



Appropriating formative assessment in Mathematics: reported practices and professional narratives of secondary school teachers

Alejandra Balbi ^a , Micaela Bonilla ^b , Andrea Figarola ^a , Juan Fraile ^{c,*} ,
Cindy Mels ^a

^a Universidad Católica del Uruguay, Montevideo, Uruguay

^b Stanford University, Stanford, USA

^c Universidad Francisco de Vitoria, Madrid, Spain

ARTICLE INFO

Keywords:

Formative assessment
Mathematics education
Teacher professional development
Self-regulated learning
Teacher role
Pedagogical change

ABSTRACT

Formative Assessment (FA) has demonstrated positive effects on learning, yet a gap remains in understanding how it takes shape in specific disciplines such as mathematics. Grasping this situated dimension is essential for advancing theoretical and practical frameworks on how teachers appropriate and re-interpret formative assessment principles under real classroom conditions. This study examines how secondary mathematics teachers construct professional narratives about FA after participating in a professional development program. We conducted a qualitative analysis of 213 meaning units derived from open-ended responses to a final survey completed by 88 teachers and triangulated these data with 135 excerpts from four focus groups involving 24 of the participating teachers. Through inductive and comparative coding, we identified 14 FA practices organized into four dimensions: (1) a culture of dialogue and peer learning, (2) supporting self-regulation, (3) feedback loops, and (4) tasks and dialogues that make mathematical thinking visible. Focus group analysis also revealed two emergent dimensions: reconfiguration of the teacher's role and emotional tensions associated with pedagogical change. The findings show that, in mathematics, FA is appropriated not as a set of discrete techniques but as an interdependent system of practices that reshapes classroom interaction, learning regulation, and the roles of teachers and students. The study contributes evidence on the situated appropriation of FA in mathematics and suggests that its implementation transforms not only classroom practices, but also teachers' understandings of teaching, learning, and their professional role.

1. Introduction and theoretical framework

Two decades have passed since the most influential formulations of the five key strategies of formative assessment were formulated: setting and sharing learning goals, generating feedback that drives progress, fostering peer collaboration, activating students as agents of their own learning, and collecting evidence of learning (Black & Wiliam, 2009; Leahy et al., 2005). Complementarily, other authors expanded this field by proposing models that conceptualize formative assessment as a cyclical and continuous process of collecting, interpreting, and using evidence to regulate teaching and learning (Cizek et al., 2019; Popham, 2008). Both ways of conceptualizing formative assessment have led to proposing FA practices as a continuum between more structured, formal practices and others that are

* Corresponding author at: Universidad Francisco de Vitoria, Carretera Pozuelo-Majadahonda km. 1.8, 28223 Pozuelo de Alarcón, Madrid, Spain.
E-mail address: juan.fraile@ufv.es (J. Fraile).

more informal, spontaneous and situated (Ruiz-Primo & Kroog, 2018). Although these conceptualizations have made a significant contribution to the field, a gap remains in the literature regarding how the continuum of formative assessment practices is understood and narrated by teachers within specific disciplinary contexts. In particular, there is limited empirical evidence on how teachers interpret, articulate, and construct professional accounts of these ideas in relation to their teaching, which is especially relevant given the documented need to strengthen the connection between teacher-education policies and teachers' own professional meaning-making (Allen & Penuel, 2015).

Against this background, the aim of this study is to examine how mathematics teachers construct professional narratives around formative assessment after participating in a professional development initiative based on the five strategies mentioned above. Rather than treating teachers' responses as a simple catalog of techniques, we analyze how they describe, interpret, and give meaning to formative assessment practices in relation to everyday mathematics teaching.

1.1. Assessment for learning and teaching

Although formative assessment is often understood as a mean to optimize student learning, it is also conceived as a strategy that transforms teaching into a process of shared inquiry between teachers and students. From this perspective, formative assessment becomes a teaching strategy that structures pedagogical interaction around generating, interpreting, and using evidence to guide instructional decisions (Popham, 2008; Ruiz-Primo, 2011; Wiliam, 2011).

Ruiz-Primo (2011) expands this conception by introducing the idea of formative assessment as a continuous, multilevel process in which evidence of learning—informal or formal, immediate or cumulative—is gathered and interpreted to serve pedagogical decision-making. From this viewpoint, formative assessment does not occur at a single moment but is embedded in the everyday gestures of classroom life: the questions the teacher asks, the explanations students construct, the reformulations and real-time adjustments that occur. The key lies in using the information obtained to modify teaching and learning while improvement is still possible.

Finally, Popham (2008) defines formative assessment as an intentional, planned process of using evidence, whose value depends less on the instruments used than on the teacher's ability to interpret the information and act on it. From this perspective, the focus is not on data collection but on feedback that guides both teacher and student toward the next steps in learning. Popham emphasizes that assessment is truly "formative" only when the information gathered leads to an instructional or cognitive change.

Taken together, these contributions shape a conceptual framework that conceives formative assessment as a situated professional practice—one that requires the teacher's interpretive judgment to make sense of evidence and translate it into instructional decisions. Today we know which strategies constitute formative assessment and how they should operate in theory; however, far less is known about how mathematics teachers understand and interpret these principles in their everyday practice. This gap motivates the present study, which seeks to explore how teachers themselves narrate and represent their assessment practices and make sense of formative assessment within the specific context of mathematics teaching.

1.2. Formative assessment in mathematics education

In recent years, a growing body of work has argued that formative assessment strategies do not transfer directly across disciplines; rather, they acquire distinctive forms shaped by the ways of knowing, reasoning, and participating that each field fosters. This line of discussion has driven a shift from a "generalist" view of formative assessment toward one that conceives it as a deeply situated practice, anchored in the epistemologies and instructional demands of each discipline (Surma et al., 2025). A recent systematic review of formative assessment training programs reinforces this turn: 60.87% of the studies analyzed adopted subject-specific approaches, showing a growing preference for disciplinary-relevant frameworks (von Hagen et al., 2025).

In mathematics, this trend is pronounced. The systematic review by Maskos et al. (2025), which examined 45 studies published between 2015 and 2023, concludes that while formative assessment is associated with improvements in teaching and learning, its effects are notably heterogeneous. The magnitude and consistency of outcomes depend on the type of mathematical content, teachers' repertoires of practice, and the duration and coherence of implementation. Programs that integrated specific mathematical tasks, opportunities to analyze students' reasoning, and contextualized feedback showed far more robust effects than those based on general approaches. By contrast, short interventions or those misaligned with curricular content tended to produce null or inconsistent results. The review also highlights a persistent limitation: the frequent lack of detail about the concrete practices teachers enact and the mechanisms through which these practices influence mathematical learning. This absence, the authors argue, constitutes an "implementation gap" in empirical research on formative assessment in mathematics.

Critical perspectives support this debate. Bennett (2011) and Dunn and Mulvenon (2009) caution that the impact of formative assessment depends not only on the quality of its implementation but also on the disciplinary domain in which it is embedded. Formative assessment does not manifest uniformly across curricular areas; its effectiveness is deeply tied to the type of reasoning each discipline requires and to the forms of intellectual participation it fosters (Bennett, 2011). In mathematics, this involves interpreting evidence that reveals students' mathematical thinking, distinguishing between conceptual and procedural errors, and offering feedback that supports progress in conceptual understanding (Jacobs et al., 2010). This is work that demands specialized pedagogical judgment grounded in knowledge of mathematical content and reasoning processes (Salamanca et al., 2023).

This complexity is also reflected in Dayal's (2021) study, which examined how two mathematics teachers incorporated formative assessment after participating in a portfolio-based training program. Despite receiving the same intervention, the teachers showed markedly different levels of adoption: while one was able to integrate practices aligned with the formative approach—such as

clarifying success criteria, promoting reflection, and fostering peer work—the other maintained a transmission-oriented and control-centered model of teaching. Dayal concludes that teachers' beliefs about teaching and assessment act as interpretive filters that shape how they adopt and make sense of formative assessment. Only when teachers internalize formative assessment as a tool for supporting learning—rather than as a technical procedure—do more interactive, dialogic, and learning-oriented practices emerge.

Thus, recent literature suggests that the key challenges in the effective implementation of formative assessment in mathematics lie not in the absence of strategies, but in how teachers interpret them, make sense of them, and weave them into their disciplinary practice (Andersson & Palm, 2017; Dayal, 2021; Maskos et al., 2025). Despite this growing body of research, there remains a limited number of qualitative studies that examine in depth how mathematics teachers themselves interpret, internalize, and enact formative assessment within their everyday classroom practice. While systematic reviews and conceptual analyses have documented heterogeneous effects and emphasized the importance of disciplinary alignment, less is known about the concrete meanings teachers attribute to formative assessment and the ways these meanings shape their instructional decisions in mathematics classrooms. Addressing this gap is particularly relevant given the specialized nature of mathematical reasoning and the pedagogical judgment it requires.

1.3. Professional development in formative assessment

Although formative assessment is widely valued by teachers, the literature consistently shows that its effective implementation in classroom practice remains complex and challenging. Key difficulties include the persistent tension with summative assessment and grading cultures, the time demands involved in providing high-quality feedback and creating opportunities for students to revise and improve their work, and the challenge of interpreting evidence of learning beyond right–wrong answers and transforming it into adaptive instructional decisions (Lamberg et al., 2020). In response to these challenges, a growing body of research highlights the central role of sustained professional development in supporting teachers' enactment of formative assessment practices, showing that brief or strategy-focused training is insufficient. Empirical studies indicate that effective programs are characterized by extended duration, engagement with authentic classroom evidence, structured reflection on instructional decision-making, and repeated opportunities to experiment and refine practice in real teaching contexts, thereby increasing the likelihood that what is learned in professional development is meaningfully translated into classroom practice (Gotch et al., 2021; von Hagen et al., 2025).

1.4. The present study

Previous research by our team showed that mathematics teachers who participated in an intensive 80-hour professional development program demonstrated significant improvements in the implementation of formative assessment practices, both overall ($d = 0.838$) and in specific dimensions such as collaboration ($d = 0.640$), feedback ($d = 0.691$), and self-assessment ($d = 0.606$), with medium-to-large effect sizes (Balbi et al., 2025). These improvements were reflected in a greater emphasis on formative feedback, peer assessment, and opportunities for learning-oriented reflection—changes that were also perceived by the students themselves.

However, while the quantitative results are robust, they do not explain how teachers made sense of the formative assessment practices promoted in the training, how they adapted them to mathematics teaching, or how these practices became part of their professional narratives. As noted earlier, generalist approaches to formative assessment are insufficient to capture its conceptual complexity and discipline-specific character (Andersson & Palm, 2017; Dayal, 2021; Maskos et al., 2025). Understanding this complexity requires qualitative approaches that give voice to teachers and illuminate their pedagogical decision-making.

1.5. Aim and research question

This study adopts a qualitative, interpretive approach to examine how mathematics teachers construct professional narratives about formative assessment after participating in a professional development program. We focus on how teachers describe and make sense of formative assessment in their everyday classroom decisions. Rather than identifying or cataloging formative assessment practices, our interest lies in understanding how teachers represent, interpret, and give meaning to these practices through their accounts of mathematics teaching.

Guided by this purpose, the study addresses the following research question: How do mathematics teachers construct professional narratives about formative assessment after participating in a professional development program? By analyzing these narratives, the study seeks to understand how formative assessment principles are interpreted, appropriated, and discursively constructed in relation to the specific demands of mathematical thinking and the realities of classroom teaching. The evidence generated is intended to inform the design of professional development initiatives that foster sustainable, context-sensitive changes in formative assessment.

2. Method

2.1. Research design overview

This study employs a qualitative, interpretive design within a naturalistic paradigm. This approach was selected to move beyond the measurement of the program's effectiveness documented in a previous quantitative phase and to develop an in-depth, situated understanding of how mathematics teachers interpret and enact formative assessment in their specific teaching contexts. The research process was discovery-oriented and drew on procedures commonly associated with this tradition (Charmaz, 2008)—such as inductive

coding, constant comparison, and iterative movement between data and emerging categories.

Data were generated through two complementary methods: an online post-course survey including an open-ended question and a series of semi-structured focus groups with course participants. The survey invited all teachers ($N = 88$) to briefly describe formative assessment practices they considered representative of their everyday work in mathematics, providing a broad overview of strategies in use. Focus groups were then conducted with a purposive subsample of teachers to elicit richer, collectively negotiated accounts of how these practices were implemented, adapted, and justified in specific classroom and school settings. Combining individual written self-reports with group discussions allowed us to triangulate perspectives and to capture the complexity of formative assessment as a situated professional practice rather than a fixed set of techniques.

2.2. Participants and context

2.2.1. Context and researcher positionality

The study was conducted in the context of an intensive 80-hour professional development program on formative assessment for secondary school mathematics teachers in Uruguay, distributed over 16 weeks (approximately 5 h per week) (Balbi et al., 2025). The program was explicitly designed to strengthen teachers' assessment practices around five key strategies of formative assessment, integrating theoretical input with classroom-based tasks and opportunities to analyze students' mathematical thinking.

Authors had prior experience working with in-service mathematics teachers and shared a professional commitment to formative assessment as a tool for supporting learning rather than only certifying achievement. This background informed an interpretive lens focused on understanding how teachers make sense of formative assessment principles within their disciplinary practice.

The first two authors were also involved in the design and facilitation of the professional development program. To manage the potential influence of these dual roles, the team adopted a reflexive stance throughout the study, discussing assumptions in analytic meetings and attending to alternative interpretations of the data. The professional development course and all data collection instruments were administered in Spanish, the participants' working language.

This study was approved by the Ethics Committee of the Universidad Católica del Uruguay (file 200723c).

2.2.2. Participants

A total of 88 Uruguayan secondary school mathematics teachers (81.82% female) completed the professional development course and constituted the pool of participants for the study. Most participants (77.27%) held degrees in mathematics education, which indicates a high level of subject-specific pedagogical preparation and familiarity with disciplinary ways of reasoning and assessing. Teachers taught in the first three years of lower secondary education in Uruguay (equivalent to 7th, 8th and 9th grade), with 42 teaching 7th, 30 teaching 8th, and 38 teaching 9th grade. The majority (72.7%) taught seven groups or fewer annually, with an average of 5.76 classes ($SD = 2.58$, range: 1–14). Class sizes averaged 23.63 students ($SD = 5.44$), and 90% of classes included between 16 and 29 students. This configuration suggests conditions that may be compatible with the individualized feedback and interactive practices emphasized in formative assessment.

Regarding workload, teachers reported an average of 19.53 weekly direct teaching hours ($SD = 10.45$), mainly in the public sector.

In terms of professional experience, the majority (70%) had four years or less of teaching practice, indicating a predominantly early-career cohort.

In addition to the survey completed by all 88 teachers, in-depth qualitative data were generated through four focus groups with 24 volunteers from this same cohort.

2.2.3. Sampling strategy and recruitment

Five weeks after the completion of the course, four focus groups were conducted with a total of 24 secondary school mathematics teachers (six per group). All participants provided informed consent prior to participation.

A purposive sampling strategy was adopted to balance heterogeneity and homogeneity across dimensions relevant to the study. Heterogeneity was sought in terms of institutional governance (public, private, and vocational secondary schools), geographic region, and grade level taught (Grades 7, 8, and 9), to capture a range of contextual perspectives. At the same time, groups were composed to be relatively homogeneous with respect to years of teaching experience. This decision was intended to promote equitable participation and reduce potential asymmetries associated with markedly different professional trajectories.

Table 1
Composition of the Focus Groups.

Focus Group (FG)	Participant Profile
FG1 ($n = 6$ / invited = 14)	Teachers from public, private, and vocational secondary schools, from the Central, Northern Littoral, and Northeastern regions of the country, teaching in Grades 7, 8, and 9. This group comprised teachers with limited professional experience (< 8 years).
FG2 ($n = 6$ / invited = 20)	Teachers from public, private, and vocational secondary schools, from the metropolitan area, the eastern region, and the southern littoral, teaching in Grades 7, 8, and 9. The group consisted of teachers with intermediate professional experience (10–14 years).
FG3 ($n = 6$ / invited = 16)	Teachers from public, private, and vocational secondary schools, predominantly from the metropolitan area and the littoral region, teaching in Grades 7, 8, and 9. This group included teachers with extensive professional experience (15–30 years).
FG4 ($n = 6$ / invited = 11)	Teachers from public, private, and vocational secondary schools, from the Central, Eastern, Northern Littoral, and Northeastern regions of the country, teaching in Grades 7, 8, and 9. The group was composed of teachers with intermediate professional experience (10–20 years).

Participation was voluntary, and invitations were sent by email without financial incentives. The final group size (six participants per focus group) was considered appropriate to facilitate in-depth discussion and balanced interaction. Data collection concluded when discussions displayed rich and recurrent patterns, with subsequent sessions contributing nuance rather than substantially new categories, consistent with thematic saturation for this design. Table 1 summarizes the final composition of the focus groups.

2.2.4. Researcher-participant relationship

The professional development program was presented to teachers primarily as a training opportunity to strengthen formative assessment in mathematics and participation in the research study. To minimize potential power imbalances and ensure that participation did not affect teachers' evaluation in the course, focus groups were scheduled after the training process and associated assessments had concluded. Participants were assured that their contributions would be anonymized and that no individual school or teacher would be identifiable in publications.

2.3. Data collection

Data were collected using a two-phase sequential design that combined an open-ended survey with semi-structured focus groups.

2.3.1. Phase 1: open-ended survey

The first phase involved a self-administered online survey completed immediately after the professional development course. The questionnaire included both closed items and three open-ended questions related to formative assessment. The present study focuses specifically on the post-course prompt: "Name up to three formative assessment practices you use in your classroom". This open-ended question was designed to capture, in teachers' own words, a concise snapshot of the strategies they themselves considered most representative of their everyday work, without constraining their responses through predefined categories or forced-choice formats. The survey was completed individually by teachers via Google Forms, at a time and place of their convenience. Participation in the research component was voluntary, and responses had no implications for course grades or certification.

Across the 88 participating teachers, this item yielded 213 distinct response units. For the purposes of analysis, each unit was defined as a segment of text (typically a sentence or clause) that described a single formative assessment practice.

2.3.2. Phase 2: semi-structured focus groups

To deepen our understanding of how these reported practices were actually enacted and interpreted in classroom settings, we conducted four semi-structured focus groups five weeks after the course ended.

A purposive sample of 24 teachers (six per group), described above, was invited to participate. Sessions were held on Zoom and lasted approximately 60–90 min. Only the moderator and the participating teachers were present during each focus group. Each focus group was moderated by a member of the research team not involved in assessing teachers' coursework.

The focus groups followed a semi-structured protocol designed to guide the conversation while allowing space for emergent themes. The protocol covered three central domains: (a) teachers' experience of the professional development course, (b) the specific formative assessment practices they had implemented in their mathematics lessons, and (c) barriers and enabling conditions for implementation in their schools. The current analysis focuses on the second domain, probing how teachers adapted and integrated formative assessment strategies into the specific demands of the mathematics curriculum and the realities of their classrooms. Questions were open-ended (e.g., "Can you describe a recent lesson in which you used [X]?"", "How did you decide to use this strategy in this particular topic?"), and follow-up probes were tailored to the flow of each discussion. At the beginning of each session, the moderator emphasized the voluntary nature of participation, the non-evaluative character of the space, and the confidentiality of the information shared, with the aim of fostering an honest, dialogic atmosphere in which both successes and difficulties could be discussed.

2.3.3. Recording and data transformation

With participants' informed consent, all focus group sessions were video-recorded online and transcribed verbatim. During transcription, identifying information (e.g., names of teachers, students, or schools) was removed or replaced with pseudonyms to protect confidentiality. Survey responses were exported from the online platform into a spreadsheet and anonymized prior to analysis.

Field notes were taken during and immediately after each focus group to document contextual nuances, group dynamics, and initial analytic impressions. These notes were later used to contextualize the transcripts and to support reflexive consideration of the researchers' role in shaping the discussions.

2.4. Data analysis

We conducted an inductive qualitative thematic analysis drawing on procedures commonly associated with constructivist grounded theory (Charmaz, 2008), such as iterative coding, constant comparison, and the progressive refinement of categories, with the aim of constructing a categorical framework of teachers' formative assessment practices. Our objective was to understand how secondary mathematics teachers describe and make sense of formative assessment after participating in the professional development program.

The analysis was carried out in three sequential stages, moving from a broad categorization of reported practices to a deeper interpretation of how these practices were enacted and interpreted in context. Data management and coding were conducted using

NVivo software, which enabled the organization of codes and the systematic retrieval of excerpts associated with each category.

2.4.1. Stage 1: inductive analysis of survey responses

The first stage focused on teachers' responses to the post-course survey question. The unit of analysis was defined as a unit of meaning: a distinct segment of text (typically a sentence or clause) conveying a complete idea about a formative assessment practice.

The first and third authors independently conducted open coding of all units of meaning, working line by line and staying as close as possible to the language used by the teachers. After independent coding, the authors compared their code lists, discussed discrepancies, and agreed on a harmonized coding scheme. During this process, similar codes were merged, refined, or split as necessary, resulting in a stable set of categories.

The final set of categories summarized the survey responses in terms of distinct formative assessment practices.

2.4.2. Stage 2: analysis of focus group data

Focus group transcripts were analyzed using a hybrid deductive–inductive approach. The 14 survey-based categories from Stage 1 served as a provisional framework. Focus group excerpts were mapped onto these categories when they described, illustrated, expanded, or contextualized one of the formative assessment practices already identified in the survey data. Coding decisions were based on the overall meaning of each excerpt. When an excerpt addressed a dimension of teachers' experience that was not adequately captured by the existing categories, it was coded as emergent. Excerpts were also treated as emergent when their meaning cut across several categories without fitting clearly within a single practice-based code.

Across the focus group data, 98 excerpts were mapped onto the 14 survey-based categories, expanding and refining the understanding of those practices, while 37 excerpts were initially coded as emergent because they were not fully captured by the preliminary framework. After this initial coding, excerpts within each category were compared to identify recurring meanings, variations, and contextual conditions associated with each practice. Mapped excerpts were used to deepen the interpretation of the 14 formative assessment practices, while emergent excerpts were retained for Stage 3, where they were analyzed inductively to identify themes not captured by the initial practice-based framework. This procedure made it possible to examine both the continuity between data sources and the aspects of teachers' accounts that extended beyond the initial categorization.

Both the mapped and emergent excerpts were further examined to identify broader analytical themes that described dimensions associated with the implementation of formative assessment practices.

2.4.3. Stage 3: review and integration of emergent excerpts

The first author reviewed the excerpts initially coded as emergent, re-examining each fragment in light of the analytical themes developed in Stage 2 and the original categorical framework from Stage 1. This process allowed for the refinement of the thematic framework and ensured that emergent excerpts were properly contextualized relative to the predefined categories.

Through iterative comparison across survey responses and focus group excerpts, the 14 formative assessment practices were subsequently organized into four broader pedagogical dimensions that represented different ways of integrating formative assessment into teaching and learning processes. The excerpts initially coded as emergent were further reviewed and consolidated into two cross-cutting dimensions —*Reconfiguration of the Teacher's Role* and *Emotional Tensions Associated with Pedagogical Change*— which were treated analytically as transversal to the full set of identified practices.

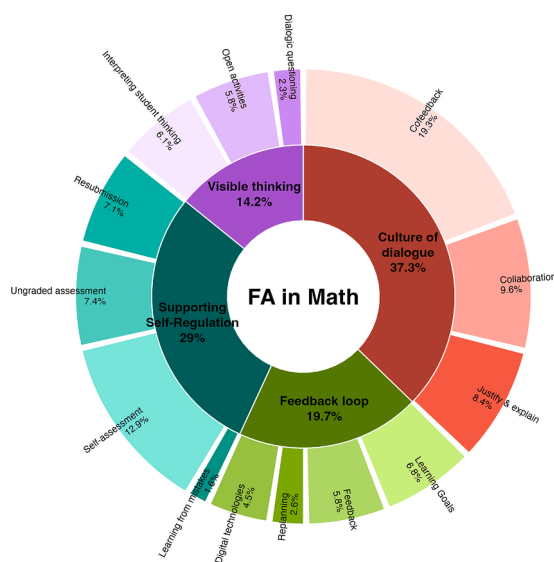


Fig. 1. Configuration of formative assessment practices.

Note. Percentages are calculated using the 311 units of analysis coded to the formative assessment practices.

3. Results

A total of 348 units of analysis (distinct segments of text conveying a complete idea about a formative assessment practice) were identified in the corpus. Fourteen formative assessment practices that teachers reported incorporating into their mathematics lessons after participating in the professional development program accounted for 311 of these units. These practices were organized into four broader dimensions: (1) *dialogue and peer learning culture*, (2) *supporting self-regulation*, (3) *feedback loops*, and (4) *tasks and dialogues that make mathematical thinking visible*. Fig. 1 summarizes this analytical organization. In addition, the focus groups revealed two cross-cutting emergent dimensions—*reconfiguration of the teacher's role* and *emotional tensions associated with pedagogical change*—represented by the remaining 37 units of analysis.

3.1. Dimension 1. Dialogue and peer learning culture

This dimension integrates three practices: cofeedback, collaboration, and justifying and explaining reasoning. Together, these practices describe classrooms where learning is no longer organized exclusively around teacher–student interaction but increasingly relies on student–student interactions that promote the exchange of ideas, argumentation, and reflection on the procedures used to solve mathematical problems.

Within this dimension, cofeedback emerges as one of the most frequently reported practices. Teachers described a range of activities in which students exchange their work, review their peers' solutions, and discuss criteria for improving their responses (Teacher (T)_16; T_65; Focus Group (FG) 1–2; FG4–1). In many cases, these dynamics involve a shift in the teacher's role from directly validating answers to facilitating dialogue among students. As one teacher explained: "*Trying to move away from the role of being the one who says whether something is right or wrong, leaving it more open for dialogue among them so they can exchange ideas and reach an agreement about what is correct and what is not*" (FG1–1). Through these processes, teachers observed that students developed greater motivation for learning, driven both by the opportunity to improve their own work and by their interest in examining and responding to the work of their peers (FG1–3; FG-2).

Peer interactions are also supported by practices that encourage students to explain and justify their mathematical reasoning, whether in pairs, small groups, or whole-class discussions. Teachers mentioned tasks that require students to argue for their procedures, compare different solution strategies, or justify the feedback they provide on a classmate's work (T_11; T_24; T_93).

Finally, several teachers noted that these practices involve explicitly teaching students how to collaborate, recognizing that working in groups does not automatically lead to productive interactions. In the focus groups, some teachers described how they began dedicating time to structuring collaboration through deliberate group composition or by establishing norms for providing useful feedback to peers (FG3–1; FG3–2; FG3–3; FG4–2). As one teacher noted, organizing groups sometimes requires "*stepping out of the comfort zone*" when forming teams that encourage meaningful exchanges among students (T_38). Together, these practices contribute to creating a classroom climate in which discussing, explaining, and revising one's own ideas and those of others becomes a central part of learning mathematics, fostering more participatory and reflective ways of engaging with mathematical thinking.

3.2. Dimension 2. Supporting self-regulation

This dimension brings together four practices: resubmission, self-assessment, learning from mistakes, and assessing without grading. Together, these practices describe an approach to assessment in which learning is understood as a gradual process of improvement, where students have opportunities to revise their work, reflect on their own learning, and use error as a source of understanding.

Within this dimension, the practice of resubmission amplifies the impact of feedback by providing new opportunities for improvement. Teachers described several mechanisms that allow students to revise their work, such as repeating an activity based on the feedback received (T_20), reworking written assessments before their final grading (T_59), or completing tasks in successive stages that lead to an improved final product (T_62). In all cases, revision was conceived as a genuine learning opportunity: "*allowing them to rework some tasks... seeing that error is part of the process*" (T_56).

In the focus groups, several teachers mentioned that they began offering more opportunities for students to redo or improve their work. For example, some reported encouraging students to repeat an activity in the following class or to bring a corrected version of a task to review together (FG1–1). Others described how, after an assessment, students can review their errors, reformulate their answers, and submit a new version of the task (FG1–4; FG2–1). In several cases, teachers observed that these second opportunities foster greater commitment and responsibility among students, who begin to perceive assessments not as a final judgment but as part of their learning process (FG3–1; FG4–1).

A second practice relates to self-assessment. Teachers described various strategies aimed at encouraging students to reflect on how they are learning, identify their progress, and make decisions to improve. These practices include asking questions about how students feel about the activities, what strategies they are using, or what they need to improve (FG2–3). In some cases, teachers incorporated specific moments to review the learning process, inviting students to think about what they had learned during the week, what difficulties they encountered, or what steps remained ahead (FG3–1; FG3–3).

In the focus groups, several teachers noted that promoting self-regulation involves helping students take responsibility for their own learning and develop greater autonomy. As one teacher expressed, the goal is for students to "*know where we are going so that they can self-evaluate and see how they feel about certain contents or skills*" (FG2–4). Likewise, some teachers highlighted that these practices not only foster individual reflection but also collaboration among peers, as students can support one another in understanding the

content and revising their work (FG2–2; FG4–6).

A third practice concerns learning from mistakes, using them as a starting point for students to reflect on the reasoning involved in problem solving. Rather than simply pointing out incorrect answers, activities are oriented toward exploring why a particular error occurred and what ideas or interpretations underpin it (T_6; T_16). In the focus groups, some teachers noted that this approach requires students to develop a more analytical perspective on errors, interpreting them as indicators of mathematical thinking rather than merely mistakes to be corrected (FG2–1).

Finally, several teachers mentioned the practice of assessing without grading as a strategy to keep the learning process open. Instead of immediately assigning a numerical grade, some teachers choose to provide qualitative comments that guide the revision of the work and allow students to improve their tasks before the evaluation is closed. As one teacher noted, they now avoid “*grading numerically... and instead provide oral or written comments highlighting achievements and suggesting how to improve*” (T_18). Others reported prioritizing written or oral comments without including a grade, so that feedback can be used to revise and improve students’ work before the final evaluation (T_12; T_40; T_59).

3.3. Dimension 3. Feedback loops

This dimension brings together four interrelated practices: replanning, clarifying goals and success criteria, feedback, and the use of digital technologies. Together, these practices describe a process in which feedback is no longer understood as an isolated moment but rather as a loop that connects the definition of goals, the interpretation of evidence, the feedback provided to students, and the pedagogical decisions made on the basis of that information.

Within this dimension, the first practice refers to replanning based on evidence of learning. Teachers described situations in which they use evidence gathered in activities or assessments to adjust instruction, modify the pace of teaching, or revise the sequence of content. For example, some mentioned reviewing assessments to identify recurring errors and deciding to “*modify the annual planning because I realized I had to go back to a topic*” (T_29). In the focus groups, several teachers noted that they began to view the records of classroom activities as evidence of both students’ learning and their difficulties, using this information to decide which content to review or where to place greater emphasis (FG1–1; FG1–2). In some cases, this process also involved reflection on their own teaching practice, evaluating whether the decisions taken were supporting learning or needed to be adjusted (FG2–1).

A second practice concerns the clarification of goals and success criteria. Teachers described different strategies for sharing with students the purpose of activities and the expected learning outcomes. These include explaining goals at the beginning of a class or a unit (T_22), presenting lists of knowledge and skills to be developed (T_38), or using rubrics and checklists that make evaluation criteria explicit (T_35). In the focus groups, several teachers reported that they began to share learning objectives more explicitly, recognizing that they had not always done so before (FG2–1; FG1–1). This practice enables students to understand the direction of classroom work and to reflect on their own progress. As one teacher explained, sharing the objective involves showing “*where we are going and what we want to achieve*”, preventing students from carrying out activities without understanding the purpose of the process (FG1–2). Likewise, some teachers mentioned the use of rubrics or checklists before assessments so that students know what is expected of them and can anticipate the criteria by which their work will be evaluated (FG3–1; FG4–1; FG4–3).

A third practice relates to feedback oriented toward learning. Teachers described feedback aimed at understanding students’ thinking and promoting improvement in their work. This feedback can take different forms—oral, written, individual, or collective—and seeks to provide specific comments that help students identify what they have achieved, what they need to review, and how they can continue progressing (T_46; T_48; T_50). In the focus groups, teachers noted that they began paying greater attention to the timing and quality of the feedback they provide, recognizing that “*not every feedback is useful*” for supporting learning (FG1–1). Some also described feedback as a cyclical process that includes reviewing errors, reformulating responses, and generating new opportunities for improvement (FG2–1). In this sense, feedback is understood as part of a continuous process that involves both teachers and students, as well as their peers (FG3–1).

Finally, several teachers mentioned the use of digital technologies as mediators of feedback. Digital tools such as quizzes on educational platforms, interactive applications, or student response systems were used to provide rapid feedback and collect evidence about students’ learning (T_2; T_12; T_60). These tools allow, for example, students to receive automatic feedback on their responses or to review their attempts during an activity. In the focus groups, some teachers noted that the course allowed them to discover new digital tools and begin using them in their classes, even when they had previously felt uncomfortable with technology (FG2–1). Others mentioned the use of educational platforms to provide immediate feedback during activities, which facilitates collecting evidence of learning in real time and adjusting instruction based on that information (FG3–1).

Together, these practices configure a feedback loop in which learning goals, classroom evidence, the comments offered to students, and pedagogical decisions are dynamically connected, allowing teaching and learning to be continuously adjusted to the needs that emerge in the classroom.

3.4. Dimension 4. Tasks and dialogues that make mathematical thinking visible

This dimension integrates three practices: open-ended tasks, dialogic questioning, and teacher interpreting students’ thinking. Together, these practices describe classrooms in which tasks and interactions make students’ mathematical thinking visible, providing teachers with opportunities to observe their reasoning, interpret their ideas, and understand how they approach problem solving.

Within this dimension, open-ended tasks occupy a central place. Many teachers reported using low-floor, high-ceiling tasks, open-ended problems, or prompts that allow multiple solution paths (T_6; T_22; T_73). These tasks were valued because they enable teachers

to observe how students justify their procedures, compare strategies, and discuss different possible solutions. At the same time, teachers highlighted that this type of activity allows students with different levels of performance to participate in mathematical activity. As one teacher noted, these tasks make it possible to identify “whether I can continue deepening the concept or whether it is necessary to reinforce what has already been worked on” (T_22). In this sense, open tasks function as windows into students’ mathematical thinking, as they allow access not only to the results obtained but also to the reasoning processes that support them.

These tasks are closely connected to the teacher’s role in generating dialogue around mathematical ideas. Teachers described situations in which they encourage students to explain their procedures and discuss different ways of solving a problem (T_11; T_73). These exchanges offer teachers opportunities to understand how students are interpreting the concepts addressed in class.

Finally, teachers emphasized the importance of developing professional noticing to interpret evidence of students’ thinking. This involves paying closer attention to the comments, errors, and interactions that occur during class in order to understand ongoing learning processes. In the focus groups, several teachers mentioned that they began listening more carefully to what students say while working on tasks, recognizing that previously they tended to speak too much and did not always give themselves time to observe and listen to how students were thinking (FG1–1). Other teachers described creating informal records to follow students’ progress across lessons, considering not only the results obtained but also aspects such as participation, collaboration, and the ways students respond to difficulties (FG1–2; FG1–3). Some also noted that interpreting these signals requires training and reflection, as it involves learning to recognize patterns in students’ comments and errors and using this information to guide instructional decisions (FG2–1; FG2–2).

Together, these practices position classroom tasks and dialogues as key mechanisms for making students’ mathematical thinking visible, enabling teachers to gather evidence from students’ reasoning and interactions and use that information to support learning.

3.5. Emergent dimension: Reconfiguration of the teacher’s role

One of the emergent findings refers to a reconfiguration of the teacher’s role in the classroom. Participants described a process of revising their own practices and beliefs about teaching, which involved moving from a role more focused on controlling the class, transmitting content, and validating answers, toward a position more oriented to facilitating learning, listening to students, and supporting their reasoning.

In several interventions, teachers noted that they began to question the central position they had traditionally occupied in classroom dynamics. Some acknowledged that, prior to the professional development experience, they tended to monopolize speaking time and assumed that they were the ones who had to lead the activity continuously. As one teacher expressed, when reflecting on her practice she realized that “*the one who spoke the most in class was me*”, which led her to try to listen more to her students and give them a greater role in constructing mathematical ideas (FG1–2). Similarly, other teachers mentioned that they learned to “*step out of the position of being the one who says whether something is right or wrong*”, creating spaces where students discuss, argue, and validate their own responses (FG1–5).

This change also involved partially relinquishing control of the classroom and accepting more open working dynamics. Some teachers noted that they began to feel more comfortable with group work, even when it involved more movement or conversation among students, interpreting this “noise” as a sign that the class was active and that students were engaged in their work (FG1–3).

This dimension also captured reflections related to teachers’ professional identity. Several participants reported revisiting assumptions about their own practice, questioning routines that they had naturalized over time. Some described this experience as an opportunity to “critique their own practice” and recognize that there were still aspects they could improve (FG2–1). Others emphasized that exchanging experiences with colleagues and accessing conceptual frameworks grounded in theory and scientific evidence strengthened their confidence in certain practices they had previously carried out intuitively, but could now support with stronger theoretical foundations (FG3–1).

Finally, some teachers summarized this change as a transition from a “judge” to a “coach” role in the classroom. Rather than focusing on evaluating or certifying students’ performance, the teacher’s role shifts toward accompanying students’ learning processes (FG4–1). Taken together, these reflections suggest that implementing formative assessment practices not only introduces new instructional strategies but also promotes a deeper redefinition of the teacher’s role within classroom dynamics.

3.6. Emergent dimension: emotional tensions associated with pedagogical change

A second emergent dimension refers to the emotional tensions that accompany processes of pedagogical change. In the focus groups, several teachers described that incorporating formative assessment practices not only involved learning new instructional strategies, but also navigating processes of uncertainty, discomfort, and questioning deeply rooted professional habits.

In particular, some teachers noted that modifying the traditional dynamics of the classroom initially generated feelings of anxiety or loss of control. For example, one teacher explained that when she allowed students to work more autonomously, she felt uneasy and even guilty, because she had the impression that “they were doing everything and I was just standing there doing nothing” (FG1–2). These emotions appear to be linked to previous conceptions of the teacher’s role, in which the visible activity of the teacher is associated with effective teaching.

Similarly, several teachers mentioned that it was challenging to accept more open classroom dynamics, where group work involves greater movement and conversation among students. In some cases, this required revising beliefs about what constitutes an “orderly” or “controlled” classroom, gradually learning to interpret these interactions as part of the learning process (FG1–1; FG1–3).

Another source of tension mentioned by teachers relates to pressures associated with time, planning, and expectations of achievement. Some participants highlighted that implementing more open practices requires recognizing that students progress at

different paces and that activities may require more time than initially planned. This situation often led teachers to reconsider decisions regarding the selection of content and the pacing of the course (FG1–5). In this sense, pedagogical change also involved learning to make decisions based on evidence of student learning, even when this meant modifying the original planning (FG1–4).

In addition, several teachers pointed out that the transition toward more formative forms of assessment generated tensions related to the strong cultural emphasis on grading. Some described difficulties in shifting students' focus away from grades and toward the learning process, noting that this change requires time and persistence. As one teacher expressed, encouraging students to focus on the process rather than the grade was "a constant struggle" throughout the year (FG2–2).

Taken together, these reflections suggest that the implementation of formative assessment practices not only entails pedagogical transformations but also involves emotional and professional processes of adaptation, in which teachers confront uncertainties and revisit their beliefs about teaching and learning.

4. Discussion

This study shows that, after participating in a professional development program on formative assessment, mathematics teachers not only reported incorporating specific practices into their lessons, but also reconfiguring the way they understand teaching, learning, and classroom participation. The fourteen practices identified, organized into four dimensions, together with the two emergent dimensions, suggest that formative assessment is appropriated in mathematics not as a repertoire of isolated techniques, but as an interdependent system of practices that reshapes classroom interactions, the regulation of learning, and the positions occupied by teachers and students in mathematical activity.

At the outset, we expected the five strategies of [William's \(2011\)](#) model to emerge rather directly in the practices reported by teachers, since these strategies were introduced in the professional development program. Indeed, our findings show that these strategies are present in the fourteen practices identified. The fact that teachers' accounts were produced after their participation in the program is therefore central to how these findings should be read. From this perspective, the narratives analyzed here offer insight into how teachers began to integrate what they had encountered in the program into their own pedagogical reasoning. This corresponds to the intermediate stages of [Desimone's \(2009\)](#) theory of action, in which professional development is expected to shape teachers' knowledge, skills, attitudes, and beliefs, and then support changes in instructional practice. Rather than providing a neutral description of pre-existing practices, these accounts show which formative assessment principles teachers appropriated, how they adapted them, and how they connected them to the specific demands of mathematics teaching. This is particularly relevant given that effective professional development is not only a matter of transmitting strategies, but of supporting teachers' capacity to interpret classroom situations and respond flexibly to students' thinking ([Kennedy, 2019](#)). In this sense, the findings can be understood as evidence of a process of professional sense-making, in which formative assessment was not described simply as a set of techniques to be applied, but as a way of rethinking mathematical learning, student participation, and pedagogical decision-making.

Therefore, the main contribution of the study does not lie merely in confirming that what was taught in the professional development program appears in teachers' narratives, but in showing that those ideas were appropriated, adapted, and reinterpreted. When formative assessment is reinterpreted within the context of mathematics teaching, these strategies no longer appear as isolated elements and instead become articulated into broader dimensions that point to a deeper pedagogical transformation.

In this sense, the findings support the view of formative assessment as a continuous process of generating, interpreting, and using evidence to guide pedagogical decisions ([Ruiz-Primo, 2011](#)), while also extending this idea by showing how that principle is enacted through an architecture of situated practices. The four dimensions identified — dialogue and peer learning culture, supporting self-regulation, feedback loops, and tasks and dialogues that make mathematical thinking visible — suggest that formative assessment does not simply modify particular assessment strategies, but introduces a different way of understanding how learning occurs and how it should be supported in the mathematics classroom.

4.1. Dimension 1. Dialogue and peer learning culture

The first dimension, centered on a culture of dialogue and peer learning, highlights that formative assessment takes on a particularly powerful expression when the classroom is no longer organized exclusively around teacher–student interaction and is also sustained by exchanges among students. Within this framework, practices such as cofeedback, collaboration, and the justification of reasoning not only expand the sources of feedback available, but also make argumentation and the revision of procedures constitutive parts of mathematical learning. This interpretation is consistent with [Biton's \(2025\)](#) study, which shows that peer assessment can foster deep learning by helping students recognize knowledge they did not initially perceive in themselves, and with [Patchan et al. \(2022\)](#), who found that peer feedback can be as effective as teacher feedback when students are exposed to multiple perspectives from their classmates. Our findings reinforce this line of research and further suggest that such practices require explicit teaching: collaboration, argumentation, and the provision of useful feedback do not emerge automatically, but form part of a classroom culture that teachers need to build deliberately.

4.2. Dimension 2. Supporting self-regulation

The second dimension points to a continuous and self-regulated conception of learning, in which assessment does not close the process, but prolongs it. The postponement of grading, opportunities for resubmission, self-assessment, and work on error appeared as conditions that sustain improvement and turn assessment into an input for cycles of revision, re-elaboration, and dialogue between

teachers and students. In the context of mathematics teaching, this logic is consistent with evidence linking peer feedback practices to greater clarity about learning goals, closer monitoring of one's own work, and greater enjoyment of the activity (Vattøy & Gamlem, 2024). Taken together, our findings suggest that self-regulation does not emerge as an individual skill detached from teaching, but rather as a pedagogical construction sustained by recurring opportunities for revision, reflection, and improvement.

4.3. Dimension 3. Feedback loops

The third dimension shows that formative assessment becomes consolidated when goals, feedback, replanning, and the use of evidence are articulated in a genuine feedback loop. This finding speaks to the literature that understands formative assessment as a cyclical, regulatory process rather than a one-off event (Ruiz-Primo, 2011), and as a continuous sequence oriented by goals, evidence, and subsequent pedagogical action (Maskos et al., 2025). Within this logic, making goals and criteria explicit did not operate as a rigid prescription, but as a shared reference point that oriented students' participation and sustained the regulation of their work. This result is consistent with international evidence on self-regulation and goal clarity (Panadero, 2023), while also adding an important nuance for mathematics education. Faced with the difficulty — previously documented by our own team — of sharing goals in a discipline where many objectives emerge during the development of the task itself (Balbi et al., 2022), teachers appeared to resolve this tension flexibly. Rather than eliminating that disciplinary specificity, they integrated it through the use of rubrics, checklists, and meta-cognitive questions, allowing goals to function as supports for feedback without abandoning a problem-solving approach or the didactic value of error.

4.4. Dimension 4. Tasks and dialogues that make mathematical thinking visible

The fourth dimension underscores the role of tasks and dialogue as mechanisms for making students' mathematical thinking visible. In particular, teachers described the use of problem-solving approaches, open-ended tasks, and multiple solution paths as resources that allow them to gather evidence about how students are reasoning while solving problems. This finding is in line with Olivares (2024), who shows that teachers value open approaches because of their capacity to promote critical thinking, collaboration, and autonomy. Our findings extend this idea by suggesting that such tasks not only support richer forms of learning, but also make it easier to respond to the diversity of attainment levels present in the classroom. By enabling different strategies, forms of participation, and mathematical productions, these tasks avoid reducing students with greater difficulties to closed or overly directive formats and make visible capabilities that might remain unnoticed in single-answer tasks. In this regard, the findings of Ron-Ezra and Levenson (2025) further support the view that open tasks can reveal valuable forms of reasoning even among students with learning difficulties. From this perspective, formative assessment may also be understood as a potentially inclusive practice.

At the same time, making students' thinking visible does not in itself guarantee appropriate pedagogical decisions. Here, our findings resonate with Parodi et al. (2024), who argue that noticing and interpreting signs of students' thinking is only one part of the process, since deciding how to act pedagogically on that basis requires an even more complex form of expertise. Similarly, Leatham et al. (2015) conceptualize mathematically significant pedagogical opportunities as classroom moments in which student thinking can be used to support engagement with important mathematical ideas. The accounts provided by our participants suggest that recognizing and acting on such opportunities involves a professional sensitivity that is progressively refined through everyday classroom interaction, as teachers learn to distinguish between routine actions and moments that open possibilities for deeper mathematical discussion. Consequently, professional development should not be limited to supporting the design of open tasks, but should also include the development of teachers' capacity to interpret such opportunities and respond to them pedagogically.

4.5. Emergent dimensions: reconfiguration of the teacher's role and emotional tensions associated with pedagogical change

Finally, the two emergent dimensions indicate that the incorporation of formative assessment practices mobilizes an identity-related and emotional reconfiguration of teachers' work. The convergence between the redefinition of the teacher's role and the emotional tensions associated with pedagogical change shows that implementing formative assessment does not simply involve incorporating new strategies, but revisiting deeply rooted assumptions about control, authority, grading, and responsibility for learning. In this respect, our findings are in dialogue with literature showing that reforms are reinterpreted through processes of professional sensemaking (Coburn, 2001), that such processes often activate vulnerability and uncertainty (Kelchtermans, 2005), and that emotions are constitutive of the way teachers experience and negotiate change in practice (Zembylas, 2003). From this perspective, professional development in formative assessment should not be limited to teaching strategies; it should also support the working through of the emotional and professional tensions that emerge when the focus shifts from teaching and grading toward listening, interpreting, accompanying, and progressively ceding greater protagonism to students.

4.6. Implications for formative assessment in mathematics classrooms

Taken together, these findings have specific implications for formative assessment in mathematics classrooms. First, they suggest that formative assessment should not be understood only as the use of general assessment strategies, but as a way of organizing classroom activity around the elicitation, interpretation, and use of students' mathematical thinking. Teachers need to attend to, interpret, and respond to students' mathematical reasoning as central to instructional decision-making. In this sense, problem-solving approaches, open-ended tasks, and multiple solution paths are especially relevant because they generate rich evidence of how students

reason, make connections, and structure conceptual understanding.

Second, the findings show that this evidence becomes pedagogically useful when it is embedded in a classroom culture of dialogue, peer interaction, feedback, and revision. When students are invited to justify strategies, compare solution paths, revise their work, and respond to feedback, assessment becomes part of the mathematical activity itself. These formative dialogues create valuable opportunities to revisit misconceptions, make connections between mathematical ideas, clarify meanings, and revise emerging conceptual understandings in real-time. Peer assessment and self-regulated learning are central mechanisms through which students learn to use feedback, monitor their own understanding, and participate in the progressive refinement of mathematical ideas.

Third, implementing formative assessment in mathematics requires a shift in the teacher's role. Teachers are not positioned primarily as evaluators or providers of correct procedures, but as interpreters of students' thinking and facilitators of dialogue, revision, and shared responsibility for learning. This repositioning can also expand students' agency, as they are invited to participate more actively in interpreting, revising, and advancing mathematical ideas. This highlights the contribution of the study to mathematics education by showing how formative assessment can support not only monitoring and feedback, but also mathematical reasoning, conceptual understanding, participation, and teacher professional change.

4.7. Limitations

This qualitative study has several limitations that should be considered when interpreting the findings. First, the study is situated in a single national context—an intensive professional development program for secondary mathematics teachers in Uruguay—and uses a non-probabilistic sample composed of teachers who completed the course and agreed to participate in the research. This configuration limits the generalizability of the results; rather than claiming representativeness, our goal is to offer an in-depth, contextually grounded account that may be transferable to similar settings. In line with contemporary discussions of qualitative research quality, the contribution of this study lies in its “situated” and “contextual” resonance, rather than in broad population-level inference (Maxwell, 2013).

Second, the sample is predominantly composed of early-career teachers: 70% reported four years of teaching experience or less. This profile is relevant because novice teachers may be especially willing to experiment with new approaches such as formative assessment and may experience professional development differently from more experienced colleagues. The practices identified here therefore reflect how an early-career cohort appropriated formative assessment in mathematics; patterns might differ in more experienced or more resistant teacher populations, or in systems with different assessment traditions.

Third, the study was conducted five weeks after the completion of the professional development program. While this allowed for initial experimentation, it does not guarantee the long-term sustainability of these practices. Literature on teacher change suggests that deep pedagogical transformation is a slow, non-linear process often marked by an initial “implementation dip” or gradual adaptation (Fullan, 2015). Consequently, the findings reflect an early stage of appropriation, and it remains unknown how these practices might evolve or stabilize over a full academic year.

4.8. Future research directions

The findings of this study open several avenues for future research aimed at bridging the gap between theoretical frameworks and classroom reality in mathematics education. First, to overcome the limitations of self-reporting, future studies should incorporate ethnographic or video-based observational designs. Using, for example, qualitative video analysis would allow researchers to examine the “fidelity of implementation” (O'Donnell, 2008) and, more importantly, the quality of the interactions during moments of implementation of formative assessment practices and, particularly, error analysis.

Second, given that formative assessment is a co-constructed process, it is imperative to include the student voice. Future inquiries should investigate how students perceive and utilize these new feedback practices. Specifically, research using the framework of “student feedback literacy” (Carless & Boud, 2018) could illuminate whether the shift in teachers' practices translates into greater student agency and self-regulation in mathematics, or if students continue to interpret these activities through a performance-oriented lens.

Third, longitudinal studies are needed to track the trajectory of teacher change over longer periods. It would be particularly relevant to investigate the mechanisms that support the sustainability of situated formative assessment, such as the role of professional learning communities or department-level leadership in schools (Vangrieken et al., 2017).

5. Conclusion

This study shows that, when appropriated by mathematics teachers through professional development, formative assessment does not operate as a set of isolated strategies, but as an interdependent system of practices that reorganizes teaching, learning, and participation in the classroom. The four dimensions identified delineate a pedagogical architecture centered on dialogue, the visibility of students' mathematical thinking, feedback, and self-regulation, while the two emergent dimensions show that this process also involves a reconfiguration of the teacher's role and the emotional experience of pedagogical change. Taken together, these findings suggest that promoting formative assessment in mathematics requires not only providing teachers with pedagogical strategies, but also creating conditions that support deep, situated, and professionally meaningful processes of change.

Funding

(1) Agencia Nacional de Investigación e Innovación [FSED_2_2019_1_156641]. (2) Universidad Francisco de Vitoria, Call for Research Educational Innovation 2025 [UFV2025-75].

CRediT authorship contribution statement

Alejandra Balbi: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Micaela Bonilla:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrea Figarola:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Data curation, Conceptualization. **Juan Fraile:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Cindy Mels:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare to not have any conflict of interest regarding this manuscript.

References

- Allen, C. D., & Penuel, W. R. (2015). Studying teachers' Sensemaking to investigate teachers' Responses to professional development focused on new standards. *Journal of Teacher Education*, 66(2), 136–149. <https://doi.org/10.1177/0022487114560646>
- Andersson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and Instruction*, 49, 92–102. <https://doi.org/10.1016/j.learninstruc.2016.12.006>
- Balbi, A., Bonilla, M., Curione, K., Ibarra, A., & Menese, P. (2025). Impact of a professional development program in formative assessment for mathematics teachers. *Professional Development in Education*, 51(1), 7–22. <https://doi.org/10.1080/19415257.2024.2433051>
- Balbi, A., Bonilla, M., Otamendi, M. A., Curione, K., & Beltrán-Pellicer, P. (2022). Formative assessment and mathematics education: The perspective of In-service mathematics teachers. *Acta Scientiae*, 24(6), 236–269. <https://doi.org/10.17648/acta.scientiae.7043>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in education. Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Biton, Y. (2025). Learning mathematics through peer assessment: "how can we assess something that we ourselves don't know how to solve?". *Eurasia Journal of Mathematics, Science and Technology Education*, 21(1), Article em2557. <https://doi.org/10.29333/ejmste/15794>
- Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Charmaz, K. (2008). Constructionism and the Grounded Theory Method. In J. A. Holstein, & J. F. Gubrium (Eds.), *Handbook of Constructionist Research* (pp. 397–412). The Guilford Press.
- Cizek, G. J., Andrade, H. L., & Bennett, R. E. (2019). Formative assessment: History, definition, and progress. In H. L. Andrade, R. E. Bennett, & G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 3–19). Routledge.
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145–170.
- Dayal, H. C. (2021). How teachers use formative assessment strategies during teaching: Evidence from the classroom. *Australian Journal of Teacher Education*, 46(7), 1–21. <https://doi.org/10.14221/ajte.2021v46n7.1>
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189X08331140>
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, 14(7). <https://doi.org/10.7275/jg4h-rb87>
- Fullan, M. (2015). *The new meaning of educational change*. Teachers' College Press.
- Gotch, C. M., Poppen, M. I., Razo, J. E., & Modderman, S. (2021). Examination of teacher formative assessment self-efficacy development across a professional learning experience. *Teacher Development*, 25(4), 534–548. <https://doi.org/10.1080/13664530.2021.1943503>
- Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, 41(2), 169–202. <https://doi.org/10.5951/jresmetheduc.41.2.0169>
- Kelchtermans, G. (2005). Teachers' emotions in educational reforms: Self-understanding, vulnerable commitment and micropolitical literacy. *Teaching and Teacher Education*, 21(8), 995–1006. <https://doi.org/10.1016/j.tate.2005.06.009>
- Kennedy, M. M. (2019). How we learn about teacher learning. *Review of Research in Education*, 43(1), 138–162. <https://doi.org/10.3102/0091732X19838970>
- Lambert, T., Gillette-Koyen, L., & Moss, D. (2020). Supporting teachers to use formative assessment for adaptive decision making. *Mathematics Teacher Educator*, 8(2), 37–58. <https://doi.org/10.5951/mte-2019-0005>
- Leahy, S., Lyon, C., & Thompson, M., & William, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63(3), 18–24.
- Leatham, K. R., Peterson, B. E., Stockero, S. L., & Van Zoest, L. R. (2015). Conceptualizing mathematically significant pedagogical opportunities to build on student thinking. *Journal for Research in Mathematics Education*, 46(1), 88–124. <https://doi.org/10.5951/jresmetheduc.46.1.0088>
- Maskos, K., Schulz, A., Oeksuez, S. S., & Rakoczy, K. (2025). Formative assessment in Mathematics education: A systematic review. *ZDM - Mathematics Education*, 57(4), 679–693. <https://doi.org/10.1007/s11858-025-01696-x>
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach*. SAGE Publications, 3rd ed.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, 78(1), 33–84. <https://doi.org/10.3102/0034654307313793>
- Olivares, D. (2024). A Socio-Constructivist perspective on problem-solving approaches in mathematics: Perceptions of future primary education teachers. *International Journal of Learning, Teaching and Educational Research*, 23(9), 220–241. <https://doi.org/10.26803/ijlter.23.9.12>
- Panadero, E. (2023). Toward a paradigm shift in feedback research: Five further steps influenced by self-regulated learning theory. *Educational Psychologist*, 58(3), 193–204. <https://doi.org/10.1080/00461520.2023.2223642>
- Parodi Escobal, S., Ochoviet Filgueiras, C., & Lezama Andalon, J. (2024). Mirada profesional del futuro profesor en torno al signo igual. *Enseñanza de las Ciencias Revista de investigación y experiencias didácticas*, 42(1), 43–63. <https://doi.org/10.5565/rev/ensciencias.5806>

- Patchan, M. M., Rambo-Hernandez, K. E., Deitz, B. N., & McNeill, J. (2022). Using peer assessment to improve middle school mathematical communication. *Journal of Educational Research*, 115(2), 146–160. <https://doi.org/10.1080/00220671.2022.2074948>
- Popham, W. J. (2008). *Transformative assessment*. ASCD.
- Ron-Ezra, M., & Levenson, E. S. (2025). Exploring how students with learning disorders engage with open arithmetic tasks using a mathematical creativity lens. *Educational Studies in Mathematics*, 120, 587–608. <https://doi.org/10.1007/s10649-025-10426-9>
- Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37(1), 15–24. <https://doi.org/10.1016/j.stueduc.2011.04.003>
- Ruiz-Primo, M. A., & Kroog, H. (2018). Looking closely at mathematics and science classroom feedback practices: Examining artifacts, students' Products, and teachers' Communications. Eds. In A. A. Lipnevich, & J. K. Smith (Eds.), *The cambridge handbook of instructional feedback* (pp. 191–218). Cambridge University Press.
- Salamanca, E., Medina-Castellano, K. A., Wilhelm, J. A., & Fisher, M. H. (2023). How teaching experience and physics and mathematics content knowledge impact professional noticing skills of STEM graduate students. *School Science and Mathematics*, 123(7), 387–397. <https://doi.org/10.1111/ssm.12568>
- Surma, T., Vanhees, C., Wils, M., Nijlunsing, J., Crato, N., Hattie, J., Muijs, D., Rata, E., William, D., & Kirschner, P. A. (2025). How knowledge matters. Eds. In T. Surma, C. Vanhees, M. Wils, J. Nijlunsing, N. Crato, J. Hattie, D. Muijs, E. Rata, D. William, & P. A. Kirschner (Eds.), *Developing curriculum for deep thinking: The knowledge revival* (pp. 5–35). Springer. https://doi.org/10.1007/978-3-031-74661-1_2.
- Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review. *Teaching and Teacher Education*, 61, 47–59. <https://doi.org/10.1016/j.tate.2016.10.001>
- Vattøy, K. D., & Gamlem, S. M. (2024). Students' experiences of peer feedback practices as related to awareness raising of learning goals, self-monitoring, self-efficacy, anxiety, and enjoyment in teaching EFL and mathematics. *Scandinavian Journal of Educational Research*, 68(5), 904–918. <https://doi.org/10.1080/00313831.2023.2192772>
- von Hagen, A., Balbi, A., Bonilla, M., & Arbildi, C. (2025). What do we mean when we talk about teacher professional development in formative assessment? A systematic review. *International Journal of Educational Research*, 131, Article 102586. <https://doi.org/10.1016/j.ijer.2025.102586>
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Zembylas, M. (2003). Emotions and teacher identity: A poststructural perspective. *Teachers and Teaching*, 9(3), 213–238. <https://doi.org/10.1080/13540600309378>