

Using an AI-Based Chatbot to Support the Learning of Algebraic Factoring

Mirelle Zavala Amezcu & Mario Sánchez Aguilar

Instituto Politécnico Nacional, CICATA Unidad Legaria

Abstract

This article reports on the design, implementation, and evaluation of FACTY, a GPT-based chatbot developed to support undergraduate students in learning algebraic factoring. The study involved ten first-year engineering students who interacted with FACTY over four stages, including diagnostic and final assessments, autonomous practice, and post-intervention interviews. Guided by a framework of feedback levels in mathematics instruction, the analysis examines the nature of FACTY's responses and their relationship to student learning outcomes. Findings indicate that the chatbot predominantly provides process-level and self-regulation-level feedback, both of which are acknowledged in the literature as critical for fostering deep and autonomous learning. Two case studies illustrate contrasting results: one student achieved substantial improvement, while another made limited progress due to foundational gaps in algebra. Interview data revealed positive perceptions of FACTY's constant availability, adaptability, and nonjudgmental interaction style. The study concludes that FACTY can serve as a complement to classroom instruction, particularly when integrated with teacher oversight and used to promote critical engagement with AI-generated feedback.

Keywords: Artificial intelligence (AI) in mathematics education; Algebraic factoring; Feedback in mathematics learning

1 Introduction

Algebraic factoring is a fundamental mathematical skill that is widely addressed in secondary school curricula and undergraduate education. Mastery of this topic is not only expected as a basic algebraic competency, but it is also essential for success in advanced university courses, such as calculus, linear algebra, differential equations, and number theory. Beyond academic success, factoring plays an important role in the development of engineers and other mathematically oriented professionals. It underpins the ability to model, simplify, and transform mathematical expressions that arise in diverse areas such as control systems, signal processing, structural analysis, and numerical simulations. Recognizing and manipulating algebraic structures enables the analysis of system behavior, optimization of solutions, and effective communication within interdisciplinary teams.

Although computational tools such as Computer Algebra Systems (CAS) can automate factoring, a solid conceptual understanding of the factoring process remains crucial. Students must be able to verify symbolic results, detect errors, and make informed choices about the computational techniques they employ. Manual factoring also reinforces algebraic fluency, supporting the interpretation of symbolic outputs and the debugging of mathematical models in real-world applications. For these reasons, fostering both procedural proficiency and conceptual understanding of factoring is important for preparing students to engage with the mathematical demands of their future practice.

Despite its continuous presence throughout students' academic trajectories, factoring remains one of the areas that students find most challenging. Several mathematics educators have noted this persistent difficulty, reporting that many students enter university without a solid understanding of algebra (Burhanzade & Aygör, 2015; Chung, 2012; Kilgore & Caprano, 2010; Olivar et al., 2018). Part of this challenge can be attributed to traditional teaching methods that prioritize procedural work over conceptual understanding and fail to consider differences in students' learning paces, styles, and trajectories. This situation often results in academic underperformance, frustration, and demotivation, particularly in educational programs where factoring is a core algebraic skill.

In response to this issue, we have developed a pedagogical proposal using an artificial intelligence (AI) tool specifically designed to support the learning process of algebraic factoring. The tool is a chatbot named FACTY, trained with mathematical content and mathematics education research findings. This chatbot can provide interactive and personalized feedback tailored to students' academic needs. The aim of this article is to report on the development and implementation of this pedagogical proposal, focusing on the types of feedback FACTY provides to a group of undergraduate students through their interactions with the chatbot while working on algebraic factoring tasks.

This study is conceived as an exploratory pilot investigation, based on a small cohort of undergraduate students and an analysis of selected cases, aimed at examining the nature of AI-generated feedback in algebraic factoring. To identify the types of feedback FACTY provides, we draw on an established characterization of feedback levels in mathematics education. This framework, which distinguishes between different functions and depths of feedback, is presented in the following section.

2 Levels of Feedback in Mathematics Instruction

Feedback has been identified as one of the most influential interventions in learning (Hattie & Timperley, 2007). However, its effectiveness varies depending on the type, timing, and level at which it is directed. Söderström and Palm (2024) expand on this perspective through a systematic review of research literature on feedback in mathematics education published between 2012 and 2021. This review aimed to identify the feedback characteristics addressed in the reviewed publications, paying attention to which features have received considerable attention and which have been overlooked. Based on this literature review, the authors identified four primary levels of feedback: task-level, process-level, self-regulation-level, and personal-level. Each level serves a distinct function:

- **Task-level feedback:** Feedback at this level indicates whether a learning task has been understood or completed correctly. It focuses on the student's product, such as the accuracy of a response. It typically states whether an answer is correct or incorrect, but does not necessarily explain how to improve it.
- **Process-level feedback:** This level highlights the approach, strategy, or steps the student uses to complete a task. It guides the approach to the task, often suggesting alternative strategies or procedures for solving it.
- **Self-regulation-level feedback:** Feedback at this level targets the student's metacognitive skills, focusing on how they monitor and regulate their actions and learning through self-assessment and autonomous decision-making. It encourages students to evaluate their work and manage their learning independently.
- **Personal-level feedback:** involves general evaluations of the student's characteristics or work. It is often unrelated to the specific details of the task and tends to be limited to general judgments (e.g., "good job"), which usually do not contribute to deep learning.

Research confirms that effective feedback targets process and self-regulation levels, fostering understanding and learner autonomy (Hattie & Timperley, 2007; Söderström & Palm, 2024).

3 Programming and Training of the FACTY Chatbot

FACTY is a customized adaptation of the ChatGPT-4 language model, built on the Generative Pre-trained Transformer (GPT) architecture and restricted to mathematical content, with a specialization in algebraic factoring. Its development began with the systematic collection of 565 school-level factoring exercises from two well-known textbooks in Mexico—*Matemáticas Simplificadas* (Aguilar et al., 2009) and *Álgebra* (Baldor, 2009)—organized into seven types of algebraic factoring (common factor, grouping of terms, difference of squares, perfect square trinomials, trinomials of the form $x^2 + bx + c$, trinomials of the form $ax^2 + bx + c$, and sum or difference of cubes) and three levels of difficulty (basic, intermediate, and advanced). The exercises were initially encoded in LaTeX and then converted into JSON format to ensure accurate interpretation by the AI system.

The construction of FACTY took approximately six months and was carried out manually. The process required no specialized hardware and no advanced programming expertise. Instead, it relied on: (a) the creation of a structured database of exercises and common student errors, (b) the drafting of prompts to guide ChatGPT-4's reasoning, and (c) the use of Python scripts to read databases and Excel files. While some basic technical skills—such as handling JSON files—were needed, the overall approach remains accessible to educators or educational departments with minimal coding experience. The only recurring cost is a ChatGPT Plus subscription, which provides access to GPT-4; any academic institution with such access, a curated exercise set, and organized feedback can replicate FACTY without having to train a new language model from scratch.

A distinctive feature of FACTY's training was the integration of a database of frequent student errors in factoring, compiled through a review of specialized mathematics education literature (e.g., Abore, 2020; Burhanzade & Aygör, 2015; Chung Wing Hong, 2012; Kilgore & Capraro, 2010; Olivari et al., 2018). This database enables FACTY to recognize common error patterns and provide adaptive and timely feedback. FACTY supports progressive skill development by starting students with basic problems and advancing toward more complex ones that demand higher algebraic mastery. It monitors the accuracy and consistency of responses within each factoring type. Once a student solves at least three consecutive exercises of the same type without error, the system either offers problems of greater complexity or introduces a new factoring case at a higher difficulty level. If the student struggles, FACTY reverts to the previous level. This mastery-based progression allows the system to “know” when to move forward or slow down, adapting the learning path to each student's performance. In this way, FACTY functions both as an automated evaluator and as a pedagogical agent, guiding learners toward conceptual understanding while identifying and addressing their algebraic errors.

4 Implementation of the Pedagogical Proposal

The didactic experience involved a convenience sample of 10 second-semester undergraduate students, all in the first year of their engineering programs at a public university in southeastern Mexico. The group consisted of regular students, meaning none were enrolled in a remedial program, and all were progressing normally within their cohort. Participants, aged between 19 and 24 years, volunteered to take part and signed an informed consent form prior to the start of the experience. All had previously completed a standard preparatory mathematics course that included algebraic factoring topics. Despite having followed a similar academic trajectory, students displayed heterogeneous levels of prior knowledge in algebraic manipulation, as evidenced by their diagnostic test performance, which ranged from no correct responses to near-complete mastery of the assessed factoring tasks.

The instructional experience was structured in four stages. The first stage involved a face-to-face session during which a diagnostic test of 14 exercises was administered (Figure 1).

Figure 1*Initial diagnostic test administered during Stage 1*

Instructions: Solve the following factoring problems.

1. $x^2 + x$	7. $x^2 - 2x + 1$	13. $1 - a^3$
2. $a^4 + a^3 - a^2$	8. $1 - 2a^3 + a^6$	14. $64 + a^6$
3. $m^2 + mn + mx + nx$	9. $x^2 + 7x + 10$	
4. $2y^3 - 6ay^2 - y + 3a$	10. $12 - 8n + n^2$	
5. $x^2 - 1$	11. $3x^2 - 5x - 2$	
6. $4a^4 - 9b^2c^2$	12. $16m + 15m^2 - 15$	

Note. The test includes 14 exercises covering seven types of factoring at basic and intermediate levels.

These exercises covered the seven types of factoring, with one basic-level and one intermediate-level item for each case. During this session, the article's first author manually reviewed the diagnostic exams, marking each item as correct or incorrect. The exams were scanned into .jpg format and returned to the students. Using these images, students began interacting with FACTY by submitting the .jpg file of their test to the chatbot for analysis. Based on the analysis of each test, FACTY generated automated feedback on the errors made, helping identify the specific topics each student needed to reinforce during the following week (see Figures 2 and 3 for an example).

Figure 2*Example of an interaction with FACTY: Uploading a scanned test and obtaining error-specific feedback*

Can you give me feedback on my diagnostic exam?

Exercise 1: $x^2 + x$

- Expected factorization type: Common factor
- Student's answer: $x(x + 1)$
- ✗ Error: None
- 🟦 Feedback: Excellent! The student correctly identified and extracted the common factor x .
- ✅ Correct answer: $x(x + 1)$
- 💡 Tip: Always check each term for a shared factor. Great start!

Note. The student uploads a scanned diagnostic exam. FACTY analyzes each exercise and provides targeted feedback.

Figure 3*Continued interaction with FACTY: Feedback on additional exercises*

Exercise 2: $a^4 + a^3 - a^2$

- Expected factorization type: Common factor
- Student's answer: $a^2(a^2 + a - 1)$
- Error: None
- Feedback: Very good! The student factored out the greatest common factor a^2 , leaving a correct trinomial.
- Correct answer: $a^2(a^2 + a - 1)$
- Tip: Double-check by multiplying your result to confirm accuracy.

Exercise 3: $m^2 + mn + mx + nx$

- Expected factorization type: Grouping
- Student's answer: $m(m + n + x) + nx$
- Error: Incorrect grouping and partial factorization.
- Feedback: The student began grouping but did not reach a fully factored expression. Correct grouping leads to two binomials:
 $(m + n)(m + x)$
- Correct answer: $(m + n)(m + x)$
- Tip: When grouping, ensure both parts yield the same binomial factor.

Note. FACTY identifies errors and provides tips for improvement on each factoring exercise.

The second stage consisted of a week of independent work with FACTY from home. During this period, students interacted with the chatbot, focusing on the types of factoring they had struggled with. FACTY provided immediate, personalized feedback tailored to each student's responses, allowing them to progress at their own pace and concentrate on procedures that required further attention. The students shared some chat transcripts generated during these interactions with the article's authors, similar to those in Figures 2 and 3.

The third stage was a second face-to-face session during which a final assessment of 14 exercises was administered. While the final test included the same seven types of factoring as the diagnostic test, the exercises' wording was different to avoid repetition and better assess students' understanding of the factoring process. The fourth and final stage consisted of individual interviews with the participating students, in which they were asked to share their perceptions of using FACTY as a resource for learning algebraic factoring. These interviews were audio-recorded for later transcription and analysis.

Across the group as a whole, the comparison between diagnostic and final assessments revealed varied trajectories. Most participants showed some improvement in their factoring performance after the period of autonomous interaction with FACTY, although the magnitude of the gains varied considerably among students. While several students moved from low or intermediate initial performance to solving a larger proportion of the final assessment correctly, others exhibited only modest progress. This variability in outcomes motivated the analysis of two contrasting cases presented in the following sections to understand better how different types of feedback may support—or fail to support—students' learning processes.

4.1 Analytical Strategy

The implementation of the pedagogical proposal reported in this article is part of a broader research project that explores the potential relationship between the types of feedback provided by FACTY and the development of algebraic factoring skills in students who receive such feedback. For this purpose, we selected students who had not demonstrated strong performance on the diagnostic test administered during the first stage to examine whether their performance improved in the final assessment of the third stage, after having interacted with FACTY and receiving its feedback. In this article, we report on the cases of two students: Guadalupe, who answered 6 out of 14 items correctly on the diagnostic test, and Leonardo, who did not answer any items correctly. Both names are pseudonyms used to protect the participants' identities.

These two cases were purposefully selected as analytically informative and contrasting trajectories within the same instructional context. Both students exhibited low initial performance in the diagnostic assessment, yet their subsequent interactions with FACTY and their learning outcomes diverged substantially. This contrast allowed us to examine how different types of feedback functioned under similar instructional conditions but in relation to distinct levels of prior knowledge.

The interactions that Guadalupe and Leonardo shared with the authors at the end of the second stage of the implementation were analyzed and synthesized in a spreadsheet. This spreadsheet does not include verbatim transcripts. Instead, it classifies excerpts from their interactions according to the type of feedback provided by FACTY (task-level, process-level, self-regulation-level, and personal-level). The spreadsheet also includes a column summarizing the students' reflections on their experience with FACTY, as expressed during individual interviews in the fourth stage, where they shared their perceptions of FACTY as a learning resource for algebraic factoring.

The identification of feedback levels was guided by the framework proposed by Söderström and Palm (2024). Excerpts from the shared interactions were examined qualitatively to determine whether FACTY's responses primarily addressed task completion, procedural guidance, self-regulation prompts, or personal evaluations. The analysis focused on the presence of defining features of process-level and self-regulation-level feedback, such as explicit guidance on strategies, prompts for reflection, and invitations to self-check or regulate one's work. This analytical process was descriptive and interpretive, and did not aim to quantify the frequency of feedback types, but rather to characterize the pedagogical function of the feedback provided in the selected cases.

5 Types of Feedback Provided by FACTY

Through the analysis of Guadalupe's and Leonardo's interactions with FACTY, we found that the chatbot primarily provided feedback at two levels: the process-level and the self-regulation-level. The following examples illustrate how FACTY's feedback aligns with these two levels.

5.1 Process-Level Feedback

In one of her interactions, Guadalupe attempted to factor an expression using grouping, but failed to correctly identify the coefficients' greatest common divisor (GCD). FACTY responded with the following message:

"Remember that the common factor should include both numbers and letters. Use prime divisors to find the GCD of the coefficients."

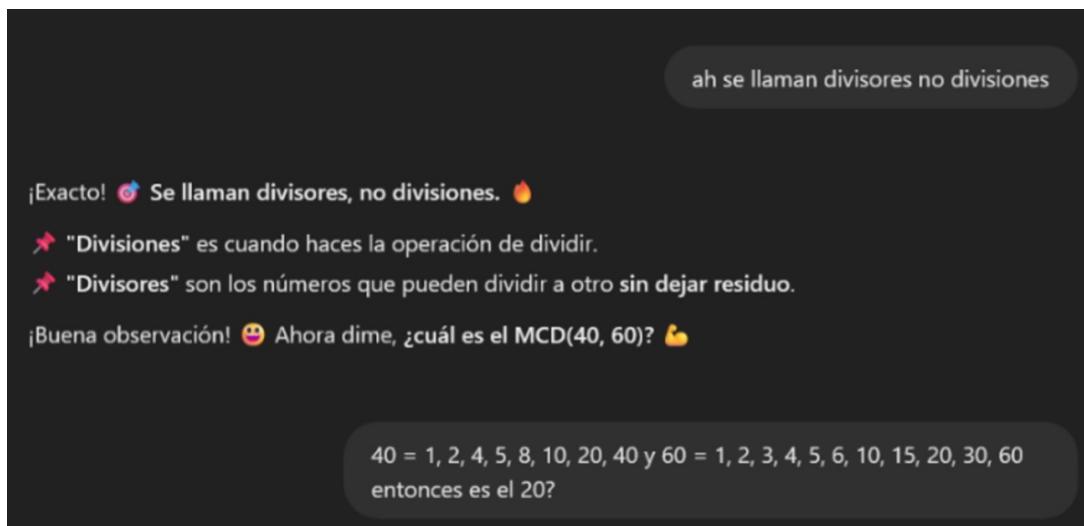
This message exemplifies process-level feedback because it goes beyond simply stating whether an answer is correct or incorrect—as task-level feedback would do—and instead guides the student in

refining the procedure she is using. Specifically, FACTY reminds Guadalupe of a key operational principle in factoring by common factor: that the common factor must include numerical coefficients and literal variables in the terms. In addition, the chatbot suggests a specific strategy—using prime divisors to identify the GCD—that contributes to a deeper understanding and more accurate execution of the method. This kind of guidance focuses on *how* the task is carried out, which is the hallmark of process-level feedback according to Söderström and Palm (2024).

Later in the same sequence, Guadalupe struggled with the concept of prime numbers. FACTY detected that she was confusing the terms *divisions* and *divisors*, which affected her understanding of prime factorization. Rather than merely correcting the error, FACTY offered a conceptual clarification that helped her grasp the distinction in context (Figure 4).

Figure 4

FACTY providing process-level feedback on the concepts of divisors and division



Note. Original interaction in Spanish. Translation: STUDENT: “Oh, they are called divisors, not divisions.” FACTY: “Exactly! They are called divisors, not divisions.” “Divisions’ refers to performing the operation of dividing.” “Divisors’ are the numbers that can divide another number without leaving a remainder.”

This feedback is also at the process-level, as it not only corrects a misunderstanding but deepens the student’s conceptual grasp of a key idea involved in factoring. It helps Guadalupe refine her approach and promotes more meaningful learning.

5.2 Self-Regulation-Level Feedback

In one of his interactions, Leonardo received feedback on a procedural mistake. When he asked FACTY what his error was, the chatbot replied:

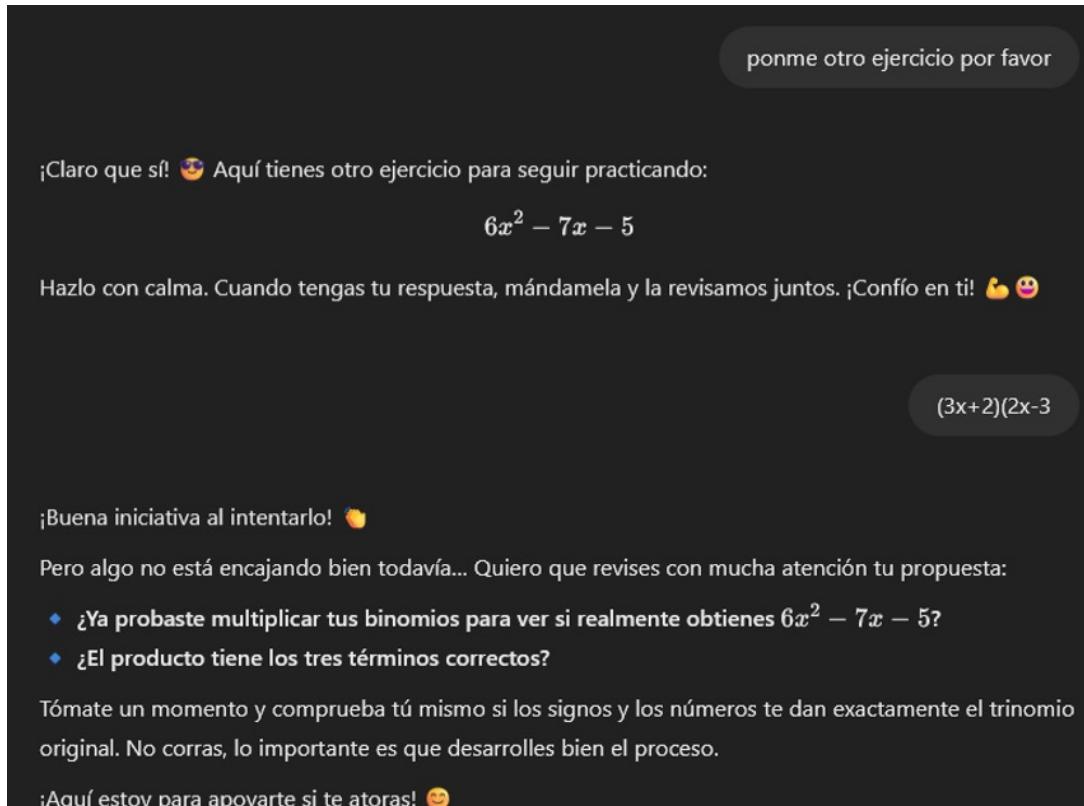
“Where do you think the mistake was this time? Try reviewing your steps one by one. You can use the previous example as a reference.”

Phrases like “Where do you think the mistake was?” and “Try reviewing your steps one by one” encourage students to reflect autonomously on their performance. Rather than providing a direct answer, FACTY prompts self-questioning and refers the student to prior examples, promoting cognitive self-regulation strategies—the richest level of feedback (Söderström & Palm, 2024).

In another interaction, Leonardo submitted his proposed factoring for verification. Rather than simply confirming whether it was correct, FACTY prompted him to check the result himself by multiplying the binomials and reviewing the signs and terms (see Figure 5).

Figure 5

FACTY prompting a student to verify the correctness of a factoring solution



Note. Original interaction in Spanish. Translation: STUDENT: “Give me another exercise, please.” FACTY: “Of course! Here’s another exercise to keep practicing: $6x^2 - 7x - 5$. Take your time. When you have your answer, send it to me and we’ll review it together. I trust you!” STUDENT: “ $(3x + 2)(2x - 3)$ ” FACTY: “Good initiative in trying! But something is still not quite right... Please check your proposal carefully: (1) Have you tried multiplying your binomials to see if you really get $6x^2 - 7x - 5$? (2) Does the product have all three correct terms? Take a moment to check whether the signs and numbers match the original trinomial exactly. Don’t rush—what matters is that you develop the process well. I’m here to support you if you get stuck!”

Again, rather than delivering the solution, FACTY fosters metacognitive behavior by prompting the student to reflect, verify, and redirect his learning. This type of feedback promotes self-assessment and mathematical reasoning, encouraging students to take ownership of the learning process and develop autonomy in their mathematical thinking.

These examples illustrate how analyzing feedback through a multi-level lens allows us to characterize the kind of mathematical support FACTY provides. Such analysis is valuable for understanding students’ difficulties with algebraic factoring and evaluating the pedagogical quality of this AI-based tool. In this sense, feedback analysis serves a dual purpose: it informs our understanding of students’ learning processes and helps assess the didactic potential of intelligent educational technologies like FACTY.

6 Students' Perceptions of FACTY

This section presents selected excerpts from individual interviews conducted during the fourth stage of the implementation. Overall, students expressed positive perceptions of FACTY as a learning tool, highlighting its adaptability, constant availability, and ability to provide immediate, judgment-free feedback. Pseudonyms are used to protect participants' identities. While most experiences were favorable, one case—presented in Section 6.1—illustrates a critical system failure that underscores the importance of preserving teacher oversight when integrating AI-based feedback tools into classroom practice.

“I no longer have to wait for my teacher to reply to my questions by email—if they even reply. At night, I ask FACTY, and it always helps me.” — Christian

“FACTY adapts to my level. If I need it to explain more about a topic, it actually does.” — Guadalupe

“FACTY challenged me to move on to advanced levels. I thought I was good at factoring, but it made me realize I had to go further.” — Apolonio

“I’m too shy to ask questions in class. FACTY doesn’t judge me, so I can ask freely.” — Manuel

These testimonials suggest that FACTY can be a valuable pedagogical complement to classroom instruction. Students emphasized the benefit of receiving immediate support outside of class, which enhances the continuity of learning beyond the classroom's physical space. FACTY's constant availability and capacity to support autonomous interaction were especially appreciated by students who preferred to work at their own pace or had already mastered basic content and sought more advanced challenges. Additionally, some students highlighted that FACTY provided a safe, judgment-free space to ask questions, particularly beneficial for those who feel too shy or anxious to participate actively during in-person lessons.

6.1 Alan's Case: A System Limitation in Practice

During the implementation phase, one particular case exposed the limits and risks of relying exclusively on an AI-based feedback system. Alan, one of the study participants, completed the diagnostic factoring test. As part of the protocol, the test was first reviewed manually by the teacher, who marked correct and incorrect answers on the physical document before scanning it and uploading it to the FACTY system for feedback generation. However, due to a temporary handwriting recognition failure—caused by a maintenance period on the OpenAI platform—FACTY was unable to process the image of the test. Instead of generating feedback based on Alan's handwritten responses, the system accessed an internal database of pre-stored correct answers and produced entirely positive feedback, stating that the student had solved all exercises correctly. This was evidently inaccurate, as the manually reviewed physical exam showed that Alan had answered only two items correctly.

The contrast between the teacher's review and the AI-generated feedback revealed a critical system flaw: at that time, there was no safeguard to prevent feedback from being issued in the absence of a successful reading of the test. While it cannot be determined with certainty that this incident directly caused the student's withdrawal from the study, the fact remains that Alan did not participate in any subsequent stages. This experience underscores the importance of maintaining teacher judgment as the

guiding principle in the use of educational technologies—particularly when there is a risk of erroneous automation that may undermine students' perception and motivation.

From an ethical standpoint, this case highlights the need for accuracy checks and human oversight when AI-based systems are used for educational feedback. In this study, student data were handled with informed consent and anonymization, and the limitations of automated feedback were communicated transparently to participants, reinforcing the teacher's role as the final authority in interpreting and validating AI-generated responses.

7 Conclusion

The implementation of our pedagogical proposal and the analysis of the feedback provided by FACTY suggest that this AI-powered tool can support the learning of algebraic factoring among undergraduate students. The chatbot complements traditional instruction, offering immediate and personalized support when teachers are unavailable. By delivering feedback primarily at the process and self-regulation levels—identified in the literature as the most effective for promoting autonomous and deep learning—FACTY contributes to students' mathematical development.

The performance of Guadalupe and Leonardo in the final assessment offers evidence of this potential. Guadalupe showed significant improvement, increasing her score from 6 correct answers on the diagnostic test to 12 on the final test. This progress reflects the benefits of tailored feedback aligned with her learning needs. In contrast, Leonardo moved from 0 to 2 correct answers, representing a modest gain. However, his limited progress was constrained by foundational gaps in prior knowledge, particularly in basic algebraic operations such as the multiplication of literal expressions. Although FACTY provided appropriate feedback, the student lacked the prerequisite knowledge to benefit from the intervention within the study's timeframe.

These cases illustrate that while FACTY can be an effective support tool, its impact depends on students' prior knowledge and readiness to engage with feedback. Furthermore, using FACTY requires a critical and reflective attitude on the part of the learner. As with any technology, its responses may not always be accurate or complete, and students should be encouraged to question, verify, and discuss the chatbot's output with instructors when needed. The findings reported here should be interpreted within the scope of an exploratory pilot study. The small sample size and the focus on two contrasting cases do not allow for generalization of learning outcomes; instead, they provide insights into how different types of AI-generated feedback may function in students' learning processes. In this sense, the study serves as a proof of concept that informs future implementations.

Some limitations of this study should be acknowledged. First, the absence of a control or comparison group prevents attributing observed learning gains exclusively to the use of FACTY. Second, the study was conducted with a small convenience sample drawn from a single institutional setting, which constrains the transferability of the findings to other educational contexts. Finally, participation was voluntary and limited to a short intervention period, which may have influenced students' engagement and the scope of observed learning outcomes. These limitations underscore the need for future replications with larger, more diverse samples, comparative designs, and more extended implementation periods.

Integrating FACTY into the mathematics classroom represents an innovative pedagogical approach combining artificial intelligence with differentiated instructional feedback principles. Its architecture allows for adaptive and timely feedback that can enhance the learning experience, particularly for students who benefit from working at their own pace, receiving individualized guidance, and developing self-regulatory learning strategies.

AI Use Statement

During the preparation of this manuscript, we used OpenAI's ChatGPT-4 model exclusively to improve the style and grammar of selected portions of the text. All AI-generated content was reviewed and edited by the authors, who retain full responsibility for the final content. No conceptual or analytical sections were generated by AI.

Teacher Implementation Box: Guiding and Monitoring AI-Supported Practice

Setup. FACTY is intended to complement, not replace, classroom instruction. Teachers may introduce the chatbot after an initial diagnostic assessment to help students identify specific algebraic topics requiring reinforcement during autonomous practice.

Oversight. Instructors are encouraged to periodically review selected student–chatbot interactions or request brief reflections from students on how feedback influenced their problem-solving strategies. This helps identify persistent misconceptions or potential misunderstandings arising from AI-generated responses.

Quick Tips. Students should be explicitly instructed to verify FACTY's feedback by checking results algebraically and to consult the teacher when feedback appears unclear or inconsistent. Maintaining teacher oversight is particularly important when AI tools are used for assessment-related tasks or when technical failures may occur.

References

Abore, N. (2020). *Factoring in algebra through a flipped classroom* [Master's thesis, State University of New York]. SUNY open access repository. <http://hdl.handle.net/20.500.12648/6207>

Aguilar, A., Valapai, F., Gallegos, H., Cerón, M., & Reyes, R. (2009). *Matemáticas simplificadas* [Simplified mathematics]. Pearson Educación.

Baldor, A. (2009). *Álgebra* [Algebra]. Compañía Editorial Ultra.

Burhanzade, H., & Aygör, N. (2015). Difficulties that students face during factorization questions. *Procedia – Social and Behavioral Sciences*, 191, 859–864. <https://doi.org/10.1016/j.sbspro.2015.04.140>

Chung, W. H. P. (2012). A case study on teaching and learning of quadratic factoring. *Hong Kong Association for Mathematics Education*, 33, 47–63. http://www.hkame.org.hk/new_html/uploaded_files/magazine/33/503.pdf

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>

Kilgore, K. E., & Capraro, M. M. (2010). A technological approach to teaching factorization. *Journal of Mathematics Education*, 3(2), 115–125. <https://journalofmathed.scholasticahq.com/article/90428>

Olivar, S., Flores, W., & Alvarado, F. (2018). Errores algebraicos en tareas de descomposición factorial por estudiantes universitarios en Nicaragua [Algebraic errors in factoring tasks by university students in Nicaragua]. *Revista Electrónica de Conocimientos, Saberes y Prácticas*, 1(1), 9–27. <http://portal.amelica.org/ameli/jatsRepo/305/3051215002/>

Söderström, S., & Palm, T. (2024). Feedback in mathematics education research: A systematic literature review. *Research in Mathematics Education*, 26(1), 1–23. <https://doi.org/10.1080/14794802.2024.2444321>



Mirelle Zavala Amezcua is a graduate student at CICATA Unidad Legaria, Instituto Politécnico Nacional, Mexico City. Her research focuses on integrating artificial intelligence tools into mathematics education, with particular emphasis on personalized feedback systems for algebraic learning.



Mario Sánchez Aguilar is a Professor in the Mathematics Education Program at CICATA Unidad Legaria, Instituto Politécnico Nacional, Mexico. He holds a Ph.D. from Roskilde University, Denmark. His research focuses on technology in mathematics education and teacher professional development. He has supervised over 35 graduate students and serves on the editorial boards of several mathematics education journals.