, Gulab, and Aslam (2022) will be followed. These include: (1) identifying the objectives of the interview, (2) formulating the initial questions, (3) validating through a pilot test, and (4) elaborating the final version. This process will ensure the coherence of the instrument with the research objectives and the relevance of the questions.

Within the framework of the **Comparative Case Study** proposed by Stake (1995) and Yin (2018), semi-structured interviews make it possible to obtain in-depth and comparable information across different cases, facilitating both the identification of common patterns and the exploration of contextual particularities. Likewise, as highlighted by Ragin (2014) in his proposal of qualitative comparative analysis, this type of instrument is essential to capture causal configurations and nuances that are difficult to identify through exclusively quantitative techniques.

For this study, semi-structured interviews will be conducted with all ten fifth grade students (5 from Case A and 5 from Case B). The interview guide will be tailored to explore the distinct experiences of each group. For Case A (AI-integrated), questions will focus on their interactions with tools like ChatGPT and Google Gemini, exploring perceived benefits and challenges for their bilingual learning. For Case B (traditional), parallel questions will explore their experiences with more conventional pedagogical and e-learning materials used in

their “English lab”. This parallel design is essential for achieving our comparative objective, allowing us to contrast perceptions across the two cases and identify patterns unique to AI integration (Ragin, 2014). Each interview is expected to last approximately 15-20 minutes and will be conducted in a comfortable setting to encourage open responses from the children.

Instrument 2: Classroom Observations (Non-Participant)

Non-participant classroom observation will serve as a second data collection instrument. This technique enables the researcher to document classroom practices and interactions without direct involvement in the teaching-learning process, thereby reducing bias and minimizing disruption (Cohen, Manion, & Morrison, 2018).

According to Angrosino (2007), classroom observations provide valuable insights into the natural context of learning, allowing researchers to capture behaviors, routines, and interactions that may not surface during interviews. Non-participant observation is particularly useful for analyzing patterns of teacher–student interaction, the use of technological tools, and the dynamics of bilingual learning environments.

As Stake (1995) suggests in the context of case study research, observation enriches the data by adding a direct dimension of evidence that complements participants’ self-reports. Field notes and observation protocols will be systematically recorded to ensure consistency and rigor in data collection (Creswell & Poth, 2018).

Observations will be conducted in both classrooms to provide a direct comparison of teaching and learning dynamics. In Case A, the observation protocol will focus specifically on documenting:

· How students' access and use AI tools (e.g., ChatGPT, Google Gemini) during activities.

· The nature of student-student and student-teacher interactions around these tools.

· The types of tasks where AI is deployed and the pedagogical purpose it serves.

In the same way, observations in Case B will document:

· How students utilize traditional e-learning materials and resources in the "English lab".

· The interaction patterns and task structures in the absence of generative AI. This direct comparative observation is fundamental for addressing our objective of exploring the nuances of classroom practice, providing tangible evidence of differences and similarities between the two scenarios (Stake, 1995).

Instrument 3: Think-Aloud Protocols

The third instrument will be the use of **Think-Aloud Protocols**, a method in which participants verbalize their thoughts in real-time while engaging in a specific task. As Ericsson and Simon (1993) explain, think-aloud methods are designed to access participants' cognitive processes by externalizing what would otherwise remain internal reasoning. This technique has been widely used in educational and language learning research to investigate problem-solving strategies, reading comprehension, and metacognitive awareness (Bowles, 2010).

According to Charters (2003), think-aloud protocols are particularly valuable in understanding not just what learners do, but how they process information while performing a

task. By verbalizing their thought processes, learners provide insights into decision-making, comprehension difficulties, and the strategies they use to complete tasks.

Each instrument in this study is designed to address different aspects of the research question. The survey measures students' self-perceptions of their use of AI tools, focusing on how often they rely on them, their confidence in completing tasks with or without AI, and their attitudes toward language learning and cultural identity. This provides quantitative indicators of the role AI plays in their learning habits and beliefs.

The classroom observation seeks to document how students interact with AI tools during language activities, as well as how teachers guide or regulate their use. Special attention is given to signs of cultural engagement (e.g., references to real-world contexts, cultural nuance in language use) and cognitive processes (e.g., problem-solving strategies, independent versus AI-assisted work). This allows the study to capture authentic practices that may not surface in self-reports.

The semi-structured interview provides deeper qualitative insights into students' experiences. It explores how they perceive AI's influence on their ability to think critically, understand cultural elements of language, and develop bilingual skills. It also uncovers feelings of trust, discomfort, or reliance on these tools, helping to explain the gap between observed classroom behavior and students' personal reflections.

Together, these instruments triangulate data by combining self-reported perceptions, observed practices, and personal narratives. This design ensures a comprehensive answer to

the research question, shedding light on how AI affects both the cultural and cognitive dimensions of bilingual learning.

The application of this protocol will be carefully designed to be comparable yet context appropriate. Students from both Case A and Case B will be given a similar language learning task (e.g., writing a short paragraph, translating a sentence, researching a topic). The key difference will be the tools they are instructed to use:

· Case A students will perform the task using their available AI tools (e.g., ChatGPT) and will verbalize their thought process as they formulate prompts, interpret results, and integrate them into their work.

· Case B students will perform the parallel task using their customary traditional resources (e.g., dictionaries, known e-learning platforms) and will similarly verbalize their strategies. This method will provide unparalleled, real-time data on the cognitive and metacognitive strategies employed by each group, directly illuminating the impact of AI on the thinking processes of these young bilingual learners.

Research Project Timeline: Qualitative Comparative Study on AI in Bilingual Classrooms

Phase	Task	Description	Instruments / Participants	Timeline
Phase 1: Planning & Setup	1.1 Finalize Design & Secure Consent	Refine the Theory of Change (QCA Stage 1), finalize all data collection instruments, and collect signed assent forms (QCA Stage 2).	Research Team, School Admin, Parents	Week 1
	1.2 Pilot Test Instruments	Conduct a quick pilot test of the interview and	1-2 volunteer students	End of Week 1

		think-aloud protocols to ensure clarity and make final adjustments.		
Phase 2: Data Collection	2.1 Conduct Classroom Observations	Perform non-participant observations in both classrooms to document the learning environment and tool usage.	Case A & B classrooms; Observation Protocol	Weeks 2-4
	2.2 Conduct Semi-Structured Interviews	Schedule and conduct individual interviews with all 10 students concurrently with the observation period.	10 students (Case A & B); Interview Guide	Weeks 2-3
	2.3 Conduct Think-Aloud Protocols	Schedule and complete the think-aloud sessions with all 10 students.	10 students (Case A & B); Think-Aloud Tasks	Week 4
Phase 3: Analysis & Interpretation	3.1 Transcribe & Prepare Data	Transcribe all audio recordings and organize field notes immediately as data collection concludes.	Interview & Think-Aloud recordings; Field Notes	Week 5
	3.2 Qualitative Coding & Analysis (QCA Stages 4 & 5)	Systematically code all data to identify key factors. Begin cross-case analysis to compare patterns between Case A and Case B.	All collected data	Week 6
	3.3 Interpret Findings & Revise Theory (QCA Stage 6)	Synthesize the results from the analysis, draw conclusions, and revise the initial Theory of Change based on the evidence.	Analysis results	Week 7
Phase 4: Reporting	4.1 Draft & Finalize Report	Write, review, and complete the full research report, including methodology, findings, and conclusion.	N/A	Week 8

This project is expected to generate meaningful outcomes at different levels of the educational community. For students, it aims to promote improved learning by fostering critical thinking, intercultural awareness, and more effective engagement with AI tools. For teachers, it provides new pedagogical strategies and resources that integrate technology while

maintaining a human-centered approach to language teaching. Finally, the study contributes to strengthening its bilingualism policy by aligning technological innovation with cultural and cognitive development, ensuring that AI integration supports long-term educational goals.

PRELIMINARY ANALYSIS

This section aims to briefly analyze and show the intial perceptions from the findings as well as the methods utilized during the implementation phase, which were compiled in relation to exploring the impact of Artificial Intelligence (AI) on the cognitive and cultural dimensions of students in bilingual learning environments.

Based on the research questions posed, the findings obtained through classroom observations, semi-structured interviews, and think-aloud protocols will be examined in order to identify significant relationships in the studied context, such as patterns and trends; describe students' perceptions; examine cognitive effects (such as vocabulary acquisition and metalinguistic awareness); and investigate the cultural and linguistic identity impacts on fifth-grade elementary school students. Nevertheless, the information previously gathered during the implementation phase reveals some initial nuances that will be presented in this preliminary analysis.

For data processing, an analytical approach based on a qualitative-comparative methodological design was chosen. The data will be systematically organized and compared between case A (group with AI integration) and case B (group without AI), using methodological triangulation as the validation mechanism.

This process allows for the categorization of the information collected through the aforementioned data collection methods, thus aligning it with the cognitive, affective, and cultural dimensions that were established in the theoretical framework.

The observations from both groups reveal contrasting learning dynamics shaped by the presence or absence of AI tools. In the AI-integrated group (5A), students used Gemini and Copilot as sources of linguistic and cultural input, experiencing initial difficulties with prompt formulation and comprehension but gradually developing autonomy, fluency, and clarification strategies. Their engagement increased when interacting with visual or culturally rich content, although the AI's speed, mixed-language output, and occasional inaccuracies required continuous teacher mediation. Meanwhile, the non-AI group (5B) relied on traditional resources and the teacher as the primary linguistic model, resulting in structured participation, strong peer collaboration, and accurate language production, though motivation decreased during extended oral activities.

Cognitively, the AI group displayed more spontaneous metalinguistic awareness as they negotiated meaning, reformulated prompts, and managed confusion arising from linguistic ambiguities or cultural content. Their learning aligned with research on AI-supported scaffolding and personalized cognitive support. In contrast, the non-AI group engaged in deliberate, teacher-guided metalinguistic reflection focused on vocabulary, grammar, and pronunciation, benefiting from predictable routines and social scaffolding. Both groups meaningfully compared cultural practices from South Korea and Colombia, though the AI group simulated cultural encounters through conversational prompts, while the non-AI group consumed culture through stories, videos, and teacher explanations.

Affectively, students in the AI group showed high levels of curiosity, excitement, and occasional anxiety, with the AI functioning as a motivating yet unpredictable partner that lowered the affective filter for some learners. The non-AI group displayed more stable emotional responses, enjoying hands-on and collaborative tasks but losing interest during longer analytical discussions. Overall, AI-enabled interactions promoted spontaneous speaking and risk-taking but offered inconsistent cultural depth and limited corrective feedback; traditional approaches fostered accuracy, collaboration, and clear modeling, though they lacked novelty and sustained engagement for prolonged oral production.

Some preliminary findings regarding the cultural impact when integrating artificial intelligence into 5th-grade English classrooms in Colombia presented some challenges, requiring a careful balance between technological advantages and significant ethical considerations, such as the potential for cultural erosion, privacy concerns, and exacerbating social disparities. While these tools offer the promise of personalized learning experiences, they frequently embody a "Global North" perspective, which could inadvertently marginalize the unique local identity of Colombian students. Furthermore, there are considerable concerns regarding the protection of minors' data under Colombia’s Law 1581. Beyond the technological aspects, the digital divide poses a risk of AI becoming an exclusive resource, thereby widening the gap between privileged and rural educational institutions. This could also lead to the replacement of the invaluable emotional support provided by educators with an impersonal algorithm, potentially fostering cognitive dependence. To ensure an ethical evolution, it is crucial that AI functions as a "co-pilot," empowering students' authentic voices rather than contributing to linguistic homogenization.

Taken together, the analysis suggests that AI enhances engagement, autonomy, and intercultural exposure but requires strong teacher mediation due to cognitive overload, technical inconsistencies, and superficial cultural delivery. Traditional resources provide structure, clarity, and reliable linguistic modeling but may struggle to maintain student motivation during extended oral tasks. Both approaches exhibit complementary strengths: AI supports conversational spontaneity and bilingual identity validation, while traditional instruction ensures accuracy and sustained collaboration. A balanced integration of AI tools, teacher guidance, and interactive pedagogical tasks appears to offer the most effective learning environment for A2-level learners.

DATA TRIANGULATION

Triangulation is understood here in accordance with Flick (2018) and Saldaña (2021) as the systematic convergence of data from multiple qualitative instruments to produce robust, multi layered interpretations of a shared phenomenon.

AI integration in bilingual classrooms is not merely a technological shift but a pedagogical and sociolinguistic event with consequences for students' cognitive processes, cultural identities, and language development. As García et al. (2021) documented, AI-generated language may be grammatically accurate yet culturally narrow, potentially limiting authentic intercultural communication. At the same time, properly scaffolded AI use has been shown to support personalized learning and lower affective barriers for second language learners (Xia et al., 2024; Ma, 2024).

The triangulation presented here draws on three instruments applied across two fifth-grade groups at Colegio Fundación Fernando Borrero Caicedo (Facatativá, Colombia) during October and November 2025: (1) semi-structured student interviews codified through a 35 code matrix (Codification Matrix); (2) non-participant classroom observations conducted across six sessions coded along four analytical dimensions (Classroom Observation Coding); and (3) think-aloud protocols (TAP) conducted with six students from the AI-integrated group across six analytical domains (TAP Coding). Together, these instruments address the study's general objective which is to explore the impact of AI integration on the cognitive and cultural dimensions of bilingual learners, and its three specific objectives (SO1: AI tool usage patterns; SO2: cognitive and language development; SO3: comparative impact of AI versus traditional environments).

Theoretically, the analysis is anchored in three interrelated frameworks. First, it draws on Vygotsky's sociocultural theory and the concept of the Zone of Proximal Development (ZPD), as interpreted for bilingual contexts by Wang (2023) and Cummins (2021), to examine how AI and human interlocutors function as mediating agents of cognitive development. Second, it engages with translanguaging theory as articulated by García (2009) and García and Li Wei (2014), who conceptualize the bilingual learner's full linguistic repertoire as a dynamic, integrated resource rather than a collection of separate language systems. Third, it applies Byram's (1997) model of intercultural communicative competence (ICC) and Deardorff's (2006) process model of intercultural competence to analyze the depth and quality of students' cultural engagement across both instructional modalities.

The analysis is organized thematically, with codes from all three instruments converging around eight major themes, each directly linked to one or more specific objectives. A summary

convergence matrix, a comparative discussion, an explicit limitations section, and pedagogical implications conclude the document.

Methodological Approach to Triangulation

This study adopted a qualitative comparative case study design (Yin, 2018; Stake, 1995; Ragin, 2014) with two embedded cases: Case A (Group 5A, AI-integrated) and Case B (Group 5B, traditional bilingual instruction). Methodological triangulation was employed as the primary validation strategy, defined by Denzin (1978, as cited in Flick, 2018) as the combination of multiple methods to study the same phenomenon, thereby compensating for the limitations of any single instrument.

The three instruments were designed to capture complementary facets of the research question. The semi-structured interview captured students' selfreported perceptions, experiences, and linguistic practices. The non-participant classroom observation documented authentic interactional behaviors and pedagogical dynamics. The think-aloud protocol provided real time access to the cognitive and metacognitive processes activated during AI interaction. This design follows Creswell and Poth's (2018) argument that combining selfreport, observation, and verbal protocol data produces a more complete picture of learning phenomena than any single method alone.

To ensure replicability of the classification system used throughout this analysis, convergence across instruments is operationally defined according to three explicit criteria. Strong convergence is assigned when all three instruments independently document the same pattern through different methodological means, for example, when interview codes,

observational field notes, and TAP transcripts each provide distinct evidence pointing to the same interpretive conclusion without contradiction. Partial convergence is assigned when two instruments document the pattern clearly and the third either adds a qualifying nuance, introduces a contextual distinction, or captures a related but non-identical phenomenon, for instance, when interviews and observations agree on a finding while TAP data reveal an important internal variation within that finding. Complementary convergence is assigned when the instruments address distinct but theoretically compatible facets of the same phenomenon, that is, when each instrument contributes a non redundant dimension that the others do not capture, and together they produce a fuller picture than any could individually. These criteria were developed deductively from the triangulation literature (Flick, 2018; Saldaña, 2021) and applied consistently by the research team across all eight themes.

Codes from each instrument were mapped onto eight thematic categories derived deductively from the theoretical framework and inductively from the data. For each theme, evidence from all three instruments is presented sequentially, followed by a convergence and divergence analysis. Participant identifiers follow the original coding conventions: 5A-P1 through 5A-P7 for AI-group interviewees; 5B-P1 through 5B-P6 for non-AI interviewees; P1 through P6 for TAP participants (all from Group 5A).

Instrument	**Group(s)**	**Code System**	**Primary Objective(s)**
Semi-structured interviews (Codification Matrix)	5A (n=7), 5B (n=6)	35 codes across 9 categories (A–I)	SO1, SO2, SO3

Instrument	Group(s)	Code System	Primary Objective(s)
Classroom observations	5A (n=10), 5B (n=10)	4 dimensions: COG, CULT, AFF, LING	SO1, SO2, SO3
Think-aloud protocols (TAP)	5A only (n=6)	6 domains: COG, META, CULT, AFF, AI, AUT	SO1, SO2

Table 1. Summary of the three instruments, participant groups, coding systems, and specific objectives addressed (see appendix for coded instruments).

Codification Overview: Instruments and Categories

Before presenting the triangulated findings, it is useful to map the coding categories of each instrument onto the shared thematic architecture of this analysis. Table 2 presents this alignment and makes explicit which codes from each instrument contribute to each theme.

Triangulation Theme	Interview Codes (Matrix)	Observation Codes	TAP Codes
1. AI Tool Usage Patterns (SO1)	A1–A6	COG (prompt use), AFF (novelty)	AI-1–AI-4, AUT-1–AUT-3
2. Cognitive Dev. & Language Acquisition (SO2)	B1–B6	COG, LING	COG-1–COG-5, META-2

Triangulation Theme	Interview Codes (Matrix)	Observation Codes	TAP Codes
3. Critical Thinking & Metacognition (SO2)	C1–C4	COG (metalinguistic)	COG-3–COG-4, META-1–META-6
4. Cultural & Intercultural Dimensions (SO2, SO3)	D1–D4	CULT	CULT-1–CULT-4
5. Affective Engagement & Motivation (SO2, SO3)	E1–E4	AFF	AFF-1–AFF-4
6. Teacher Mediation & Scaffolding (SO1, SO3)	F1–F3	COG, AFF (mediation)	(indirect: AI-3, AUT context)
7. Language Identity & Translanguaging (SO2, SO3)	G1–G3	LING (code-switching)	META-4, AUT-3
8. Limitations & Critical Assessment (SO1, SO2)	H1–H4	AFF (anxiety), COG	AI-1–AI-3, AFF-2–AFF-3

Table 2. Cross-instrument code alignment by triangulation theme.

Triangulated Findings

The following eight thematic sections present evidence from the Codification Matrix (interviews), Classroom Observations, and Think Aloud Protocols, interpreted considering the theoretical frameworks described above. Each section synthesizes relevant codes and evidence across instruments, identifies points of convergence, divergence, or complementarity according

to the operational criteria defined in the methodological section, and situates the findings within the broader theoretical literature. Student excerpts and observational field notes are reproduced as recorded in the source instruments. Spanish language excerpts are followed by researcher translations in square brackets.

Theme 1: AI Tool Usage Patterns (SO1)

Evidence from the Codification Matrix (Interviews)

Interview data reveal a varied spectrum of AI tool use across the seven students in Group 5A. All seven participants demonstrated Code A1 (Conversational AI Use), initiating dialogues with Google Gemini or Microsoft Copilot for language practice and informational inquiry. Six of seven reported Code A2 (AI as Vocabulary Scaffold), using the tools to look up, pronounce, and contextualize new vocabulary. Five students evidenced Code A3 (AI for Writing Support), and four referenced Code A4 (AI for Pronunciation and Accent), using the tools to practice spoken English and simulate interactions with speakers from other countries, including Korean interlocutors consistent with the unit's cultural theme. Three students described Code A5 (Home AI Use), indicating that engagement extended beyond the classroom into autonomous home practice, and two documented Code A6 (Peer-Mediated AI Use), in which AI served as a shared resource for collaborative problem-solving. Group 5B showed none of these AI specific codes, relying instead on books, teacher led instruction, peer collaboration, and multimedia resources (Categories I1–I4), establishing a clear instrumental divergence between the two cases at the level of resource utilization.

Evidence from Classroom Observations

The observational data corroborate and extend the interview findings across three sessions. In Session 1, students interacted with Google Gemini to cocreate a storybook, with the teacher modeling prompt formulation before students contributed their own ideas. The observer recorded [COG] basic prompt formulation and [AFF] initial shyness that gave way to growing participation as the session progressed. In Session 2, both Gemini and Copilot were employed, and their differential affordances became salient: Gemini's visible text transcript facilitated comprehension, while Copilot's faster speech rate without a transcript generated confusion and a compensatory reliance on board posted expressions. One student negotiated AI language output directly: 'Can you speak in English please?' (Observer Field Notes, Session 2, Group 5A). By Session 3, students demonstrated [COG] growing self-regulation in AI interaction and used board expressions more autonomously to structure their queries. This cross session trajectory, from basic prompts toward autonomous interaction, was identified as Pattern 1 (Cognitive Autonomy Progression) in the observational synthesis.

Evidence from Think-Aloud Protocols (TAP)

The TAP data provide the most granular account of AI tool interaction. AI-3 (Absence of Corrective Feedback) was present across all six TAP participants, the single most consistent cross participant pattern in the dataset. Neither Gemini nor Copilot provided corrective feedback on any linguistic error produced during any recorded session. AI-1 (Technical Failure) was documented in P1, P4, and P5, with conversations terminating abruptly or producing out of context content. AI-2 (Language Switch to Korean) disrupted the interactions of P1 and P2, requiring students to redirect the AI. Code AUT-2 (Reactive Posture) was dominant across the corpus: most students responded to AI stimuli without voluntarily initiating questions,

suggesting that the AI's conversational structure tended to position students as respondents rather than initiators.

Convergence and Divergence Analysis

The three instruments converge strongly on the finding that AI tools function primarily as vocabulary and pronunciation scaffolds (A2, A4 in interviews; LING observations; COG-2 in TAP), and that their use requires substantial teacher mediation from the outset. There is also strong convergence on the differential impact of Gemini versus Copilot: interview references to language switching bugs (H2), observational records of Copilot related anxiety (AFF, Session 2), and TAP patterns of AFF-2 concentrated in Copilot users (P4, P5, P6) all point to the same technical limitation. A complementary divergence exists between the interview data, which captures home use (A5) and peer-mediated use (A6) and the observational and TAP data, which are bounded to the classroom context. This divergence enriches rather than contradicts the overall picture: AI use extends into students' broader learning ecology beyond the classroom.

Theme 2: Cognitive Development and Language Acquisition (SO2)

Evidence from the Codification Matrix (Interviews)

Interview data document vocabulary acquisition in both groups, though through different mechanisms. In Group 5A, all seven participants evidenced Code B1 (Vocabulary Expansion), six reported Code B3 (Speaking Confidence), and five referenced Code B4 (Memory and Retention). In Group 5B, B1 was present in five of six students and B4 in four, suggesting that vocabulary development and memory consolidation are achievable through both AI-supported and traditional modalities. Code B6 (Academic Performance Impact) showed greater prevalence in Group 5A (4/7) than Group 5B (2/6), though this should be read as a perceptual finding rather

than a measured outcome, since it reflects students' self-assessment rather than standardized evaluation.

Evidence from Classroom Observations

Observational data point to differential patterns of language acquisition across cases. In Group 5A, the linguistic dimension [LING] shows a progression in oral production across sessions, from collaborative reading of an AI generated storybook (Session 1) to individual sustained oral practice in front of the AI (Session 3). Students drew on board posted expressions to negotiate meaning and rephrase queries, developing [COG] growing self-regulation in AI interaction. Group 5B reached its peak oral production in Session 2, where role play with Korean character flashcards generated sustained English conversation across all students. The observational record for Group 5B also documents rich metalinguistic correction events, the thirteen/thirty pronunciation distinction, the 'childrens/children' morphological error, and the omission of the third person singular -s, each addressed immediately and explicitly by the teacher. This pattern, designated Pattern 2 (Effectiveness of Traditional Role Play for Oral Production) in the observational synthesis, suggests that human mediated interaction may generate more reliable metalinguistic correction than AI interaction in its current form.

Evidence from Think-Aloud Protocols (TAP)

The TAP data illuminate the cognitive strategies activated during AI interaction. COG-1 (Contextual Inference) was documented in P1 and P4, as students derived meaning from contextual cues without knowing specific vocabulary items. COG-2 (Prior Knowledge Retrieval) was the most frequently coded cognitive strategy, appearing in P2, P3, P5, and P6 as students drew on autobiographical information to produce content. COG-4 (Comprehension Monitoring) was documented in P1 and P2, who detected gaps and formulated requests to fill

them, a strategy consistent with Vygotsky's ZPD and indicative of active cognitive engagement. COG-5 (Processing Difficulty) appeared in P3, P4, P5, and P6, reflecting the cognitive load imposed by fast AI speech, unexpected language switching, and unfamiliar Korean vocabulary.

A particularly important TAP finding concerns META-2 (Linguistic Mirroring): all six participants showed evidence of unconsciously replicating syntactic or lexical patterns from the AI's output in their own subsequent responses. This constitutes implicit acquisition through comprehensible input, consistent with Krashen's input hypothesis, and suggests that AI interaction facilitates incidental vocabulary uptake and structural imitation even in the absence of explicit feedback. The sole documented instance of verifiable lexical acquisition in the TAP corpus occurred with P5, who retained and reproduced the Korean pronunciation of 'tomato' following the AI's introduction of the term.

Convergence and Divergence Analysis

All three instruments converge on vocabulary expansion and growing speaking confidence in both groups (B1, B3 in interviews; LING observations; COG-2 and META-2 in TAP). They diverge, however, on the mechanism and quality of acquisition. TAP data reveal that AI interaction tends to produce implicit, incidental acquisition without explicit corrective scaffolding (AI-3 universal). Observational data document richer metalinguistic correction events in Group 5B, where teacher feedback addressed morphology, phonology, and syntax with immediacy and precision. Interview data align with both patterns: Group 5A students associate AI most strongly with speaking confidence and academic self-perception, while Group 5B students attribute progress more to reading fluency and direct teacher guidance.

Theme 3: Critical Thinking and Metacognitive Processes (SO2)

Evidence from the Codification Matrix (Interviews)

Interview data document metacognitive activity in both groups. In Group 5A, five of seven students described Code C1 (Error Identification), recognizing cases of incorrect or confusing AI output and responding critically rather than accepting it uncritically, a form of evaluative thinking that extends beyond receptive AI use. Six of seven evidenced Code C2 (Self-Correction Strategies), rephrasing queries or seeking clarification when encountering difficulty. Code C4 (Problem Solving Independence) appeared in five of seven students. In Group 5B, C2 was strongly present (5/6), and Code C3 (Metacognitive Reflection) was evident in four of six students, indicating that traditional instruction also generates meaningful metacognitive activity, particularly through teacher guided reflection on linguistic form and task performance.

Evidence from Classroom Observations

Observational evidence for metacognition is particularly clear in Sessions 1 and 3 of Group 5A. In Session 1, students questioned the accuracy of an AI generated image, wondering why the Korean teacher appeared to be instructing from the back of the classroom—demonstrating critical engagement with AI output [COG]. In Session 3, students actively identified vocabulary they did not understand ('mug', 'slides') and applied board posted phrases to rephrase queries, evidencing [COG] metacognitive awareness through active meaning negotiation. In Group 5B, Session 2 produced the richest metacognitive correction events of the entire study: the teacher's immediate addressing of thirteen/thirty, 'childrens/children', and third person -s omissions prompted [COG] morphological awareness and explicit linguistic reflection. The use of analogies (explaining 'cabbage' via comparison with 'lettuce' as context for kimchi) also activated [COG] contextual vocabulary clarification.

Evidence from Think-Aloud Protocols (TAP)

The TAP data illuminate both the potential and the limits of AI mediated metacognition. META-3 (Self-Correction) was largely absent from the entire TAP corpus: no participant spontaneously detected and corrected their own linguistic errors during AI interaction. This finding directly converges with AI-3 (universal absence of corrective feedback): without an interlocutor who signals error, the conditions for monitoring and self-repair are substantially undermined. META-5 (Clarification Request) was documented in P1, P2, P3, and P4, confirming that students did deploy metacognitive strategies to address comprehension gaps. META-4 (Code-Switching) appeared in P2 and P6 as a meaning negotiation strategy: P2 redirected the AI toward English, and P6 used Spanish to retrieve an unknown English term ('how do you say fluido?'). META-1 (Lexical Awareness) emerged in P1's reformulation of questions about kimchi preparation, indicating explicit reflection on word choice and the limits of available vocabulary.

Convergence and Divergence Analysis

Convergence is partial across instruments. All three confirm that both groups engage in metacognitive activity, but the instruments diverge on the type. Interview codes C1–C4 and TAP meta codes show that AI group students' direct metacognitive attention toward evaluating and redirecting AI output, while non AI students focus metacognition on linguistic form guided by teacher feedback. This distinction is also captured in the observational data: Group 5A's metacognition is triggered by AI inconsistencies (language switching, inaccurate images), while Group 5B's is triggered by teacher correction events. The critical divergence concerns depth: TAP data suggest that AI mediated metacognition tends to remain at the level of comprehension monitoring (COG-4) rather than linguistic self-correction (META-3), whereas teacher mediated

metacognition in Group 5B addresses specific grammatical structures with documented immediate precision.

Theme 4: Cultural and Intercultural Dimensions (SO2, SO3)

Evidence from the Codification Matrix (Interviews)

The interview data document substantive intercultural engagement in both groups, centered on the unit's South Korea Colombia comparative theme. However, before analyzing the findings by instrument, it is necessary to establish a fundamental theoretical distinction that runs throughout the entire discussion: the difference between superficial cultural access and genuine intercultural development. Drawing on Byram's (1997) model of Intercultural Communicative Competence, this distinction can be articulated in terms of two central components: *savoir être*, understood as the attitudinal disposition of openness, curiosity, and willingness to suspend one's own cultural judgments, and *savoir comprendre*, defined as the ability to interpret documents or events from another culture, explain them, and relate them to one's own cultural referents.

The mere access to cultural information, such as naming foods or practices from another country, constitutes a descriptive level that does not necessarily activate either of these two components. Spontaneous intercultural reflection, by contrast, presupposes the simultaneous mobilization of both, the student not only recognizes the difference but interprets it, compares it, and adopts a genuinely open attitude toward it. This distinction is decisive for reading the data presented below.

Code D1 (Korea Cultural Encounter) was present in six of the seven students in Group 5A and in three of the six students in Group 5B, indicating that AI mediated contact with

Korean culture was more widespread in the experimental group, although both groups accessed Korean cultural content. However, the mere presence of Code D1 does not imply intercultural development: in terms of Byram (1997), it represents in most cases a contact situated at the level of *savoir*, factual knowledge about another culture, without guaranteeing the activation of *savoir comprendre* or *savoir être*. Code D2 (Intercultural Curiosity) appeared in five of seven 5A students and four of six 5B students, a comparatively small difference suggesting that curiosity about foreign cultures is not exclusively generated by AI tools.

Code D3 (Global Awareness) was present in six of seven 5A students and five of six 5B students, confirming that both instructional modalities activated global perspectives. The AI specific Code D4 (AI as Cultural Mediator), which captures students' use of AI to simulate cross cultural encounters and cultural role play, appeared in five of the seven students in Group 5A and was inapplicable to Group 5B, constituting the most distinctive differential between groups.

From the perspective of intercultural competence models, this code expresses a capacity potentially aligned with *savoir apprendre/faire*, the ability to discover and interact with what is culturally new (Byram, 1997), although its actual effectiveness depends on the level of pedagogical scaffolding available, as underscored by Klimova and Chen (2024) in their systematic review of the impact of AI on the development of intercultural communicative competence.

Evidence from Classroom Observations

The observational data reveals a qualitative distinction in cultural engagement across groups that the interview data alone do not fully capture, and which is particularly pertinent in

light of the theoretical difference between superficial cultural access and spontaneous intercultural reflection.

In Group 5A, Session 1 produced (CULT) access to authentic cultural content about Korean dishes and the capital city through AI, although the observer noted that cultural access was fluid but superficial, conditioned by the quality of the prompts formulated. This finding is consistent with what Klimova and Chen (2024) report, noting that AI tools offer significant benefits for personalizing cultural learning, but that relevant challenges persist in terms of cultural sensitivity and depth of intercultural experience when explicit pedagogical scaffolding is absent.

In terms of Byram (1997), the access observed in these sessions operates predominantly at the level of *savoir*, accumulation of cultural data, with no consistent evidence of *savoir comprendre* activation, that is, students did not systematically link that information to their own Colombian cultural reality, nor did they interpret it in terms of shared or divergent meanings.

In Session 2, AI mediated cultural content was further limited by the tool's language barrier, Korean vocabulary mixing, which further reduced the depth of cultural exploration. Session 3 generated interactive access to information about Korean festivals, but explicit comparative reflection was limited due to the individual structure of the activity, which did not foster the confrontation of perspectives that characterizes genuine intercultural development.

Group 5B, by contrast, generated deeper and more spontaneous cultural comparisons across all three sessions, more frequently evidencing the components of *savoir être* and *savoir comprendre* in Byram's (1997) terms. In Session 1, students spontaneously compared food,

schedules, and daily practices between Colombia and Korea with greater elaboration than Group 5A, revealing an active attitude of openness toward cultural difference rather than mere recognition of it.

In Session 2, a student independently noted that Korean children remove their shoes before entering the classroom, a discovery made without teacher prompting, which generated spontaneous comparative cultural reflection: precisely the kind of relational interpretation that defines *savoir comprendre* in Byram's model.

In Session 3, the study's most salient cultural finding emerged which is the discovery that Christmas and Easter are not celebrated in South Korea generated genuine surprise and extended cultural discussion. This type of reaction, in which students suspend their own cultural presuppositions to approach a different value system, constitutes a direct expression of the *savoir être* described by Byram (1997) as the attitudinal foundation of all intercultural competence.

This observational pattern, identified as Pattern 4 (Cultural Depth: Videos vs. AI), supports the conclusion that teacher mediated, video enriched instruction generated greater intercultural development, not merely cultural access, than AI mediated interaction in the absence of explicit pedagogical scaffolding.

Evidence from Think-Aloud Protocols (TAP)

The TAP data present the most critical perspective on AI's cultural mediation capacity and offer the most direct evidence for distinguishing between cultural exploration and genuine intercultural development. Code CULT 4 (Absence of Intercultural Engagement) dominated

the TAP corpus, appearing in P2, P3, P5, and P6. This finding is theoretically significant: even when the AI introduced Korean cultural content (AI 4), students generally did not respond with intercultural curiosity or comparative reflection, indicating that mere exposure to cultural information does not automatically activate any of the components in Byram's (1997) model.

In the absence of *savoir être*, the attitudinal disposition toward openness and curiosity, the cultural content supplied by AI does not become intercultural experience but inert information. This limitation is consistent with the warnings issued by Klimova and Chen (2024), who conclude that the impact of AI tools on intercultural competence depends critically on human supervision and the pedagogical design surrounding their use.

Code CULT 2 (Cultural Knowledge Acquisition) was documented in P1, P4, and P5, but in each case the acquisition was superficial, limited to naming Korean foods or practices without exploring their cultural significance, which corresponds to a descriptive level of *savoir* with no evidence of *savoir comprendre*. The one notable exception was P1, the only participant in the TAP group to express genuine intercultural motivation: *"I feel excited, I would like to meet with a Korean boy, a real Korean boy and I would like to be in this country one day of my life"* (P1, TAP Session 2).

This expression evidences the simultaneous activation of *savoir être*, openness and aspiration toward the cultural other, and what Byram (1997) terms critical cultural awareness, in so far as the student does not merely receive information about Korea but constructs a genuine desire for intercultural encounter. This divergent case, coded as CULT 3, was identified as a key triangulation case, suggesting that AI interaction can serve as a catalyst for

intercultural aspirations in some learners, although this capacity depends on individual factors that extend beyond the tool itself.

Theoretical Interpretation: Intercultural Communicative Competence

To interpret these findings with greater theoretical depth, this analysis draws on Byram's (1997) model of intercultural communicative competence (ICC) and Deardorff's (2006) process model of intercultural competence. Byram (1997) identifies five interdependent components of ICC: knowledge (savoirs), attitudes (savoir être), skills of interpreting and relating (savoir comprendre), skills of discovery and interaction (savoir apprendre/faire), and critical cultural awareness (savoir s'engager). Applied to this dataset, the findings reveal a consistent pattern: both instructional modalities activated knowledge acquisition (facts about Korea) and early attitudinal engagement (curiosity, surprise), but neither consistently reached the higher order components of ICC, skills of interpretation, discovery in interaction, or critical cultural awareness, without deliberate pedagogical scaffolding.

AI mediated interaction in Group 5A tended to remain within the savoir (knowledge) dimension of Byram's model: students learned cultural facts through prompting but rarely engaged in the comparative reflection (savoir comprendre) or critical evaluation (savoir s'engager) that characterize deeper intercultural competence. The TAP dominance of CULT-4 is consistent with this interpretation; cultural content was received but not processed interculturally in the absence of explicit scaffolding. Group 5B's spontaneous cultural discoveries (shoes at the classroom door, the absence of Christmas in Korea) more closely approaches Byram's savoir comprendre dimension, as students related new cultural information to their own reference framework and articulated meaningful comparisons.

Deardorff's (2006) process model adds a valuable sequential lens. The model positions attitudes (respect, openness, curiosity) as the necessary foundation for knowledge and comprehension skills, which in turn enable adaptive communicative behavior across cultural contexts. In both groups, attitudinal engagement was evident, interview Codes D1–D3 and observational [CULT] records confirm genuine curiosity about Korean culture. The challenge identified by the triangulation is the transition from attitude to skill: neither the AI interaction (which lacked structured comparative prompts) nor the traditional sessions (which were culturally rich but primarily receptive) systematically guided students toward the interpretive and reflective skills that Deardorff places at the model's output stage. This gap points to the need for explicitly designed intercultural tasks, reflection protocols, guided comparison activities, or cross-cultural dialogue structures, in both AI-integrated and traditional bilingual classrooms.

Within the Colombian context, these findings carry additional significance. Colombia's National Bilingualism Program frames English proficiency as a pathway to global citizenship, yet the intercultural dimension of that citizenship remains underspecified in most school curricula (García, 2009). The data from both groups suggest that intercultural awareness can emerge organically through culturally themed units, particularly when students encounter surprising cultural differences that challenge their existing assumptions, but that the depth of that awareness depends on the quality of pedagogical mediation rather than on the presence or absence of technology.

Convergence and Divergence Analysis

The three instruments converge on the finding that both groups accessed Korean cultural content (D1, D3 in interviews; CULT codes in observations; AI 4 in TAP). They

diverge significantly, however, in the depth and spontaneity of that engagement, which refers directly to the theoretical distinction established at the outset of this theme: accessing cultural information is not equivalent to developing intercultural competence. Drawing on Byram's (1997) model, the interview data reveal comparable presence of *savoir*, cultural knowledge, across groups (D2), while observational data show that Group 5B more frequently evidenced *savoir comprendre* and *savoir être*, relational interpretation and spontaneous attitudinal openness. The TAP data, in turn, indicate that AI interaction, without explicit cultural scaffolding, rarely generates the sustained intercultural reflection that these components require. This convergent divergence, agreement on the breadth of cultural access, disagreement on the depth of intercultural development, is theoretically significant and aligns with the findings of Klimova and Chen (2024), who conclude that while AI personalizes and enriches cultural exposure, its contribution to the development of genuine intercultural competence remains limited without active pedagogical supervision. This finding is consistent with the emphasis of Cummins (2021) and Wang (2023) on culturally responsive instructional design as a necessary condition for technological tools to activate processes of intercultural formation rather than mere cultural information transmission.

Theme 5: Affective Engagement and Motivation (SO2, SO3)

Evidence from the Codification Matrix (Interviews)

Motivation was strongly present in both groups. Code E1 (Intrinsic Motivation) was evidenced in all seven Group 5A students and five of six Group 5B students, making it one of the most broadly distributed codes in the matrix. Code E2 (Gamification Engagement) appeared in all seven 5A and all six 5B students, confirming that game based tools such as Kahoot, Blooket, and Bamboozle function as a shared motivational resource across both instructional

contexts. Code E4 (AI Novelty Effect)—the perception that AI makes learning more interesting, was present in six of seven Group 5A students but partially in only one Group 5B student, documenting the motivational premium that AI novelty generates in the AI group. Code E3 (Future Oriented Motivation) appeared at nearly equal rates across both groups (4/7 and 4/6), indicating that aspirations related to travel, career, and cross-cultural communication are not differentially generated by AI exposure.

Evidence from Classroom Observations

Observational data reveal that while Group 5A maintained more consistent motivation across all three sessions, Group 5B showed higher peak motivation in specific activity types. In Session 1, both groups demonstrated [AFF] high motivation and active participation, Group 5A energized by technological novelty and co-creation, Group 5B by hands-on craft production and peer collaboration. In Session 2, Group 5B reached its motivational peak (Pattern 2): the role-play activity with Korean flashcards generated 100% participation, sustained laughter, and intense social interaction, all coded [AFF] high engagement, while Group 5A showed generally positive affect alongside documented [AFF] specific anxiety with Copilot. In Session 3, Group 5A showed [AFF] greater enthusiasm compared to prior sessions, while Group 5B exhibited [AFF] a noticeable decline in motivation during extended oral production, dependent on teacher incentives to sustain participation.

Evidence from Think-Aloud Protocols (TAP)

The TAP data provide a nuanced view of affective engagement at the individual level. AFF-1 (Positive Affect) was documented in P1 and P2, P1 expressing excitement about Korean culture and P2 demonstrating enthusiastic fluency. AFF-2 (Anxiety/Nervousness) was concentrated in Copilot users (P4, P5, P6), confirming the observational and interview

convergence on Copilot specific affective difficulty. AFF-3 (Frustration) appeared in P5 and P6, the former expressing difficulty with the AI's speed ('I can't do it very very good'), the latter frustrated at not being able to speak as fluently as the AI. AFF-4 (Communicative Confidence) was documented in P2 and P3, who maintained sustained fluency and confident tone throughout their interactions. These distinctions, only visible at the individual TAP level, reveal tool specific affective responses that aggregate interview data do not capture: Gemini generated more positive affect and confidence, while Copilot generated more anxiety and frustration.

Convergence and Divergence Analysis

Strong convergence across all three instruments confirms that both instructional modalities generate high motivation levels, and that gamification is a shared motivational resource regardless of AI integration. The instruments diverge on the nature and stability of motivation: interview and TAP data suggest that AI novelty generates a motivational premium in Group 5A (E4, AFF-1), while observational data show that Group 5B's peak motivation occurs in socially interactive, physically collaborative activities (Session 2 role-play). This divergence suggests that AI based motivation may be partially tool dependent and novelty driven, while collaborative motivation in traditional settings derives from social belonging and peer interaction, a distinction with implications for the sustainability of engagement across different task types and over extended instructional periods.

Theme 6: Teacher Mediation and Pedagogical Scaffolding (SO1, SO3)

Evidence from the Codification Matrix (Interviews)

Teacher mediation was strongly evidenced across both groups. Code F1 (Teacher as AI Facilitator) was present in all seven Group 5A students, who referenced the teacher's role in guiding AI interaction, posing questions, and structuring AI-student exchanges. Code F2 (Teacher-Centered Instruction) was the dominant pedagogical mode in Group 5B (6/6), while in Group 5A it was shared with AI co-facilitation. Code F3 (Scaffolded Feedback) documents that both AI and teacher were cited as sources of feedback in Group 5A, while Group 5B cited the teacher exclusively. These patterns align with Vygotsky's notion of mediated learning, in both cases, a more knowledgeable agent (teacher, AI, or both) structures the learner's engagement with the task, but the tools through which mediation operates differ markedly.

Evidence from Classroom Observations

Observational data establish teacher mediation as the most consistent cross-case pattern of the entire study, designated Pattern 3 (Indispensable Role of the Teacher in Both Cases) in the observational synthesis. In Group 5A, teacher mediation was essential for modeling AI tool use, reducing initial anxiety, and verifying comprehension of AI output. Without continuous teacher presence and intervention, many AI interactions would have stalled or produced confusion: in Session 1, initial shyness was overcome through teacher guidance; in Session 2, the teacher redirected students when Gemini produced confusing Korean vocabulary. In Group 5B, teacher mediation was the primary mechanism for linguistic correction, cultural contextualization, and oral production scaffolding across all three sessions. This cross-case consistency confirms the theoretical position articulated by Cummins (2021) and Wang (2023):

teacher guidance is not supplementary but structurally necessary in bilingual learning environments, regardless of the technological resources available.

Convergence and Divergence Analysis

All three instruments converge unequivocally on the indispensability of teacher mediation (Strong convergence). The TAP data add an indirect but important dimension: the universal absence of corrective feedback from AI (AI-3) means that in Group 5A, the teacher remains the only agent capable of providing explicit linguistic correction, reinforcing the observation that AI redistributes rather than reduces the teacher's mediating role. The divergence between instruments concerns the nature of that mediation: in Group 5A, teacher mediation is primarily technical and metacognitive (modeling tool use, managing AI errors, verifying output quality); in Group 5B, it is primarily linguistic and cultural (correcting grammar, contextualizing cultural content, sustaining oral production). Both functions are necessary, and neither is reducible to the other.

Theme 7: Language Identity and Translanguaging (SO2, SO3)

Theoretical Framework: Translanguaging Space

Language identity and translanguaging codes reveal comparable bilingual awareness across both groups, challenging any assumption that AI integration uniquely promotes bilingual identity development. Code G1 (Code Switching Awareness) was present in five of seven Group 5A students and five of six Group 5B students, an effectively identical rate.

Code G2 (Spanish as Learning Bridge) apeared in four of seven 5A students and five of six 5B students, indicating a slight advantage for the non AI group in using L1 strategically to support L2 comprehension. Code G3 (Preference for English Immersion) appeared equally

across both groups (3/7 and 3/6). These patterns suggest that bilingual identity development and translanguaging strategies are not differentially generated by AI exposure and may instead reflect students' broader sociolinguistic orientation toward English.

This convergence across groups is interpretable through Cummins' (1979) linguistic interdependence hypothesis, which holds that competence developed in the L1 constitutes a shared cognitive foundation that underlies L2 processing, regardless of the instructional environment. Within this framework, the strategic use of Spanish as a bridge is not a phenomenon generated by AI, but rather an expression of a common underlying proficiency shared by both groups, rooted in their shared bilingual experience within the context of the Colombian National Bilingualism Programme (Ministerio de Educación Nacional, 2004), a policy that, while advancing English learning in formal education, has tended to privilege this language over the recognition of students' full linguistic repertoire (De Mejía, 2006).

Evidence from the Codification Matrix (Interviews)

Observational evidence for translanguaging appeared in both cases. In Group 5A, students occasionally used Spanish to scaffold their AI queries when lacking English vocabulary, a practice the observer noted as consistent with [LING] linguistic problem solving and [COG] metalinguistic awareness of one's own limitations. In Group 5B, the teacher's use of Spanish analogies (cabbage/lettuce to explain kimchi) explicitly modelled translanguaging as a pedagogical tool. Both cases thus demonstrate translanguaging practices, though in different directions: in Group 5A, translanguaging bridges L1 to the AI; in Group 5B, it bridges L1 to new cultural lexical content. These distinct directions can be understood through the concept of the translanguaging space proposed by García and Wei (2014), who define this

space as one in which multilingual speakers mobilize their full linguistic repertoire in creative and critical ways, transcending the boundaries imposed by institutionally named languages. In both groups, the classroom functions as a translanguaging space: in Group 5A, mediated by technology; in Group 5B, mediated by teaching practice.

Evidence from Classroom Observations

Observational evidence for translanguaging appeared in both cases, though in different forms. In Group 5A, students occasionally used Spanish to scaffold their AI queries when lacking English vocabulary, a practice coded [LING] as linguistic problem-solving and [COG] as metalinguistic awareness of one's own limitations. This represents a student-initiated translanguaging space in which the learner draws on the full bilingual repertoire to maintain communicative continuity with the AI. In Group 5B, the teacher's use of Spanish (analogies explaining 'cabbage' by comparing it with 'lettuce' as context for understanding kimchi) explicitly modeled translanguaging as a pedagogical strategy, creating a teacher led translanguaging space consistent with García and Li Wei's (2014) description of educationally designed translanguaging practices.

Evidence from Think-Aloud Protocols (TAP)

The TAP data document the most explicit instances of strategic translanguaging in the corpus. META 4 (Code-Switching) appeared in P2 and P6 with qualitatively distinct motivations. P2 used English meta-commentary to redirect the AI (“Can you speak in English please?”), a form of metalinguistic self regulation. P6 inserted Spanish strategically when lacking an English vocabulary item (“how do you say fluido?”), as a pragmatic communicative strategy. Code AUT 3 (Strategic Resource Use) was documented in both

cases, confirming that code switching was employed as a functional communicative resource rather than as a failure of bilingual competence.

These TAP patterns align with García's (2009) conceptualization of translanguaging as a flexible and purposeful deployment of the full linguistic repertoire, rather than a deficit, and with García and Wei's (2014) clarification that translanguaging practices go beyond the language systems endorsed by the state, revealing the bilingual speaker's linguistic agency.

From a critical bilingualism perspective, the practices documented in the TAP are not merely compensatory strategies: they are acts situated within power relations between languages, through which students reclaim the epistemic value of their L1 within an environment that symbolically privileges English (Cummins, 2000).

Convergence and Divergence Analysis

All three instruments converge on the presence of translanguaging in both cases, with comparable rates of code switching awareness and L1 bridge use across groups (interview codes G1/ G2; observational LING patterns; TAP META 4 and AUT 3). The instruments diverge on the directionality and intentionality of translanguaging: TAP data reveal conscious, strategic use in Group 5A students; observational data document teacher modelled use in Group 5B; and interview data suggest that both groups recognize and engage with the Spanish/English interface of their bilingual identity

This convergent picture supports a translanguaging pedagogy framework applicable to both AI integrated and traditional bilingual classrooms, and is consistent with García and Wei's (2014) translanguaging space as an analytical framework that allows the practices of

both groups to be read not as isolated phenomena, but as expressions of a situated critical bilingualism, one that, in the Colombian context, operates in tension between an educational policy oriented toward English as a language of global prestige and the reality of classrooms, where Spanish remains the primary vehicle of meaning, identity, and belonging (De Mejía, 2006; Cummins, 2000)

Theme 8: Limitations and Critical Assessment of AI Tools (SO1, SO2)

Evidence from the Codification Matrix (Interviews)

The interview data document a critically aware relationship between Group 5A students and the AI tools they used. Code H1 (AI Inaccuracy/Error) was present in five of seven students, who reported cases of incorrect or confusing AI output. Code H2 (AI Language Switching Bug) appeared in three of seven students, documenting frustration when Gemini unexpectedly switched to Korean mid conversation. Code H4 (Desire for AI Improvement) was documented in five of seven students, who proposed specific changes to the tools, evidence of critical agency rather than passive acceptance. Code H3 (Over Reliance Concern) was implied in two responses, suggesting emerging awareness that AI use may reduce the cognitive effort required for language tasks. None of these codes were applicable to Group 5B, reinforcing the instrument-specific character of AI critique data.

Evidence from Classroom Observations

Observational evidence of AI limitations is most concentrated in Session 2 of Group 5A. The observer documented Gemini mixing English and Korean vocabulary, generating student confusion; Copilot speaking at a rate that exceeded students' processing capacity without providing a visible transcript; and students increasingly relying on board expressions as a compensatory linguistic strategy when AI interaction broke down. The observer designated this

Pattern 5 (Differential Anxiety Toward AI Tools), concluding that tool design characteristics—specifically speech rate and transcript availability, directly impacted the affective and cognitive dimensions of learning, and that these characteristics should be treated as independent methodological variables in future AI in education research.

Evidence from Think-Aloud Protocols (TAP)

The TAP data provide the most systematic account of AI limitations as experienced in real time. AI-3 (Absence of Corrective Feedback) was universal across all six TAP participants. AI-1 (Technical Failure) affected P1, P4, and P5, with the AI terminating conversations abruptly and disrupting knowledge acquisition. AI-2 (Language Switch to Korean) disrupted the interactions of P1 and P2. The TAP document's own Section 4 explicitly predicted convergence with interview data on error fossilization (AI-3) and with observational data on teacher refraining from correction during AI interactions, creating what the TAP analysis characterized as a zero feedback environment for linguistic form.

Convergence and Divergence Analysis

All three instruments converge strongly on the identification of AI limitations. Interview codes H1–H4 document student perceived errors, bugs, and improvement proposals; observational data record specific tool dysfunction events across sessions; and TAP data capture the real time cognitive and affective impact of those disfunctions. The three instruments also converge on the differential usability of Gemini versus Copilot, with Gemini consistently generating lower anxiety and better comprehension across all data sources. No divergence of substance exists across instruments on this theme. The strength of the convergence, six of six TAP participants, multiple observational sessions, and five of seven interview participants all

corroborating the same finding, establishes the absence of corrective feedback as the most robust and consequential limitation of AI as a language learning interlocutor in this study.

Cross-Instrument Convergences and Divergences

Table 3 presents a summary convergence matrix synthesizing the degree of agreement across instruments for each triangulation theme. Convergence type follows the operational definitions established in the methodological section: Strong (all three instruments independently document the same pattern through different methodological means); Partial (two instruments agree clearly and the third adds a qualifying nuance or internal distinction); Complementary (instruments address distinct, non-redundant aspects of the same phenomenon without contradiction).

Theme	Instruments Converging	Convergence Type	Key Integrated Finding
1. AI Tool Usage Patterns (SO1)	Matrix + Obs + TAP	Strong	AI tools scaffold vocabulary and pronunciation; teacher mediation is essential; Copilot generates more difficulty than Gemini across all instruments.
2. Cognitive Dev. & Language Acquisition (SO2)	Matrix + Obs + TAP	Partial	Both groups develop vocabulary; AI produces implicit acquisition via linguistic mirroring; traditional instruction produces richer metalinguistic correction through teacher feedback.

Theme	Instruments Converging	Convergence Type	Key Integrated Finding
3. Critical Thinking & Metacognition (SO2)	Matrix + Obs + TAP	Partial	Both groups show metacognition; AI activates evaluative metacognition (checking AI output); traditional instruction activates corrective metacognition (grammar, phonology). Self correction (META-3) largely absent in AI group.
4. Cultural & Intercultural Dimensions (SO2, SO3)	Matrix + Obs + TAP	Partial (divergent on depth)	Both groups engage culturally at the knowledge-attitude level (Byram, 1997); AI provides interactive but superficial access; traditional instruction generates deeper comparative reflection more closely approaching savoir comprendre competencies.
5. Affective Engagement & Motivation (SO2, SO3)	Matrix + Obs + TAP	Partial	Both groups highly motivated; AI novelty generates a motivational premium in 5A; peak motivation in traditional contexts occurs in social/collaborative activities; Copilot generates tool-specific anxiety.
6. Teacher Mediation (SO1, SO3)	Matrix + Obs + TAP (indirect)	Strong	Teacher mediation is indispensable in both cases. Its function shifts from technical/metacognitive in 5A to linguistic/cultural in 5B. AI redistributes rather than

Theme	Instruments Converging	Convergence Type	Key Integrated Finding
			reduces the teacher's mediating role.
7. Language Identity & Translanguaging (SO2, SO3)	Matrix + Obs + TAP	Complementary	Translanguaging present in both groups at comparable rates; student initiated in 5A (TAP), teacher-modeled in 5B (observation). Both cases create distinct translanguaging spaces (García & Li Wei, 2014).
8. AI Limitations & Critical Assessment (SO1, SO2)	Matrix + Obs + TAP	Strong	Universal absence of AI corrective feedback across all instruments; Gemini superior to Copilot for A2-level EFL learners; students demonstrate critical AI evaluation capacity.

Table 3. Cross-instrument convergence matrix by triangulation theme.

Discussion

Four interpretive claims emerge from the convergence patterns documented above. Each is grounded in the multi instrument evidence and situated within the study's theoretical framework.

First, AI integration in the bilingual classroom functions as a motivational and scaffolding amplifier, but not as a replacement for human pedagogical mediation. The consistency of Code F1 (all seven interview participants), Pattern 3 (all six observational sessions), and the indirect evidence from AI-3 (universal absence of corrective feedback in TAP)

collectively establish that AI tools expand the range of linguistic and cultural input available to students while leaving the critical scaffolding functions (correction, contextualization, and metacognitive guidance) to the teacher. This is consistent with Cummins' (2021) argument for sustained teacher involvement in bilingual settings and with Ma's (2024) finding that AI tools, without deliberate pedagogical framing, fail to address the affective and cultural complexity of bilingual learners.

Second, the quality and depth of intercultural learning depend more on pedagogical design than on the presence or absence of AI tools. Group 5B's consistently deeper and more spontaneous cultural comparisons, documented in the observational synthesis (Pattern 4), in the TAP dominance of CULT-4, and in the comparable interview rates for D2 and D3, indicate that videos, physical artefacts, and teacher mediation can produce richer intercultural reflection than AI interaction not structured around explicit intercultural objectives. Mapped onto Byram's (1997) ICC model, neither group consistently reached the higher order competencies of savoir comprendre or savoir s'engager, and Deardorff's (2006) process model highlights a shared gap: both groups developed intercultural attitudes and some knowledge, but neither modality systematically cultivated the interpretive and critical skills that define deep intercultural competence. This is not a critique of AI as such, but an argument for explicit intercultural task design across both modalities.

Third, the universal absence of corrective linguistic feedback from AI represents the most consequential limitation identified by this triangulation. AI-3's presence across all six TAP participants converges with observational evidence of uncorrected errors in Group 5A (META-6 codes) and with the asymmetry in feedback quality between the two cases. Group 5B's observational sessions document rich corrective events targeting morphology, phonology, and

syntax, each producing measurable metalinguistic awareness. This asymmetry aligns with Lee and Warschauer's (2023) research on corrective feedback in second language acquisition, suggesting that conversational AI tools in their current form should be systematically complemented by teacher orchestrated correction in AI-integrated classrooms, not treated as self-sufficient interlocutors.

Fourth, tool design variables (speech rate, transcript availability, and language output stability) are not neutral for A2-level EFL learners. The convergence of interview Code H2, observational Pattern 5, and TAP AFF-2 (concentrated in Copilot users) establishes a clear evidence base for this conclusion. Future AI in education implementations should evaluate tool specific affordances in relation to learners' proficiency levels and cognitive processing capacity, rather than treating all conversational AI tools as interchangeable instructional resources. For young EFL learners at A2 level, tools with visible transcripts and predictable language output appear substantially more appropriate than those without.

Regarding translanguaging, the triangulation affirms that the translanguaging space (García & Li Wei, 2014) exists in both instructional modalities but is shaped differently by each. AI interaction creates learner initiated translanguaging spaces in which students draw on Spanish strategically to maintain communicative continuity (TAP META-4, AUT-3). Traditional instruction creates teacher designed translanguaging spaces in which L1 is deployed as an explicit comprehension scaffold (observational LING data). Both forms are pedagogically legitimate and consistent with Cummins' (2021) CUP model, yet neither was systematically theorized or expanded by the teaching practice documented in this study. There is scope for deliberate translanguaging pedagogy in both contexts.

Study Limitations

This study presents several methodological and contextual limitations that must be acknowledged before drawing broader conclusions from the findings. The most significant concerns the size and scope of the sample. Both participating groups comprised a small number of students (Group 5A: n = 7 for interviews, n = 10 for observations; Group 5B: n = 6 for interviews, n = 10 for observations), and the think-aloud protocols were conducted exclusively with six participants from Group 5A, which means that real time cognitive data are not available for Group 5B. This asymmetry limits the comparability of TAP evidence across cases and prevents a fully symmetrical triangulation for Themes 3, 5, and 7. Moreover, all participants belonged to a single private institution in Bogotá, with similar socioeconomic profiles and English proficiency levels (A2). The findings therefore describe a specific, bounded educational context rather than AI integration in Colombian bilingual education broadly, and should not be generalized beyond comparable institutional settings without further research.

A second set of limitations concerns the instruments themselves and the conditions of data collection. The think-aloud protocol, while powerful for accessing cognitive processes in real time, introduces a reactivity effect: verbalizing thought may alter the very processes being captured (Ericsson & Simon, 1993; Bowles, 2010), particularly for young learners who may not have fully developed metacognitive vocabularies. The classroom observations were bounded to three sessions per group, a timeframe that may not capture the full trajectory of change over a longer instructional period, including, for instance, whether the AI novelty effect documented in Group 5A (Code E4) would persist, diminish, or transform with sustained AI exposure. Similarly, the interview data reflect students' self-perceptions at a specific point in time, which

may not correspond precisely to their actual learning behaviors. Finally, the study does not include independent measures of language proficiency gain, intercultural competence development, or long term AI reliance, which means that the cognitive and cultural claims are grounded in qualitative evidence rather than in longitudinal or comparative assessment data. These limitations do not diminish the validity of the triangulated findings within their proper scope, but they do indicate the boundaries within which those findings should be interpreted and the directions in which future research could productively extend this work.

Conclusions and Pedagogical Implications

This triangulation has synthesized evidence from three qualitative instruments to produce an integrated interpretation of how AI integration shapes the cognitive and cultural dimensions of bilingual learners in a private Colombian institution. The conclusions are organized according to the study's three specific objectives.

With respect to SO1 (Ways in Which AI Tools Are Used), the triangulation confirms that fifth grade students in Group 5A engaged with AI as vocabulary scaffolds, writing assistants, pronunciation models, cultural interlocutors, and, in some cases, peer mediated collaborative resources. Use extended beyond school hours in three of seven cases. However, interaction was predominantly reactive (AUT-2) rather than initiative driven, and the quality of AI engagement was substantially conditioned by prompt formulation capacity, a skill that developed progressively across the three observed sessions, suggesting that sustained AI use, with deliberate instructional support, can cultivate greater learner autonomy over time.

With respect to SO2 (Influence on Cognitive Development, Language Acquisition, and Critical Thinking), AI integration produced self-reported gains in speaking confidence (B3), vocabulary expansion (B1), and error identification (C1). The TAP data add that implicit acquisition through linguistic mirroring (META-2) occurred consistently across AI users, even without explicit feedback. The central limitation, absent corrective feedback (AI-3 universal), means that the self-monitoring and repair processes associated with deeper metalinguistic processing (META-3) were not systematically activated. Traditional instruction, by contrast, generated richer grammatical metacognition and more accurate language production through consistent teacher correction, confirming the continued importance of human feedback in second language development.

With respect to SO3 (Comparative Impact), the triangulation reveals complementary rather than competing strengths. AI integration offers conversational spontaneity, motivational novelty, interactive cultural exposure, and student initiated translanguaging. Traditional instruction offers deeper intercultural reflection through human storytelling and multimedia, reliable corrective feedback, peak motivation in collaborative activities, and clearer linguistic modeling. Neither modality alone constitutes an optimal bilingual learning environment. A deliberate integration that draws on both, structured around intercultural tasks aligned with Byram's (1997) ICC model, supported by translanguaging pedagogy (García & Li Wei, 2014), and anchored by teacher-mediated feedback, appears to offer the most effective framework for A2-level bilingual learners in the Colombian private school context.

The following pedagogical implications are drawn from this triangulation and organized according to the strength of their empirical support:

Implications supported by strong convergence (all three instruments):

- AI tool selection in bilingual classrooms should be governed by explicit learner-accessibility criteria: visible transcripts, stable language output, and speech rates appropriate to the proficiency level of the target learners. For A2-level students, tools such as Gemini, which provide readable text alongside audio, are more appropriate than Copilot style interfaces without this scaffolding feature.
- Teacher orchestrated corrective feedback must be maintained as a deliberate pedagogical practice in AI integrated classrooms. AI tools in their current conversational form cannot substitute for explicit linguistic correction, and the absence of such feedback creates conditions for error fossilization in language development.
- Teacher mediation should be conceptualized as a non negotiable element of AI-integrated instruction rather than as a transitional support to be phased out. The teacher's mediating function changes alongside AI integration (from primary knowledge transmitter to technical facilitator and cultural contextualizer) but it does not diminish.

Implications supported by partial convergence (two or more instruments, with some qualification):

- Prompt formulation should be explicitly taught as a metacognitive and linguistic skill, sequenced progressively across instructional units, and not assumed as an intuitive

capacity. The S1→S2→S3 autonomy progression observed in Group 5A suggests that with explicit modeling and sustained practice, students develop more elaborated prompt strategies over time.

- Intercultural tasks aligned with Byram's (1997) ICC model should be designed as a deliberate layer of AI-integrated instruction. Without explicit prompts for cultural comparison, reflection, and critical analysis, AI interaction tends to remain within the factual knowledge dimension of intercultural competence, without generating the interpretive or reflective skills associated with deeper cultural understanding.
- Translanguaging should be recognized and cultivated as a pedagogical resource in both AI-integrated and traditional bilingual classrooms. Rather than treating L1 use as an obstacle to English development, teachers can leverage the translanguaging space, whether student initiated (as in Group 5A) or teacher modeled (as in Group 5B), as a bridge between students' linguistic repertoires and the target language.
- Collaborative, role play, and game based activities (shown to generate peak motivation and oral production in Group 5B's Session 2) should be maintained as high engagement task types in bilingual pedagogy, regardless of whether AI tools are also present. These activities appear to generate qualitatively different forms of social motivation and linguistic risk taking than AI mediated tasks.

This triangulation ultimately affirms the study's foundational premise: AI should function as a pedagogical co-pilot that extends students' access to language and culture without displacing the corrective precision, cultural depth, and relational quality of human instruction.

The most effective bilingual learning environment documented here is neither the one with the most AI nor the one entirely without it, but the one in which AI tools, teacher expertise, intercultural task design, and collaborative engagement are purposefully integrated around shared cognitive and cultural learning goals.

Contributions to Bilingual Learning Environments

One of the central contributions of this research to bilingual learning environments lies in the reconceptualization of the teacher's mediating role within AI-integrated classrooms. Across the triangulated data sources, semi-structured interviews, non-participant classroom observations, and think-aloud protocols, a consistent and unambiguous pattern emerged: regardless of whether Artificial Intelligence (AI) tools were present, teacher mediation proved to be structurally indispensable to the quality of bilingual learning. This finding challenges simplified narratives that frame AI as a replacement for human instruction and instead positions it as a pedagogical copilot whose effectiveness is contingent on sustained, skilled teacher involvement.

In the AI integrated group (Group 5A), teacher presence was essential at every stage of the learning process: from modeling prompt formulation in the initial sessions, to redirecting students when conversational AI tools produced unexpected language switching or culturally shallow output, to verifying the accuracy of information retrieved from tools such as Google Gemini and Microsoft Copilot. Critically, the think-aloud protocol data revealed that AI tools, in their current conversational form, provided no corrective linguistic feedback whatsoever

across all six participating students. This universal absence of corrective feedback (coded as AI-3) means that without an active teacher to address grammatical, phonological, and morphological errors, the conditions for sustained metalinguistic development are substantially undermined. The practical implication for bilingual learning environments is therefore direct: AI integration does not reduce the necessity of teacher-mediated correction; it redistributes and redefines the forms that mediation must take.

This contribution is grounded theoretically in Vygotsky's (1978) concept of the Zone of Proximal Development (ZPD) and its application to bilingual contexts as articulated by Cummins (2021) and Wang (2023). Within the ZPD framework, cognitive growth in a second language requires the presence of a more knowledgeable agent who can scaffold learners beyond their current level of independent performance. In AI-integrated bilingual classrooms, the AI can serve as a partial mediating agent, offering vocabulary exposure, pronunciation modeling, and cultural information, but it cannot fulfill the corrective, relational, and culturally sensitive dimensions of scaffolding that a trained educator provides. The data from Group 5A's classroom observations illustrate this clearly: teacher intervention was the primary mechanism through which AI generated confusion was transformed into a productive learning moment, and through which students' growing autonomy in prompt formulation was shaped over the three observed sessions.

The contribution this finding makes to bilingual learning environments is therefore twofold. First, it offers an evidence-based framework for understanding what AI can and cannot do within bilingual pedagogical spaces. AI tools, particularly those with accessible transcripts and stable language output such as Gemini, can effectively lower affective barriers, expand

students' exposure to authentic target language input, and generate the kind of motivational novelty that sustains engagement across sessions. However, they cannot replace the immediate, linguistically precise, and contextually responsive feedback that characterizes expert bilingual teaching. Second, and more constructively, this research contributes to bilingual learning environments a model for rethinking the distribution of pedagogical functions between human teachers and AI systems. Rather than treating these as competing agents, the evidence calls for a complementary model in which AI handles input diversification and interactive cultural exposure while teachers retain ownership of corrective feedback, intercultural scaffolding, and metacognitive guidance. This reallocation of functions, grounded in evidence from a real Colombian bilingual classroom, constitutes a transferable pedagogical architecture for other institutions navigating the integration of AI into their bilingual curricula.

This study also contributes to bilingual learning environments a critical awareness of the importance of AI tool selection as a pedagogical decision, not merely a technological one. The differential outcomes observed between students using Gemini and those using Copilot, the former generating higher confidence and lower anxiety, the latter producing cognitive overload and affective difficulty in A2 level learners, demonstrate that not all AI tools are pedagogically equivalent in bilingual contexts. Bilingual educators and curriculum designers must therefore evaluate AI tools not only on the basis of their linguistic capabilities but on their accessibility, speech rate, cultural neutrality, and scaffolding affordances relative to the specific learner population. This represents a meaningful addition to the professional knowledge base of bilingual learning environment design, contributing practical and replicable criteria for ethical and effective AI integration.

This research leaves to the field of bilingual learning environments a clear and empirically supported proposition: effective AI integration is not a matter of substituting technology for pedagogy, but of designing bilingual spaces in which technology and teacher expertise are purposefully aligned around shared cognitive and cultural learning goals. The teacher's role does not diminish in an AI integrated bilingual classroom, it evolves, and its evolution demands deliberate professional preparation, critical tool literacy, and a sustained commitment to the corrective, intercultural, and relational dimensions of bilingual education that no algorithm has yet been able to replicate.

Within the Colombian context specifically, this research addresses a gap that national bilingual education policy has left largely unexamined. The Programa Nacional de Bilingüismo (Ministerio de Educación Nacional, 2004) established English proficiency as a strategic national goal, yet its implementation frameworks have focused primarily on measurable linguistic outcomes, with limited attention to the cognitive and cultural dimensions of language learning in technologically mediated environments. This study, conducted in a private institution in Facativá, generates empirical evidence from a context that reflects the conditions under which AI tools are already being adopted across Colombian bilingual schools, often without structured pedagogical guidance. The finding that AI integration, in the absence of deliberate intercultural scaffolding, tends to reproduce cultural homogenization rather than foster genuine intercultural development carries direct implications for how Colombian institutions interpret their bilingual mandate. As De Mejía (2006) has argued, authentic bilingualism in the Colombian context cannot be reduced to the acquisition of English as a prestige language; it requires the active recognition and cultivation of students' full linguistic and cultural identities. The data from both groups in this study confirm that this deeper dimension of bilingual education does not emerge

automatically from exposure to AI tools, it depends on the quality of pedagogical mediation surrounding them. This research therefore contributes to the Colombian field a concrete, evidence based argument for rethinking how AI is incorporated into bilingual curricula: not as a mechanism of linguistic standardization, but as a resource that, when guided by critically trained educators, can open spaces for intercultural reflection that the national bilingualism agenda has yet to fully address.

Situating AI within Classical Didactic Knowledge and Pedagogical Innovation in Educational Reform

AI and the Classical Didactic Knowledge of the Teacher

A further and necessary contribution of this research lies in positioning AI integration within the classical didactic knowledge framework that has historically defined the professional identity of the teacher. Classical didactics understands teaching as a structured, intentional act in which the teacher's professional authority resides not in technological access but in the irreplaceable capacity to adapt, mediate, and respond, cognitively, culturally, and affectively, to the particular needs of specific learners. This framework is the necessary lens through which the findings of this study must be interpreted. When Claro Narváez, Díaz Díaz, and Román Pérez (2024) argue that teacher training and professional development are critical when using AI, and that the teacher's role cannot be overshadowed in the integration of new technologies into bilingual learning, they are articulating precisely this classical premise: that the professional judgment of the educator is the irreducible foundation upon which any technological integration must rest.

The findings of this study confirm that premise empirically. The data reveal that AI tools such as Google Gemini and Microsoft Copilot were unable to provide corrective linguistic feedback (AI-3), failed to adapt to the A2 proficiency level of the learners, and produced cultural content that remained superficial in the absence of teacher scaffolding. These limitations are not incidental to the technology, they are structural. They reflect the fact that AI tools, in their current conversational form, lack the capacity to transform content into learnable forms for specific populations, to identify and repair learner errors in real time, and to respond with the cultural sensitivity and relational attunement that define professional bilingual teaching. As Wang (2023) establishes in her framework of cognitive dimensions in bilingual education, the mental processes activated through culturally situated language learning, metalinguistic awareness, cognitive flexibility, and intercultural reasoning, require a mediating agent who understands both the discipline and the learner. AI tools can stimulate those processes through input exposure and interactive practice, but they cannot orchestrate them with the precision that sustained bilingual development demands. It was not the AI tool that transformed learning moments in Group 5A; it was the teacher who knew when to intervene, how to reframe AI output, and how to use technological mediation as a bridge to deeper linguistic and intercultural engagement.

In classical didactic terms, AI therefore occupies the role of a sophisticated instructional resource, interactive and motivationally rich, but epistemically subordinate to the teacher's professional judgment. The data from Group 5A consistently illustrate this: AI tools engaged students' attention through novelty (E4: AI Novelty Effect) and provided extensive vocabulary and pronunciation exposure (codes A2, A4), but they failed to verify comprehension

systematically, provide corrective feedback (AI-3), or adapt to students' Zone of Proximal Development (ZPD) without deliberate teacher orchestration (Vygotsky, 1978; Cummins, 2021). The classical didactic sequence, from attention and engagement, through guided practice, corrective feedback, and consolidation was never fully activated by the AI alone; it was activated by the teacher working alongside the AI. This insight carries a direct implication for teacher preparation: training teachers to use AI tools technically is insufficient. Teachers must be trained to exercise their classical didactic judgment in relation to AI output, developing the capacity to evaluate the pedagogical quality of AI generated content, identify the gaps that AI leaves in the learning sequence, and design instructional responses that compensate for those gaps. The Colombian bilingual classroom studied here offers an empirical model of this expanded professional competence in action.

AI and Pedagogical Innovation in the Field of Educational Reform

Beyond its relationship to classical didactics, AI integration in bilingual classrooms must also be situated within the broader field of pedagogical innovation and educational reform. Colombia's Programa Nacional de Bilingüismo (Ministerio de Educación Nacional, 2004) established English proficiency as a national strategic goal, yet as this study demonstrates, and as De Mejía (2006) has long argued, authentic bilingualism in the Colombian context requires far more than measurable linguistic outcomes. It requires the active recognition of students' full linguistic and cultural identities, the development of intercultural competence, and the cultivation of the higher order cognitive processes that Wang (2023) identifies as central to meaningful bilingual education. Within this reform framework, AI has been widely presented

as a catalyst for personalization, equity, and 21st century competency development. This study's evidence both supports and critically qualifies that claim.

The data confirm that AI tools can contribute meaningfully to the goals of bilingual educational reform: they expand students' access to authentic target language input, support personalized interaction patterns, lower affective barriers for certain learners (AFF-1 in the TAP corpus), and generate the motivational novelty (E4) that the reform agenda associates with student centered learning. These findings are consistent with the empirical literature reviewed in this study, including Xia, Shin, and Kim's (2024) documentation of AI's capacity to deliver culturally adaptive, personalized bilingual instruction, and Ma's (2024) analysis of AI's role in enriching academic language learning and cultural sensitivity for bilingual students. In the Colombian reform context, where the Programa Nacional de Bilingüismo has prioritized proficiency outcomes but provided limited guidance on the pedagogical architecture through which those outcomes should be achieved, AI tools represent a concrete and accessible mechanism for enriching classroom practice.

However, the triangulated data also reveal a significant tension between the reform promise of AI-as-innovation and the pedagogical realities documented in this study. AI interaction in Group 5A remained predominantly reactive (AUT-2: Reactive Posture), intercultural engagement rarely exceeded the factual knowledge level of Byram's (1997) ICC model (CULT-4: Absence of Intercultural Engagement dominant in the TAP corpus), and critical linguistic self-monitoring (META-3: Self-Correction) was largely absent during AI mediated tasks. These findings demonstrate that AI does not automatically deliver the pedagogical innovations that reform agendas promise. As Edinoh, Onah, and Ogunbowale

(2025) observe in their analysis of AI in educational contexts, the transformative potential of AI depends critically on the institutional and pedagogical structures within which it is embedded, and where those structures are not deliberately redesigned to accommodate AI's specific affordances and limitations, technology adoption risks producing superficial change rather than genuine pedagogical renewal. In Group 5A, AI tools were integrated into an existing bilingual curriculum without restructuring the didactic sequence around the specific demands of conversational AI. The result was that the tools amplified existing strengths, vocabulary exposure, cultural curiosity, motivational engagement, without structurally addressing the gaps in corrective feedback, metalinguistic depth, and intercultural reflection that the reform agenda requires.

The contribution this research makes to pedagogical innovation in the context of bilingual educational reform is therefore both critical and constructive. Critically, it provides empirical evidence that AI integration does not constitute innovation in the absence of intentional pedagogical redesign. Constructively, it offers a model, grounded in the ZPD framework (Vygotsky, 1978; Cummins, 2021), the intercultural competence literature (Byram, 1997; Klimova & Chen, 2024), and the translanguaging tradition (García, 2009; García & Wei, 2014), for what that redesign must include: explicit prompt formulation pedagogy, AI complementary corrective feedback protocols, intercultural task design that moves learners from factual cultural knowledge toward the interpretive and reflective competencies of Byram's higher order ICC components, and translanguaging approaches that legitimize students' full linguistic repertoires. Crucially, none of these elements represent a departure from classical didactic knowledge; they represent its deliberate expansion into a new technological context. When AI integration is understood in this way as an expansion of, rather than a departure from,

the professional knowledge of the bilingual teacher, it becomes genuinely innovative: not because it replaces pedagogical expertise, but because it creates new conditions under which that expertise can operate with richer input, broader cultural reach, and greater personalization than traditional instructional settings alone can provide.

In sum, this study contributes to the field of bilingual learning environments a theoretically grounded and empirically supported argument for positioning AI neither as a revolutionary break from classical pedagogical tradition nor as a neutral technological supplement to unchanged instructional practice, but as a pedagogical resource whose transformative potential is realized only when it is deliberately integrated within the professional knowledge, ethical commitment, and institutional structures that define teaching as a human, culturally situated, and intellectually rigorous practice. The reform of Colombian bilingual education through AI integration is both possible and necessary. It is also demanding. It requires of bilingual teachers not less knowledge but more: not only the classical didactic expertise that has always defined the profession, but also the critical technological literacy to evaluate, guide, and humanize the tools that an increasingly AI mediated educational world places at their disposal.

APPENDIX

Classroom observation:

Observation 5A

https://usantotomaseduco-my.sharepoint.com/:w:/g/personal/jose_sanchezazu_usantotomas_edu_co/Eb-KcBBIOYFFldnbRN32PS8BuGE3ZPrh6PJyGl68ys8mOw?e=Aef5zG

Observations 5B:

https://usantotomaseduco-my.sharepoint.com/:w:/g/personal/jose_sanchezazu_usantotomas_edu_co/EfTnAwEcsMRKi0Xb9d2F9V4BEgnw_amkj6V1IYfvLZaDeg?e=7IeqmB

Think aloud protocols:

https://usantotomaseduco-my.sharepoint.com/:w:/g/personal/jose_sanchezazu_usantotomas_edu_co/EdUntPfjueFIhnpH7d9AEWgBnbJwtG1YgJchbwCsvtmeNw?e=l3NM58

Semi-structured interviews:

interview 5A:

https://usantotomaseduco-my.sharepoint.com/:w:/g/personal/jose_sanchezazu_usantotomas_edu_co/EZRZZZdreyRJh1nGo5o6qMABeUKvzMIh0Rkf9tdcBa8Ekw?e=oeanXE

Interview 5B:

https://usantotomaseduco-my.sharepoint.com/:w:/g/personal/jose_sanchezazu_usantotomas_edu_co/ERfzpGAztzREsxcxggCRm-QB9YuWChQhKRaqjqYMV128pA?e=qHnCJl

Classroom observation coding:

https://usantotomaseduco-my.sharepoint.com/:w:/g/personal/yuryyepes_usta_edu_co/IQCyNfuopcqGQKjK29halpBdAVrfGSkh_GTGrebjMzfHOtg?e=xYRea7

Semi structured interviews coding:

https://usantotomaseduco-my.sharepoint.com/:w:/g/personal/yuryyepes_usta_edu_co/IQD1TocekxKAQIqmaDdz5ZnXARg6wfnlKRrnTgfYPLN11bA?e=D48V1t

Think aloud protocols coding:

https://usantotomaseduco-my.sharepoint.com/:w:/g/personal/yuryyepes_usta_edu_co/IQAlA7RMfAu4S61bNg4rKcdMASCJ2VStmO4DcjXs34C_8XQ?e=t5pqKG

References

Angrosino, M. (2007). *Doing ethnographic and observational research*. SAGE. https://doi.org/10.4135/9781849208932

Baker, C., & Wright, W. E. (2017). *Foundations of bilingual education and bilingualism* (6th ed.). Multilingual Matters.

Barhoi, A., Gupta, L., Kumar, V., & Kaushik, S. (2025). Artificial intelligence (AI). In S. Kaushik & V. Kumar (Eds.), *Emerging trends in artificial intelligence and machine learning* (pp. 45–67). Springer.

Bialystok, E. (2018). Bilingual education for young children: Review of effects and consequences. *Annual Review of Applied Linguistics*, *38*, 81–95. https://doi.org/10.1017/S0267190518000020

Bowles, M. A. (2010). *The think-aloud controversy in second language research*. Routledge.

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

Charters, E. (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education Journal*, *12*(2), 68–82. https://doi.org/10.26522/brocked.v12i2.38

Chubaryan, A., & Vardanyan, M. (2024). Exploring the cognitive dimensions of language acquisition. *Armenian Folia Anglistika*, *20*(1), 13–24. https://doi.org/10.46991/AFA/2024.20.1.13

Claro Narváez, M. F., Díaz Díaz, C. D., & Román Pérez, K. J. (2024). *Incorporating artificial intelligence and gamification in EFL education: Strategies for enhancing teaching and learning outcomes of EFL teachers*. Universidad Santo Tomás. https://repository.usta.edu.co/items/a99cb423-3937-4036-a6fc-20f8db34ebb2

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge. https://doi.org/10.4324/9781315456539

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*(2), 222–251. https://doi.org/10.3102/00346543049002222

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Multilingual Matters.

Cummins, J. (2021). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Modern Language Review*, *77*(2), 143–152. https://doi.org/10.3138/cmlr.77.2.143

De Mejía, A.-M. (2006). Bilingual education in Colombia: Towards a recognition of languages, cultures and identities. *Colombian Applied Linguistics Journal*, (8), 152–168. https://doi.org/10.14483/22487085.176

Edinoh, K., Onah, A. C., & Ogunbowale, O. A. (2025). Exploring the challenges of artificial intelligence in basic education in Nigeria. Zenodo. https://doi.org/10.5281/zenodo.14849231

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

Fontana, A., & Frey, J. H. (1994). Interviewing: The art of science. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 361–376). SAGE.

García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Wiley-Blackwell.

García, O., Johnson, S. I., & Seltzer, K. (2017). *The translanguaging classroom: Leveraging student bilingualism for learning*. Caslon.

García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan. https://doi.org/10.1057/9781137385765

Guzman Johannessen, B. G. (2019). *Bilingualism and bilingual education: Politics, policies and practices in a globalized society*. Springer.

Hovy, D., & Spruit, S. L. (2016). The social impact of natural language processing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 591–598). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-2096

Kallio, H., Pietilä, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, *72*(12), 2954–2965. https://doi.org/10.1111/jan.13031

Klimova, B., & Chen, J. H. (2024). The impact of AI on enhancing students' intercultural communication competence at the university level: A review study. *Language Teaching Research Quarterly, 43*, 102–120. https://doi.org/10.32038/ltrq.2024.43.06

Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. SAGE.

Lee, H., & Warschauer, M. (2023). The effects of machine translation on vocabulary acquisition in second language writing. *Language Learning & Technology*, *27*(1), 45–62. http://www.lltjournal.org/item/1234

Lee, S. K. (2003). Exploring the relationship between language, culture, and identity. *GEMA Online Journal of Language Studies*, *3*(2), 64–76. https://journalarticle.ukm.my/731/

Ma, B. (2024). *The impact of artificial intelligence tools on bilingual students in U.S. education: A study on academic language-learning, cultural sensitivity, and inclusiveness* [Master's thesis, University of Washington]. University of Washington, College of Education.

Ministerio de Educación Nacional. (2004). *Programa Nacional de Bilingüismo Colombia 2004–2019: Inglés como lengua extranjera, una estrategia para la competitividad*. República de Colombia.

Naz, N., Gulab, F., & Aslam, M. (2022). Developing semi-structured interview guide: A systematic approach. *Corporate Social Responsibility and Sustainability Review*, *4*(1), 20–30. https://doi.org/10.26710/cssr.v4i1.170

Nguyen, T. H. (2022). The impact of AI tutors on code-switching in bilingual children. *Bilingualism: Language and Cognition*, *25*(4), 789–801. https://doi.org/10.1017/S136672892200045X

Pink, S. (2015). *Doing sensory ethnography* (2nd ed.). SAGE.

Ragin, C. C. (2014). *The comparative method: Moving beyond qualitative and quantitative strategies*. University of California Press.

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Stake, R. E. (1995). *The art of case study research*. SAGE.

United Nations Educational, Scientific and Cultural Organization. (2023). *Atlas of the world's languages in danger* (3rd ed.). UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000370465

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Wang, A.-L. (2023). *Understanding bilingualism, bilinguality, and bilingual education in an era of globalization*. IGI Global. https://www.igi-global.com/book/understanding-bilingualism-bilinguality-bilingual-education/292477

Xia, Y., Shin, S.-Y., & Kim, J.-C. (2024). Cross-cultural intelligent language learning system (CILS): Leveraging AI to facilitate language learning strategies in cross-cultural communication. *Applied Sciences*, *14*(13), Article 5651. https://doi.org/10.3390/app14135651

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.