



BRILL

IMPLEMENTATION AND REPLICATION STUDIES IN
MATHEMATICS EDUCATION 3 (2023) 1–24

IRME

brill.com/irme

Study of the Effects of a Mathematics Teacher Education Program: What Went Wrong?

Andrés Pinzón | ORCID: 0000-0002-6463-8383

Facultad de Educación, Universidad de los Andes, Cra. 1 No. 18a-12,
La Candelaria, Bogotá 111711, Colombia

Corresponding author, e-mail: aa.pinzon364@uniandes.edu.co

Pedro Gómez | ORCID: 0000-0001-9929-4675

Facultad de Educación, Universidad de los Andes, Cra. 1 No. 18a-12,
La Candelaria, Bogotá 111711, Colombia

argeifontes@uniandes.edu.co

Abstract

If we want to understand what works in studies of teacher education programs, we also need to understand what does not work. In this article, we discuss why a study evaluating the effects of an education program on implementation practices yielded unexpected results. Interviews with a sample of teacher graduates from the program revealed that the program did have effects on implementation practices that were not evident in the original study. These effects are in the form of increased student participation, teamwork and the conception of error as opportunity. The instrument and procedures of the original study did not allow these effects to be seen.

The impact sheet to this article can be accessed at [10.6084/m9.figshare.22339567](https://doi.org/10.6084/m9.figshare.22339567).

Keywords

implementation practices – in-service teachers – mathematics teacher education

1 Introduction

Teacher professional development is an important component of the quality of education (Desimone et al., 2002; Guskey, 2002). Without an idea of what is working and why, it is difficult to implement evidence-based professional development programs that can be improved based on that evidence (Darling-Hammond et al., 2017). Studies on evaluation of teacher education programs often report on successful case studies. It is therefore important to determine whether there are cases where the expected results are not achieved, in order to learn from them. To understand why a program did not have a lasting impact, researchers should consider following up on less successful outcomes (Tirosh et al., 2015).

When evaluating teacher education programs, it is not easy to address all the factors that may influence their effectiveness. For example, the initial training of the participant, their motivation, the colleagues they work with, the school they work in and even the design of the program (Opfer & Pedder, 2011; Kennedy, 2016; Minor et al., 2016). For that reason, in the evaluation of these programs, successful results, or at least the expected results, are not always obtained. However, unsuccessful or unpublished results are hardly ever given importance. One possible reason for this may be the phenomenon of publication bias: papers showing positive results are more likely to be published and research reporting unsuccessful results is more likely to be archived (Antonakis, 2017). If we want to understand what works in studies of education programs, we also need to understand what does not work, so that we can interpret research findings in relation to the full range of research in the area.

We organise this article into six sections. In the first section, we describe what we mean by implementation practices. In the second section, we present a teacher education program and the results obtained in an evaluation of its effects on implementation practices. In the third section, we present the research question. In the fourth section, we describe the method developed in this study. In the fifth section, we present the results. Finally, we present a discussion and the implications of the study.

2 Implementation Practices

Implementation practices refer to the teacher's implementation of his or her planning in the classroom. This implies a transformation of what is planned, since the teacher's performance must adapt to the complexity of the events that occur in the classroom throughout a session (Cho, 1998, p. 21). In terms of

Remillard and Heck, it is the enacted curriculum: it is an emergent curriculum that the teacher builds with the students: it is expressed in the interactions between teacher and students around the tasks of each lesson (Remillard & Heck, 2014). Therefore, if teacher education programs are to contribute to the improvement of participants' implementation practices, they must help the teacher to improve teacher-student interactions around tasks, and to adapt to the complexity of classroom events.

On the other hand, among the possible ways of implementing the mathematics curriculum, two main approaches are suggested in the literature: student-centred and teacher-centred. Student-centred lessons and reactions can be observed, for example, when a teacher asks students to explain, present and justify their answers, or when the teacher asks and guides students as they discover whether the answers are valid or not. On the other hand, when the lesson development and reactions are teacher-centred, the teacher is the one who clarifies, explains, evaluates or shows whether the answers are valid or not, and he/she provides direct feedback on the answers (Son, 2013, 2016).

In the following, we describe a mathematics teacher education program that aims to provide opportunities for teachers to develop competences to design, implement and develop the curriculum (Gómez, 2018). We also present the main results of the original study with which we wanted to evaluate the effects of the program on planning, implementation and assessment practices.

3 Professional Development Program for Mathematics Teachers

The professional development program under study is the Master's degree in Mathematics Education at the Universidad de los Andes (Bogotá, Colombia). It is a master's degree focused on pedagogical content knowledge for practicing secondary school mathematics teachers. Teachers who participate in the program do so voluntarily and must pass a selection process to be accepted. Many of these teachers receive financial support from their employer, which further motivates them to improve their teaching practices. Additionally, they have the support of their schools.

The program consists of four semesters. In each semester, two consecutive courses are taken and each course lasts nine weeks. In total, the program includes 920 hours of professional development. The second author is the designer of the program and both authors participated in the program as responsible for one course each. In this program, teachers are expected to develop in-depth pedagogical knowledge of mathematical content oriented to support their decisions for lesson planning, implementation, and evaluation.

For implementation practices, trainee teachers are expected to develop knowledge, skills and attitudes to:

- establish learning expectations and identify learning constraints,
- foresee the performance of students when approaching tasks,
- design, analyse and reformulate learning tasks that contribute to the achievement of learning expectations and the overcoming of learning limitations,
- design and use instruments and procedures for collecting, coding and analysing information to evaluate learning and teaching,
- implement a curriculum design,
- evaluate the relevance, efficacy and efficiency of the curriculum design and its implementation, and
- produce a new curriculum design based on this evaluation (Gómez, 2018).

The content of each module refers to an aspect of the didactical analysis model (Gómez, 2002, 2007, 2018). This model proposes a conceptualization of the activities that the mathematics teacher undertakes to plan, implement and evaluate didactical units. By describing the procedure that a teacher is expected to carry out when designing, implementing, and assessing a didactical unit, this model describes, in a systematic and sequential way, the actions of an ideal teacher. The didactical analysis model is configured around four analyses, corresponding to the dimensions of the curriculum, which form a cycle: subject matter, cognitive, instructional and performance.

At the beginning of the program, participants are organised in groups of three or four people. Each group works on a specific mathematical topic. Each group is accompanied throughout the two years of the program by a tutor. The topics are selected according to the courses that the trainee teachers expect to be in charge of in the second year of their training, in order to be able to implement their curriculum design proposal. Teachers are expected to align their subject matter with the school's curricular guidelines. Within the context of Colombian schools' autonomy (Torres & Duque, 1994), the program allows teachers to design their own curriculum, rather than imposing one from the program or the Ministry of Education. Each group carries out a cycle of didactical analysis on their topic throughout the program. At the end of the first five courses, the groups produce a didactical unit design addressing the conceptual structure, representation systems and phenomena of the mathematical concept; learning expectations and constraints; anticipation of schoolchildren's performances when approaching tasks; task sequences that contribute to the achievement of learning expectations and overcoming learning constraints; and finally, assessment of learning and teaching (planning). Between the fifth and sixth module, the groups focus on the implementation of the didactical unit. There, the focus is on the interactions among students and between teacher and students (implementation). Between the sixth and seventh

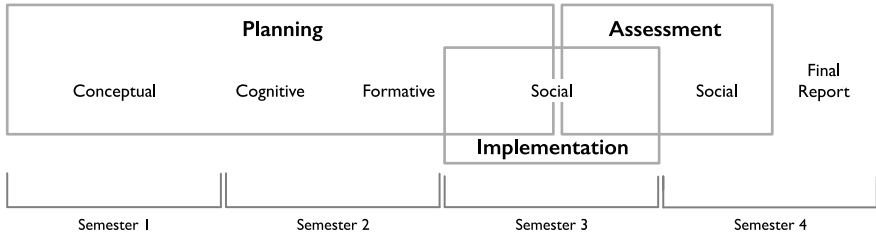


FIGURE 1 Teachers' practices and the cycle of didactical analysis in the professional development program

module, they focus on the collection and analysis of information emerging from the implementation (evaluation) and, in the last module, on the production of the report of the overall experience. Figure 1 shows how the three practices and the four analyses that make up the didactical analysis model are addressed over the two years of the professional development program.

Each module is configured around four activities. Over the two weeks of an activity, each group produces a draft of the activity, which they send to their tutor. The tutor comments on the draft. Based on these comments, the group produces the final document of their work and presents it to their peers, trainer and tutors. Students comment on and critique their peers' presentations. Each student must produce, for each activity, a document of comments and critiques of the work and presentation of a group other than his or her own. As can be seen, this training program is based on the conceptual framework of the social theory of learning, community of practice and interdependent learning (Wenger, 1998). Group work systematically promotes the processes of negotiation of meanings among group members in each of the activities. The tutor's comments and peer critiques (both in the final comments and in the presentations) foster these meaning negotiation processes by generating doubts, raising differences of opinion, and requiring a solution to the problems raised by the tutor. In addition, individuals build their own identity (and develop their knowledge) by contributing to the group's work, and by commenting on and criticizing the work of a group other than their own.

4 Evaluation of the Program's Impact on Teachers' Practices

In a first study conducted by us (Pinzón & Gómez, 2023), we sought to establish the effects of the program on the trainees' planning, implementation and evaluation practices. To obtain evidence of these effects, the teachers who graduated from the program responded to questions in a questionnaire inquiring about their practices with a particular group of students, in a lesson and in

a specific mathematics topic. The two authors designed the questionnaire. We selected self-reports as a data collection instrument because they are a powerful tool for collecting information from large groups of teachers (Uysal, 2012). These instruments turn out to be as reliable as observations when teachers are asked about specific topics of their classroom practices. In fact, Desimone (2009, p. 190) has pointed out, on the basis of several studies, that “surveys that ask behavioural and descriptive, not evaluative, questions about the teachers’ professional development experiences and teaching have been shown to have good validity and reliability.”

To answer the questionnaire, we asked teachers to take the most recent class as a reference, so that they could answer about concrete facts and perform an introspection of their performance (Varela & Shear, 1999). However, it is inevitable that they may not have remembered everything requested in the questionnaire or that their answers may have a social acceptability bias. The authors made decisions in this regard when ideal responses were perceived in the questionnaires. For example, if a questionnaire was not answered by at least 80% or a teacher’s percentage difference exceeded two standard deviations in any of the practices, it would not be considered for the results. The planning questions addressed the documents that the teacher had in mind for the design of the lesson — such as the curriculum and official documents (use of curriculum documents); the conceptual structure — concepts, procedures, representation systems and phenomena of the mathematical concept (conceptual dimension); the learning expectations, the possible errors and difficulties, and the forecast of the different ways of solving each task (cognitive dimension); the possible actions of students and teacher, the selection of resources and materials, and the selection and sequencing of tasks (formative dimension); and finally, the planned assessment in terms of the purpose and form in which it will be carried out (social dimension). The implementation section focused on three specific issues: teacher’s reactions to students’ unanticipated strategies, teacher’s reactions to errors, and interactions. The third and last part inquired about the teacher’s use of the information collected for the purpose of assessing their students and evaluating their implementation. In Figure 2, we describe the organisation of the questionnaire based on the three sections outlined. The questions on implementation practices included in the questionnaire were as follows.

Reactions to students’ unexpected strategies

- Were there any students who solved a task in a different way than you explained it in class?

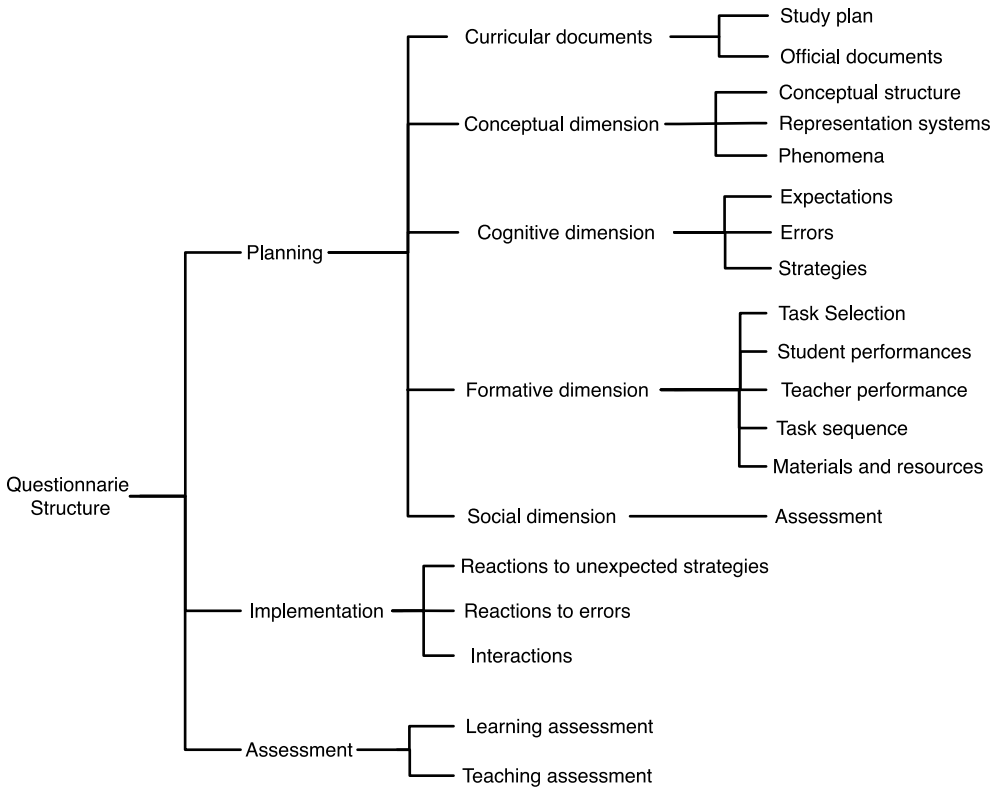


FIGURE 2 Structure of the questionnaire

- If you answered yes to the previous question, how did you react to this situation?

Reactions to unforeseen errors

- Did your students make errors or get stuck when trying to solve the tasks?
- If you answered yes to the previous question, think of one such occasion and explain what you did when you realised that one or more students were making an error or were blocked.

Interactions

- Think about the total amount of time that was spent on the topic. How much of the time were you explaining or giving examples?
- What proportion of the time were students working on the tasks or activities you set?

- How much of the time were you and the students sharing answers or results, or discussing those answers or results?
- What proportion of the time did students work individually?

These questions were formulated with the aim of identifying teachers who, according to the forms of implementation practices and the purposes of the program, moved from teacher-centred to student-centred practices.

The information needed for this first study was collected through a questionnaire that was answered by the participants between 2014 and 2018. Participation of trainee teachers was voluntary and had no impact on their evaluation. The responses were collected at the beginning of the program (entrance), and six months after finished (exit). We got answers from 51 of these teachers (this represents 71% of mathematics teachers who completed the professional development program). We considered valid those questionnaires that had at least 80% of the complete information. Those questionnaires in which we identified non-response biases were discarded. In total, we identified 44 entrance/exit valid questionnaires. Of these 44 valid questionnaires, it was necessary to eliminate a questionnaire from a teacher who claimed to have participated simultaneously in another training program. Therefore, 43 questionnaires were considered valid for the purposes of this study. This group of 43 teachers has the same characteristics as the original group of 51 teachers, as shown in Table 1.

The responses of the teachers were coded by text segments, each representing a single described action. Each code was assigned a weighting, which reflected the relative importance of the responses according to the expectations of the program. In the case of the implementation section, we assigned the following weightings: reactions to students' unexpected performance (30%), reactions to unforeseen errors and blockages (30%), and interactions (40%).

Once the measure of each response was established, we were able to compare the two responses of the same teacher to a question and establish the percentage difference between the two. In the case of the questions under the

TABLE 1 Participant's characteristics (percentages)

	Women	5 or more years of experience	Working in public schools	With bachelor's degrees
Answers (51)	49%	86%	82%	82.3%
Valid questionnaires (43)	50%	88%	79.5%	84%

headings “Reactions to unexpected student actions” and “Reactions to unforeseen errors”, the comparison was made if the teacher answered yes to the first question (otherwise the question was omitted from the overall assessment). Finally, given that the same subjects answered the input and output questionnaire, we used the Student’s *t*-test for paired data to determine whether or not the differences were statistically significant and, if they were, we calculated the effect size with Cohen’s *d* statistic.

The effects of the professional development program were evident in planning practices (use of curriculum documents, anticipation of errors, and selection and sequencing of tasks) and assessment practices (of student learning and teaching). However, although statistically significant differences were found for the implementation practice, their effect was minimal. Below, we summarize the main results obtained in the original study on the effects on teachers’ curricular practices.

4.1 Curricular Practices

In Table 2 we present results for the curricular practices.

We observe that there is a statistically significant difference in planning and assessment practices and that, when evaluating the practices as a whole, we obtain an increase of 5.46 percentage points and an effect of 0.52. This is evidence of an effect considered optimal in the curricular practices of the teachers who participated — effects considered optimal for Cohen’s *d* should be close to 0.5 (Gertler et al., 2017). However, we can observe that, although there is a statistically significant difference in implementation practices, its effect is close to zero. In the following, we present more detail on implementation practices.

TABLE 2 Results of the practices

Variables	Mean*	Standard deviation	<i>P</i> -value for a tail	Rejection of H_0 **	Effect***
Planning	5.03	15.39	0.01	Yes	0.46
Implementation	2.8	9.6	0.027	Yes	0.06
Assessment	11.19	26.3	0.002	Yes	0.49
Total practices	5.46	9.6	0.001	Yes	0.52

* Percentage point difference.

** $\alpha = 0.05$, equivalent to a confidence interval of 95%.

*** Cohen’s *d* effect.

4.2 *Implementation Practices*

In Table 3 we present the disaggregated results of the implementation practices.

The results in Table 3 show that the program has a statistically significant effect on how teachers react to students' errors with an increase of 7.5 percentage points and an effect of 0.15. Similarly, in interactions, there is an increase of 3.92 percentage points and an effect of 0.28. However, in the reactions to unplanned strategies of the students, we have a negative difference, although not statistically significant, of -3.38 percentage points. And, when assessing the whole section on implementation practices, we find a slight increase of 2.8 percentage points, which represents a statistically significant difference, but with a small effect of 0.06. The education program appears to have no effect on teachers' reactions to unanticipated strategies and very small effects on reactions to errors and interactions, which in turn leads to very low effects on implementation practices.

It is important to note that part of the information we sought to obtain from the implementation practices depended on the occurrence of situations of reactions to errors or unplanned strategies. Not all teachers provided us with information in this respect: 13 teachers reported situations of unplanned strategies and 35 reported situations of errors. Therefore, comparisons of situations of reactions to errors or unplanned strategies were made with a small number of the participating teachers. However, if reactions to unintended strategies are omitted from the implementation variables, the mean differences increase to 5.46 percentage points, and these differences are statistically significant, but the effect is still low (0.27).

We expected the participants, at the end of the program, to be able to move from teacher-centred lessons and reactions to student-centred lessons

TABLE 3 Results of implementation practices

Variable	Mean*	Standard deviation	P-value	Rejection of H_0 **	Effect***
Reactions to unforeseen strategies	-3.38	15.09	0.07	No	-
Reactions to errors	7.5	27.28	0.035	Yes	0.15
Interactions	3.92	15.22	0.045	Yes	0.28
Total implementation	2.8	9.6	0.027	Yes	0.06

* Percentage point difference.

** $\alpha = 0.05$, equivalent to a confidence interval of 95%.

*** Cohen's d effect.

and reactions. We constructed the questions in the implementation practice questionnaire and set the weights we gave to those questions for this purpose. However, as we have mentioned, the study shows a small effect on these questions, even when the question on unintended strategies is omitted.

5 Research Question

Despite the literature warning that it is unlikely to see changes in teachers' practices as a result of participating in a professional development program (Gregoire, 2003), our initial study found statistically significant effects on planning and evaluation practices. Additionally, as the teachers participated voluntarily and received support from their schools, our observations and perceptions suggest that the program likely had positive effects on graduates' implementation practices, like the effects seen on their planning and evaluation practices. However, as is evident from the results presented above, we did not obtain such results in our study. The differences between our expectations and the results obtained can be explained on the basis of two hypotheses: (a) the program has no effect on graduates' implementation practice or (b) the instrument and procedures we used in our study did not allow us to measure such effects. In this article, we try to establish which of these two hypotheses is valid. Thus, in this research we approach the identification of 'effects' produced by a professional development program in the classroom practices. This is an important and recurring theme in implementation research in mathematics education: Studies that focus on identifying effects of innovative teacher professional development programs and interventions (Koichu et al., 2021), and that this is precisely the area where our work is located and to which it contributes.

6 Method

In this section, we describe the method we used to address the research question we formulated in the previous section. We present the teachers involved, the data collection, the coding of the data and the analysis of the data.

6.1 *Participating Teachers*

The participants were seven graduate mathematics teachers who participated in the original study: three women and four men. We identified the teachers with the fictitious names Alejandro, Nelly, Nancy, Ramiro, Samuel, Tomás and

Yadira. The teachers were selected based on their results in the original study. We selected two teachers whose results showed significant favorable differences in their implementation practices (Alejandro and Ramiro, on average +18.61 percentage points) and five teachers whose results showed significant unfavorable differences (Nelly, Nancy, Samuel, Tomás and Yadira, on average -8.3 percentage points). During the first half of 2021, we contacted them by telephone for their participation in this study and they participated voluntarily. Nancy and Samuel graduated in 2016, Alejandro and Tomás in 2017, and Nelly, Ramiro and Yadira in 2018. Ramiro worked in a private school and the other six teachers worked in public schools in Bogotá and Cundinamarca in Colombia. All teachers, except Ramiro, had a degree in mathematics teaching and, all of them had more than 8 years of experience as mathematics teachers.

6.2 *Data Collection*

We collected the information through interviews. Although the researchers were educators in the program, at the time of the interview there was no student-teacher power relationship that could bias the responses. Participation was voluntary and the participants were clear that their answers and opinions would have no repercussions for them. We interviewed each teacher individually. These interviews were semi-structured, audio-recorded and lasted on average 15 minutes. The interview design was guided by the focus of the study: changes in implementation practices. The interview included an introductory stimulus for the teacher: "Please tell us how you perceive that your classroom practice has changed as a result of participating in the master's program". Based on how the teacher responded to this initial question, we asked more specific questions. Some of these questions were as follows: "Do you perceive that your classroom performance has changed?", "Why?", "Do you perceive that you interact differently with students?".

In contrast to the first study, in which we asked them about issues they had realized in a recent class, and from that information we made inferences about changes in their implementation practices, this time we asked the teachers directly about the changes they consciously perceived as a result of their participation in the program. This implies that the changes reported in the interviews are not those that they express as happening habitually in their lessons.

6.3 *Coding of Information*

We recorded the audio interviews and transcribed them. We called episodes those text segments of the transcripts that addressed issues related to the focus of the study. We analysed the transcripts according to a process of content analysis (Mayring, 2015), based on grounded theory (Corbin & Strauss, 1990).

The conceptual framework of the didactical analysis model, the focus of the study, the structure of the interview and the evidence itself guided the identification of the codes we assigned to the episodes we identified.

We were able to classify the episodes into three categories related to (a) increased student participation, (b) teamwork and (c) error as opportunity. For each category, we identified codes that characterise them. For example, for the category “Increased student participation” the codes assigned were reduced time spent on explanations, increased time for student interaction, consulting students about their perception of the tasks performed, and flexibility to modify planning. This is the case of the episode “Now, I ask them what they know about the concept and I begin to develop the topics based on that prior knowledge” whose assigned code is increase of time for student interaction. For the category “teamwork” the codes assigned were Increased group work and teamwork with assigned roles. For the category “error as an opportunity” the codes assigned were increased attention to errors, establishing errors, error aids, and change of attitude towards errors. This is the case of the episode “I propose situations in which there is an error, so that they can identify it”, whose assigned code is to establish the errors.

Finally, the researchers reviewed the coding process to ensure that each of the selected episodes was labelled with the appropriate codes.

6.4 *Analysis of the Information*

We identified statements and reflections on the effects of the program on teachers’ implementation practices and organised them into the previously described categories and codes. In the following section, we present specific statements and provides verbatim statements from the participants to support them.

7 *Changes in Implementation Practices*

As mentioned above, we organised the perceptions of the teachers interviewed into three categories: increased student participation, teamwork and error as opportunity. We describe these categories below.

7.1 *Increased Student Participation*

Two teachers told us that, as a result of participating in the education program, they had provided more space in their lessons for their students’ participation. To this end, they modified their implementation practices by implementing three types of strategies: (a) reducing the time devoted to explanations and

increasing the time for student interaction; (b) consulting students about their perception of the tasks developed; and (c) being more flexible in modifying their planning, according to students' difficulties and interests.

Tomás is an example of a teacher who declared having reduced the time spent on explanations and increased the time for student interaction:

I try to make the class more dynamic. It's not just a lecture. Before [participating in the program], I explained all the content before [the students] worked on the activities. Now, I ask them what they know about the concept and I start to develop the topics based on that prior knowledge, and this motivates them more to participate in the class.

Tomás explicitly stated how the program implied changes in his practice. He acknowledged that, before participating in the program, his class followed a traditional scheme whereby he first explained the topic and then asked his students to solve the activities. After participating in the program, Tomás seems to focus more on students' prior knowledge in order to promote their learning through class participation based on it.

Alejandro commented that, in order to know the students' perception of the class,

one aspect that has become stronger in my classes, having participated in the program, is formative assessment, involving them [the students] more in what they learn and how they learn it. I now ask them at the end of the class what they found most difficult or interesting about the tasks they did.

For her part, Nancy told us that

I take time out of class to tell me what they learned, what we didn't do well, who wants to summarise what we did. It's not just about them getting information from me, it's about them being aware of what they are doing and learning.

Yadira also stated that "now I take the affective dimension more into account. When we do the evaluation [of the class], I ask the students how they felt, what they liked, what they didn't like, what can be improved". These teachers introduced, on the occasion of their participation in the program, a new strategy in their implementation practice: motivating students' participation and

involvement in their learning by asking them to reflect on their experience in each session.

Finally, Tomás, referring to his flexibility in modifying his planning, stated that

the teacher's time and the teacher's interests may differ from what the student wants or can. I plan my objectives for a class, but they are not always achieved as one wants [when implementing]. That's why [after participating in the program] I leave the eagerness to fulfil a content [planning] and attend to the needs and difficulties that arise from them [the students]. A single task is not enough, they sometimes need a variety of tasks to understand a certain topic: exercising algorithms, manipulating materials, discussing with classmates, watching a video, etc.

Tomás explicitly stated a change in his implementation practice that is a consequence of a change in his concerns. He moved from being concerned about content coverage to being concerned about his students' learning, in particular their needs and difficulties. He became aware that he must be more flexible with time management in class in order to give opportunities for students to participate and express their difficulties.

7.2 *Teamwork*

The teachers highlighted the effects of the program on the grouping of students in their classes and the respective teamwork of those groups. Nancy stated that "I have included time for group work, which I didn't do before. [Before participating in the program,] my priority was individual work". Yadira commented that she has "balanced the time to develop the tasks [after participating in the program]. I increased the time for teamwork and now half [of the time] is individual work and the other half is group work". Samuel and Nelly said that they have decided that "the groups have defined roles" and that these roles should be rotated to promote leadership. This is the case of Samuel, who said "the groups are no longer organised by them [the students], but now [after participating in the program] I assign roles to each one and they must rotate so that each one leads a task".

We interpret the above perceptions of teachers in two ways. Firstly, and as a result of participating in the program, teachers recognised the importance of students' interaction in their learning. The teacher no longer seems to focus on transmitting information to be remembered and repeated. Instead, learning takes place when students interact with each other, assuming different roles in solving activities. Secondly, the teachers told us that this change in

attitude is a consequence of their experience of participating in the program: they themselves learned to reach agreements on the solution they proposed for the activities of the program. This is the case of Alejandro, who said “I learn together with my students. We organise the room into a round table and each group presents what they have done. We all discuss by looking at each other” and Nelly, who told us that she learns together with her students:

Just as we had to work in groups in the program, with all the differences in ways of thinking, I try to keep the groups going throughout the [academic] period and help each other [the students] to learn.

7.3 *Error as a Learning Opportunity*

Teachers reported issues associated with seeing students’ errors as an opportunity in their lessons. After participating in the program, (a) they are more attentive to their students’ errors, (b) they seek to establish those errors, (c) they design and implement their performance according to the errors and (d) they have a different attitude towards them.

Nelly, stating that “I am more attentive to the students’ errors. I don’t move on to the next topic, if we haven’t dealt with the most frequent errors”, showed a change in her implementation practice as a result of participating in the program. She hinted that, before the program, she did not take into account students’ errors in her classroom performance. Now, she considers that the errors students make should be addressed in class.

For his part, Tomás indicated that he is not only attentive to the errors his students make, but that, in addition, he encourages them to reflect on situations in which the error appears. He stated that

I ask more questions to find out about the errors. I don’t wait for them [the students] to necessarily make the errors, but I propose to them situations where the error is, so that they can identify it ... with the master’s degree, I learnt that we should not take anything for granted. If a student doesn’t show errors in class, it doesn’t mean that they don’t have them or won’t have them later on.

Ramiro showed us that change in his practice is not restricted to identifying the error: It is necessary to provide support for students to overcome it. He stated that

My classroom practice is now focused on how to help students overcome their difficulties. My focus is no longer on one way of solving a task, but

on looking at what learning paths there are and what errors are possible. As I have a small group of students, I try to personalise learning. I start the lessons with a diagnosis of prior knowledge. Then I give specific help for specific errors. Not everyone learns in the same way. And the students perceive it like this: “it’s that [teacher] Ramiro cares a lot until you learn the subject”.

Finally, Samuel, with his intervention, evidenced the change that these teachers had during the program in relation to the error of their students. They no longer see error in a negative way; they see it as a learning opportunity:

I no longer see them as making in errors, but as “incurring” in errors.¹ That is, I no longer punish the error but see it as an opportunity for them to learn. I am more attentive and I make the most of it [for learning].

Similarly, Nancy stated that

one thing that stands out a lot [from participating in the program] is the handling of error as something that no longer distresses the students and the teacher. They can express their concerns and recognise where they are going wrong without feeling judged. And for my part, I apply the aids provided and, if one aid doesn’t work, I look for alternatives.

8 Results

The changes in implementation practice reported by the teachers interviewed suggest that the program did influence their implementation practices. However, as we showed above, these effects are not apparent in the results of our original study. This leads us to believe that the instrument we designed and implemented at the time did not allow us to establish the effects of the program. Therefore, informally and with data from a small sample, we reject our first hypothesis. We then present some conjectures as to why the second hypothesis is confirmed, i.e., that the instrument and procedures we used in our original study did not allow us to measure these effects. We organise these reflections into the same categories we used in the previous section.

¹ In Spanish, there are two verbs associated with error: *cometer* and *incurrir*. “Cometer” (to make) has a negative sense: sins and crimes are committed. “Incurrir” (to incur) has a positive sense.

8.1 *Increased Student Participation and Concern for Student Learning*

When referencing the questionnaire used in the original study, it is important to note that while the variable “Interaction” took into account questions related to student participation in lessons, it did not specifically inquire about the extent to which teachers consult with students about tasks or make adjustments to the curriculum to address their difficulties and interests. As demonstrated in Table 2, the instrument yielded statistically significant differences and positive effects. However, to improve the instrument in this aspect, we recommend including the following questions in future studies:

- Did the students participate in the selection of the tasks developed? If so, please describe their level of participation.
- Did you make any modifications to the planning to address the difficulties and interests of the students? If so, please describe the changes made.

8.2 *Teamwork*

The instrument we used in our original study did not ask about how teachers promote learning in the classroom. As can be seen from the responses of the teachers interviewed, it would be possible to obtain information in this regard by including questions such as the following in the initial questionnaire.

- Did you organize the teamwork or give indications for its organization? If you answered yes to the previous question, how were the roles in the groups assigned?

8.3 *Error as Opportunity*

The questionnaire used in the original study asked about one particular aspect of teacher practice in relation to student errors: how the teacher acts when he/she identifies that one or more students make an error. However, the instrument did not allow us to inquire about other aspects of the teacher’s attitude towards error that the teachers interviewed revealed to us (being attentive to error, identifying it and seeing it as part of the learning process) and which are concretized in the idea of seeing error as an opportunity to learn. Therefore, the instrument can be improved in this aspect by including the following questions.

- Did you inquire about errors that students did not make? If you answered yes to the previous question, how did you do it?

When a student made a mistake:

- What was your first thought as to the source of that error?
- Did you consider more than one aid to overcome the error?

9 Discussion

To understand what works in studies of teacher education programs, we also need to understand what does not work. In this article, we present a study in which we analyse why, in the results of an evaluation of the effects of a training program on implementation practices, we did not get the effects we expected, i.e., “something went wrong.” The training program is based on the ideas of community of practice and interdependent learning (Wenger, 1998), which promotes a high degree of autonomy in the participating teachers. Hence, we agree with Karsenty (2021) that the more the teacher education program is autonomous, the more likely it is that the project engage challenges regarding the upscale of settings, but on the other hand there are more potential for the success of the upscale of program values. Interviews with a sample of teacher graduates from the program revealed that the program did have effects on implementation practices that were not evident in the original study. In this study, we confirm the hypothesis that the results obtained in the original study, in which the effects were close to zero, are due to the fact that the instrument and procedures used did not allow for measuring those effects, and we reject the hypothesis that the program had no effect on the implementation practice of the graduates.

We are surprised that, despite the fact that the teachers interviewed highlighted the effect of the program on their implementation practices (in relation to giving more participation to their students in their classes), the results found, although positive, have a small effect. When we designed the questionnaire, we specified the purposes of the program in three aspects: reactions to unplanned strategies, reactions to errors and proportion of time in interactions. However, what we observed in the original study is that this way of specifying the program’s expectations is partial and does not allow us to see the full effects of the program. For example, the questions on interactions only allow us to see some aspects that we would expect teachers to show in their practices (time for explanations and spaces for discussion of results). The vision of learning that underlies the design of the program implies that it is the students who construct knowledge by interacting with each other. Therefore, group work and providing spaces for students to express their opinions and perceptions are important for the program. However, we constructed a questionnaire with which we aimed to show whether teachers move from giving more time to explanations, proposing only individual work and little interaction, to teachers spending less time in front of the board and more time with

students discussing their results. In other words, we focused on whether the teacher was becoming less “traditional”. We did not go into detail about what we expected the teacher to do in terms of learning opportunities for their students (in the context of learning together and discussing with each other), nor how students were expected to participate by expressing their opinions and perceptions of learning.

We recognize that teacher learning and changes in curricular practices are dependent on at least three subsystems: the teacher, the school in which she or he works, and the training program (Opfer & Pedder, 2011). In the case of the teachers that participated in this program, they voluntarily participated, and they went through a rigorous admission process, in which it was assured that they had the appropriate attitudes towards learning and teamwork for a suitable performance in the program. All teachers were working in schools that participated in the same public policy and had the endorsement of their school principal to carry out the implementation of their program’s work. Therefore, we can guarantee that these two subsystems, teachers and schools, are similar among participants, and changes in practices can be associated with participation in the training program. We do not ignore that there might be other factors or circumstances, unknown to us, that might also have influenced the results, like teachers’ attitudes and beliefs (Gregoire, 2003).

We expected the results of the original study to show that, as a result of participating in the program, teachers’ implementation practice would be more student-centred than it was at the start of the program. The results we obtained in that study (with small positive differences and small effect) do not indicate this effect of the program. However, when we conducted the interviews in this study, we found that teachers report that this “centre shift” has indeed occurred. Their testimonies bear witness to this. This is the case of Ramiro when he said “My classroom practice is now focused on how to help students overcome their difficulties”; of Alejandro when he stated that “I learn together with my students [...] We all discuss while looking at each other”; or of Tomás when he mentioned that “I left the desire to comply with a content [planning] and I attend to the needs and difficulties that arise from them [the students]”. This evidence shows that the questionnaire could be improved, as it failed to capture this reality.

We cannot be sure with statistical certainty that the program has an effect on the implementation practices of the graduates. To do so, it would be necessary to modify the original questionnaire with questions that would address the issues proposed by the teachers interviewed and allow us to establish whether, for those questions and in a scheme such as the one used in the original study, we can find statistically significant differences. Likewise, this

new study confirmed that it is necessary to complement the information from the questionnaire with random interviews to teachers to have new sources of information and to be able to contrast them. We recognize that the differences established between one lesson report and another may have multiple causes other than the direct effects of the program. However, when contrasted with an interview with the teachers, some assumptions can be confirmed or rejected.

We acknowledge that one limitation of this study is the sample selected for the interviews. This sample is not entirely representative of all graduates from the program, and it was based on the voluntary participation of those contacted. Despite this, we made sure to include graduates from different cohorts, who in the first study obtained both positive and negative results in their implementation practices. The information collected through these interviews revealed actions and practices of the teachers that were not previously addressed in the original questionnaire, which has allowed us to identify areas of improvement in the questionnaire. However, we recognize that due to the characteristics of the sample, we cannot be certain whether adding these questions to the questionnaire and applying it to a new sample of teachers would have a positive impact on the program's implementation practices. For these reasons, we do not claim that these results are generalizable. However, they do provide existence proof (Schoenfeld, 2000) of certain aspects of teachers' practices that were not initially addressed in our study questionnaire. These results serve to shed light on previously unexplored areas of teachers' practices. Another limitation is that, given the number of teachers participating in the first study, it was not possible to contrast the information given by them with classroom observations, since this requires a considerable amount of resources. However, the interviews allow us to glimpse effective changes in implementation practices since teachers refer to them in their own words and with concrete examples, six months after the end of the program.

This study gives us perspectives for future lines of research. On the one hand, we intend to continue with the adaptation of the questionnaire to include questions that allow us to address the issues proposed by the teachers interviewed and, subsequently, to establish whether, in a scheme such as the one used in the original study, comparing their practices before and after participating in the education program, we find statistically significant differences and effects considered optimal in the literature. On the other hand, the possibility remains open of complementing this study with others that discriminate which profile of teachers obtain greater effects in their curricular practices. These results could vary according to the years of experience, initial training and the type of educational institution in which the teacher works — public or private — or its location — rural or urban. Lastly, given the limitations

of resources when working with large groups of teachers, it would be beneficial for future studies to compare the teachers' reports with those of the students, with student questionnaires. This would provide a more comprehensive understanding of the implementation practices in question and help to identify any discrepancies or disparities between the perspectives of teachers and students.

Acknowledgements

This study was carried out with the support of the Francisco José de Caldas Fund (Colciencias, Colombia), within the framework of the research program 64639, call for projects 808 of 2018, project "Acometer la inequidad a través de la educación: matemáticas en la educación básica y media" [Addressing inequity through education: mathematics in basic and secondary education], 2019–2021.

References

- Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly*, 28(1), 5–21. <https://doi.org/10.1016/j.leaqua.2017.01.006>.
- Cho, J. (1998). *Rethinking curriculum implementation: Paradigms, models, and teachers' work*. Annual meeting of the American Educational Research Association, San Diego, CA.
- Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1), 3–21.
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.
- 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/https://doi.org/10.3102/0013189X08331140>.
- Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24(2), 81–112. <https://doi.org/https://doi.org/10.3102/01623737024002081>.
- Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2017). *La evaluación de impacto en la práctica* [Impact assessment in practice]. BID — Banco Mundial.

- Gómez, P. (2002). Análisis didáctico y diseño curricular en matemáticas [Didactical analysis and curriculum design in mathematics]. *Revista EMA*, 7(3), 251–293.
- Gómez, P. (2007). *Desarrollo del conocimiento didáctico en un plan de formación inicial de profesores de matemáticas de secundaria* [Development of pedagogical knowledge in a preservice secondary mathematics teacher education program]. Departamento de Didáctica de la Matemática de la Universidad de Granada. <http://funes.uniandes.edu.co/444/>.
- Gómez, P. (Ed.). (2018). *Formación de profesores de matemáticas y práctica de aula: conceptos y técnicas curriculares* [Training of mathematics teachers and classroom practice: concepts and curricular techniques]. Universidad de los Andes.
- Gregoire, M. (2003). Is it a challenge or a threat? A dual-Process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, 15(2), 147–179. <https://doi.org/10.1023/A:1023477131081>.
- Guskey, T. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership*, 59(6), 45–51.
- Karsenty, R. (2021). Implementing professional development programs for mathematics teachers at scale: What counts as success? *ZDM — Mathematics Education*, 53(5), 1021–1033. DOI: 10.1007/s11858-11021-01250-01255.
- Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, 86(4), 945–980. <https://doi.org/https://doi.org/10.3102/0034654315626800>.
- Koichu, B., Aguilar, M. S., & Misfeldt, M. (2021). Implementation-related research in mathematics education: the search for identity. *ZDM — Mathematics Education*, 53(5), 975–989. <https://doi.org/10.1007/s11858-021-01302-w>.
- Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. In A. Bikner-Ahsbals, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education. Examples of methodology and methods* (pp. 365–380). Springer. https://doi.org/10.1007/978-94-017-9181-6_13.
- Minor, E. C., Desimone, L., Lee, J. C., & Hochberg, E. D. (2016). Insights on how to shape teacher learning policy: The role of teacher content knowledge in explaining differential effects of professional development. *Education Policy Analysis Archives/ Archivos Analíticos de Políticas Educativas*, 24(61), 1–30. <https://doi.org/10.14507/epaa.24.2365>.
- Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, 81(3), 376–407. <https://doi.org/10.3102/0034654311413609>.
- Pinzón, A., & Gómez, P. (2023). *Effects of a professional development program on teachers' curricular practices* [Manuscript submitted for publication]. Universidad de los Andes.

- Remillard, J., & Heck, D. (2014). Conceptualizing the curriculum enactment process in mathematics education. *ZDM — Mathematics Education*, 46(5), 705–718. <https://doi.org/10.1007/s11858-014-0600-4>.
- Schoenfeld, A. H. (2000). Purposes and methods of research in mathematics education. *Notices of the American Mathematical Society*, 47(3), 641–649.
- Son, J.-W. (2013). How preservice teachers interpret and respond to student errors: Ratio and proportion in similar rectangles. *Educational Studies in Mathematics*, 84(1), 49–70. <https://doi.org/https://doi.org/10.1007/s10649-013-9475-5>.
- Son, J.-W. (2016). Moving beyond a traditional algorithm in whole number subtraction: Preservice teachers' responses to a student's invented strategy. *Educational Studies in Mathematics*, 93(1), 105–129. <https://doi.org/10.1007/s10649-016-9693-8>.
- Tirosh, D., Tsamir, P., & Levenson, E. (2015). Fundamental issues concerning the sustainment and scaling up of professional development programs. *ZDM — Mathematics Education*, 47(1), 153–159. <https://doi.org/10.1007/s11858-015-0665-8>.
- Torres, J. C., & Duque, H. (1994). El proceso de descentralización educativa en Colombia. *Revista Colombiana de Educación*, 29, 7–50. <https://doi.org/10.17227/01203916.5364>.
- Uysal, H. H. (2012). Evaluation of an in-service training program for primary-school language teachers in Turkey. *Australian Journal of Teacher Education*, 37(7), 14–29. <https://doi.org/10.14221/ajte.2012v37n7.4>.
- Varela, F. J., & Shear, J. (1999). First-person methodologies: What, why, how. *Journal of Consciousness studies*, 6(2–3), 1–14.
- Wenger, E. (1998). *Communities of practice. Learning, meaning, and identity*. Cambridge University Press.