MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2023, 19(4), em2245
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/13057

Peer assessment processes in a problem-solving activity with future teachers

Patricia de-Armas-González [1*], Josefa Perdomo-Díaz [1], Diana Sosa-Martín [1]

[1] Universidad de La Laguna, Tenerife, SPAIN

Received 02 December 2022 ▪ Accepted 02 March 2023

Abstract

Assessing problem-solving remains a challenge for both teachers and researchers. With the aim of contributing to the understanding of this complex process, this paper presents an exploratory study of peer assessment in mathematical problem-solving activities. The research was conducted with a group of future Secondary mathematics teachers who first were asked to individually solve an open-ended problem and then, to assess a classmate's answer in pairs. We present a study of two cases involving two pairs of students, each of whom assessed the solution of a third classmate. The analysis was carried out in two interrelated phases: (a) individual solutions to the mathematical problem and (b) the peer assessment process. The results show that, in both cases, the assessors were strongly attached to their own solutions, which directly influenced the assessment process, focused on aspects that involve the general problem-solving process and the results. The main difference between the evaluation processes followed by the two pairs lies in the concept of assessment. While the first pair focuses on assessing the resolution process and errors, the other focuses its discussion on giving a numerical grade.

Keywords: evaluation, mathematics, peer assessment, problem-solving, teacher training

INTRODUCTION

Among the wide diversity of processes that take place in teaching and learning contexts, evaluation plays a fundamental role, allowing students to ascertain their learning progress and helping teachers make decisions about teaching processes (Silver & Mills, 2018).

Assessment is a complex process that involves many aspects, and it can be carried out in a variety of ways. However, traditional tests continue to be the main assessment method used, in particular by mathematics teachers (e.g., Chanudet, 2016; Nieminen & Atjonen, 2022). This could be related to the type of training that teachers have received regarding evaluation processes. In fact, one of the concerns of pre-service teachers is the lack of practical experience with assessment (Zevenbergen, 2001). In this regard, peer assessment offers an excellent setting to train future teachers in the process of evaluating, giving them the opportunity to analyze assessment under the perspective of both the student and the teacher.

As many studies point out, participating in peer assessment instances yields benefits both for the assessor and for the assessed peer (Custodia et al., 2015). For example, it enhances assessed peers' responsibility (Lavy & Shriki, 2014; Mogessie, 2015; Topping, 2009), it allows them to have an objective view of how the assessor will review their work (Mogessie, 2015), and it improves their learning (Mogessie, 2015; Topping, 2009). When peer assessment is carried out in a mathematical context, specific benefits are obtained, such as an expanded knowledge of mathematical concepts (Beaver & Beaver, 2011; Zevenbergen, 2001) or improved tools and strategies for solving problems (Beaver & Beaver, 2011). While, when peer assessment is carried out with mathematics pre-service teachers, they also acquire useful benefits for the development of their professional competences. In addition to becoming aware of the complexity of assessment (Zevenbergen, 2001), they have the opportunity to realize that the same mathematical content can be conveyed effectively in various ways (Lavy & Shriki, 2014; Zevenbergen, 2001),

This work is part of the doctoral thesis of Patricia de-Armas-González.

✉ parmasgo@ull.edu.es **(*Correspondence)** ✉ jperdomd@ull.edu.es ✉ dnsosa@ull.edu.es

Contribution to the literature

- Little is known about the peer assessment process itself. What process does a student follow to evaluate a peer's work? What does he/she look for? We address these issues with a methodological design based on case studies, which allows us to delve into the main characteristics of both the resolution of an open-ended problem and the peer assessment process.
- The results of this study may contribute to increasing the existing knowledge about the complex processes of assessment, especially when the evaluation is carried out on skills.
- According to Bakker et al. (2021), this will be one of the main eight themes that mathematics education research should focus on in next decade, together with teacher professional development, among others.

and to gain experience in recognizing valid arguments (Beaver & Beaver, 2011).

In the assessment process, the evaluator must identify important aspects of the task, make judgments about the quality of the response, identify possible errors in the production, measure the performance of students and interpret evidence of learning (Arnal-Bailera et al., 2018; Goos, 2014). However, little is known about the peer assessment process itself. What process does a student follow to evaluate a peer's work? What does he/she look for? In this paper, we present an exploratory study, carried out with secondary school mathematics teachers in training, whose main objective is to analyze the process they follow when assessing a classmate's solution to a problem. The general question that guides the research is: What characterizes the process that these future teachers follow when they evaluate the resolution of a problem given by a classmate?

We address this question with a methodological design based on case studies, which allows us to delve into the main characteristics of the peer assessment process and to have a deeper knowledge of this process (Creswell, 2012).

This research belongs to a scarcely explored field in which we try to identify what processes pre-service secondary teachers follow to assess open-ended problems when they do not receive any guidelines. According to Ukobizaba et al. (2021) there is a serious need for more research on assessment to improve problem-solving. Research found in problem-solving assessment mainly focus on investigating the methods used to assess (e.g., Cárdenas et al., 2016; Nieminen & Atjonen, 2022), but not the processes followed to evaluate. Likewise, those found in peer assessment verify its benefits (e.g., Lavy & Shriki, 2014; Mogessie, 2015) but do not observe the evaluation process or how their own resolution takes part in it. Few of these investigations are developed with pre-service teachers, despite the important role that assessment plays in teacher training. The results of this study may contribute to increasing the existing knowledge about the complex processes of assessment, especially when the evaluation is carried out on skills. According to Bakker et al. (2021), this will be one of the main eight themes that mathematics education research should focus on in the next decade, together with teacher professional development, among others.

ASSESSMENT AND PEER ASSESSMENT

Evaluating is often associated just with assigning a grade. However, it comprises much more. Assessment also allows for obtaining qualitative information about students' learning. This is useful for both teachers and students, helping students to identify where they are in their learning, where they should focus their work and the best way to do it (Nortvedt et al., 2016) and helping teachers to decide the actions they should take in their approach (Shahbari & Abu-Alhija, 2018). Depending on the purpose for which the evaluation is carried out, we will be facing a summative or formative assessment. Summative assessment goal is to certify the level of achievement or skills a person has gained, while formative assessment focuses on how the students' learning process is advancing, how well they are acquiring the objectives, and use this information to reorient teacher's action and the students' learning (Silver & Mills, 2018; Wiliam & Thompson, 2007). Thus, formative assessment gives to the evaluation the character of an instrument to monitor student learning and to improve it (Suurtam et al., 2010).

Peer assessment is one of the models of evaluation that contributes to giving this process a formative perspective. Its main characteristic is that students of the same level review the work of a peer to determine its quality (Topping, 2009). As we mentioned in the introduction of this paper, peer assessment can have positive effects on students' mathematical learning, expanding their knowledge of mathematical concepts or improving their tools for solving problems (Beaver & Beaver, 2011; Zevenbergen, 2001). Moreover, involving students in peer assessment activities makes them protagonists and responsible for their own learning (Black & Wiliam, 2009; Lavy & Shriki, 2014; Mogessie, 2015), and this is one of the key points for an assessment to be formative (Wiliam & Thompson, 2007).

One of the main characteristics that distinguishes peer assessment from other assessment models is the type of feedback the students receive, which comes from another student. This could be positive for different reasons. On one hand, teachers often do not have enough time to provide students as detailed feedback as a peer would offer (Sadler & Good, 2006). On the other hand, peers' feedback could be more understandable for students because they use the same language (Seifert & Feliks, 2018). However, there are some not-so-positive aspects related to peer assessment that should be kept in mind. For example, some students distrust peer assessment because they feel it is less accurate than teacher assessment and sometimes feedback provided by a peer is unclear and it is not useful to assess students to improve their productions or their learning (Seifert & Feliks, 2018).

In general, there is some consensus among researchers in pointing out the positive and not so positive aspects of peer assessment. However, there is one point on which the debate remains open: the assessment criteria. Most research on peer assessment provides assessment criteria (Beaver & Beaver, 2011; Custodia et al., 2015; Lavy & Shriki, 2014; Zevenbergen, 2001). However, some studies maintain that in some cases it is better not to provide assessment criteria (Jones & Alcock, 2014; Wyatt-Smith et al., 2010), such as when assessing as a group and trying to reach a consensus. As Seifert and Feliks (2018) point out, although giving a rubric makes assessment easier for students, reducing gaps in the assessment, a rubric also can limit students' own considerations and additional and interesting viewpoints.

In teacher training context, peer assessment acquires an especial interest and importance, as we have already mentioned. Assessment is a usual professional activity for teachers so being involved in peer assessment allows them to learn about this process (Sadler & Good, 2006; Seifert & Feliks, 2018). Nevertheless, some future teachers do not feel able to provide constructive and accurate assessment for their peers and are not comfortable criticizing other students' responses (Seifert & Feliks, 2018). In this research we will try to find out what may be some possible causes of this, analyzing what training teachers do when they evaluate their peer resolution to a mathematical problem and how they use their own mathematical knowledge to analyze the work of their peers and to give them useful feedback. Moreover, the study does not include the provision of evaluation criteria to prospective teachers in order to try to capture the most authentic information about future teachers' viewpoints (Seifert & Feliks, 2018).

PROBLEM-SOLVING AND ASSESSMENT

Since the publication of Polya's (1945) book, *How to solve it*, problem-solving has become one of the main lines of research in mathematics education (e.g., Felmer et al., 2019; Liljedahl et al., 2016). The large amount and diversity of research has generated a range of meanings associated with the terms "problem" and "problem-solving" that sometimes detract from the clarity of the studies conducted (Mason, 2015). At the present time, there is a broad consensus that "problem" refers to a mathematical task that generates in someone a sense of problematicity, of not knowing a direct way to solve it and desire to find it (Mason, 2015; Schoenfeld, 1992). Thus, for a mathematical task to be a problem, it must have the potential to generate an intellectual challenge to the person trying to solve it, in a way that enhances its mathematical development, promotes its conceptual understanding and mathematical reasoning, and the competence to communicate mathematical ideas (Cai & Lester, 2010). The task must also arouse the interest of the person trying to solve it (Szetela & Nicol, 1992) and make him/her feel capable of facing it (Felmer & Perdomo-Díaz, 2016).

Problem-solving is one of the main processes involved in mathematical thinking (Drijvers et al., 2019; Schoenfeld, 1992), but it is also fundamental in technology, engineering, biology, physics, or medicine. Industries are exhibiting a tendency to perform routine tasks in an automated way, increasingly valuing their employees' problem-solving skills (Chan & Clarke, 2017). For this reason, making students competent problem solvers has become one of the main objectives at all educational levels. As a result of these new demands together with research contributions, several countries have introduced changes in their mathematics curricula, giving problem-solving a greater presence (e.g., Singapore, Spain, and the USA). However, incorporating problem-solving in the classroom is a complex process. In this process, it is important to consider three questions: what, when and how (Mason, 2016; Olson & Knott, 2013).

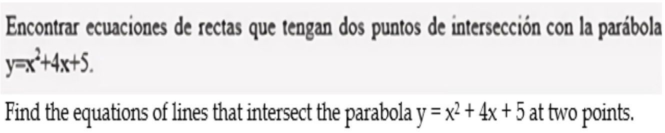
One of the challenges of incorporating problem-solving in classrooms is its assessment. Many teachers use traditional methods to assess problem-solving, which mainly assesses the application of an algorithm to solve the activity, using written tests with questions taken from textbooks or activities carried out previously with students (Cárdenas et al., 2016; Nieminen & Atjonen, 2022; Szetela & Nicol, 1992). This type of evaluation focuses exclusively on the final result. However, not specific aspects of problem-solving are assessed, such as the search of strategies, the existence of multiple solutions or reflection on the solution obtained and the strategy used (Felmer & Perdomo-Díaz, 2016), although they are aspects that teachers consider important to assess (Cárdenas et al., 2016). Arnal-Bailera et al. (2018) found that mathematics university teachers notice five main aspects when assessing problem-solving tasks:

(a) procedure: understood as the general solving process (for example, how the problem is approached, what cases are studied, what representations are used …),

(b) calculations: referring to specific steps of the procedure (for example, set up equations, solve

Table 1. Peer assessment problem-solving workshop general description

Activity	Description	Grouping	Materials
Problem-solving	Solve an open-ended problem	Individual	Paper & pencil
Peer assessment	Discuss & assess a peer solution of problem solved	Random pairs	A third peer's resolution to open-ended problem
Self-assessment	Assess their own resolution	Individual	Own's resolution to open-ended problem
Final discussion	Discussion about general impressions of workshop	Whole group	-


Encontrar ecuaciones de rectas que tengan dos puntos de intersección con la parábola $y=x^2+4x+5$.

Find the equations of lines that intersect the parabola $y = x^2 + 4x + 5$ at two points.

Figure 1. Problem statement & translation (Source: Authors' own elaboration)

them, perform arithmetic operations, calculate derivatives …),

(c) errors,

(d) exposition, interpreted as the accuracy of the reasoning (for example, justifications in words of the steps taken, of the solutions obtained …), and

(e) result.

We analyze if future secondary mathematics teachers notice those problem-solving characteristics and if there are other attributes that they look at.

We focus on the process of assessing open-ended problems, defined as those that can be solved in more than one way, have multiple solutions, or can be expanded by changing the initial conditions (Chan & Clarke, 2017).

In particular, in this research we decided to use a problem that allows infinite solutions and different solution methods. This choice is intended to generate the possibility that the participants present different resolution methods and solutions, so that they have to evaluate answers with different mathematical characteristics to those given by themselves.

THE PRESENT STUDY

The research was carried out with a group of students from the University of La Laguna (Spain) enrolled in the master's program for secondary and high school mathematics teachers during the 2019-2020 academic year. Participants were the 16 students who attended a course called "mathematics learning and teaching". This course was divided into three modules. The aim of the first module was to introduce students to general aspects related with mathematics education; in the second and third modules these aspects were deepened, presenting them in particular situations of compulsory and post-compulsory secondary education.

Research took place in the first module of the course, in the context of a peer assessment workshop designed to generate some discussions about both evaluation and problem-solving. At the beginning of the workshop, participants were asked to individually solve an open-ended problem. Then, they were randomly paired to assess a third student's solution, using their own assessment criteria. The resolution was assigned randomly to each pair, and it was anonymous to guarantee the objectivity of the evaluation.

We addressed the general research question through the following specific ones:

1. What aspects of problem-solving (operation, reasoning, results, …) do future teachers focus on when evaluating a peer's solution?
2. In what ways do prospective teachers use their own resolutions to evaluate the resolution of their peers?

MATERIALS AND METHODS

In this section we briefly describe the peer assessment workshop, as well as the data collection and analysis processes. The research was conducted using a descriptive and exploratory approach based on case studies. The case study is used to explore the intrinsic characteristics of each one of the cases to understand the evaluation processes involved in this research, so it is not considered an instrumental case because a general understanding is not intended (Stake, 1995). We consider that this method is appropriate to address the proposed research problem since it allows us to attend to the questions from a qualitative perspective, offering detailed information about the processes observed on each of the cases studied (Creswell, 2012).

Peer Assessment Workshop

The peer assessment workshop was divided into four parts, each of them with focus on different activities: individual problem-solving, peer assessment, self-assessment, and a final discussion (**Table 1**). We will just present a more in-depth description of the two first parts of the workshop (individual problem-solving and peer assessment) because those are the ones that are related to our research questions. Nonetheless, **Table 1** shows an overview of the whole workshop design.

The first part of the workshop consisted of a mathematical activity, where each of the 16 students were asked to individually solve an open-ended problem (**Figure 1**), just using paper and pencil. We chose an open-ended problem because the variety of results and ways of solving, together with the lack of

Evalúa la resolución del problema que se te entrega en documento aparte. Argumenta de forma detallada dicha evaluación.

(No escribir sobre el documento de resolución del problema)

Assess the problem resolution you receive on a separate document. Support your assessment in detail. (Do not write on the document with the problem solution).

Figure 2. Peer assessment document & translation (Source: Authors' own elaboration)

experience confronting non-routine problems, would enrich the assessment process.

Open-ended problems can be classified as non-routine problems in the sense that solvers do not know a previously established procedure to solve them (Díaz et al., 2020), so it was likely to encounter difficulties finding solutions and its assessment would be a greater challenge for the future teachers.

In the same vein of making the evaluation a challenge, we chose a problem with a topic in which students usually present difficulties, the functions, specifically, involving a quadratic function (Amaya & Medina, 2013; Díaz et al., 2020). Facing an answer with different mathematical characteristics to those given by themselves and involving a topic generally problematic, would force them to reflect on its evaluation, allowing us to notice these aspects of the resolution to which they give more importance or how attached they are to their own resolution when evaluating.

The problem was submitted to an internal validation process. Each researcher solved the problem individually, confirming that it could not be easily solved by a trivial procedure, that different methods could be used to obtain solutions (algebraically, graphically, ...) and that a variety of answers were obtained. This first individual activity is important because it offers the students the opportunity to think about the problem situation before evaluating a peer's answer and it provides the necessary material to carry out the later peer assessment.

For the second part of the workshop, researchers selected eight of the individual resolutions randomly. We chose a random selection since there was not enough time between the first and the second part of the workshop to analyze each resolution for deliberate choice. Then, we divided the students into eight pairs, also randomly. Random groupings increase mobility of knowledge and engagement in tasks (Liljedahl, 2014), which would encourage future teachers' engagement with assessment, improving the analysis.

Afterwards, each pair of students was given the anonymous solution of a third partner, and the peer assessment instrument (**Figure 2**), where we asked them to evaluate the resolution received, with no further indication and no evaluation criteria. Giving them more indications or criteria to evaluate, distanced from our objective of analyzing what aspects of the resolutions they consider when evaluating problem-solving and what they took as a reference to assess.

Table 2. Participants in each case studied and their roles

Case	Assessors	Assessed
Case 1	Ana & Peter	Robert
Case 2	Robert & Daniel	Julia

The discussion held by each assessment pair was audio-recorded in order to obtain the maximum detail of information about their peer assessment process.

Two of the authors of this paper were present during the peer assessment workshop, one as the teacher, and the other one as researcher. During the problem-solving and peer assessment activities, they just intervened to clarify doubts about the statement of tasks and take care of technical aspects related to the audio-recording of the pairs' discussions. Peer assessment workshop design did not allow to evaluate the whole students' solutions, however, this aspect lacks relevance in the current research since the objective was not to provide them feedback of the evaluation received.

The Cases Studied and the Analysis Process

The analysis relied on the study of two cases. Each case involved three participants: two students who, together, evaluated a partner's solution to the mathematical problem (the assessors), and the student whose solution was evaluated (the assessed). Criteria used to select the two cases between the eight pairs were: that the solutions assessed in each case were clearly different, there exist also differences between the assessors' profiles of each case, and that data collected let to obtain richer information to be analyzed.

Participants in each case and their role are indicated in **Table 2**. Case 1 was chosen because the assessors had an intense debate during the peer evaluation process, which could enhance the analysis. This couple was the one that spent the longest time arguing about their partner's resolution. The assessors of this case had different profiles between them: Ana has a Mathematics degree, earned last year, and did not have any teaching experience, while Peter has a statistic degree, obtained almost ten years ago, and has teaching experience. The solution assessed in this case presents an algebraic point of view of the problem.

Case 2 was chosen because the two assessors had similar profiles between them, both have a mathematics degree earned in the last two years and some teaching experience, and because the assessed solution presents a geometric point of view of the problem. By selecting these two cases, the circumstance occurs that one of the assessors of the second case (Robert) was the student assessed in case 1, which helps to simplify the analysis and the presentation of the results. Data analyzed for each case consist of three individual problem resolutions, presented by the two assessors and the

Figure 3. Ana's solution extract with a view of general resolution process followed & translation (Image of the resolution made by Ana)

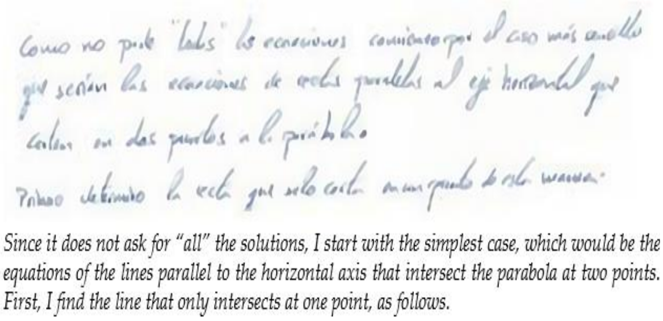


Figure 4. Peter's solution extract with a view of general resolution process followed & translation (Image of the resolution made by Peter)

assessed of each case, and the recording of the discussion held by the assessors during the peer evaluation activity.

In each case, the analysis was divided into two phases. Firstly, we analyzed the three individual problem solutions of each case. The interest was to study how they had solved the problem, what type of solutions they were looking for and what method they used to find them. To systematize this analysis, we established four categories of analysis: *general process*, *cases studied*, *representations used*, and *solutions found*. The general process category and the solution category correspond to those found by Arnal-Bailera et al. (2018) (procedure and result, respectively). The cases studied category and the representations category have been added by the authors considering the characteristics of the problem posed. The first of them has to do with the open nature of the problem, as it can be solved in multiple ways, analyzing the cases studied helps us to enhance the information about the steps followed during the resolution to reach the solution. And the second one has to do with the quadratic function involved in the problem. Often, students are instructed to graph functions during lessons (Mangwende & Maharaj, 2018), so we thought it important to analyze the representations used when solving the problem to know more about how attached they were to a solving procedure for a particular type of problem. This first analysis reveals the starting point of the evaluators when they assess their classmate's resolution. Knowing how they solved the problem allows us to analyze their assessment process in more detail.

Secondly, we analyze the transcript of the recorded discussion of the assessors during the peer assessment activity. In this phase we took two variables into account: the solution of the problem to which the assessors referred while discussing (their own or that of the evaluated peer) and the aspect of the resolution they were discussing about. To analyze this second variable, we established four categories of analysis, combining the categories used in phase 1 with those used by Arnal-Bailera et al. (2018):

- *General process*: Comments about the steps followed during the resolution procedure (for example, how the problem is approached, cases studied, representations used, …).
- *Operation*: Comments on a specific task within the general process (for example, perform arithmetic operations, set up equations, solve them, calculate points of the parabola, …).
- *Reasoning*: Comments about the explanations given by the solver (for example, justifications of the steps taken, the solutions obtained, …).
- *Result*: Comments about the results presented by the student.

The analysis of the peer assessment process was structured by dividing the audio-recording transcript into episodes based on the four categories defined.

RESULTS

Analysis of Case 1

In this section we present the analysis of the case involving Ana and Peter as assessors, and Robert as assessed. We present the two phases of the analysis separately.

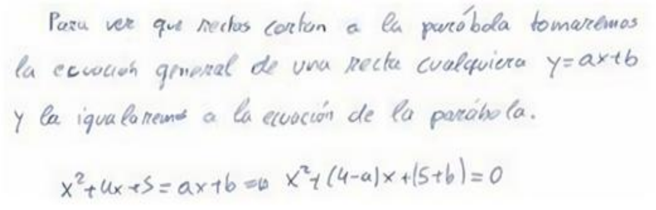
Phase 1: Individual resolutions to the problem

In this analysis, we first present the main characteristics of the two assessors' solutions in terms of the four categories indicated in the methodology section (*general process*, *representations used*, *cases studied*, and *solutions found*), and then we compare those characteristics with the assessed solution' one.

The resolutions of the problem presented by the two assessors have some similarities and differences. First, focusing on the *general process* it can be observed that both assessors use a geometric reasoning, but while Ana builds a graphical representation of the situation (**Figure 3**), Peter studies particular cases (**Figure 4**).

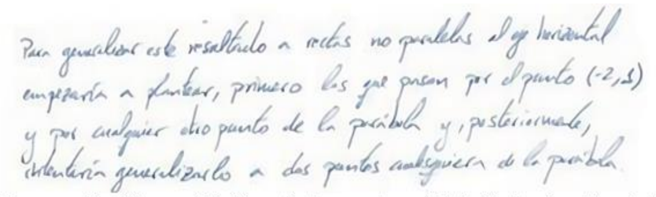
The assessed, Robert, uses an algebraic approach, assumes a solution and analyses its properties (**Figure 5**).

Figure 3, **Figure 4**, and **Figure 5** also let us observe the *cases studied* by the assessors, both analyze horizontal lines, although Ana only considers these cases, and Peter



To see which lines intersect parabola, let us take the general equation of any line y=ax+b, and set it equal to the equation of the parabola.

Figure 5. Robert's solution extract with a view of general resolution process followed & translation (Image of the resolution made by Robert)



To generalize this result to lines that are not parallel to the horizontal axis, I would first consider those that go through the point (-2, 1) and through any other point on the parabola, and then I would try to generalize it to any two points on the parabola.

Figure 6. Peter's explanation of how to study non-horizontal lines & translation (Image of the resolution made by Peter)

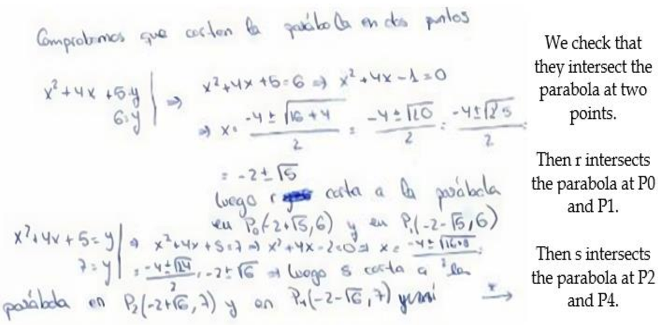


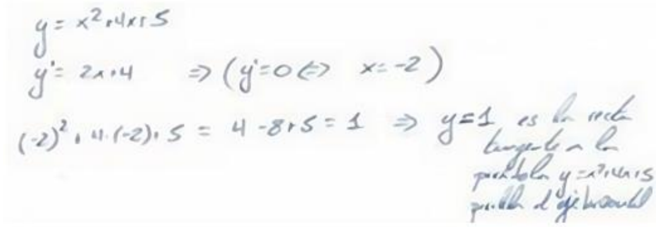
Figure 7. Example of Ana's use of symbolic representation & translation (Image of the resolution made by Ana)

includes a strategy to generalize to other types of lines (**Figure 6**). Robert, for his part, studies the most general case from the beginning.

Regarding the mathematical *representations used*, the assessors, Ana and Peter, mainly use symbolic representations, although at different times and for different purposes: Ana uses it to obtain information that allows her to draw the parabola (**Figure 3**) and to check her solutions (**Figure 7**), while Peter combine symbolic and verbal representations, using the latter to indicate different parts of their resolution process and the former to carry it out (**Figure 8**).

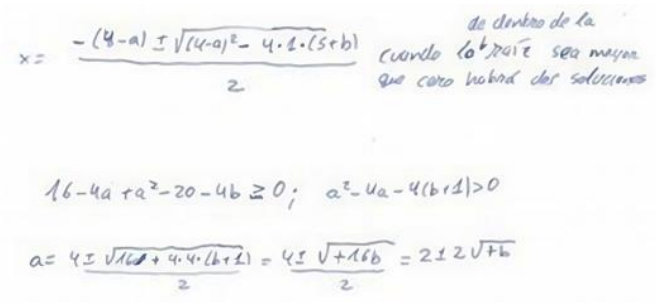
Robert, as Peter, also combines symbolic and verbal representations and uses the symbolic one for the same purposes than him (**Figure 9**).

Finally, in relation to the *solutions founded* by each member of the assessor pair, Ana only presents two particular solutions (**Figure 3**) and a generalization



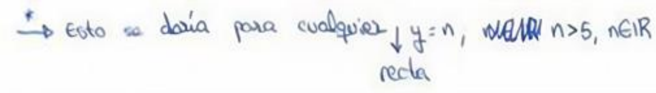
y=1 is line tangent to parabola $y=x^2+4x+5$ parallel to horizontal axis.

Figure 8. Example of use of symbolic representation by Peter & translation (Image of the resolution made by Peter)



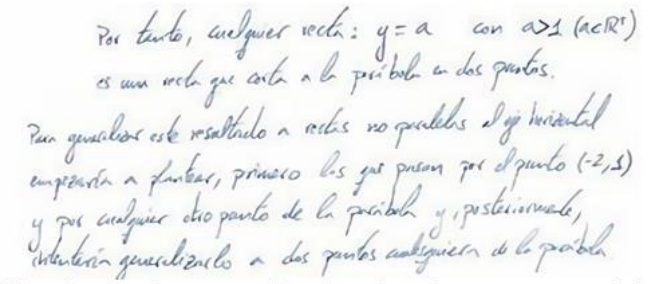
When term inside square root is greater than zero, there will be two solutions.

Figure 9. Example of use of symbolic representation by Robert & translation (Image of the resolution made by Robert)



This holds for any line y=n, n>5.

Figure 10. Ana's generalization for horizontal lines & translation (Image of the resolution made by Ana)



Therefore, any line y=a with a>1 is a line that intersects parabola at two points. To generalize this result to lines that are not parallel to horizontal axis, I would first consider those that go through point (-2, 1) & any other point on parabola, & then I would try to generalize it to any two points on parabola.

Figure 11. Solutions provided by Peter & translation (Image of the resolution made by Peter)

based on her graphical representation of the parabola (**Figure 10**).

Her solution does not show all the parallel lines that intersect the parabola at two points, due to a mistake with the representation of the vertex of the parabola, missing the lines between the point that she represents as vertex and the true vertex of the parabola.

Peter first considers all the solutions that pass through the vertex of the parabola, then he notes the existence of other solutions that do not pass through the vertex, but he did not indicate them (**Figure 11**).

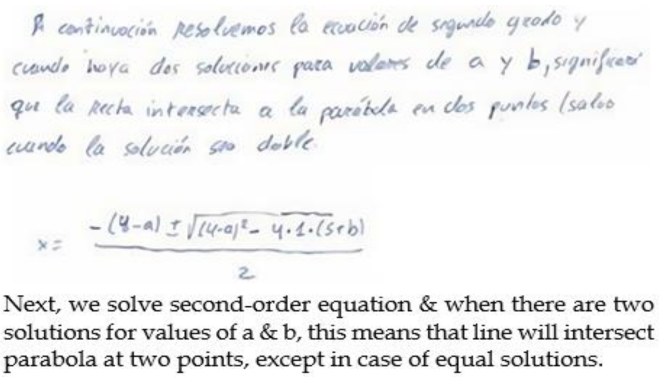


Figure 12. Robert's solution & translation (Image of the resolution made by Robert)

Robert, unlike Ana and Peter, does not study particular solutions. He tries to find conditions that a and b must satisfy so that the line y=ax+b and the parabola intersect at two points (**Figure 12**).

However, the solutions given are incorrect, since he makes calculation errors at the start of the process that carry on to the end.

Table 3 summarizes the main results of this first phase of the analysis of case 1, which will be taken into consideration in the next phase.

Phase 2: Peer assessment process

In this section, we present the analysis of transcript of discussion between Ana and Peter during assessment process, in terms of the four categories indicated in the methodology (*general process*, *operation*, *reasoning*, and *result*). We identified a total of 19 episodes. Most of them could be classified into one of the four categories above; for those who could not be classified in any of them, we created the category *other*.

The two categories with the highest number of episodes are *general process* and *result*, followed by *other*, *operation* and *reasoning* (**Table 4**). This reflects that the discussion among the evaluators focused mainly on the overall process and on the results, with some isolated reference to operations and reasoning.

The three episodes classified as *other* have to do with aspects such as the problem statement, the assessors' self-perception as problem solvers, and the assessment process itself (**Table 5**).

In many of the episodes, the assessors not only referred to the evaluated solution, but also commented on their own solutions. So, we complemented the previous analysis with the review of the solution of the

Table 3. Summary of the first phase of the analysis for case 1

	Assessor (Ana)	Assessor (Peter)	Assessed (Robert)
General process	Make a diagram (graphical representation of parabola).	Examine particular cases.	Assume a solution is available & determine its properties.
Cases studied	Horizontal lines.	Horizontal lines. Lines that pass through vertex. Lines that pass through any two points on parabola.	General equation of a line.
Representations used	Graphical & symbolic.	Verbal and symbolic.	Verbal & symbolic.
Solutions found	y=6; y=7; & y=n, with n>5	y=a, with a>1; Family of lines that pass through vertex (-2, 1), except y=1 & x=2.	y=ax+b, with $a \in (-\infty, 2-\sqrt{8}) \cup (2+\sqrt{8}, \infty)$ & b=0 y=ax+b, with $a \in (-\infty, 2-2\sqrt{b}) \cup (2+2\sqrt{b}, \infty)$ & b>0

Table 4. Number of episodes in each category (case 1)

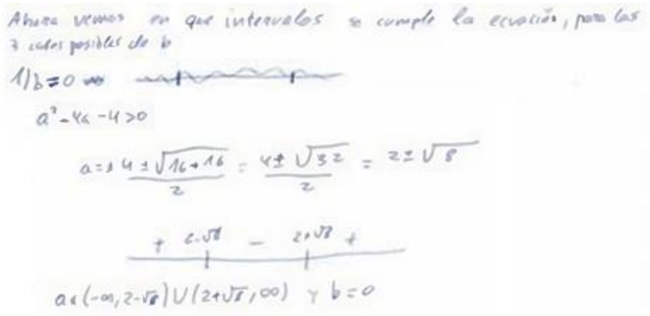
	Number of episodes	Solution to which they refer		
		Assessed (Robert)	Assessor (Ana)	Assessor (Peter)
General process	7	6	5	3
Operation	2	2	0	1
Reasoning	2	2	1	0
Result	5	5	1	3
Other	3	-	-	-
Total	19	15	7	7

Table 5. Episodes in the *other* category-1

Ana: And, truth be told, it [problem statement] does not tell you to find all of them. It tells you to find equations for line. **Peter:** Equations of lines. Two would be enough. That was my first thought.
Peter: I'm pretty rusty with this. I thought it would involve solving problems like those in high school. But this is a bit beyond me. **Peter:** I just did not feel capable of solving it any other way, but of course.
Ana: But do we have to give it a grade or something? **Peter:** No, no, no. We do not have to correct it, do we? Well, in principle, I would not assess it because I do not know what I have to assess it on, if I have to assign it a grade, or say if it was on the right track.

Table 6. Episodes in the *general process* category-1

Peter: I did something very simple, but it has nothing to do with this. I found an infinite number of solutions that were all parallel to the horizontal axis.
Ana: Me too. I stopped there. Because no more would fit.
Peter: I say that because this approach for me is not … I wanted to do that and then generalize. But I did not get to generalize, I do not know if it was because I did not have enough time or […] That one is a bit beyond me, plus I just do not see it. I'm not sure that was it. It does not sound right to me but I'm not saying it's not.
Ana: It's just that ... by not representing anything graphically, I think it's much harder.



Now we see intervals, where equation holds for three possible values of b

Figure 13. Robert's partial solution & translation (Image of the resolution made by Robert)

problem to which the assessors referred while discussing (their own or that of the evaluated peer).

Table 4 shows that, in most episodes, Ana and Peter refer to more than one resolution, commenting on their own responses on several occasions. This was specially observed in the *general process* episodes. Only in two of the seven episodes of this category the assessors were focused on Robert's resolution. In one episode, Ana refers only to her solution, while in the other four, Ana and Peter are comparing their own solutions with the assessed one (**Table 6**).

This fragment also reflects a certain affinity of the assessors to their own resolution process, which at times hinders their assessment. In the episodes of the remaining categories, Ana and Peter continue using their own resolutions as a reference. In one of the two *operation* and *reasoning* episodes and in three of the five *result* episodes, they compare their solutions with Robert's one. The existence of these episodes allowed us to observe the difficulties the assessors encounter to understand the operations that Robert performs and the way he expresses his results. An example of this fact can be seen in the following fragment from the transcript where the assessors are trying to understand the meaning of the expression that Robert obtains in his first attempt to solve the equation (**Figure 9**).

Ana and Peter do not seem to realize that Robert goes on to consider different cases (b=0 and b<0) to solve the inequality he had posited (**Figure 13**), which increases their doubts (**Table 7**).

Later in their conversation, the assessors discuss the meaning of the results obtained by Robert for the case b=0, which also leaves them with doubts (**Table 8**).

After this discussion, Peter attributes the problems he is having understanding Robert's reasoning to the lack of an explanation, as we can see in the following excerpt (**Table 9**).

Analysis of Case 2

We continue presenting the analysis of the case involving Robert and Daniel as assessors, and Julia as assessed. As in the previous case, we will distinguish the two phases of the analysis.

Phase 1: Individual resolutions to the problem

Following the same scheme as in Case 1, in this phase we first present the main characteristics of the two assessors' solutions in terms of the categories *general process*, *cases studied*, *representations used*, and *solutions found*, and then we compare those characteristics with the assessed solution' one. As in the other studied case, some similarities and differences were identified. First, the *general processes* followed by the assessors are similar (**Figure 5** and **Figure 15**).

Table 7. Episode in the *operation* category-1

Peter: No but … a depends on b and b depends on a, surely. But the point is that he solves it again as if what he had inside the radical is a second-order equation [He is describing what is shown in **Figure 9**]. Well, actually an inequality. He solves for the value of a, and he's left with a function of b. And then he does an analysis based on the possible values of b. And he only analyzes one specific case. That is, in reality when he only analyzes the case b=0, that is, a=2. When b=0, which is the case, he says, is it just me or a=2 and that's it?
Ana: a=2, yes.
Peter: And all this ...
Ana: Ah, but then?
Peter: Of course. He goes back. He goes back because here he concludes. And now he says well, now b=0, okay. Then a=2. He always goes back. That's what I do not understand.
Ana: But of course, what I do not think he considered is that this still has to be satisfied. Yes, yes, it is satisfied because when b=0, a=2, then this has a solution

Table 8. Episode in the *result* category-1

Peter: What I'm saying is, what he is doing with y=ax+b and suddenly saying b is equal to zero, is that he is taking all the lines that pass through (0, 0). Because he's left with y=ax. The parabola has ... You calculated it more or less like I did. Its shape is more or less like this. Actually, what he is going to calculate is all the lines that pass through (0, 0), the family of lines that more or less would be ... From here, that one only intersects it once, that would not work; all the ones in the middle of the family of lines that pass through here until the tangent line, around here at some point, well, it's an asymptote actually ... (**Figure 14**). [...] **Peter:** Yes, I suppose the a ... What he just got here: when b=0, what he's saying is that the a has to be from minus infinity to two minus the square root of 8, this must be ... [They are referring to the part of Robert's document shown in **Figure 10**]. I doubt this interval is that large. Only a few "as" are going to be valid, the ones that, the other way, the ones that are here, at some point, if the a is smaller, none of them intersect until it gets large enough to come around the other way to here. I do not know if you understand what I mean.

Table 9. Episode in the *reasoning* category

Peter: If he had tried to explain in words what he was doing, what he was trying ... we could follow his reasoning. But at some point, I cannot follow it and I do not know why he's doing the things he's doing.

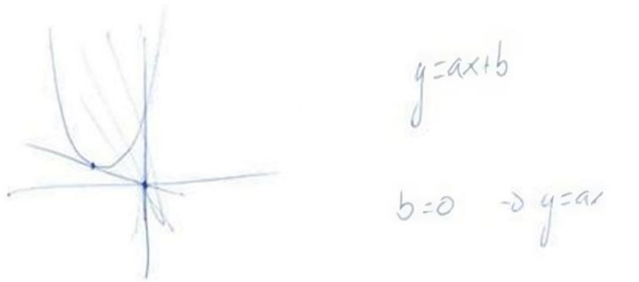


Figure 14. Drawing Peter uses to support his reasoning (Image of the resolution made by Peter)

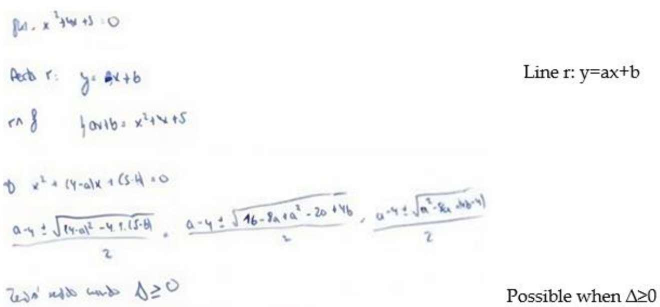


Figure 15. Daniel's solution extract with a view of general process followed & translation (Image of the resolution made by Daniel)

Both assume that there is a solution to the problem and analyze its properties, while Julia, the assessed, examines particular cases (**Figure 16**).

Moreover, Robert and Daniel solve the problem from an algebraic perspective, while Julia only needs to use arithmetic tools.

In terms of the *cases studied*, the assessors try to find the general properties of the lines that satisfy the conditions of the problem (**Figure 5** and **Figure 15**). Neither of them was successful with this strategy.

However, given the impossibility of solving the desired equation, Daniel chose to look for particular solutions (**Figure 17**), something that his partner did not do. The assessed, Julia, only studies two particular solutions (**Figure 16** and **Figure 18**).

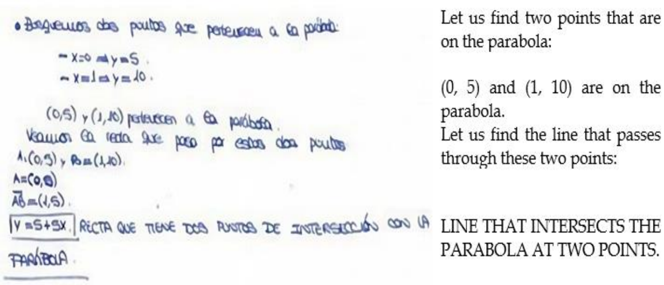


Figure 16. Julia's solution extract with a view of general process followed & translation (Image of the resolution made by Julia)

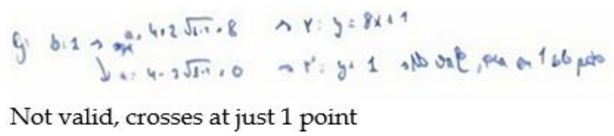


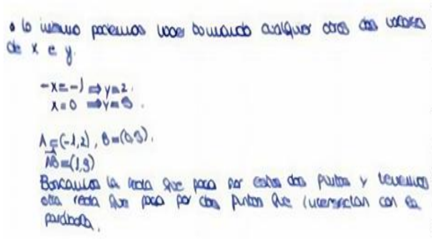
Figure 17. Daniel's solution & translation (Image of the resolution made by Daniel)

About the mathematical *representations used*, the three students mainly use the symbolic representation throughout the resolution process. Daniel combines the use of the symbolic system with the graphical one, to represent the parabola and to check his solutions (**Figure 19**).

Robert and Julia combine the symbolic representation with the verbal one to explain their resolution process (**Figure 5** and **Figure 18**), and in the case of Robert, also to justify some of the steps performed (**Figure 9** and **Figure 12**).

Regarding the *solutions found*, neither of the two evaluators presented correct answers. Only Julia provides correct solutions, corresponding to the two particular lines shown above.

Table 10 summarizes the main results of the first phase of the analysis of case 2, which will be taken into consideration in the next phase.



We can do the same thing by taking any two other values of x & y.

We look for the line that passes through these two points & we have another line that passes through two points that intersect parabola.

Figure 18. Second case studied by Julia & translation (Image of the resolution made by Julia)

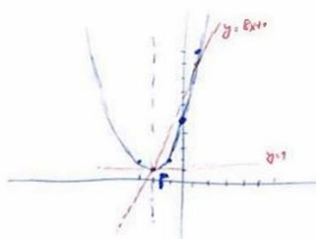


Figure 19. Daniel's graphical representation (Image of the resolution made by Daniel)

Table 10. Summary of the first phase of the analysis for case 2

	Assessor (Daniel)	Assessor (Robert)	Assessed (Julia)
General process	Assume a solution is available & determine its properties.	Assume a solution is available & determine its properties.	Examine particular cases.
Cases studied	General equation of a line.	General equation of a line.	Line that passes through points (0, 5) & (1, 10) of the parabola. Line that passes through points (0, 5) & (-1, 2) of parabola.
Representations used	Graphical & symbolic.	Verbal & symbolic.	Verbal & symbolic.
Solutions found	y=8x+1	y=ax+b, with $a \in (-\infty, 2-\sqrt{8}) \cup (2+\sqrt{8}, \infty)$ & b=0 y=ax+b, with $a \in (-\infty, 2-2\sqrt{b}) \cup (2+2\sqrt{b}, \infty)$ & b>0	y=5+5x. Although she does not give its expression, she provides line that passes through points (0, 5) & (-1, 2) & whose directing vector is (1, 3).

Table 11. Number of episodes in each category (case 2)

	Number of episodes	Solution to which they refer		
		Assessed (Julia)	Assessor (Daniel)	Assessor (Robert)
General process	8	7	3	4
Operation	1	1	0	0
Result	6	4	1	1
Other	6	-	-	-
Total	21	12	4	5

Table 12. Episodes in the *other* category-2

Daniel: Because of course, the statement says, "find the equations of lines."
Robert: The statement can be misinterpreted no matter what you do. And if it's "equations of lines," how many do you have to find for your answer to be right?

Phase 2: Peer assessment process

The analysis of the transcript obtained from Daniel and Robert's assessment process discussion yielded a total of 21 episodes. The category that covers the highest number of episodes is *general process*, followed by *result* and *other*. There was just one episode about *operation* (**Table 11**).

In this case, no comments were identified in the *reasoning* category, which could be associated with the fact that the document assessed lacked explicit explanations.

The six episodes classified as *other* have to do with the problem statement (five episodes) or the assessment process itself (one episode). Regarding the problem statement, the assessors have questions about the "appropriate" number of lines that had to be provided to consider that the problem is solved (**Table 12**).

About the assessment process, Robert admits not being sure that they have to assess, as the excerpt shows (**Table 13**).

As in case 1, we complemented the previous analysis with the review of the solution of the problem to which the assessors referred while discussing (their own or that of the evaluated peer).

Table 11 shows that assessors refer to their own resolutions just where they discuss the *general process* to solve the problem.

In three of these episodes, the assessors only refer to the assessed resolution, while in the four remaining episodes of this category, the assessors compare their own *general process* with Julia's. Most of these episodes refer to the fact that Julia, the student being assessed, only studies particular solutions to the problem (**Table 14**).

Table 13. Episodes in the *other* category-3

Robert: We are assessing this without knowing what we have to do. Without us having done it.
Daniel: I do not know what I would do.
Robert: Because if I assign an exercise and I know what answer they have to find, it's easier for me to assess it. If I get to this point, I do this ... If they do something else, it depends on what you want, I just do not know.

Table 14. Episode in the *general process* category-2

Robert: What strikes me is that she is only looking for one line, she did not look for more.
[...]
Daniel: Yes, that is, she wrote the equation of the parabola. She took two points and, using the equation of the line that passes through two points, she found that line. So, sure, that way you make sure it intersects it at two points.

Table 15. Episode in the *general process* category-3

Robert: We're supposed to grade it. What would you give it? A 5 [out of 10]?
Daniel: Well, I did not know how to finish it.
Robert: I did not finish it either. But I tried something else.
Daniel: Well, he found one, right? So, more or less ... That's enough for a 5.
Robert: Ok, then.
Daniel: And then, the idea of doing the same thing over and over again is fine in the sense that visually, you can get an idea of all the lines that intersect it. But he does not give an equation.
Robert: It's impossible. It would take you forever.
Daniel: It's impossible.
Daniel: That is, he does not characterize the lines that intersect the parabola twice. He only gives examples. And you can give infinite examples, but you're not giving all of them.
[…]
Daniel: For me, a 10 would require finding a general equation for all the lines, and I would give an 8 or 7 to people who come close, you know?

Table 16. Episode in the *result* category-2

Robert: Yes, that case b less than 0 would have failed ... Well, some of the cases ...
Daniel: Yes.
Robert: To me, at least, it depended on the value of b.

Table 17. Episode in the *general process* category-4

Daniel: If that person had done that [referring to Robert's process in **Figure 10**], I would grade it higher. But they really only found one and gives instructions for finding "as many as you want," but does not ...
Robert: You can take infinite points, but that's not elegant.
Daniel: I would give it a 5. Do you agree?
Robert: OK, sure.

It generates a discussion between the evaluators about what grade to assign to their partner's answer. In this discussion, the evaluators continually compare the general resolution process that they used, searching for general solutions, with that used by their partner (**Table 15**).

In the episodes of the remaining categories, the assessors mainly focus on Julia's resolution, with just two references to their own work, related with the *results* obtained by each one. For example, in the following excerpt, the assessors are discussing if the characteristics of a solution should have to be graded with seven or eight points, and Robert put their own solution as example (**Table 16**).

This comparison between their own *general process* and their *solutions* with those of Julia, could be indicative that the assessors have formed a bond with their own solutions (**Table 17**).

Finally, the only one comment classified in the *operation* category (**Table 18**) corresponds to an episode in which Robert expresses some doubts about the second part of Julia's resolution (**Figure 18**).

DISCUSSION

The general objective of this research was to analyze how prospective secondary mathematics teachers assess their classmate's solutions to a mathematical open-ended problem. Our interest in this issue lies in two facts. Firstly, assessment is one of the main teacher professional tasks, so it must be part of teachers training. Secondly, traditional tests are still the most evaluation method used by mathematics teachers (e.g., Chanudet, 2016; Nieminen & Atjonen, 2022), so if we intend this to change, teacher training must include alternative methods to assess students' learning.

Table 18. Episode in the *operation* category-2

Robert: I do not understand this part very well. He took two points, joined them and created the line?
Daniel: Of course. He took two points on the parabola.
Robert: Ok, yes. He created the line from the points.
Daniel: Yes
Robert: One point and the direction vector? It's not. Is there another method? I do not get it.
Daniel: No, no. Like the first one. What happens is he's saying: well, I'll take any point on the parabola and the line that passes. Well, he takes two points on the parabola, and the line that passes through those two points. That is, he or she explains [what to do] it but does not give the equation of the line that passes through those points.

Table 19. Comparison of the number of episodes in each category of case 1 & case 2

	Case 1	Case 2
General process	7	8
Operation	2	1
Reasoning	2	0
Result	5	6
Other	3	6
Total	19	21

Table 20. Episode in the *reasoning* category in case 1

Ana: It's just that ... by not representing anything graphically, I think it's much harder.

One way to introduce future teachers into the analysis and reflection about evaluation is through peer assessment activities, which offer interesting opportunities for teacher training, making future teachers aware of the complexity of the assessment process at the same time that offer them information about their own subject knowledge (Beaver & Beaver, 2011; Lavy & Shriki, 2014; Zevenbergen, 2001).

In this study, we were interested in identifying particular characteristics of the process that future secondary mathematics teachers follow to evaluate their peers' resolution to a mathematical open-ended problem when they have not been specifically training for that. The two research questions were: What aspects of the problem-solving do future teachers focus on when evaluating a peers' solution? And, in what ways do prospective teachers use their own resolutions to evaluate the resolution of their peers? The answers to these questions provide information on which aspects of problem-solving are considered important by future teachers. They also reveal information about future teachers' conception about evaluation, if they associate this process only with the assignment of a grade or with a formative perspective (Silver & Mills, 2018; Wiliam & Thompson, 2007). Finally, the answers to these questions allow us to know what the basis on which future teachers is build their evaluation when they have not yet received specific training on the evaluation process. All this knowledge is especially useful since it allows establishing a baseline on which to build the professional knowledge of these future teachers about the evaluation process.

Related to the first research question, results show that in the two analyzed cases (**Table 19**, collected from **Table 4** and **Table 11**), the category in which a greater number of episodes was found was the *general process*. The second most common issue assessors looked at was the *result*. The two cases also have in common that, during the discussion held to evaluate the work of their partner, they made little reference to *operations* and, when they did, it was to show their doubts about some part of the mathematical procedure carried out by their colleagues.

One of the differences between the two cases is that just case 1 assessors referred to the mathematical *reasoning* presented by their classmate. They complain about the lack of explanations given from the assessed peer (although what really happens is that they do not identify them) and how difficult it is to understand the answer as it does not incorporate graphic reasoning (**Table 20**).

These results reflect that difficulty in understanding the resolution given by another person directly intervenes in the assessment process, which confirms that assessors' mathematical knowledge is one of the variables that intervene in the quality of the assessment (Cáceres & Chamoso, 2015).

The categories used in the analysis (*general process*, *operation*, *reasoning* and *result*), defined from those indicated by Arnal-Bailera et al. (2018), were especially useful for most of the episodes identified during the peer assessment process, in the two cases studied. Nevertheless, the analysis also revealed episodes that did not correspond to any of these categories, and which were classified in the *other* category. In the two assessing pairs we find a few episodes in this category (**Table 19**). In both cases this comments are related to the problem statement (**Table 5** and **Table 12**) and to the concept of assessment (**Table 5** and **Table 13**), and in case 1, we also find comments related to the self-perception as problem solvers (**Table 5**).

Regarding the problem statement, the two assessor pairs had doubts about the number of lines that should be indicated for the problem to be solved. Case 1 interpreted the expression "find lines" as "more than one", while case 2 interpreted it as "all of them", and this conditioned their evaluations. This reflects the role of the type of mathematical task in the assessment process.

Episodes about participants' self-perception as problem solvers (**Table 5**) refers to the fact that students do not usually face mathematical situations that go beyond the direct application of an algorithm in which it is necessary to look for strategies or different solutions (Felmer & Perdomo-Díaz, 2016). The use of an open-ended problem in this research was essential since it allowed the future secondary teachers to deal with different strategies to solve the problems and to obtain different solutions. Peer assessment activity led future teachers the opportunity to observe different ways of dealing with the same problem, enriching their problem-solving strategies and their mathematical knowledge (Beaver & Beaver, 2011; Lavy & Shriki, 2014; Zevenbergen, 2001).

Finally, about future teachers' concept of assessment, three of the four assessors associated it with a method for assigning a grade, something quite common, as pointed out by Shahbari and Abu-Alhija (2018). Just Peter, one of the assessors of case 1, established a difference between assessment and grading (**Table 5**). Providing assessment criteria could have helped avoid the future teachers' questions related to the assessment process. However, we consider that not indicating evaluation criteria was a success in the context of this study, since it led us to observe participants' own points of view (Jones & Alcock, 2014; Seifert & Feliks, 2018).

Related to the second research question, the analysis of the two cases showed that future teachers' do not limit themselves to reviewing the resolution to be evaluated. They also consider their own responses to the mathematical problem. Both pairs of assessors show a certain degree of attachment to their own solutions. In case 1, assessors mention their own solutions 14 times, comparing them with that of the evaluated in episodes of the four categories from Arnal-Bailera et al. (2018), but mainly in episodes about general process and result (**Table 4**). Case 2 assessors refer to their own solution in nine moments, all of them referring to the general process of resolution except two, referred to the result (**Table 11**). In most of these episodes, where the assessors compare their responses with the assessed ones, future teachers were trying to understand the mathematical expression that their peer gave as solution (case 1) or valuing the fact that their peer presented only particular solutions to the problem (case 2). These results show the influence of the attachment to one's own solution in the peer assessment processes, already mentioned in other studies (e.g., Cárdenas et al., 2016).

The continuous references by the assessors to their own resolutions allowed us to observe that one of the variables that can influence the peer evaluation process is the difference between the assessors' resolutions and that being evaluated. Phase 1 of the analysis showed that, in both cases, the two assessors solved the problem in a similar way and that their resolution differed considerably from that they had to assess (**Table 3** and **Table 10**). This result is not the outcome of an intentional action, since the assessing pairs were randomly grouped, and their solutions were not checked before grouping.

To sum up, results indicate the following main characteristics for the peer assessment process in the two cases analyzed: Assessors look at their peer resolution, but also consider their own one, becoming to show some grade of attachment to their work. During the evaluation, the assessors refer to the four categories by Arnal-Bailera et al. (2018), i.e., *general process*, *operation*, *reasoning* and *result*, but also to other issues such as the problem statement and the concept of assessment. Finally, there is some evidence that variables such as the type of task whose resolution will be evaluated, the difference between the assessors' resolution and the assessed one, and the assessors' conception of assessment could influence the peer evaluation process.

We have been able to observe all this thanks to the methodological design chosen to carry out this study. Opting for a case study has allowed us to in-depth analyze the peer assessment process followed by the members of each case. This decision also introduces some limitations in the research, mainly related to particularities of each case like the type of resolution that each future teacher studied present and the differences between them. Nevertheless, we believe that the in-depth study of these two cases led us to expand the set of categories to analyze the assessment process, particularly in peer assessment settings, and to generate a scheme of analysis that could be interesting for other researchers.

Results of this study can also be useful to design mathematics teachers training courses with focus on evaluation, for both initial training and teacher professional development. Ayalon and Wilkie (2021) found in their research that pre-service teachers, after participating in a course with focus on formative assessment, became aware of the need to distinguish between different levels of quality of responses to a task and acquired confidence in designing mathematical tasks and assessment criteria. While Watson (2000) points out that teachers, even receiving assessment training, may be involved in situations of inequality, because of problems of observation, perspective, interpretation, and expectation. In addition, assessing with evidence and critically is also a competence that teachers in training must acquire, so it is important to study whether pre-service teachers are prepared to develop evaluation adequately. Assessment will be a fundamental part of teachers' professional activities, which is why it is important to receive specific training on it. Even more after the COVID-19 pandemic in which teachers had to create new ways of assessment, with an emphasis on how to make a distance assessment and how to do it effectively (Bakker et al., 2021). So, more care needs to be taken in assessment training, since teacher assessment contributes to students' learning, their mathematical attainment, and their future forecast.

Author contributions: All authors have sufficiently contributed to the study and agreed with the results and conclusions.

Funding: This study was partially supported by the Project ProID2021010018, from Research and Innovation Strategies for Smart Specialization (Canary Islands Government, ERDF 2014-20).

Ethical statement: Authors stated that all participants were over 18 years old and their participation was voluntary. Authors further stated that the study does not require ethics committee approval since no personal data were analyzed and pseudonyms are used in this paper.

Declaration of interest: No conflict of interest is declared by authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Amaya, T., & Medina, A. (2013). Dificultades de los estudiantes de grado once al hacer transformaciones de representaciones de una función con el registro figural como registro principal [Difficulties of eleventh grade students when making transformations of representations of a function with the figural register as main register]. *Educación Matemática* [*Mathematics Education*], *25*(2), 119-140.

Arnal-Bailera, A., Muñoz-Escolano, J. M., & Oller-Marcén, A. M. (2018). Análisis de las anotaciones realizadas por profesores al calificar pruebas escritas de matemáticas [Analysis of the annotations made by teachers when grading written math tests]. In L. J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, A. Aguilar-González, P., Alonso, F. J. García García, & A. Bruno (Eds.), *Investigación en educación matemática XXII* [*Research in mathematics education XXII*] (pp. 131-140). SEIEM.

Ayalon, M., & Wilkie, K. J. (2021). Investigating peer assessment strategies for mathematics pre-service teacher learning on formative assessment. *Journal of Mathematics Teacher Education, 24*, 399-426. https://doi.org/10.1007/s10857-020-09465-1

Bakker, A., Cai, J., & Zenger, L. (2021). Future themes of mathematics education research: An international survey before and during the pandemic. *Educational Studies in Mathematics, 107*, 1-24. https://doi.org/10.1007/s10649-021-10049-w

Beaver C., & Beaver. S. (2011). The effect of peer assessment on the attitudes of pre-service elementary and middle school teachers about writing and assessing mathematics. *IUMPST: The Journal, 5*, 1-14.

Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31. https://doi.org/10.1007/s11092-008-9068-5

Cáceres, M. J., & Chamoso, J. M. (2015). La evaluación sobre la resolución de problemas de matemáticas [The assessment on solving math problems]. In L. Blanco, J. A. Cárdenas, & A. Caballero (Eds.), *La resolución de problemas de matemáticas en la formación inicial de profesores de primaria* [*The resolution of mathematics problems in the initial training of primary school teachers*] (pp. 225-241). Universidad de Extremadura.

Cai, J., & Lester, F. (2010). *Why is teaching with problem-solving important to student learning?* National Council of Teachers of Mathematics.

Cárdenas, J. A., Blanco, L. J., Guerrero, E., & Caballero, A. (2016). Manifestaciones de los profesores de matemáticas sobre sus prácticas de evaluación de la resolución de problemas [Mathematics teachers' statements about their problem-solving assessment practices]. *Bolema, 30*(55), 649-669. https://doi.org/10.1590/1980-4415v30n55a17

Chan, M. C. E., & Clarke, D. (2017). Structured affordances in the use of open-ended tasks to facilitate collaborative problem-solving. *ZDM Mathematics Education, 49*, 951-963. https://doi.org/10.1007/s11858-017-0876-2

Chanudet, M. (2016). Assessing inquiry-based mathematics education with both a summative and formative purpose. In P. Liljedahl, & M. Santos-Trigo (Eds.), *Mathematical problem-solving: Current themes, trends, and research* (pp. 177-207). Springer. https://doi.org/10.1007/978-3-030-10472-6_9

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.

Custodia, E., Márquez, C. & Sanmartí, N. (2015). Aprender a justificar científicamente a partir del estudio del origen de los seres [Learn to justify scientifically from the study of the origin of beings]. *Enseñanza de las Ciencias* [*Science Education*], *33*(2), 133-155. https://doi.org/10.5565/rev/ensciencias.1316

Díaz, V., Aravena, M., & Flores, G. (2020). Solving problem types contextualized to the quadratic function and error analysis: A case study. *EURASIA Journal of Mathematics, Science and Technology Education, 16*(11), em1986. https://doi.org/10.29333/ejmste/8547

Drijvers, P., Kodde-Buitenhuis, H., & Doorman, M. (2019). Assessing mathematical thinking as part of curriculum reform in the Netherlands. *Educational Studies in Mathematics, 102*, 435-456. https://doi.org/10.1007/s10649-019-09905-7

Felmer, P., Liljedahl, P., & Koichu, B. (Eds). (2019). *Problem-solving in mathematics instruction and teacher professional development.* Springer. https://doi.org/10.1007/978-3-030-29215-7

Felmer, P & Perdomo-Díaz, J. (2016). Novice Chilean secondary mathematics teachers as problem solvers. In P. Felmer et al. (Eds.), *Posing and solving mathematical problems*. Research in Mathematics Education. Springer. https://doi.org/10.1007/978-3-319-28023-3_17

Goos, M. (2014). Mathematics classroom assessment. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 413-417). Springer. https://doi.org/10.1007/978-94-007-4978-8

Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education, 39*(10), 1774-1787. https://doi.org/10.1080/03075079.2013.821974

Lavy, I., & Shriki, A. (2014). Engaging prospective teachers in peer assessment as both assessors and assessees: The case of geometrical proofs. *International Journal for Mathematics Teaching and Learning*.

Liljedahl, P. (2014). The affordances of using visually random groups in a mathematics classroom. In Y. Li, E. Silver, & S. Li (Eds.), *Transforming mathematics instruction: Multiple approaches and practices*. Springer. https://doi.org/10.1007/978-3-319-04993-9_8

Liljedahl, P., Santos-Trigo, M., Malaspina, U., & Bruder, R. (2016). *Problem-solving in mathematics education.* Springer. https://doi.org/10.1007/978-3-319-40730-2_1

Mangwende, E., & Maharaj, A. (2018). Secondary school mathematics teachers' use of students' learning styles when teaching functions: A case of Zimbabwean schools. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(7), 3225-3233.https://doi.org/10.29333/ejmste/91679

Mason, J. (2015). When is a problem? Contribution in honor of Jeremy Kilpatrick. In E. Silver, & C. Keitel-Kreidt (Eds.), *Pursuing excellence in mathematics education* (pp. 55-69). Springer. https://doi.org/10.1007/978-3-319-11952-6_4

Mason, J. (2016). When is a problem …? "When" is actually the problem! In P. Felmer, E. Pehkonen, & J. Kilpatrick (Eds.), *Posing and solving mathematical problems* (pp. 233-285). Springer. https://doi.org/10.1007/978-3-319-28023-3_16

Mogessie, M. (2015). Peer assessment in higher education–twenty-first century practices, challenges and the way forward. *Assessment & Evaluation in Higher Education, 42*(2), 226-251. https://doi.org/10.1080/02602938.2015.1100711

Nieminen, J. H., & Atjonen, P. (2022). The assessment culture of mathematics in Finland: A student perspective. *Research in Mathematics Education*. https://doi.org/10.1080/14794802.2022.2045626

Nortvedt, G., Santos, L., & Pinto, J. (2016). Assessment for learning in primary school mathematics teaching: The case of Norway and Portugal. *Assessment in Education: Principles, Policy & Practice, 23*(3), 377-395. https://doi.org/10.1080/0969594X.2015.1108900

Olson, J. C., & Knott, L. (2013). When a problem is more than a teacher's question. *Educational Studies in Mathematics, 83*, 27-36. https://doi.org/10.1007/s10649-012-9444-4

Polya, G. (1945). *How to solve it*. Princeton University Press. https://doi.org/10.1515/9781400828678

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1-31. https://doi.org/10.1207/s15326977ea1101_1

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem-solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334-370). MacMillan.

Seifert, T., & Feliks, O. (2018). Online self-assessment and peer- assessment as a tool to enhance student-teachers' assessment skills. *Assessment & Evaluation in Higher Education, 44*(2), 169-185. https://doi.org/10.1080/02602938.2018.1487023

Shahbari, J. A., & Abu-Alhija, F. N. (2018). Does training in alternative assessment matter? The case of prospective and practicing mathematics teachers' attitudes toward alternative assessment and their beliefs about the nature of mathematics. *International Journal of Science and Mathematics Education, 16*(7), 1315-1335. https://doi.org/10.1007/s10763-017-9830-6

Silver, E. A., & Mills, V. L. (Eds.) (2018). *A fresh look at formative assessment in mathematics teaching*. The National Council of Teachers of Mathematics.

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

Suurtam, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practices, 17*(4), 399-417. https://doi.org/10.1080/0969594X.2010.497469

Szetela, W., & Nicol, C. (1992). Evaluating problem-solving in mathematics. *Educational Leadership, 49*(8), 42-45.

Topping, K.J. (2009). Peer assessment. *Theory into Practice, 48*(1), 20-27. https://doi.org/10.1080/00405840802577569

Ukobizaba, F., Nizeyimana, G., & Mukuka, A. (2021). Assessment strategies for enhancing students' mathematical problem-solving skills: A review of literature. *EURASIA Journal of Mathematics, Science and Technology Education, 17*(3), em1945. https://doi.org/10.29333/ejmste/9728

Watson, A. (2000). Mathematics teachers acting as informal assessors: Practices, problems and recommendations. *Educational Studies in Mathematics, 41*, 69-91. https://doi.org/10.1023/A:1003933431489

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-82). Erlbaum. https://doi.org/10.4324/9781315086545

Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice, 17*(1), 59-75. https://doi.org/10.1080/09695940903565610

Zevenbergen, R. (2001). Peer assessment of student constructed posters: Assessment alternatives in pre-service mathematics education. *Journal of Mathematics Teacher Education, 4*, 95-113. https://doi.org/10.1023/A:1011401532410

https://www.ejmste.com