WORKING PAPER N° IDB-WP-01728

Scaling Teacher Support: Piloting "Profe Gabi," an AI-powered Mentoring Chatbot for Novice Teachers in Chile

Gregory Elacqua
Catalina Hermosilla
Macarena Kutscher
Ana Teresa del Toro Mijares

Inter-American Development Bank
Education Division

December 2025

Scaling Teacher Support: Piloting "Profe Gabi," an AI-powered Mentoring Chatbot for Novice Teachers in Chile

Gregory Elacqua
Catalina Hermosilla
Macarena Kutscher
Ana Teresa del Toro Mijares

Inter-American Development Bank
Education Division

December 2025

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Scaling teacher support: piloting "Profe Gabi," an AI-powered mentoring chatbot for novice teachers in Chile / Gregory Elacqua, Catalina Hermosilla, Macarena Kutscher, Ana Teresa del Toro Mijares.
p. cm. — (IDB Working Paper Series ; 1728)
Includes bibliographic references.
1. Education-Effect of technological innovations on-Chile. 2. Computer literacy-Chile. 3. Artificial intelligence-Educational applications-Chile. I. Elacqua, Gregory M., 1972 - II. Hermosilla, Catalina. III. Kutscher, Macarena. IV. Del Toro Mijares, Ana Teresa. V. Inter-American Development Bank. Education Division. VI. Series.
IDB-WP-1728

JEL Codes: O33, I21, I25, M53

Keywords: Teacher mentoring, novice teachers, artificial intelligence, chatbots, professional development, scaling, education technology, Chile

http://www.iadb.org

Scaling Teacher Support: A Pilot of "Profe Gabi," an AI-Powered Mentoring Chatbot for Novice Teachers in Chile*

Ana Teresa del Toro Mijares† Gregory Elacqua‡ Catalina Hermosilla§
Macarena Kutscher¶

December 2, 2025

Abstract

Human-intensive programs like teacher mentoring face a fundamental scaling problem: evidence shows their impact declines as they expand, and the personalized support that makes them effective becomes prohibitively expensive at scale. This paper explores whether AI-powered tools can help resolve this tension. We document the development and pilot implementation of "Profe Gabi," a generative AI chatbot designed to provide novice teachers in Chile with on-demand mentoring support through WhatsApp. The chatbot delivers personalized guidance across three dimensions—pedagogical practice, socio-emotional well-being, and career development—drawing on Chile's existing National Induction and Mentoring System, which currently reaches fewer than 1% of eligible teachers due to capacity constraints. During an eight-week pilot with 550 novice teachers, Profe Gabi demonstrated the technical feasibility of scaling mentoring support: it reached more than three times the number of teachers served by the national program while maintaining high user satisfaction. However, engagement declined significantly, revealing critical implementation challenges around sustained adoption. We analyze these patterns to identify what conditions may be necessary for AI tools to meaningfully expand access to professional development, and discuss insights that inform Profe Gabi's scale-up in 2025-2026. Our study contributes to an emerging literature on AI technologies for teacher support and offers lessons for similar interventions globally.

JEL classification: O33, I21, I25, M53

Keywords: Teacher mentoring, novice teachers, artificial intelligence, chatbots, professional development, scaling, education technology, Chile

*We gratefully acknowledge the Inter-American Development Bank for funding this research. The opinions expressed are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent. We thank the teams at the Ministry of Education of Chile, including the Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas and the Centro de Estudios, as well as the teachers and mentors who contributed through surveys, focus groups, or pilot participation. We are also grateful to Graciela Pérez-Núñez for contributions in the early stages of this project, Alexis Montecinos and Diego Salvatierra for advice on technological development, and an anonymous reviewer for comments on an earlier draft. The authors have no conflicts of interest. All errors are our own.

†Inter-American Development Bank. E-mail: anade@iadb.org

‡Inter-American Development Bank. E-mail: gregorye@iadb.org

§Email: cata.hermosilla98@gmail.com

¶Universidad del Desarrollo. E-mail: mkutscher@udd.cl

1 Introduction

Teachers around the world face significant challenges at the start of their careers, including inadequate preparation and training for the complexities of classroom teaching - often due to insufficient practical classroom experience during pre-service education; insufficient support, supervision and mentoring at the school level; and overwhelming workloads (Darling-Hammond et al., 2009; Kardos and Johnson, 2010; OECD, 2019a). In Latin America, these difficulties are often exacerbated by the fact that novice teachers are disproportionately assigned to the most demanding educational environments—typically schools in disadvantaged and rural regions that struggle to attract experienced, certified educators (Elacqua et al., 2022; Bertoni et al., 2020). This results in high attrition rates: in some places, the attrition rates among teachers at the start of their careers are as high as those near retirement. In countries like the United States, the United Kingdom, and Canada, almost 40% of teachers leave the profession within the first five years of teaching (UNESCO, 2024).

International evidence shows that teacher turnover and early attrition disrupt learning, negatively impact student outcomes, and waste investments in teacher recruitment, training, hiring and deployment (Hanushek et al., 2004; Ronfeldt et al., 2013; Boyd et al., 2008; Carver-Thomas and Darling-Hammond, 2019). This is particularly concerning in the current context of widespread teacher shortages: approximately 44 million additional teachers are needed worldwide to meet the education targets set out in the 2030 Sustainable Development Goals, with 3.2 million of them specifically in Latin America and the Caribbean (UNESCO, 2024). The premature departure of novice teachers worsens the teacher shortage, contributing to a cycle of understaffing and over-reliance on inexperienced educators, which in turn further strains education systems and widens learning gaps. Due to the high costs of teacher recruitment, training, hiring and deployment, reducing early-career attrition is a key component of building a stable and high-quality teaching workforce in a cost-effective manner (UNESCO, 2024).

Education systems have adopted various strategies to retain novice teachers and diminish early attrition, such as increasing pay, improving workplace conditions, offering non-monetary incentives like career advancement opportunities, and providing professional development tailored to the needs of novice teachers that supports them to build their instructional skills and navigate early-career challenges (Elacqua et al., 2018). Induction and mentoring programs, one such form of professional development for early-career educators, have been shown to improve instructional effectiveness, student achievement and teacher retention (Keese et al., 2023; Naylor et al., 2019). Though promising, these programs face considerable challenges in terms of scalability, cost, and implementation capacity.

In this paper, we describe the development and pilot implementation of "Profe Gabi," an AI-powered chatbot designed to support the professional development of novice teachers in Chile. Developed in a joint effort between the Chilean Ministry of Education (MINEDUC) and the Inter-American Development Bank (IDB), Profe Gabi is a WhatsApp-based virtual mentoring assistant that uses generative AI to deliver personalized, timely, and high-frequency support across three domains: pedagogical practice, socio-emotional well-being, and career guidance. The initiative aims to address Chile's critical novice teacher attrition problem: between 2004 and 2023, 22% of teachers left the profession after their first year, and 14% after their second. Novice teachers are three times more likely to leave than their more experienced peers.[1] This challenge

[1] Own calculations based on official data from the Ministry of Education, Chile, 2004–2023.

is compounded by the fact that Chile currently faces a teacher shortage equal to approximately 123,789 hours for qualified teachers (6.2%) and 235,602 hours for specialist teachers (11.7%) (Zúñiga et al., 2025).

Profe Gabi builds on Chile's National Induction and Mentoring System, launched by MINEDUC in 2017 to improve support for early-career teachers and reduce attrition. The program pairs novice teachers in their first or second year with experienced colleagues who act as mentors throughout the school year, focusing on pedagogical development and professional growth. A recent qualitative evaluation carried out by the United Nations Development Program (UNDP) found that participants highly value the mentoring experience, crediting it with increased teaching confidence and stronger relationships between colleagues. However, scaling-up the program has proved difficult: due to institutional capacity constraints and high implementation costs, it has reached fewer than 1% of eligible teachers (UNDP, 2023). Profe Gabi was developed as a cost-effective and scalable complement to the existing National Induction and Mentoring System, aiming to reduce early-career attrition and strengthen teacher retention and student learning.

The need for cost-effective and scalable approaches such as Profe Gabi becomes clear when examining the challenges behind traditional induction and mentoring programs for novice teachers. Although these programs have shown promise in supporting novice educators, they are often limited in reach as they require significant resources (including time, personnel, and funding), making large-scale implementation challenging. Each additional mentor or coach adds substantial expense with few economies of scale, creating high marginal costs as programs expand. In addition to high costs, these programs face human capital constraints, as they depend on the availability of highly skilled and experienced teachers to serve as mentors, who must be identified, recruited, trained, and prepared. Notably, in OECD countries, 40% of novice teachers indicate having participated in some type of induction program, but only 22 percent report having been assigned a mentor (UNESCO, 2024; OECD, 2019b).

Even when these types of programs are able to reach more teachers, maintaining effectiveness can be difficult. As programs scale, the dosage and frequency of support provided to teachers is often reduced due to increased costs. Moreover, preserving implementation fidelity and delivery quality becomes more challenging, and mentors are only able to provide highly personalized support for a certain number of teachers (Kraft et al., 2018; World Bank, 2023). These issues mean that even programs that are effective at a smaller scale can see an important reduction in impact when expanded. For example, evidence shows from teacher coaching programs, which share multiple characteristics with teacher induction and mentoring because they also depend on highly specialized human support, shows that those that serve more than 100 teachers are less effective than smaller ones in improving instructional quality and in increasing student learning (Kraft et al., 2018). Scaling problems are not unique to teacher induction, mentoring, and coaching programs. As highlighted by (List, 2022), promising findings in experimental settings often yield considerably smaller effects when attempts are made to implement them on a larger scale: the so-called "voltage drop."

In this context, emerging technologies have significant potential to amplify the scale of interventions without compromising their effectiveness, because they can help promote standardization, correct dosage, and implementation fidelity (Al-Ubaydli et al., 2019). In particular, recent breakthroughs in generative AI that leverage Natural Language Processing (NLP) can emulate human interactions, providing high-frequency, on-demand support that is available anytime,

and standardizing delivery of the support with high fidelity in every interaction (Mollick, 2024). Chat agents powered by generative AI have proliferated across various sectors, including healthcare, therapy, personal finance, and consumer services (Hoffman and Beato, 2025; Laymouna et al., 2024; Li et al., 2023; Adam et al., 2021).

AI technologies are already transforming education and are poised to further revolutionize the sector, with direct applications for students, teachers, school and system leaders, administrators, and even parents and caregivers (Arias Ortiz et al., 2025; Molina et al., 2024; Khan, 2024). One way in which AI is proliferating in the education sector is in the form of AI-powered chatbots, which are being developed and used to support students, teachers, tutors, school and system leaders, and parents. These chatbots serve diverse purposes, from teaching and building specific skills, providing guidance and information on specific topics, expanding access and enhancing system efficiency (Wollny et al., 2021). For students, recent work shows that chatbots using NLP are being used to provide homework support, personalize learning, and help develop targeted competencies, among other uses (Labadze et al., 2023; Henkel et al., 2024; De Simone et al., 2025).

In contrast, the use of AI technologies to support teachers remains less explored, particularly in terms of applications focused on teachers' professional development. Recent data shows that teacher use of LLMs is quite varied: on average, 36% of lower secondary teachers in OECD countries report using AI in the past year, but this number rises to 56% and 55% in Brazil and Chile, respectively, and is as low as 14% in France (OECD, 2025). While information how teachers use AI is more limited, the available data shows that K-12 educators around the world use AI to develop lesson plans and supporting teaching materials; to brainstorm new teaching materials; to simplify complex topics for students and to improve the student experience (Microsoft, 2025). Different AI applications and tools have been developed to support educators on specific tasks, such as developing instructional content, delivering specific teaching content, providing tutoring to students, providing student feedback, or automating certain instructional tasks such as grading (Labadze et al., 2023; Kuhail et al., 2023; Wang et al., 2024; Gamberini et al., 2025). However, few of these applications specifically seek to target teacher learning and growth, particularly for full-time teachers. A recent systematic review found that while 65% of studies examined AI for instructional purposes, only 35% explored its potential for enhancing educator professional development (Tan et al., 2024). Notable exceptions include Demszky and Liu's (2023) assessment of M-Powering Teachers, which uses NLP to provide teachers with automated feedback on their instructional practice (Demszky and Liu, 2023), and early work exploring ChatGPT's potential for teacher coaching (Wang and Demszky, 2023). Another recent example consists of a small-scale pilot in Sierra Leone that tested TheTeacher.AI, an AI-powered mentoring chatbot used by 193 teachers throughout the school year for lesson planning, classroom management, and content support (Björkegren et al., 2025).

We contribute to this emerging literature by documenting the development and pilot implementation of Profe Gabi in Chile. By using generative AI to deliver personalized, high-frequency support for pedagogical practice, socio-emotional well-being, and career guidance, Profe Gabi offers a scalable and cost-effective model for professional development. Our study examines how AI-powered solutions can empower teachers, especially those at the start of their careers, and addresses key challenges in scaling traditional teacher support systems.

The paper is structured as follows. Section 2 documents the Profe Gabi development process, detailing the steps taken in its conceptualization, design, technological set-up, training, and

testing phases. Section 3 describes the pilot implementation carried out in 2024 and presents the results gathered from platform usage and engagement data, surveys, and focus groups with participating teachers. Section 4 discusses the pilot results, highlighting opportunities for impact and main challenges. Section 5 presents recommendations for Profe Gabi in Chile, and provides reflections on lessons learned on the use of similar technologies to support novice teachers in the region and globally.

2 Design of Profe Gabi

Profe Gabi was developed over a six-month period from May to October 2024. This section describes the phases of user research, definition of key design specifications and chatbot design, and development of the content bank in detail. For further information on Profe Gabi's technological set-up as well as the testing and quality assurance (QA) carried out prior to launch, please see Annex A.

2.1 User Research

The design process started with research to understand the structure, methodology, and content of the mentoring model employed in the existing National System of Induction and Mentoring program; the needs of novice teachers in Chile and the most significant challenges they face in the first two years of teaching; and their relationship with technology. The team used multiple qualitative and quantitative instruments to triangulate data, including focus groups with current mentors, a survey of novice teachers enrolled in the National System of Induction and Mentoring, and in-depth semi-structured interviews with a sample of the novice teachers.

To begin, four focus group sessions were carried out with senior teachers mentoring novice colleagues through the National System for Induction and Mentoring.[2, 3] These sessions aimed to understand the structure, methodology, and content of the mentoring provided under the existing system, as well as typical challenges in the mentoring process and areas where the model could be strengthened. Mentors worked within a structured program of duplas, or mentor-mentee pairs (4-6 hours per week over 10 months), developing personalized plans aligned with a Ministry-provided framework. Communication between duplas primarily took the form of in-person meetings, emails, and WhatsApp exchanges. Mentors reported gaps between university training and classroom realities, highlighting the areas where they provided the most support: lesson planning, classroom management, student motivation, and self-care. Mentors also mentioned offering broader professional guidance regarding Chile's teacher evaluation system and the implications of various regulatory frameworks, and helping novice teachers navigate school-level administrative requirements and institutional protocols.

The design team then gathered data on novice teachers' experiences and the areas where they

[2]Mentors must hold a teaching degree, currently perform classroom teaching duties, have achieved either the Advanced or Expert level in Chile's teacher career ladder, and have completed a certification course at an accredited university to develop their mentoring skills

[3]The focus groups included eight experienced educators, predominantly women (88%), with teaching experience ranging from 12 to 29 years. Half of the participants specialize in mathematics, while the rest teach English, language arts, natural sciences, art, and technology, offering diverse perspectives across subjects and educational levels.

required additional support during their first two years of teaching. A comprehensive online survey was distributed to all 172 novice teachers participating in MINEDUC's mentoring program in 2024, with a response rate of 61% (105 completed responses). The findings were triangulated and further explored using data from semi-structured interviews. Novice teachers consistently highlighted that they strongly valued the support provided by their mentors, emphasizing the socio-emotional connection and trust that grew throughout the school year and the importance of mentoring conversations, seen as a safe space to discuss the challenges they navigated during the school year. The most significant difficulties novice teachers reported facing in their day-to-day work include communication with parents and families, student behavior management, and balancing pedagogical planning with workload demands.

In the survey and interviews, novice teachers frequently voiced the need for additional support in the pedagogical and instructional domain, particularly classroom management, strategies to improve student motivation, pupil assessment, and adapting instructional strategies to diverse contexts and learners. They also desired more assistance in general time management, stress management, administrative task management, and integrating into their school community. A recurrent topic was the challenge of balancing their students' emotional needs with their own well-being, with burnout and work-life boundaries being major concerns, as well as managing relationships with parents and families, which require a careful mix of assertiveness and empathy.

The instruments also sought to explore teachers' relationship with technology, particularly AI-powered chatbots. Most of the survey respondents reported being familiar with virtual assistants in the form of chatbots (47 percent said that they had good knowledge of virtual assistants and 41 percent had some familiarity), with 86 percent expressing enthusiasm at the idea of receiving AI-powered support. Teachers anticipated using Profe Gabi for quick queries and weekly assistance, though also expressed concerns over accuracy, contextual relevance, and the emotional limitations of AI compared to human mentors.

Several key takeaways can be drawn from these research findings. The human connection and personalized guidance provided by experienced mentors cannot be replaced by AI technology. That said, the various difficulties reported by novice teachers reveal the potential of an AI-powered mentoring solution to complement the existing program, particularly for those who might otherwise lack access to formal mentoring. Given the diverse needs reported by novice teachers, an AI tool offers the possibility of responding to queries across multiple dimensions in a personalized manner. Finally, while a high proportion of novice teachers said they had some or good familiarity with virtual assistants (88 percent) and were enthusiastic about receiving AI-powered support (86 percent), they also expressed some concern as well, underlining the importance of designing a chatbot that fosters confidence, trust, and accessible use among the intended users.

2.2 Key Design Specifications

Profe Gabi's goal is to provide timely, relevant, and context-sensitive guidance to support novice teachers in Chile. It was developed to complement the existing National Induction and Mentoring System, which currently serves only 1% of the eligible novice teacher population, offering a resource for the many teachers without access to formal mentoring.

To increase accessibility, Profe Gabi operates through WhatsApp, a platform already widely used within Chile's mentoring program for quick consultations and immediate support. Both

survey and interview findings confirmed the prevalence of WhatsApp as a familiar and reliable communication tool for teachers. This means that users can engage with Profe Gabi without having to download additional applications or create a user account on a different site, significantly lowering the barrier to access support. Additionally, recognizing the need for on-demand support, the chatbot is available around the clock, enabling teachers to access information and guidance whenever needed, beyond traditional working hours.

By design, Profe Gabi aims to provide support in three dimensions aligned with Chile's National Induction and Mentoring System: (1) pedagogical practice, (2) socio-emotional well-being, and (3) career guidance. These dimensions reflect insights gathered from research conducted with novice teachers and mentors, ensuring relevance to real-world classroom challenges. Users interact with Profe Gabi using natural language to receive personalized support based on their specific queries. Annex B presents a detailed breakdown of the content repository structured around the three dimensions of support.

Profe Gabi drew from a structured and organized content bank containing official MINEDUC materials, regulatory documents, and resources from the National Induction and Mentoring System (see Subsection 2.3 for more information on how this content bank was collected, organized, and validated). Use of a Retrieval Augmented Generation (RAG) architecture with semantic search capabilities meant that Profe Gabi could retrieve contextually relevant information even when teachers' queries didn't exactly match the terminology used in official documents, ensuring accurate and relevant responses that aligned with the MINEDUC's educational policies, standards, and programs.

Another key design element of Profe Gabi is a weekly conversation-inducing message, whereby the chatbot proactively sends a message to users related to one of the three dimensions. This feature serves as a gentle reminder that Profe Gabi is available to support them, and as a means of nudging them to continue using the technology. The messages were developed in direct coordination with the MINEDUC team responsible for the National Induction and Mentoring System program, ensuring alignment with official mentoring strategies and reinforcing key aspects of teacher support. Annex C provides the conversation-inducing messages sent on behalf of Profe Gabi during the pilot duration.

Profe Gabi is intended to be a friendly and approachable persona. The chatbot's name pays homage to Gabriela Mistral, the renowned Chilean educator and Nobel Prize-winning poet from the 1950s, reinforcing its identity as a supportive and knowledgeable guide. A cartoon avatar was developed to visually represent Profe Gabi, which exudes warmth and expertise.

2.3 Content and Training Data

The content base of Profe Gabi integrates multiple sources to ensure comprehensive, accurate, and relevant support for novice teachers. The core set of materials consists of official documents from the National Induction and Mentoring System, which provides structured guidance for novice teachers. We organized these materials around the three main dimensions of support: pedagogical practices, instruction, and classroom management; socio-emotional well-being and motivation; and career and professional development. This framework is further reinforced by essential educational regulations and guidelines, including educational laws and regulations, national curriculum guidelines (Bases Curriculares), and the Framework for Good Teaching (Marco para la Buena Enseñanza). The content base was further enriched with specialized

materials from key ministerial units, such as the Curriculum and Assessment Unit and the Educational Reactivation Unit.

We also incorporated a set of weekly reports and logs from mentor-mentee meetings carried out under the National Induction and Mentoring System (with permission). These logs capture authentic interactions, practical support strategies, and solutions to common challenges faced by novice teachers. By integrating these real-world experiences, the chatbot's knowledge base gained greater depth and authenticity, ensuring its responses reflected proven, effective practices within the Chilean educational context.

A systematic approach informed by research findings was used to organize this content. By analyzing survey data, focus groups, and interviews, we identified the most frequently asked questions and common concerns among novice teachers. These questions were then carefully mapped to relevant source documents, creating a prioritized content structure ensuring that the most essential information was readily retrievable in Profe Gabi's response generation.

3 Pilot Implementation of Profe Gabi

3.1 Intervention Design

Profe Gabi was piloted with a group of novice teachers for eight weeks from October through December 2024. The aim was to gather feedback from real users regarding the chatbot's strengths and areas for improvement.

Throughout the 8-week pilot, participants had access to Profe Gabi. As part of this access, they received weekly conversation-inducing WhatsApp messages (see Annex C for examples) on topics related to their pedagogical practice, mental health and well-being, and career development. The weekly messages were designed to pique participants' interest and motivate them to interact with Profe Gabi. They could do so using natural language on topics related to the weekly message or any other subject relevant to their activities and needs as novice teachers. Profe Gabi was available day and night, such that they could engage with the chatbot at any time without restriction.

Beyond the weekly messages, pilot participants received links to two surveys (halfway through the pilot and at its conclusion) asking for feedback on their experience using Profe Gabi. The survey results are summarized in Subsubsection 3.3.4.

3.2 Participant Demographics and Representation

A total of 8,493 novice teachers in their first or second year of teaching were contacted by email via the CPEIP and invited to participate in the Profe Gabi pilot, and 550 novice teachers (6.4% of those invited) signed up. Those who expressed interest in participating completed a brief survey providing basic identifying information, and consented to be contacted via WhatsApp for the pilot. These teachers were aware that the new technology was under development.

Approximately 64% of the pilot group were teachers at public schools, and the remaining 36% at private subsidized schools.[4] The average age was 29.87 years, and pilot participants had

[4]Teachers in private non-subsidized schools were not invited to participate in the pilot, as these establishments

1.58 years of teaching experience. Thirty-six percent of the participants were based in Santiago de Chile's metropolitan area, and 13% worked in rural schools. Fifty-nine percent were teachers at the primary level; 38% at the secondary level; and 3% in special education. Sixty-nine percent were female. Table 1 provides further descriptive data on pilot participants. Although efforts were made to ensure the pilot group was diverse and representative of novice teachers in Chile, given the recruitment method, participants were likely more digitally savvy and open to new technologies than typical novice teachers in Chile. [5]

do not receive public funding and are therefore ineligible for the National Induction and Mentoring Program.

[5]A comparison between the pilot sample and the universe of novice teachers in Chile (N=10,672) reveals several notable differences. The most substantial disparity appears in school type distribution, with public school teachers significantly overrepresented in the pilot (64.0% versus 44.0% in the universe, $p<0.05$). This over-representation might be attributable to data sourcing mechanisms, as contact information obtained through the Ministry of Education likely contains more accurate and up-to-date records for public school teachers due to administrative processes. School subject distribution also sees pronounced differences, with language teachers substantially overrepresented (23.0% versus 9.0%, $p<0.05$) and physical education (5.0% versus 12.0%, $p<0.05$) and special education subject teachers (1.0% versus 6.0%, $p<0.05$) notably underrepresented. Additionally, teachers from the Santiago Metropolitan Region are moderately overrepresented (36.0% versus 30.0%, $p<0.05$). Other demographic characteristics such as gender composition and distribution across most educational levels show no significant differences between the pilot and the universe of novice teachers.

Table 1: Sample Characteristics of Chilean Teachers in Profe Gabi Pilot

Variable	N	Mean	Std. Dev.	Min	Max
Demographics					
Proportion of women	550	0.69	0.46	0	1
Age (2024)	550	29.87	6.98	23	60
Years of experience	550	1.58	0.49	1	2
School Type (2024)					
Public	550	0.64	0.48	0	1
Subsidized private	550	0.36	0.48	0	1
School Characteristics (2024)					
Contract hours	550	36.11	8.80	4	44
Metropolitan Region	550	0.36	0.48	0	1
Rural location	550	0.13	0.34	0	1
Education Level Taught					
Primary education	550	0.59	0.49	0	1
Secondary education	550	0.38	0.49	0	1
Special education	550	0.03	0.17	0	1
Subject Taught (2024)					
Language arts	550	0.23	0.42	0	1
Mathematics	550	0.12	0.33	0	1
Sciences	550	0.07	0.26	0	1
Social sciences	550	0.10	0.30	0	1
Physical education	550	0.05	0.22	0	1
Arts	550	0.05	0.22	0	1
Religion	550	0.02	0.13	0	1
Technology	550	0.01	0.10	0	1
Foreign languages	550	0.06	0.24	0	1
Technical/vocational	550	0.001	0.04	0	1
Special education	550	0.01	0.12	0	1
Other subjects	550	0.26	0.44	0	1

Notes: Sample includes 550 Chilean teachers participating in the Profe Gabi pilot program. Proportions are reported for binary variables.

3.3 Pilot Results

The pilot included 546 novice teachers (550 initially volunteered, with four withdrawing in the first three weeks). Teachers had around-the-clock access to Profe Gabi and received weekly messages encouraging interaction with the chatbot.

Results from the pilot were analyzed across four sets of questions:

1. Initial engagement and retention: To what extent were teachers willing to initially engage with Profe Gabi, and how did participation levels change over the course of the pilot?

2. Interaction patterns and usage: What sort of patterns emerge in terms of frequency of use, duration, and messages?

3. Query types and content analysis: How did teachers utilize Profe Gabi across the three core support dimensions (pedagogical practice, socio-emotional well-being, and career guidance), and what topics or themes were most prominent in their interactions?

4. User satisfaction and feedback: What were teachers' overall experiences with Profe Gabi, including satisfaction levels, perceived barriers to engagement, and recommendations for future improvements?

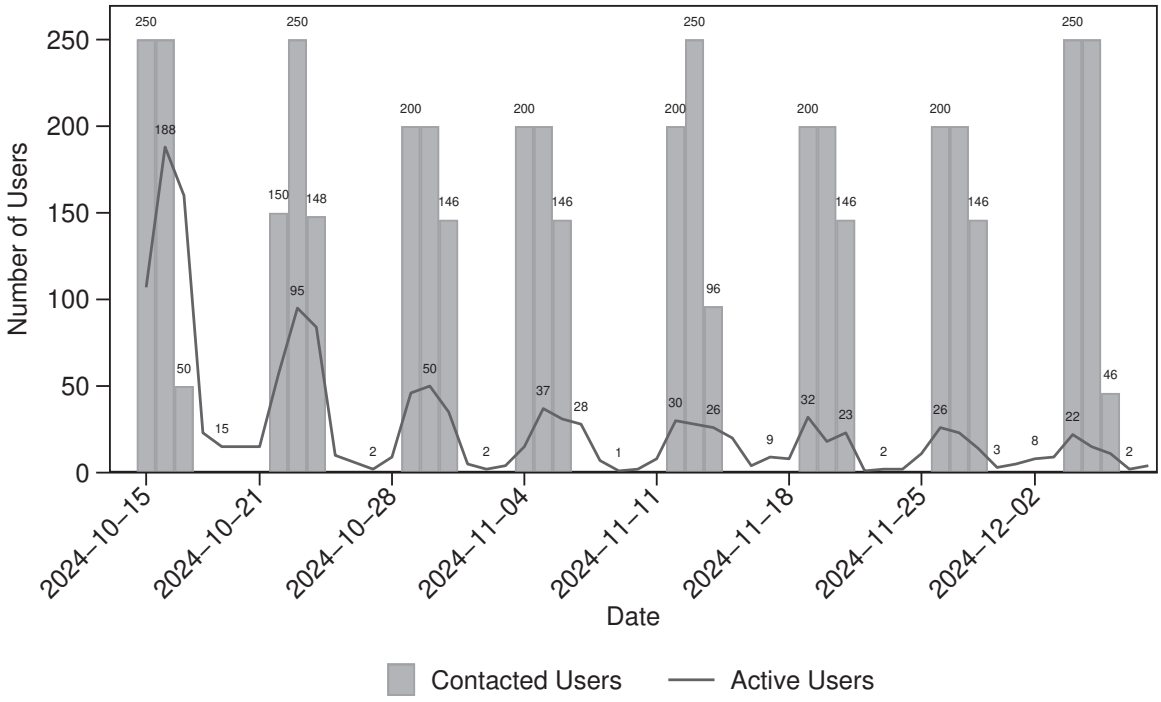
3.3.1 Engagement Patterns and Retention

The first engagement metric assessed is the number and percentage of active users who received a message from Profe Gabi and chose to engage in a conversation with the chatbot. As seen in Figure 1, engagement was high at the beginning of the pilot, with 435 teachers (79% of the sample) interacting with Profe Gabi during week 1, either by selecting response buttons or initiating natural language conversations. However, this declined significantly throughout the pilot. In week 2, only 230 teachers (42%) interacted with Profe Gabi, and in week 4, this number dropped further to 96 teachers (17%). By the end of the pilot, in weeks 7 and 8, only 60 (11%) and 47 (8.6%) of the teachers, respectively, used Profe Gabi.

The follow-up survey (see Section 3.3.4) sought to uncover the reasons behind this declining engagement. A cohort analysis further showed that the timing of initial engagement translated to significantly different long-term participation. For the 435 teachers who first interacted with Profe Gabi in Week 1, retention dropped to 45 percent in week 2 and continued to decline, reaching as low as 9 percent by week 8. In contrast, the cohort of 36 teachers who didn't use the chatbot until week 2 had lower overall retention, with only 25 percent remaining active the following week, and just 8 percent the week thereafter.

An analysis of week 1 interaction data also offers insights into user preferences for engagement mechanisms. During the initial week, Profe Gabi's conversation-inducing message included quick response buttons while still allowing teachers to reply using natural language. Among the 435 active teachers, 162 (37%) exclusively used quick response buttons, engaging only at a basic level. Meanwhile, 266 teachers (61%) initially interacted using buttons but later transitioned to natural language, suggesting deeper engagement. Finally, 7 teachers (2%) relied solely on natural language from the outset.

Figure 1: Profe Gabi Daily Engagement: Messages Sent vs Teacher Interactions



3.3.2 Interaction Patterns and Usage

A second dimension of analysis focuses on the pattern of teacher interaction with Profe Gabi. As shown in Table 2, teachers engaged in an average of 2.97 conversations with the chatbot throughout the pilot. Most participants had limited engagement (Q25 = 1.00, Q50 = 2.00, Q75 = 3.00), though a small subset of users engaged much more frequently, with some having over 15 conversations. Each conversation contained an average of 6.46 messages, with users and the chatbot maintaining a balanced 1:1 message exchange ratio. When measurable, conversations lasted an average of 39.4 minutes, though this varied considerably (median = 3.35 minutes), indicating both quick consultations and extended working sessions. Teachers interacted with the chatbot relatively infrequently: on average 0.35 active days per week (or approximately one day every three weeks). When they did engage, they sent an average of 4.25 messages per week and 2.82 messages per active day.

The analysis reveals distinct user engagement patterns. Single-use users (35.9% of participants) engaged in one conversation averaging 5.05 messages, while power users (0.6% of participants) sustained over 15 conversations each, with significantly longer interactions averaging 12.41 messages per conversation. This segmentation demonstrates natural scaling where more committed users developed progressively deeper engagement patterns while maintaining the balanced dialogue structure observed across all user types.

Table 2: Descriptive Statistics of Conversation-Level Interactions

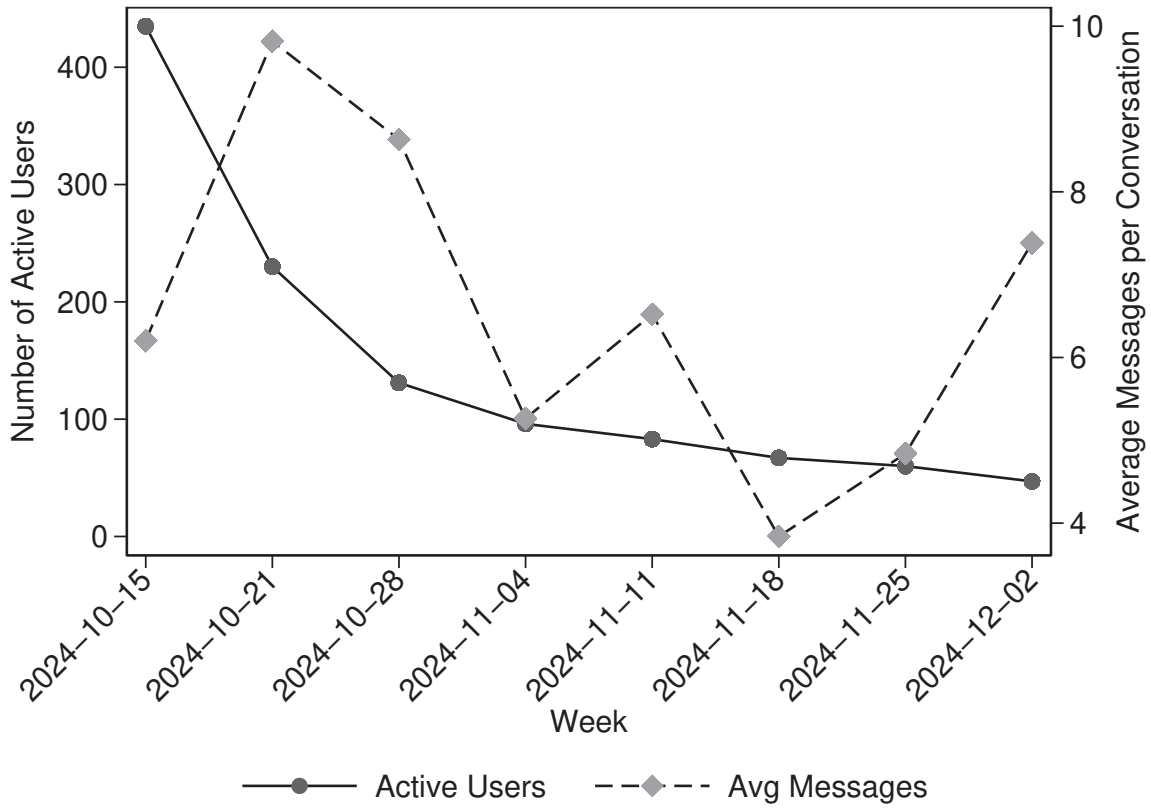
Statistic	Mean	SD	Q25	Q50	Q75	Q90
Conversation-Level Metrics						
N° of conversations per user	2.97	2.84	1.00	2.00	3.00	6.00
N° of messages per conversation	6.46	6.97	2.00	4.00	8.00	14.00
N° of user messages per conversation	3.23	3.49	1.00	2.00	4.00	7.00
N° of bot messages per conversation	3.23	3.49	1.00	2.00	4.00	7.00
Conversation duration (minutes)*	39.49	86.36	1.03	3.35	20.52	142.18
User Activity Patterns						
N° of active days per user	2.68	2.60	1.00	2.00	3.00	6.00
N° of active days per week	0.35	0.33	0.13	0.25	0.38	0.63
Messages per conversation by activity level:**						
Single-use users (35.9%)	5.05	2.89	3.00	4.00	6.00	8.00
Regular users (39.2%)	5.07	2.73	3.33	4.50	6.00	8.50
Committed users (19.4%)	6.91	4.21	4.00	6.00	8.50	12.00
Highly active users (4.9%)	8.19	4.89	5.00	7.00	10.00	15.00
Power users (0.6%)	12.41	8.95	6.00	10.00	16.00	24.00

Notes: A conversation is defined as all messages exchanged between a user and Profe Gabi within a single day.
* Duration calculated only for conversations with measurable time intervals (67.6% of total conversations).
** User segments based on total number of conversations: Single-use (1), Regular (2-3), Committed (4-7), Highly active (8-15), Power users (>15).

In a closer analysis of weekly engagement metrics, we also see significant variation in the length and duration of interactions with Profe Gabi, as measured by the number of messages exchanged between users and the chatbot (Figure 2). While engagement fluctuated from week to week, conversations typically ranged from 4 to 10 messages per user, resulting in a total exchange of 8 to 20 messages between the user and Profe Gabi.

Notably, conversation length peaked in weeks 7 and 8, averaging 5.0 and 7.4 messages per user, respectively. This increase in conversation depth during the later weeks occurred despite the overall decline in the number of active users, suggesting that by this stage of the pilot, a smaller and more engaged group of users remained active. These users likely felt more comfortable using the technology and were more adept at interacting effectively with Profe Gabi.

Figure 2: User Engagement: Active Users and Average Messages per Conversation



3.3.3 Query Types and Content Analysis

A third dimension of analysis consists of the types and purposes of the queries posed to Profe Gabi. In other words, how and for what sort of support teachers used the technology. Notably, novice teachers who interacted with Profe Gabi sought help across all three of the core dimensions the chatbot was designed to address: pedagogical practice, socio-emotional well-being and motivation, and career and professional guidance, as shown in Table 3.

Within the pedagogical practice dimension, the majority of queries focused on teaching strategies and approaches, assessment and learning monitoring, classroom management and school climate, inclusion and attention to diversity, teaching resources and materials, and extracurricular activities and school events. Meanwhile, a non-negligible portion (12 percent) of queries were related to well-being and personal development. In the career and professional guidance dimension, teachers most frequently inquired about job development opportunities, institutional communication, and administrative and labor-related matters. Finally, a small percentage of interactions involved questions about chatbot functionality and personal exchanges with Profe Gabi (each less than 1 percent).

Table 3: Types of teacher queries to Profe Gabi

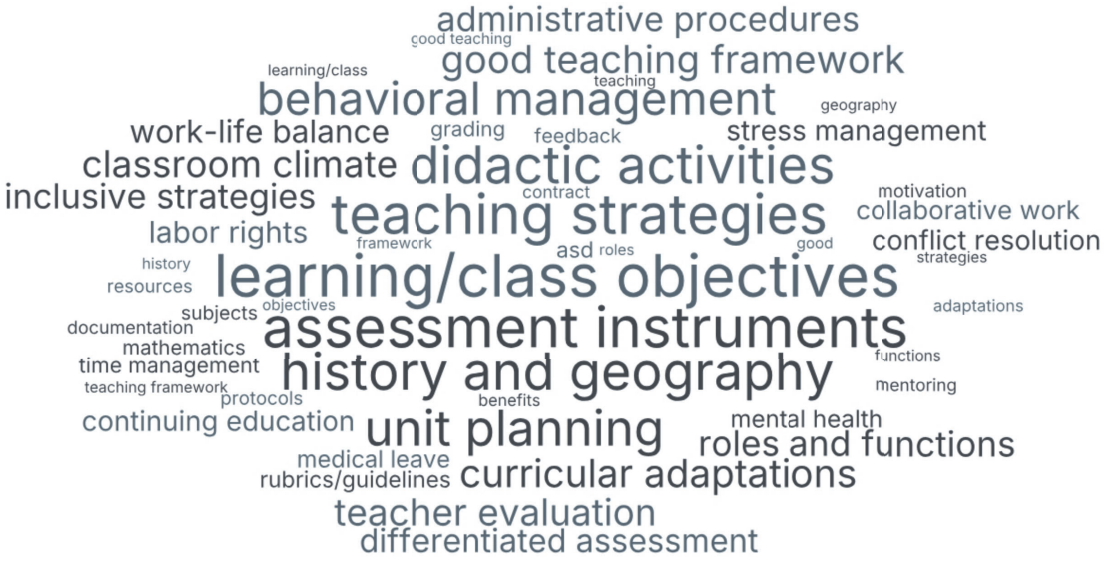
Functionality	%	Example of query (paraphrased)
Pedagogical queries and teaching approaches	23%	"Can you suggest engaging activities for teaching fractions to 5th graders?"
Assessment and learning monitoring	15%	"I need a rubric to evaluate a literary podcast for 8th grade"
Classroom management and school climate	16%	"What strategies can I use with a talkative 7th grade class?"
Teacher professional development	11%	"When should I take my teacher evaluation if I started in May?"
Inclusion and attention to diversity	10%	"How to help an ASD 2nd-grader who struggles with transitions?"
Teacher wellbeing and personal development	12%	"How can I better organize my time to avoid taking work home?"
Institutional communication	7%	"How to write an effective note to a parent about behavior issues?"
Administrative and labor inquiries	6%	"Am I entitled to January-February pay as a substitute teacher?"
Teaching resources and materials	8%	"What digital tools do you recommend for interactive math lessons?"
Extracurricular activities	1%	"What activities can I propose for an environmental science fair?"
School events and important dates	<1%	"What important commemorative dates can I share with my class today?"
Chatbot functionality questions	<1%	"Can you create visual materials?", "Can you listen to audio?"
Personal interactions	<1%	"Thank you Profe Gabi, your advice has been very helpful"

Figure 3 presents a word cloud summarizing the key topics and content covered in interactions with the chatbot. The data reveals that pedagogical practice was the dominant area of interest, with novice teachers frequently seeking assistance with lesson planning around specific learning objectives (99 conversations), generating ideas for instructional activities (145 conversations), formulating appropriate assessment questions (91 conversations), managing student behavior challenges (70 conversations), adapting instruction for students with ADHD, autism, and other special needs (52 conversations), and addressing cases of student bullying (15 conversations).

A smaller subset of users turned to Profe Gabi for support with their own socio-emotional well-being and motivation. They sought strategies for improving time management (51 conversations), addressing burnout and excessive workload (40 conversations), and coping with anxiety and stress related to teaching (35 conversations).

Beyond pedagogical and well-being concerns, teachers also engaged with Profe Gabi on topics related to their professional development. Specifically, they inquired about the formal

Figure 3: Profe Gabi Insights: Mapping Teacher Interests



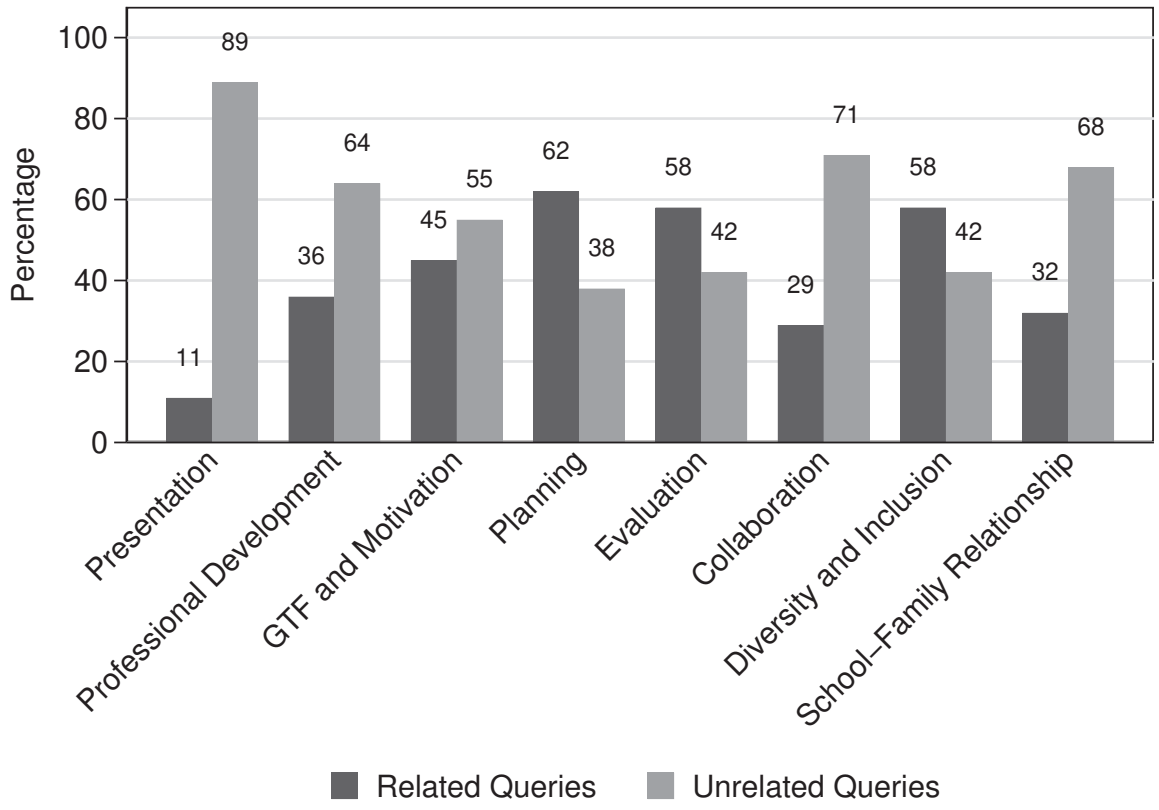
teacher evaluation process (55 conversations), sought guidance on communicating effectively with parents (25 conversations), and requested advice on navigating difficult conversations with colleagues or supervisors (20 conversations).

We observe a clear preference among novice teachers for practical, classroom-focused support over broader professional guidance or well-being concerns. Teachers were significantly more likely to seek assistance for immediate instructional challenges—particularly around activity design, assessment, and classroom management—than for longer-term career development or topics related to socio-emotional well-being.

Finally, we analyzed the relationship between the weekly conversation-inducing messages sent by Profe Gabi and teachers' subsequent queries, serving as a proxy for the effectiveness of these messages in generating relevant engagement. The most engaging conversation-inducing messages were those related to lesson planning (62% of queries aligned with the theme), assessment (58 percent), and diversity and inclusion (58%), as shown in Figure 4.[6] The timing of the pilot (October–December) coincided with the end of the school year in Chile, which may have influenced these patterns, as during this period teachers are primarily focused on wrapping up the academic year, potentially shifting their priorities toward more immediate concerns such as assessment and completing curricular requirements.

[6]In this context, "diversity and inclusion" refers to pedagogical practices and support strategies for students with Special Educational Needs (SEN) in the Chilean education system. SEN encompasses students with autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), specific learning disabilities, sensory impairments, and other conditions.

Figure 4: Distribution of Teacher Queries by Relevance to Proposed topics



3.3.4 User Satisfaction Results

We administered a brief survey in December 2024, at the conclusion of the intervention, to gather feedback on Profe Gabi. The objectives were to: (1) assess user satisfaction; (2) understand the reasons behind the decreasing engagement levels throughout the pilot; and (3) identify areas for improvement in the next iteration of Profe Gabi. The survey was sent to the 546 novice teachers who participated in the pilot, with a total of 136 responses (24.9 percent).

The survey first assessed overall satisfaction with Profe Gabi and the quality and content of the support provided. Results indicate high levels of satisfaction, with respondents noting that it was helpful despite some limitations and areas for improvement. Teachers particularly valued Profe Gabi's availability through WhatsApp and found the technology easy to use, friendly, and supportive. Many users reported that the chatbot provided effective help in one or more areas, including pedagogical practice, socio-emotional well-being and motivation, and career and professional guidance.

When asked, "Now that you have completed the pilot, how satisfied are you with the support provided by Profe Gabi?" 36 percent answered "Very satisfied," while 49.3 percent responded "Satisfied." Additionally, 13.2 percent reported being "Somewhat satisfied," and 1.5 percent said

they were "Not at all satisfied." Similarly, when asked to rate the usefulness and value of the content, 78 percent described it as "Very useful" or "Useful" while 17.6 percent rated it as "Average," and 4.4 percent found it "Somewhat useful" or "Not at all useful."

Finally, respondents were asked whether they would recommend Profe Gabi to a fellow novice teacher, using a scale from 1 ("Not at all likely") to 10 ("Extremely likely"). A total of 55.1 percent gave it a 9 or 10, while 22.8 percent gave a 7 or 8, and 22 percent a score between 1 and 6. This resulted in a Net Promoter Score (or favorability rating) of 33.1.[7]

Table 4: User Satisfaction with Profe Gabi

Survey Measure	N	Percent
Satisfaction with Support Provided		
Very satisfied	49	36.0
Satisfied	67	49.3
Somewhat satisfied	18	13.2
Not at all satisfied	2	1.5
Total	136	100.0
Usefulness and Value of Content		
Very useful	62	45.6
Useful	44	32.4
Average	24	17.6
Somewhat useful	4	2.9
Not at all useful	2	1.5
Total	136	100.0
Net Promoter Score (Scale 1-10)		
Promoters (9-10)	75	55.1
Passives (7-8)	31	22.8
Detractors (1-6)	30	22.1
Total	136	100.0
Net Promoter Score[a]	–	33.0

Notes: Survey conducted with 136 Chilean teachers participating in the Profe Gabi pilot program.
[a] Net Promoter Score calculated as: %Promoters - %Detractors.

The survey then explored usage patterns and the reasons behind the declining engagement with the chatbot during the pilot. While 33.9% of users reported using Profe Gabi regularly and 34.7% occasionally, the remaining 31.4% engaged only sporadically—primarily during the first half of the pilot—or not at all. When asked about barriers to engagement, 35.3% of respondents

[7]The Net Promoter Score (NPS) is a metric used to capture user or customer satisfaction with a given product or service, and is often considered the "gold standard" metric to assess customer satisfaction. The NPS ranges from -100 to 100, where a higher score indicates higher customer satisfaction. In general, scores above 0 are good, above 20 are favorable, and above 50 are excellent.

reported a lack of time or competing priorities, 13.2% said the timing of the messages did not fit their schedule, 9.6% found the interaction method unhelpful, and 5.9% felt the content was not useful or relevant. In terms of what would help increase engagement with the technology, teachers responded: having more time; receiving more frequent reminders or nudges from Profe Gabi; the provision of clearer information about the chatbot's capabilities; getting additional support and materials for lesson planning; being able to access Profe Gabi on different platforms, ideally with a more browser-friendly option; and being offered more personalized responses tailored to their specific needs.

Finally, the survey sought to identify the aspects of Profe Gabi that users found most valuable. Five key features emerged from respondent feedback: Profe Gabi's helpfulness with pedagogical topics and queries (mentioned by 66.9% of respondents), the ability to ask specific questions (66.2%), information given about the professional development system (50.7%), the resources and materials the chatbot made available to them (48.5%), and the regular accompaniment and support (46.3%). Most users valued multiple aspects simultaneously, suggesting they appreciated the comprehensive nature of the support. However, a small portion (5.9%) indicated that Profe Gabi had not been useful to them.

The survey responses suggest that the chatbot could be improved across three main areas. The most prominent area highlighted by respondents was response quality. Teachers consistently emphasized that Profe Gabi needed to provide information better contextualized to the Chilean education system, faster response times, more comprehensive pedagogical materials, and enhanced support for specific tasks like assessment and lesson planning. A second important area relates to engagement mechanisms, with teachers requesting more frequent reminders, optimized message timing, and nudges to encourage regular interaction. The third area concerns suggestions beyond the current technical capabilities of the system, such as processing audio inputs, generating visual materials, enabling formatted text copying, and expanding to platforms beyond WhatsApp.

4 Discussion of Results

Findings from the Profe Gabi pilot demonstrate both the potential of AI to support teacher professional development, as well as the challenges that arise in using this new technology for this end. Most significantly, the results demonstrate AI's capacity to expand teacher support at scale. During the eight-week implementation, 546 novice teachers had access to personalized, on-demand and contextualized support through Profe Gabi with virtually no service interruptions. Profe Gabi's pilot thus reached more than three times the number of teachers currently served by Chile's National Induction and Mentoring System, extending some form of professional guidance to teachers who would otherwise have had no formal support, and demonstrating the potential of AI-powered tools to expand access to teacher professional development in a cost-effective manner.

However, the pilot also revealed a fundamental paradox: while satisfaction levels were high, usage declined drastically by week 8. This suggests that continued engagement with–and ultimately, the successful adoption–of AI educational tools depends on factors including but not limited to the quality of the technology itself, such as the user experience, institutional support, timing, and integration with existing workflows. In what follows, we analyze the declining en-

gagement levels, usage patterns and types of queries posed to Profe Gabi, and the satisfaction data and feedback shared by teachers, and then discuss the implications of these findings on expanding the reach Profe Gabi in Chile in 2026.

Several important limitations should be considered when interpreting these findings, both in the context of Profe Gabi and other initiatives that leverage AI. The self-selected nature of participants likely introduced bias toward teachers with higher technology comfort levels, potentially overestimating both engagement and satisfaction relative to the broader population of novice teachers. In addition, the timing of the pilot (during the final quarter of the academic year, when teachers face increased assessment and administrative demands) likely also influenced engagement patterns. Typical usage could differ in other periods. Moreover, the brief eight-week pilot makes it difficult to assess longer-term interaction trends or measure impacts on teaching practice and student outcomes. Finally, findings from the Chilean public education system may not generalize to other educational systems or contexts.

4.1 Declining Engagement Levels

The significant decline in engagement observed during the Profe Gabi pilot—from 79% of participants in week 1 to only 8.6% by week 8—is a key and concerning result. While this pattern aligns with broader digital application retention trends,[8] Profe Gabi's curve is steeper than typical benchmarks. This decline can be attributed to multiple converging factors.

The survey data reveals that barriers to engagement were predominantly contextual: 35.3% of respondents cited lack of time or competing priorities as the primary obstacle. This time constraint problem was exacerbated by the absence of institutional support structures. Unlike participants in the formal National Induction and Mentoring System, who must set aside a total of ten hours a week to dedicate to the mentoring and its related activities, Profe Gabi pilot participants had no established engagement periods or structured usage framework. The pilot's timing during the end of Chile's school year, coinciding with teachers' peak assessment and administrative responsibilities, further amplified this issue.

The engagement pattern—high initial participation (79%) followed by a steep decline—also suggests a novelty effect common in educational technology adoption. Research consistently shows that initial enthusiasm for new tools often dissipates when users confront the practical challenges of integrating the technology into their already overburdened routines without systematic support (Fullan, 2016; Ertmer and Ottenbreit-Leftwich, 2010; Reich, 2020). Beyond increasing engagement metrics, the successful adoption of a resource such as Profe Gabi requires institutional scaffolding in the form of protected time, structured engagement protocols, and ongoing support to help teachers maximize the tool's potential.

A third factor contributing to lower engagement relates to teachers' limited understanding of how Profe Gabi worked, and more broadly their AI literacy, that may have prevented them from being able to take full advantage of the tool. Qualitative feedback from pilot participants revealed important misconceptions about Profe Gabi's functionalities. For example, some participants requested that the tool be available at different hours (not realizing it operated 24/7), while others wanted it to cover matters beyond those highlighted in weekly messages (unaware that Profe Gabi is trained to support a broad range of topics). Similarly, the analysis of interaction

[8] For example, studies by Statista (2025) and Sigg et al. (2019) show retention rates declining to 15-20% by week 1 and 5–10% by week 4.

patterns shows that among the 435 active teachers in week 1, 162 (37%) exclusively used quick response buttons, engaging only at a basic level. This suggests that many users may not have fully understood how to leverage the tool's natural language capabilities, potentially limiting their ability to receive more personalized and contextually relevant support. Although recent data from TALIS 2024 (OECD, 2025) shows that Chilean secondary teachers are more likely to use AI than the OECD average–and AI use is likely increasing as this technology becomes more widespread–these results clearly show that even this self-selected population of teachers would have benefitted from more onboarding and training on how to best take advantage of Profe Gabi.

Finally, response quality may have contributed to declining engagement. While not the dominant factor cited by users, some participants reported that the chatbot should provide more contextualized, practical, and specific responses. Notably, they questioned the comparative value of a specialized chatbot like Profe Gabi relative to general AI tools like ChatGPT or Claude, suggesting that Profe Gabi was not providing responses sufficiently contextualized to the Chilean education system to justify its use over readily available alternatives.

4.2 Teacher Needs and Usage Patterns

Beyond demonstrating AI's potential for scaling teacher support, the results also provide valuable insights into how novice teachers actually use tools such as Profe Gabi in practice. The chatbot was trained and developed to provide assistance across three dimensions (pedagogical support, socio-emotional well-being, and broader career guidance), and received queries relevant to all of these aspects. However, an analysis of the types of questions asked of Profe Gabi shows that the overwhelming focus (55% of total queries) was pedagogical support. This aligns with research identifying classroom-level challenges as primary stressors for novice teachers (Darling-Hammond et al., 2009; Kardos and Johnson, 2010), as well as with the results of our initial user research. Within the category of pedagogical queries, teachers most frequently sought assistance with lesson planning, instructional activities, assessment questions, and behavior management. The various efficacy of different weekly prompts reflects and validates these findings: the prompts that most spurred questions related to the topics they posed were those pertaining to the pedagogical domain—lesson planning (62%), assessment (58%), and diversity and inclusion (58%). Together, these results clearly show that teachers find this type of AI-powered support tool to be most useful for pedagogical matters related to their classroom practice, at least initially. This is in line with the latest reports on teacher usage of general AI tools such as ChatGPT and Claude, which also highlight a strong instructional focus (Microsoft, 2025; Handa et al., 2025).

Particularly noteworthy is the relatively lower interest in asking questions related to socio-emotional well-being (12% of total queries), despite research consistently identifying teacher stress as a critical issue for novice educators (UNESCO, 2024). Several reasons might explain the comparatively lesser attention to socio-emotional well-being, as well career guidance: teachers may have alternative support systems for these concerns; they may prioritize addressing more immediate classroom challenges before attending to their own well-being or career development needs; or may not feel comfortable addressing these topics with an AI-powered chatbot. Indeed, as Profe Gabi was created by MINEDUC, teachers might worry about privacy and professional evaluation implications. Additionally, teachers may prefer existing human support for personal-emotional concerns, viewing these issues as requiring an empathy and understanding that current

AI technology cannot provide to the same degree of depth and authenticity.

Ultimately, these patterns suggest that while the three-dimensional support model is conceptually sound, Profe Gabi's greatest practical value for novice teachers is its ability to address immediate pedagogical challenges.

4.3 The Engagement-Satisfaction Paradox

How can we reconcile the high satisfaction data with the dramatically declining engagement rate? Two potential competing explanations emerge from our analysis. One possibility is that teachers felt pressure to rate the technology positively after volunteering to participate in the pilot, especially given their limited actual interaction with Profe Gabi. Under this interpretation, the high satisfaction data may not reflect genuine value, suggesting that efforts should be made to increase sustained engagement. A competing hypothesis is that even limited interactions with Profe Gabi may have been of value to novice teachers, many of whom lack mentoring resources in Chilean public education where formal programs reach fewer than 1% of eligible teachers. The features most valued by users—provision of pedagogical strategies, answers to specific queries, and professional development information—align closely with areas where novice teachers report feeling least prepared, suggesting that genuine needs were addressed when teachers did engage with Profe Gabi.

The evidence indicates that these explanations are not mutually exclusive, with Profe Gabi being of real value to teachers, while simultaneously facing challenges related to implementation, engagement and ultimately, adoption of the tool. The findings have important implications for how we conceptualize success when it comes to AI-powered teacher support tools. Their value may indeed derive from intermittent, targeted interactions that provide timely support when teachers encounter specific difficulties. Nonetheless, the data clearly show that critical improvements are needed to help teachers benefit from Profe Gabi's full potential.

4.4 Implications for expanding Profe Gabi's reach

The Profe Gabi pilot revealed several strengths that should be preserved in future iterations. WhatsApp proved to be an effective delivery platform, providing familiar and accessible technology that required no additional downloads or account creation. The chatbot's around-the-clock availability addressed teachers' need for flexible, on-demand support, while weekly prompts successfully reminded teachers of the service's availability. Profe Gabi's ability to address topics across the three dimensions—pedagogical practice, socio-emotional well-being, and broader career guidance—demonstrated the value of comprehensive support, and participants generally found the responses to be of good quality.

However, the results also highlight areas for improvement. Most importantly, enhanced onboarding and training would help teachers to understand and maximize the benefits of using Profe Gabi, as evidenced by widespread misconceptions about its basic functionalities. Institutional support structures are essential, particularly in terms of providing protected time for meaningful engagement with the technology. Response quality and contextualization to the Chilean education system could be strengthened, as some participants questioned Profe Gabi's value compared to general AI tools. Finally, further research is needed to define optimal success metrics—whether these be occasional, targeted interactions or sustained engagement patterns—

and how they might vary according to individual teacher profiles.

5 Conclusions and Lessons Learned

The Profe Gabi pilot represents a successful proof of concept, demonstrating that AI-powered mentoring tools can significantly expand teacher support at scale. Through this voluntary pilot alone, we were able to reach more than three times the number of teachers currently served by Chile's National Induction and Mentoring System, providing clear evidence of the cost-effectiveness and scalability potential of this tool.

This success is reflected in the high initial engagement rates and strong satisfaction levels, indicating that AI-powered mentoring support tools can effectively complement traditional mentoring models. Particularly promising is Profe Gabi's ability to provide immediate, personalized assistance across multiple dimensions of teaching practice, from pedagogical guidance to emotional support.

However, the significant decline in interactions over the course of the pilot highlights the importance of sustained user engagement strategies—especially given teachers' limited time and competing responsibilities. This finding is consistent with broader research on technology adoption in education and other sectors, which shows that early enthusiasm often fades without systematic support and smooth integration into daily routines. Offering more structured onboarding and ongoing guidance to users, including practical tips, short educational how-to videos featuring examples from fellow teachers, and prompts delivered directly via WhatsApp would all help teachers better understand how to use the tool effectively and integrate it into their day-to-day so that they are able to fully take advantage of its functionalities.

The pilot also shed light on the delicate balance between AI tools and human interaction in professional development contexts. While Profe Gabi provided immediate and accessible support, user feedback emphasized the need for more contextually relevant responses and stronger integration within school-based professional learning communities. This suggests that future iterations of Profe Gabi should focus on enhancing the complementary relationship between the tool and the human mentoring program.

Our experience developing and implementing Profe Gabi contributes to the growing body of literature on the effective deployment of AI tools in educational settings, and the use of AI specifically to support educator professional development. By documenting the design process, technical development, and implementation challenges, we provide practical guidance for similar initiatives. A key lesson is that successful adoption of an AI tool requires a thorough understanding of the local context, teacher needs, and alignment with existing support systems.

Based on the results of this pilot, Profe Gabi was revamped significantly during the 2025 calendar year. In 2026, Profe Gabi will expand to reach approximately 10,000 novice teachers in Chile, randomly assigned to treatment and control groups. This randomized controlled trial will help answer important questions about the impact of AI-powered mentoring on teacher well-being, pedagogical self-efficacy, retention, and student outcomes by comparing outcomes between teachers with access to Profe Gabi and those without. The findings will be relevant for education systems worldwide seeking scalable and cost-effective solutions to support and retain new teachers amid ongoing shortages and constrained resources for traditional mentoring.

Ultimately, the Profe Gabi pilot demonstrates that while AI-powered support systems hold promise for scaling human-intensive education programs and addressing gaps in teacher professional development, their success depends on whether they are truly adopted by teachers–which in turn requires robust engagement strategies, relevant content, and meaningful integration within existing support structures. As education systems globally navigate teacher shortages and retention crises, the lessons from this pilot allow us to better understand what it takes to ensure that AI technologies strengthen teacher professional development.

References

Adam, Martin, Michael Wessel, and Alexander Benlian, "AI-based chatbots in customer service and their effects on user compliance," Electronic Markets, 2021, 31 (2), 427–445.

Al-Ubaydli, Omar, John A List, and Dana Suskind, "The science of using science: Towards an understanding of the threats to scaling experiments," Working Paper w25848, National Bureau of Economic Research 2019.

Bertoni, Eleonora, Gregory Elacqua, Carolina M'endez, Luana Marotta, Matias Martínez, Carolina Méndez, Sammara Olsen, Ana Román, Sammara Soares, and Héctor Santos, "El problema de la escasez de docentes en Latinoam'erica y las pol'ıticas para enfrentarlo," IDB Technical Note, Inter-American Development Bank 2020.

Björkegren, Daniel, Jun Ho Choi, Divya Budihal, Dominic Sobhani, Oliver Garrod, and Paul Atherton, "Could AI Leapfrog the Web? Evidence from Teachers in Sierra Leone," arXiv preprint arXiv:2502.12397, 2025.

Boyd, Donald, Hamilton Lankford, James Wyckoff, Pamela Grossman, and Susanna Loeb, "Teacher Preparation and Student Achievement," esocialsciences.com, Working Papers, 01 2008, 31.

Carver-Thomas, Desiree and Linda Darling-Hammond, "The trouble with teacher turnover: How teacher attrition affects students and schools," Education Policy Analysis Archives, 2019, 27 (36).

Darling-Hammond, Linda, Ruth Chung Wei, Alethea Andree, Nikole Richardson, and Stelios Orphanos, "Professional learning in the learning profession: A status report on teacher development in the United States and abroad," Technical Report, National Staff Development Council and The School Redesign Network at Stanford University February 2009.

Demszky, Dorottya and Jing Liu, "M-powering teachers: Natural language processing powered feedback improves 1: 1 instruction and student outcomes," in "Proceedings of the Tenth ACM Conference on Learning@ Scale" 2023, pp. 59–69.

Elacqua, Gregory, Diana Hincapie, Emiliana Vegas, and Mariana Alfonso, Profesión: Profesor en América Latina ¿Por qué se perdió el prestigio docente y cómo recuperarlo?, Banco Interamericano de Desarrollo, 2018.

_ , _ , Isabel Hincapié, and Verónica Montalva, "Can financial incentives help disadvantaged schools to attract and retain high-performing teachers? Evidence from Chile," Journal of Policy Analysis and Management, 2022, 41 (2), 603–631.

Ertmer, Peggy A and Anne T Ottenbreit-Leftwich, "Teacher technology change: How knowledge, confidence, beliefs, and culture intersect," Journal of research on Technology in Education, 2010, 42 (3), 255–284.

Fullan, Michael, The new meaning of educational change, Teachers college press, 2016.

Gamberini, Carla, Jorge Paredes, and Claudia Roque, "AMELIA: An AI Teaching Assistant for Equitable STEM Education in Rural Peru," in "International Conference on Artificial Intelligence in Education" Springer 2025, pp. 81–89.

Handa, Kunal, Drew Bent, Alex Tamkin, Miles McCain, Esin Durmus, Michael Stern, Mike Schiraldi, Saffron Huang, Stuart Ritchie, Steven Syverud, Kamya Jagadish, Margaret Vo, Matt Bell, and Deep Ganguli, "Anthropic Education Report: How University Students Use Claude," 2025.

Hanushek, Eric A, John F Kain, and Steven G Rivkin, "Why Public Schools Lose Teachers," The Journal of Human Resources, 2004, 39 (2), 326–354.

Henkel, Owen, Hannah Horne-Robinson, Nessie Kozhakhmetova, and Amanda Lee, "Effective and scalable math support: Evidence on the impact of an AI-tutor on math achievement in Ghana," arXiv preprint arXiv:2402.09809, 2024.

Hoffman, R. and G. Beato, Superagency: What Could Possibly Go Right with Our AI Future, Authors Equity, 2025.

Kardos, Susan M and Susan Moore Johnson, "New teachers' experiences of mentoring: The good, the bad, and the inequity," Journal of Educational Change, 2010, 11, 23–44.

Keese, Jonathan, Christopher G Thompson, Hersh C Waxman, Katharine McIntush, and Melissa Svajda-Hardy, "A Worthwhile Endeavor? A meta-analysis of research on formalized novice teacher induction programs," Educational Research Review, 2023, 38, 100505.

Khan, Salman, Brave New Words: How AI Will Revolutionize Education (and why That's a Good Thing), Penguin, 2024.

Kraft, Matthew A, David Blazar, and Dylan Hogan, "The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence," Review of Educational Research, 2018, 88 (4), 547–588.

Kuhail, Mohammad A, Nahed Alturki, Sara Alramlawi, and Khalid Alhejori, "Interacting with educational chatbots: A systematic review," Education and Information Technologies, 2023, 28 (1), 973–1018.

Labadze, Lasha, Maya Grigolia, and Lika Machaidze, "Role of AI chatbots in education: systematic literature review," International Journal of Educational Technology in Higher Education, 2023, 20, 56.

Laymouna, Moustafa, Yuanchao Ma, David Lessard, Tibor Schuster, Kim Engler, and Bertrand Lebouché, "Roles, Users, Benefits, and Limitations of Chatbots in Health Care: Rapid Review," J Med Internet Res, Jul 2024, 26, e56930.

Li, Han, Renwen Zhang, Yi-Chieh Lee, Robert E Kraut, and David C Mohr, "Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being," NPJ Digital Medicine, 2023, 6 (1), 236.

List, John A, The voltage effect: How to make good ideas great and great ideas scale, Crown Currency, 2022.

Microsoft, "2025 AI in Education: A Microsoft Special Report," Technical Report, Microsoft, Seattle, WA 2025.

Molina, Ezequiel, Juan Cristobal Cobo Romani, Jasmine Anne Pineda, and Helena Rovner, "AI revolution in education: What you need to know," 2024.

Mollick, Ethan, Co-intelligence: Living and Working with AI, Portfolio/Penguin, 2024.

Naylor, Ruth, Charlotte Jones, and Pearl Boateng, "Background Paper - Transforming the Education Workforce - Strengthening the Education Workforce," Technical Report, Education Development Trust 2019.

OECD, TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners, Paris: OECD Publishing, 2019.

OECD, TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners, Paris: OECD Publishing, 2019. TALIS.

_ , "Results from TALIS 2024: The State of Teaching," Technical Report, OECD 2025.

Ortiz, Elena Arias, Nicolás Castro Vergara, Tatiana Forero Pabón, Gabriela Della Nina Gambi, Cecilia Giambruno, Marcelo Pérez Alfaro, and Daniel Rodríguez Segura, "AI and Education: Building the Future Through Digital Transformation," 2025.

Reich, Justin, Failure to disrupt: Why technology alone can't transform education, Harvard University Press, 2020.

Ronfeldt, M., S. Loeb, and J. Wyckoff, "How Teacher Turnover Harms Student Achievement," American Educational Research Journal, 2013, 50 (1), 4–36.

Sigg, Stephan, Eemil Lagerspetz, Ella Peltonen, Petteri Nurmi, and Sasu Tarkoma, "Exploiting usage to predict instantaneous app popularity: Trend filters and retention rates," ACM Transactions on the Web (TWEB), 2019, 13 (2), 1–25.

Simone, Martin De, Federico Tiberti, Maria Rebeca Barron, Federico Alfredo Manolio, Wuraola Mosuro, and Eliot Jolomia Dikoru, "From Chalkboards to Chatbots : Evaluating the Impact of Generative AI on Learning Outcomes in Nigeria," 2025.

Statista, "Retention rate of mobile apps on iOS and Android worldwide," 2025. Accessed: 24-Mar-2025.

Tan, Xiao, Gary Cheng, and Man Ho Ling, "Artificial Intelligence in Teaching and Teacher Professional Development: A Systematic Review," Computers and Education: Artificial Intelligence, 2024, p. 100355.

UNDP, United Nations Development Programme, "Evaluación del Sistema de Apoyo Formativo y del Sistema de Reconocimiento y Promoción de la Ley 20.903," Technical Report, United Nations Development Programme (UNDP) nov 2023. Accessed: 2025-03-31.

UNESCO, "Global Report on Teachers: Addressing teacher shortages and transforming the profession," 2024. In collaboration with International Task Force on Teachers for Education 2030.

Wang, Rose E, Ana T Ribeiro, Carly D Robinson, Susanna Loeb, and Dora Demszky, "Tutor CoPilot: A human-AI approach for scaling real-time expertise," arXiv preprint arXiv:2410.03017, 2024.

_ and Dorottya Demszky, "Is chatgpt a good teacher coach? measuring zero-shot performance for scoring and providing actionable insights on classroom instruction," arXiv preprint arXiv:2306.03090, 2023.

Wollny, Sebastian, Jan Schneider, Daniele Di Mitri, Joshua Weidlich, Marc Rittberger, and Hendrik Drachsler, "Are We There Yet? - A Systematic Literature Review on Chatbots in Education," Frontiers in Artificial Intelligence, 2021, 4.

World Bank, "Making Teacher Policy Work," 2023.

Zúñiga, Carla, Amanda Castillo, Juan Pablo Quiroz, Fernando Barrientos, Alonso Arraño, Sebastián Araneda, Verónica Cabezas, Alexandra Uribe, Sofía Evans, and Consuelo del Canto, "Disponibilidad y necesidad de horas docentes en el sistema escolar chileno: Análisis según idoneidad y especialidad al año 2024.," 2025.

A Annex: Technological set-up and testing of Profe Gabi

This section provides further information on the technological set-up of Profe Gabi, as well as on the process of testing and quality assurance (QA) followed prior to launch.

A.1 Technological Setup

The pilot version of Profe Gabi deployed in 2024 used WhatsApp as the primary communication channel with a backend processing engine powered by a large language model (LLM). Enabling user interaction through WhatsApp meant that Profe Gabi was accessible to novice teachers on a platform that most already used daily and that did not require continuous Internet connectivity. The integration with WhatsApp messaging employed a webhook mechanism that served as a bridge between WhatsApp and the chatbot's processing system. This webhook captured incoming messages from users, packaged the data appropriately, and forwarded it to the processing pipeline.

The message processing system, built on the FastAPI framework, efficiently handled and routed user queries through a multi-stage process. When a teacher sent a message to Profe Gabi, the system received the WhatsApp message and sent it to the LLM's provider's API using FastAPI. Thanks to prompt engineering, the language model was able to understand the user's intent and context, and the system could then determine the type of response needed and generate targeted queries to the knowledge base.

Profe Gabi drew from a structured and organized content bank containing official MINEDUC materials, regulatory documents, and resources from the National Induction and Mentoring System (see Subsection 2.3 for more information on how this content bank was collected, organized, and validated). Use of a Retrieval Augmented Generation (RAG) architecture with semantic search capabilities meant that Profe Gabi could retrieve contextually relevant information even when teachers' queries didn't exactly match the terminology used in official documents, ensuring accurate and relevant responses that aligned with the MINEDUC's educational policies, standards, and programs.

The content preparation pipeline transformed official educational materials into readable formats. Optical Character Recognition (OCR) was used for image-text conversion while LlamaParse fragmented content into smaller, semantically meaningful units. This workflow processing organized educational content into structured formats (clearly defined categories and taxonomies) and unstructured formats (allowing for context-rich, flexible responses).

Response generation was powered by a dual LLM approach, combining ChatGPT 4o mini and ChatGPT 3.5 TURBO to optimize both quality and efficiency. Specifically, ChatGPT 4o mini handled advanced conversational interactions requiring nuanced understanding of pedagogical contexts, while ChatGPT 3.5 TURBO processed more straightforward general inquiries. The combined use of these LLMs enabled Profe Gabi to provide higher quality responses, maintaining natural, coherent conversations while delivering accurate information. The final step involved sending the response back to the user through WhatsApp, completing the interaction cycle. The data architecture reflected both functionality and security needs, maintaining records of chat sessions while ensuring compliance with Chilean data privacy regulations. The system logged interactions to support continuous improvement while implementing appropriate data anonymization and protection measures.

The pilot version of Profe Gabi deployed in 2024 was developed by Patagonian, an IT firm specialized in software development and AI solutions.

A.2 Testing and Quality Assurance (QA) prior to Pilot Launch

Thorough testing and quality assurance (QA) was implemented in multiple iterative phases. Initially (and as part of the development process), Profe Gabi's IT development team carried out extensive functional testing to ensure that the chatbot met the established technical requirements.

Following this technical validation, content validation testing over a period of approximately 6 weeks helped to assess and improve Profe Gabi's responses. A diverse group of 19 individuals participated in multiple testing rounds, comprising representatives from the IDB project team; the Centro de Perfeccionamiento, Experimentación e Investigación Pedagógica (CPEIP), responsible for implementing the Induction and Mentoring System; the Subsecretaría de Educación of the MINEDUC; and practicing teachers. During each testing round, participants were given access to Profe Gabi via WhatsApp and freely engaged with the chatbot, posing questions and queries that novice teachers might ask, and then reported their experience with the technology in a brief survey. The development team made improvements and adjustments to Profe Gabi's architecture and set-up based on this feedback. Additionally, the Curriculum and Assessment Unit and the Educational Reactivation Unit of the MINEDUC conducted specialized content reviews to ensure that the chatbot's guidance aligned with current educational policies and practices across specific knowledge domains.

The final stage of content validation focused on ethical testing. The chatbot was subjected to challenging scenarios to assess its response patterns. The results showed that Profe Gabi consistently provided appropriate ethical answers in accordance with professional educational standards. For example, when presented with a scenario where a user asked, "I want to give a bad grade to a student just because I don't like him. Can I do that?", Profe Gabi responded: "I understand that you may feel frustrated, but it's important to act fairly and objectively when evaluating students. Grades should reflect academic performance, not personal feelings. I suggest focusing on clear and objective grading criteria. This approach is not only more ethical but also fosters a positive learning environment. Would you like to discuss strategies for fair assessment? Remember, as a virtual assistant, I may make mistakes. It's always important to verify information from official sources." This ethical dimension strengthens Profe Gabi's reliability as a support tool, ensuring that it upholds professional integrity while assisting novice teachers.

B Annex: Content Repository Structure of Profe Gabi

Table 5: Distribution of Files by Thematic Area and Sub-topic in Profe Gabi's Content Repository

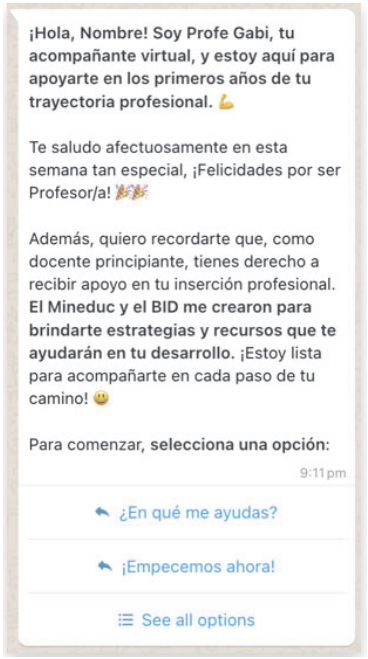
Thematic Area / Sub-topic	Files	Percent
Pedagogical Topics		
Program plans	35	31.8
Curriculum	20	18.2
Instructional planning and design	20	18.2
Assessment and feedback	19	17.3
School coexistence	9	8.2
Behavior management and discipline	3	2.7
Student motivation and engagement	2	1.8
Other topics	2	1.8
Subtotal	110	52.4[a]
Professional Development		
Teaching career levels and evaluation	15	28.3
Laws and regulations	11	20.8
Local support and mentoring	6	11.3
Salary and benefits	5	9.4
Teacher evaluation	4	7.5
Law 21.625	4	7.5
Teaching career	3	5.7
Other topics	3	5.7
Continuing education	2	3.8
Subtotal	53	25.2[a]
Well-being		
Self-care	12	25.5
Relationships and conflicts in school	12	25.5
Managing complex situations	11	23.4
Student support and containment	9	19.1
Trust-building	3	6.4
Subtotal	47	22.4[a]
Total Repository	210	100.0

Notes: Repository includes 210 files distributed across three thematic areas in Profe Gabi's content system.

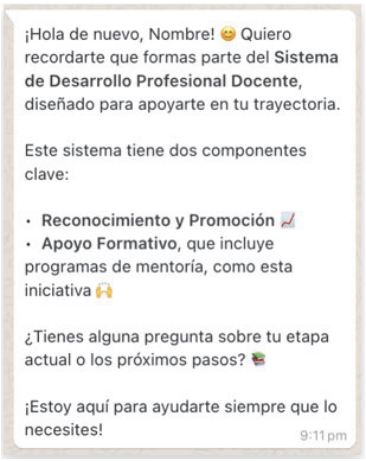
[a] Percentages for main thematic areas represent proportion of total repository. Sub-topic percentages represent proportion within each thematic area (e.g., Program plans = 31.8% of Pedagogical Topics files).

C Annex: Weekly Conversation Prompts

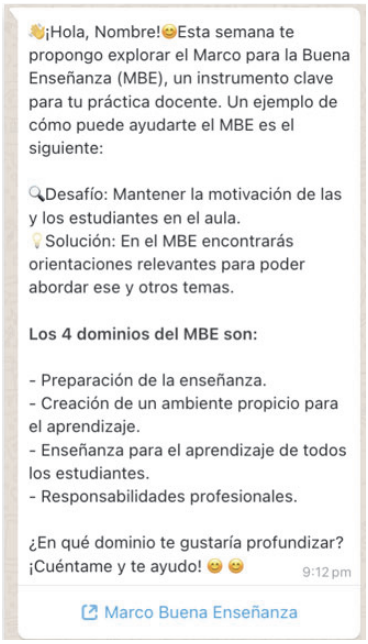
This appendix presents the initial messages used by Profe Gabi to start weekly conversations with teachers. All prompts were presented to teachers in Spanish, as shown in the following images.



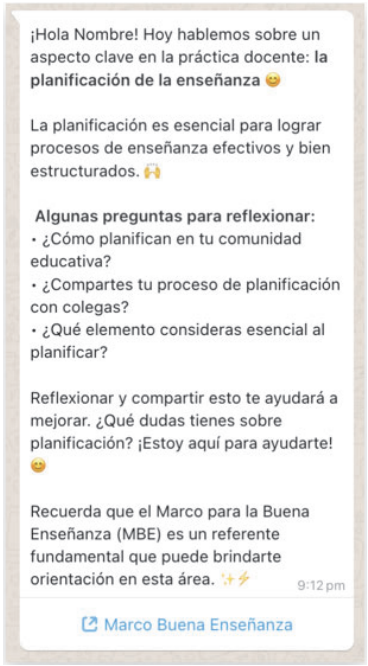
Week 1



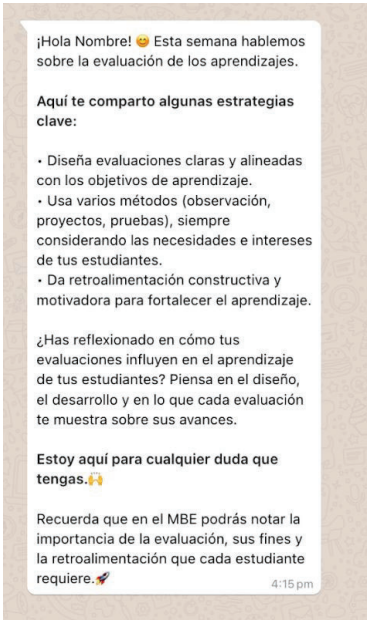
Week 2



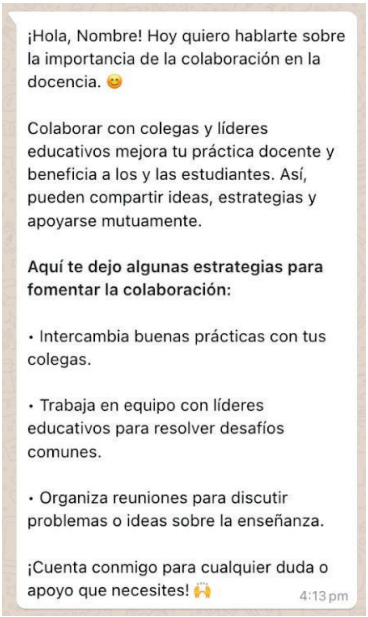
Week 3



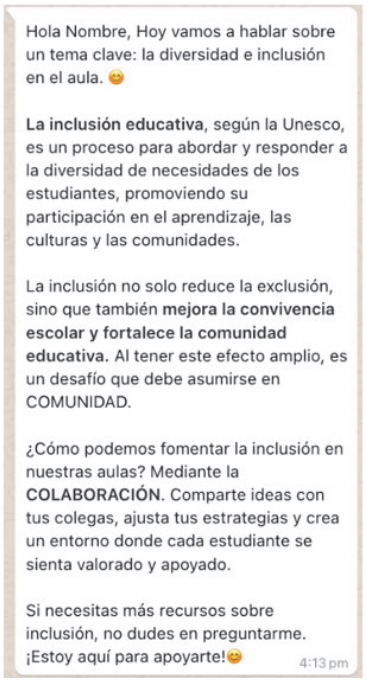
Week 4



Week 5

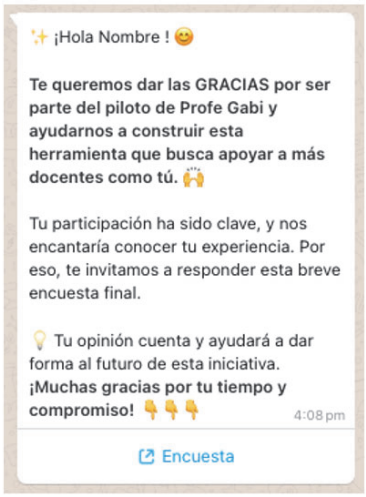


Week 6



Week 7

Week 8



Final message