PRESERVICE TEACHERS INTERPRETING AND ACTING ON STUDENTS' RESPONSES TO A PROBABILITY PROBLEM

MARÍA DEL MAR LÓPEZ-MARTÍN
University of Almería
mdm.lopez@ual.es

MARÍA BURGOS NAVARRO
University of Granada
mariaburgos@ugr.es

VERÓNICA ALBANESE
University of Granada
vealbanese@ugr.es

ABSTRACT

To ensure the learning of mathematics, teachers must be able to analyse their students' mathematical practices when solving tasks, interpret the difficulties that students encounter, and decide how to manage students' difficulties. This competence in didactic analysis and intervention allows teachers to adapt their teaching to meet individual student needs. In this paper, we analyse how preservice teachers interpreted students' responses to a task involving the proportional determination of probabilities and understanding the sample space. We propose didactic strategies that help students overcome difficulties. The results revealed the difficulties some preservice teachers experienced in adequately interpreting student responses and making informed decisions to improve learning.

Keywords: *Probability; Teacher training; Ontosemiotic approach; Didactic analysis*

1. INTRODUCTION

The incorporation of probability teaching from the earliest educational levels (Australian Curriculum, Assessment and Reporting Authority, 2014; Ministerio de Educación & Formación Profesional, 2022; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) poses a challenge for teachers who do not always have the necessary mathematical and didactic training to teach probability (Franco & Alsina, 2022; Vásquez & Alsina, 2017). Probabilistic reasoning involves the following: (a) understanding the fundamental probabilistic ideas of variability, randomness, independence, and predictability/uncertainty; (b) calculating or estimating the probabilities of events in everyday random situations; (c) using the language of chance appropriately; and (d) using arguments to prove the veracity of a probabilistic statement or the validity of the solution to the problem (Sánchez & Valdez, 2017). On the one hand, it is clear that teachers must possess these competences, but on the other hand, possessing them is not enough to ensure that probability is taught well (Sánchez & Valdez, 2017; Vásquez & Alsina, 2017). Teachers must be able to recognise the characteristics of students' probabilistic thinking and use this information to make relevant action decisions (Batanero et al., 2015; Burgos et al., 2022).

Research has indicated that some preservice teachers exhibit biases in probabilistic reasoning and that their didactic knowledge of probability could be improved (Batanero & Álvarez-Arroyo, 2024; Franco & Alsina, 2022). As a result, some teachers may encounter difficulties interpreting students' answers when solving probability problems, recognising the relevant mathematical elements in students' responses, and deciding how to act in the face of students' understanding or lack thereof (Batanero et al., 2015; Burgos et al., 2022, 2023). Research has also suggested that nonnormative proportional reasoning can lead to errors in the interpretation of probability concepts or the application of probabilities (Begolli et al., 2021; Bryant & Nunes, 2012; Van Dooren, 2014). Proportional reasoning

Statistics Education Research Journal, 24(2). https://doi.org/10.52041/serj.v24i2.819

is part of sample space analysis, proportional probability quantification, and the understanding and use of correlations (Begolli et al., 2021; Bryant & Nunes, 2012; Van Dooren, 2014). However, most of the research in this area has focused on analysing the ability to successfully solve probability comparison tasks in the context of urns (Batanero et al., 2023), with less attention paid to understanding the construction of the sample space (Hernández-Solís et al., 2021; Supply et al., 2023).

In this regard, Hernández-Solís and collaborators analysed the construction of sample spaces by primary education students (Hernández-Solís et al., 2021) and secondary education students (Hernández-Solís et al., 2024). Despite observing a reasonable intuition about the concept of a sample space in both urn and roulette contexts (Hernández-Solís et al., 2021), the authors noted that students tended not to adequately relate the sample space to the greater or lesser probability of each experimental outcome. In addition, those with higher levels of proportional reasoning achieved better results relative to constructing sample spaces (Hernández-Solís et al., 2024).

Unlike these studies, research evaluating the understanding of sample spaces in the context of teacher education has focused on situations involving the calculation and comparison of probabilities or the fairness of games (Ortiz & Mohamed, 2014; Vásquez & Alsina, 2015a, 2015b, 2017). In any case, the difficulties observed in some preservice teachers' understanding of sample spaces could limit their ability to interpret students' errors when solving problems in this context (Ortiz & Mohamed, 2014) or lead prospective teachers to consider incorrect strategies as valid when analysing students' responses (Vásquez & Alsina, 2017). We have not found studies in which preservice teachers analyse primary students' responses to tasks that aim to determine the sample space to create equal probabilities, i.e., proportionally determine the composition of an urn so that the probability of success is the same as in another urn (Supply et al., 2020). Additionally, in our work, we evaluate teachers' abilities not only to interpret but also to manage students' errors.

Thus, the aim of this study is to characterise and assess the knowledge and competences of preservice primary school teachers in:

1. interpreting student responses to a task that requires determining the composition of an urn (with a known number of possible cases) so that the probability of success is the same as in an urn where the composition is not known, but the ratio of favourable cases to unfavourable cases is known and
2. proposing strategies to help students overcome the difficulties that lead to inadequate solutions.

The interest in the proposed mathematical situation lies in the fact that it connects two essential components of probabilistic reasoning: the identification of the proportional nature of probability calculations and the understanding of the sample space (Supply et al., 2020). Knowing how future teachers identify and interpret these characteristics or their absence from the mathematical practices of students and how they decide to act on them provides new and relevant knowledge for the design of teacher training actions on probabilistic reasoning.

2. KNOWLEDGE AND COMPETENCES OF MATHEMATICS TEACHERS

To study and characterise mathematics teachers' knowledge, we adopted a model that considers teachers' knowledge and their competences for teaching mathematics: the Didactic-Mathematical Knowledge and Competences (hereinafter DMKC) model for mathematics teachers (Godino et al., 2017). In this model, it is assumed that the teacher must know the school mathematics of the educational level at which they teach (common mathematical knowledge) but must also be able to articulate this knowledge with that corresponding to some subsequent levels (extended mathematical knowledge). However, as mathematical content comes into play, didactic-mathematical knowledge is required for different facets that condition the design, implementation, and management of teaching processes for a specific mathematical subject. The model incorporates the six facets of knowledge described in Table 1.

All six facets are interconnected and part of the specialised knowledge of mathematics teachers. For instance, in the context of a given mathematical task, the teacher must mobilise the diverse meanings needed to solve the task (epistemic facet) and solve the task using various procedures and justifications tailored to the educational stage where it is applied (interactional and ecological facets). The knowledge we aim to evaluate in this work is directly linked to cognitive and interactional facets because these facets encompass the knowledge that enables teachers to interpret and attend to the mathematical thinking of their students (Godino et al., 2017).

Table 1. Facets of teacher knowledge in the DMKC model

Facet	Associated knowledge
Epistemic	Content itself, i.e., the particular way in which mathematics teachers understand and know mathematics
Cognitive	Learners' reasoning, conflicts, and errors that emerge when solving specific problems
Affective	Students' affective, emotional, attitudinal, and belief-related aspects of mathematical objects and the learning process they follow
Interactional	Mathematics teaching, task organisation, management of teacher interactions, and interventions to ensure learning progress, including identification and resolution of student difficulties
Mediational	Technological, material, and temporal resources suitable for enhancing or facilitating student learning
Ecological	Relationships between mathematical content and other disciplines, as well as curricular, cultural, and socio-professional factors that influence the processes of mathematical instruction

One of the fundamental aspects of the DMKC model is the interconnection between teachers' knowledge and competence, which is understood as the ability to address the basic didactic problems involved in the teaching and learning of mathematics and, in particular, to provide appropriate responses to real classroom situations. This competence for didactic analysis and intervention (Godino et al., 2017) involves the ability to interpret and evaluate students' solutions to mathematical tasks, to identify essential mathematical elements in mathematical practices that show students' knowledge or difficulties (cognitive analysis competence), and to make informed decisions that, starting from erroneous strategies, foster students' meaningful learning (didactic configuration management competence).

Mathematical content knowledge alone is insufficient for preservice and in-service teachers to interpret students' understanding and respond to students' knowledge or difficulties. Research such as that of Jakobsen and colleagues (2014) and Simpson and Haltiwanger (2017) emphasises the importance of characterising the didactic-mathematical knowledge teachers need to accurately analyse and assess students' responses.

In the specific context of probability, studies such as those by Batanero et al. (2015), Mohamed (2012), and Vásquez and Alsina (2015a, 2015b) assessed the knowledge of preservice primary school teachers in recognising correct answers and describing the difficulties that led students to provide incorrect responses, as well as in deciding how to help students recognise and overcome their errors. The results showed that preservice teachers did not always manage to explain the causes of errors, and the proposed solutions were not always linked to the prior evaluation of the student's response (Mohamed, 2012). Additionally, these studies highlighted the need to strengthen probabilistic reasoning (Batanero et al., 2015) and knowledge about certain events, calculation and comparison of probabilities of elementary events, and understanding event independence (Vásquez & Alsina, 2015a, 2015b) to ensure adequate competence in interpreting and managing errors in students' responses. For prospective secondary school teachers, Dröse and colleagues (2022) indicated that while the teachers could identify procedural errors, they were less adept at recognising the conceptual difficulties underlying them.

More recently, Burgos and colleagues evaluated preservice primary school teachers' knowledge and skills in interpreting students' responses to probability comparison tasks (Burgos et al., 2022) and fair games (Burgos et al., 2023), with specific attention to the proportional reasoning involved and how it was considered in their strategies to help students overcome difficulties. The results demonstrated the difficulties of the participants in justifying why they considered a primary school student's answer to be correct, in identifying possible erroneous strategies behind incorrect answers, in reflecting on proportional reasoning in these answers, and in drawing up meaningful didactic proposals to help students overcome the difficulties that generated the errors. In light of these results, the authors insisted on the need to include or reinforce in training plans the interpretation and analysis of students' answers and decision-making as a means of developing the teacher's didactic-mathematical knowledge and skills in cognitive and interactional facets.

3. METHODS

This research follows a mixed qualitative-quantitative and exploratory methodological approach. By integrating these perspectives, this study provides a comprehensive understanding of preservice teachers' (PTs) didactic-mathematical knowledge and competences related to probability. The qualitative component focuses on identifying and analysing specific patterns in how PTs interpret and evaluate students' solutions to a probabilistic task, whereas the quantitative analysis examines trends and relationships that complement and deepen the qualitative findings.

First, Section 3.1 describes the participants and the course in which the task was implemented. Next, Section 3.2 presents the task designed to assess the knowledge and competences of prospective teachers. Section 3.3 outlines the criteria for evaluating students' solutions, emphasising their alignment with established mathematical and instructional practices. Finally, Section 3.4 details the analytical approach, which combines content analysis with statistical techniques to ensure the validity and depth of the findings.

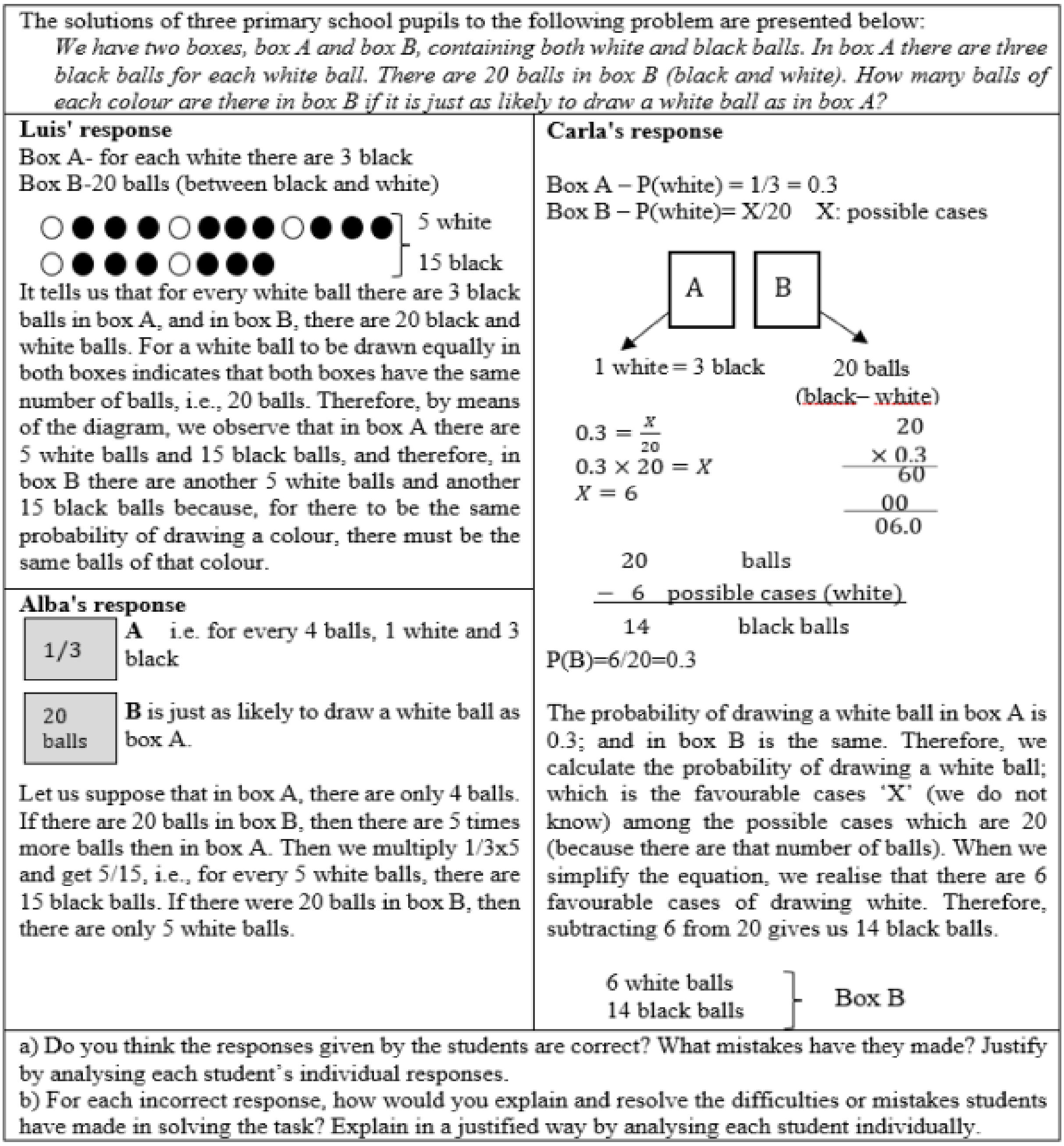
3.1. CONTEXT

In this research, 133 PT students in their third year of primary education at a Spanish university (year 2023) participated. They were enrolled in a course on teaching and learning statistics and probability. All of them had the same instructors responsible for the course (one of whom is the first author of this paper). Students received specific training on the main concepts, properties, and stochastic procedures in primary education and didactic foundations for the content, specified in cognitive aspects (learning probability, analysis of errors and difficulties) and instructional aspects (design, sequencing of tasks, materials and resources for the teaching of probability). In this paper, we analysed the responses to a specific task (Figure 1) included in the midterm assessment of the course, which had a direct impact on students' final grades for the course.

3.2. DATA COLLECTION INSTRUMENT AND A PRIORI ANALYSIS

To assess didactic-mathematical knowledge about probability in the cognitive and interactional facets, the PTs were asked to analyse the solutions of three fictitious students to a probability problem and to make decisions for the resolution of the difficulties encountered (see Figure 1).

First, the PTs were expected to reflect not only on whether the answer given by each student was correct but also on whether the justification given was adequate. Luis correctly determined the composition of box B using an equal distribution strategy. However, his justification was not entirely adequate because he assumed that there must be the same number of balls in both boxes for the probability to be the same, thereby not attending to the proportional nature of the probability calculation. To help him, it might be useful to use two boxes with different numbers of balls in proportion: one box with one black ball and three white balls and another box with four black balls and twelve white balls, for example. The probability of drawing a white (or black) ball can then be calculated so that the student recognises that the probability of drawing a white (or black) ball is the same in both boxes because the fractions that determine these probabilities are equivalent. A teacher could also start from a given composition in one urn and ask the student to determine the composition of another by establishing multiplicative relationships between the number of favourable and possible cases (double in both cases, triple in both cases, etc.).



The solutions of three primary school pupils to the following problem are presented below:

We have two boxes, box A and box B, containing both white and black balls. In box A there are three black balls for each white ball. There are 20 balls in box B (black and white). How many balls of each colour are there in box B if it is just as likely to draw a white ball as in box A?

Luis' response

Box A- for each white there are 3 black
Box B-20 balls (between black and white)

○●●●○●●●○●●● } 5 white
○●●●○●●● } 15 black

It tells us that for every white ball there are 3 black balls in box A, and in box B, there are 20 black and white balls. For a white ball to be drawn equally in both boxes indicates that both boxes have the same number of balls, i.e., 20 balls. Therefore, by means of the diagram, we observe that in box A there are 5 white balls and 15 black balls, and therefore, in box B there are another 5 white balls and another 15 black balls because, for there to be the same probability of drawing a colour, there must be the same balls of that colour.

Carla's response

Box A – P(white) = 1/3 = 0.3
Box B – P(white)= X/20 X: possible cases

A B

1 white = 3 black 20 balls (black– white)

$0.3 = \frac{X}{20}$
$0.3 \times 20 = X$
$X = 6$

20 × 0.3 = 60, 00, 06.0

20 balls
− 6 possible cases (white)
14 black balls

P(B)=6/20=0.3

The probability of drawing a white ball in box A is 0.3; and in box B is the same. Therefore, we calculate the probability of drawing a white ball; which is the favourable cases 'X' (we do not know) among the possible cases which are 20 (because there are that number of balls). When we simplify the equation, we realise that there are 6 favourable cases of drawing white. Therefore, subtracting 6 from 20 gives us 14 black balls.

6 white balls
14 black balls } Box B

Alba's response

1/3 — **A** i.e. for every 4 balls, 1 white and 3 black

20 balls — **B** is just as likely to draw a white ball as box A.

Let us suppose that in box A, there are only 4 balls. If there are 20 balls in box B, then there are 5 times more balls then in box A. Then we multiply 1/3x5 and get 5/15, i.e., for every 5 white balls, there are 15 black balls. If there were 20 balls in box B, then there are only 5 white balls.

a) Do you think the responses given by the students are correct? What mistakes have they made? Justify by analysing each student's individual responses.
b) For each incorrect response, how would you explain and resolve the difficulties or mistakes students have made in solving the task? Explain in a justified way by analysing each student individually.

Figure 1. Formulation of the proposed task

Alba obtained the correct composition for box B, assuming that there were only four balls in box A. This particularisation is possible from the point of view of the equivalence (same probabilities) of the boxes, so she could be aware of the proportional nature of the calculation of probabilities (the same probability is obtained respecting the ratio between favourable and unfavourable cases). However, she does not say so explicitly. In the representation, she expressed the ratio of white to black balls in box A and established a multiplicative relationship between this and the composition of box B ("5 times more balls than in box A"). When Alba expressed "1/3×5 and get 5/15, i.e., for every 5 white balls, there are 15 black balls," she could have made a mistake in multiplying the fraction by a number (multiplying numerator and denominator by that number). However, it is possible that she intended to use "×5" to express the ratio 5 times more; 5/15 is the fraction equivalent to 1/3 when multiplying the numerator and denominator by 5. It would be appropriate to ask Alba to explain why she can assume that there are four balls in A and discuss the connection between probability and the multiplicative relationship using "5 times more." It would also be appropriate for her to explain what she meant by "we multiply 1/3×5" and ask what the probability of getting a white ball would be in both cases and why it would be the same.

Carla's answer shows confusion between the terms "possible" and "favourable" (misuse of language, not ignorance of the concept). Before posing the proportional equation, she called the "favourable" cases "possible" when she referred to *x* as the unknown number of white balls in B, but then she identified "$x = 6$" as the number of favourable cases among the 20 possible cases in B. Although the proportional equation is derived from the equality of probabilities (which implies understanding the proportional nature in the calculation of probabilities), in the case of the probability of A, the left-hand term, she uses the ratio 1 to 3 (favourable to unfavourable) instead of 1 to 4 (favourable to possible). In the right-hand term, she applies Laplace's rule correctly, so her knowledge might be partial (she gets it right at one point and wrong at another). To help Carla, it would be necessary to identify what she understood by "favourable cases" and "possible cases," as she initially used them incorrectly. We can propose different events and ask Carla to identify the favourable cases while pointing out the unfavourable cases in relation to the set of elements of the sample space, i.e., over the total number of possible cases. It would be useful to help Carla distinguish between the possible equations to solve this problem, $\frac{1}{4} = \frac{x}{20}$ from favourable cases to possible and $\frac{1}{3} = \frac{x}{20-x}$ from favourable to unfavourable cases, and how they relate to each other through the connection between proportionality and probability.

In the a priori analysis, although there was no disagreement among the researchers, collective discussion was necessary to interpret how Alba used the multiplication of fractions or how Carla applied Laplace's rule.

3.3. CATEGORIES OF RELEVANCE

In this section, we describe the categories of relevance established for PTs' responses to analysing student responses and offering suggestions to resolve difficulties, interpreted as the degree of correspondence with the institutional didactic-mathematical practices of reference (Subsection 3.2). As in previous studies that assessed the knowledge and competences of prospective teachers to interpret students' responses and propose actions to help them overcome difficulties (Burgos et al., 2022, 2023; Vásquez & Alsina, 2015b, 2017), we defined three levels of relevance for both tasks: not relevant, partially relevant, and relevant. The goal was to identify responses that, while not inappropriate, were not entirely adequate because the professional discourse (justification of the evaluation, proposed strategies) was imprecise or incomplete. This categorisation allowed us to systematically evaluate PTs' responses according to their relevance and depth, providing a framework for both qualitative and quantitative analyses.

The PTs' responses were coded in relation to their analysis of each of the three students' work (Luis, Alba, and Carla). Based on the a priori analysis, the criteria used to establish these categories focused on the references to mathematical elements and how these elements were used to assess the correctness of the students' answers. (See Table 2.) PTs identified errors and considered their nature when proposing actions with a didactic-mathematical foundation. (See Table 3.)

3.4. QUALITATIVE AND QUANTITATIVE DATA ANALYSIS

This subsection integrates qualitative and quantitative approaches to examine how PTs interpret and evaluate student responses and their strategies to address identified errors. These approaches offer complementary perspectives. The qualitative analysis focuses on the depth and specific details of PTs' responses, examining how PTs interpret students' understanding and develop strategies to overcome difficulties. Meanwhile, the quantitative analysis highlights the overarching patterns and trends in PTs' evaluations. The following subsections provide a comprehensive explanation of each method and its implementation in the study.

Table 2. Relevance categories in student response assessment

Relevance category	Description of the category	Examples
Not relevant	The correctness of the solution is not adequately recognised (in particular, any errors are not identified when there are errors) and the justification is missing, inadequate, or inconclusive (does not allow the degree of correctness of the solution to be assessed).	*PT100*. [Luis] has correctly justified his procedure since in both boxes there must be the same number of white balls and black balls. *PT107*. Alba: The answer is not correct, she incorrectly identifies the probability, making an erroneous equivalence. *PT104*. Carla's correct answer and justification: she calculates the probability of white balls in A, as this must be equal in B she performs Laplace's rule to obtain the number of white balls in which this probability belongs. Then she subtracts these balls from the total and obtains the number of black balls. She does not fail in equiprobability.
Partially relevant	The student's response is adequately assessed as correct, but the justification contains inaccuracies or identifies only partially or ambiguously the errors made.	*PT70*. Luis arrives at a correct final answer but his justification is not. He comments that both boxes have the same number of balls so he deduces the final answer. *PT92*. Alba performs the exercise well using equivalent fractions. *PT18*. Carla: The explanation is not correct because the approach is not correct, therefore the justification given is not correct since it says that to have the same probability in both cases, box B must have 6 white balls and 14 black balls.
Relevant	The student's solution is adequately assessed and justified by specifying and interpreting errors where they exist (according to a priori analysis)	*PT1*. Luis' answer is correct because he has correctly obtained the proportion of balls in box B, but his justification is wrong, as he states that for both boxes to have the same probability of drawing a white ball, there must be the same number of balls, when this is not necessary as long as there is the same proportion. *PT14*. Alba's answer is correct because she uses the ratio 1/3 (for every white ball there are 3 black balls). She correctly calculates that there are 5 times as many balls in box B and multiplies these 5 times as many in the ratio (1×5)/(3×5)=5/15 having 5 white balls and 15 black balls. *PT121*. [Carla] does not correctly relate the number of favourable cases to the number of possible cases. Thus, she incorrectly poses the equation, and when solving it, she does not obtain the correct number of balls that should be in box B, 15 black and 5 white balls.

Table 3. Relevance categories for error management

Relevance category	Description of the category	Examples
Not relevant	Strategies to help students overcome the difficulties encountered are not proposed; strategies are presented generically without considering the particularities of the students' responses; or only actions that do not involve the content or are not adequate to correct the errors made by the student are proposed.	PT119: In this case, I would tell the student [Alba] to review the exercise and read the statement again because her justification is correct, even though her answer is incorrect. This may have occurred because she had read the statement too quickly.
Partially relevant	Strategies are proposed, although PTs do not clearly identify how they consider students' lack of mathematical knowledge or are presented in a confusing way.	PT13: With Carla, I would go over the meaning of "n" [possible cases] in Laplace's Rule because I think it was just an oversight.
Relevant	Appropriate strategies are proposed to identify how a lack of knowledge is considered in relation to the probabilistic or proportional reasoning employed.	PT2: Luis can be made to understand that the proportion of white and black balls does not imply an equal number of balls through a practical simulation, which leads to visually detect that there can be 28 balls, and that for each white ball, there are 3 black balls as well.

Qualitative analysis Through content analysis (Cohen et al., 2000), emerging categories that reflect the didactic-mathematical knowledge and skills of primary school PTs were identified to interpret and assess the responses of primary school students and propose didactic strategies to solve errors. The qualitative analysis of the data (in our case, written operational and discursive practices) involved applying predefined or emergent codes to facilitate the interpretation of the participants' (subjective, personal) meaning. The validation of the analytical process involved having multiple researchers conduct the analysis and compare their findings (Montes, 2021). Moreover, to ensure reliability within the scope of subjectivity intrinsic to this type of analysis, the content analysis relied on the a priori analysis conducted to search for references to mathematical and didactic-mathematical elements in the written work of the PTs.

First, the three researchers independently solved the task proposed to the participants. Subsequently, with the participation of an external expert collaborator, the researchers shared their analyses to reach a consensus on the errors observed in the pupils' solutions, compare and unify possible strategies to help them overcome these difficulties, and agree on relevance categories.

Next, each researcher individually conducted a content analysis of part of the PTs' responses, identifying types of responses both in the interpretation of primary students' solutions (item a) and in the error management proposals (item b). At this stage, the researchers met to share their initial response categories and identify common or complementary categories. Each researcher then assigned a degree of relevance to the PTs' responses according to the reference didactic-mathematical practices (Section 3.2) and the defined relevance categories (Section 3.3). This process (content analysis of the PTs' responses, categorisation, and assignment of relevance degree) was repeated until half of the participants' reports were analysed. At that point, the researchers met again to share any new types of responses found, discuss difficulties in applying the defined categories (refining the definition of those that caused confusion), and review the relevance assessments that raised doubts in this second round. Then, the researchers proceeded with the analysis of the remaining PTs' productions and thereafter, met once more to compare and validate the results of the last round of content analysis.

During this phase, differences in the initial evaluations were observed. These discrepancies were resolved through collaborative discussions, where researchers presented detailed justifications for their assessments, debated their perspectives, and revisited the predefined relevance categories and criteria. This iterative process included multiple rounds of sharing and validation. In cases of persistent disagreement, researchers revisited the reference didactic-mathematical practices and consulted an external expert collaborator. These discussions ensured that all evaluations adhered consistently to the predefined framework and criteria, aligning the researchers' assessments with the didactic-mathematical practices established in the study. By the end of this stage, all researchers had reached a consensus, enhancing the reliability and consistency of the categorisations. As a result of this cyclical and inductive process, the categories described in Section 4 were obtained.

Quantitative analysis The quantitative approach allowed for the identification of general trends and patterns in the data (Leavy, 2022), systematically assessing the degree of correctness of responses and the correlation between the variables of response interpretation and error management. Responses classified as not relevant were assigned a score of 0; those classified as partially relevant were assigned a score of 1, and those classified as relevant were assigned a score of 2.

This scoring framework provided a structured basis for categorising responses into three levels: not relevant, partially relevant, and relevant, as described in Table 2 and Table 3. These ordinal categories led to using nonparametric statistical methods for analyses, including Wilcoxon signed-rank tests and Kendall's Tau correlation. The Wilcoxon signed-rank test assesses the median difference between paired observations without assuming normality, making it ideal for comparing relevance scores in response interpretation and error management. To explore relationships between key variables, Kendall's Tau correlation was employed due to its suitability under the same ordinal assumptions. Furthermore, following recent methodological recommendations (Wasserstein et al., 2019), we integrated effect size interpretations and confidence intervals into the presentation of results. This approach provided deeper insights into observed trends and correlations between key variables, response interpretation, and error management. From these, the statistical analysis developed in Section 4 allowed us to determine an overall quantitative assessment of the didactic-mathematical knowledge of future teachers and explore the relationships between the PTs' competence to interpret and act on students' answers.

4. RESULTS

This section presents the results obtained from the content analysis of PTs' productions. We describe the categories of assessment of students' responses (Section 4.1) and error management (Section 4.2) found and analyse their appropriateness. We then examine the possible correlation between degrees of relevance (Section 4.3).

As indicated, to identify the response categories for the first item, the researchers analysed the references to mathematical elements used by the PTs to evaluate students' solutions and detect errors. To identify the response categories for the second item, the researchers examined the PTs' descriptions to identify the proposed didactic actions for managing the identified errors. Two difficulties arose during analyses. First, difficulties arose that required joint analysis when the PT's response did not clearly present the mathematical elements or present them in an orderly manner. Second, it was necessary to agree on how to proceed when the proposed actions appeared to fit into two different categories. To simplify the presentation of results, the researchers included the response in the didactic strategy category that was most related to or coherent with the identified error.

4.1. ANALYSIS OF STUDENTS' SOLUTIONS

The PTs had to analyse the students' answers, justifiably assessing their correctness and identifying their errors. The analysis of the PTs' reports allowed us to identify categories in the evaluation of each student's answer (Luis, L; Alba, A; Carla, C).

The information in Table 4 shows differences in the PTs' perceptions of the adequacy of the answers and justifications provided by Luis, Alba, and Carla. These disparities are most notable in the assessment of Luis's and Alba's responses. However, in both cases, PTs considered their responses (composition of the boxes) to be correct, and the difference arose when assessing their justifications.

Table 4. Distribution of frequencies and percentages of answers and justifications of PTs according to their correctness (n=133)

	Correct answer and justification	Incorrect answer and justification	Correct answer and incorrect justification	No reply
Luis	47 (35.34%)	24 (18.05%)	**59 (44.36%)**	3 (2.25%)
Alba	**61 (45.86%)**	33 (24.81%)	35 (26.32%)	4 (3.01%)
Carla	19 (14.29%)	**104 (78.19%)**	0 (0.00%)	10 (7.52%)

Note: Frequencies and percentages of correct answers and justifications are displayed in bold font.

According to Table 5, most of the PTs considered that although Luis obtained the correct composition, his justification was incorrect, arguing that the equality in the probability of events did not necessarily translate into an identical number of balls (category AL1). Among these responses, approximately two-thirds (49 out of 79) were classified as relevant because PTs adequately justified the solution, pointing out that equal probability did not depend on the number of balls and correctly recognising the proportional relationship. However, the remaining responses in this category were classified as not relevant because their justifications for Luis's solution were inadequate or inconclusive. There were no partially relevant responses in this category. All responses in category AL2, all but one in AL3 (which related Luis's iconic strategy to the representation of proportion and was rated as partially relevant), and all in AL4 were deemed not relevant because PTs incorrectly assumed that equiprobability required an equal number of balls (AL2), focused on the validity of the iconic representation but did not reflect on the mathematical elements that should support it (AL3), or considered Luis's solution incorrect solely based on his use of the iconic representation or the insufficiency of his calculations (AL4). The partially relevant responses, mainly in category AL1, demonstrated a certain level of understanding of the strategy used by Luis and recognised, albeit incompletely or ambiguously, the limitation in his justification (assuming that the same probability required the same number of balls).

Table 5. Synthesis of categories, examples, and relevance classification in PTs' evaluations of Luis's solution (n = 130 responses; 3 non-responses)

Code	Category	Description	Example(s)	Relevance*	Count (percent)
AL1	Equal probability does not imply an equal number of balls.	Luis argues that an equal probability of success in both boxes necessarily implies an equal number of balls of each colour, which is considered incorrect. The PTs point out that although Luis obtains the correct result, the justification he provides is not adequate because it is not necessary that there is the same number of balls in each box but that the proportion is maintained to guarantee equal probability.	*PT1*. Luis' answer is correct because he has correctly obtained the proportion of balls in box B, but his justification is wrong, as he states that for both boxes to have the same probability of drawing a white ball, there must be the same number of balls, when this is not necessary as long as there is the same proportion.	2 (49; 36.84)	79 (59.40)
			PT19. Luis's reasoning is correct, but the statement does not specify that there must be the same number of balls in each box. His final statement is key, as he mentions that the same probability can be achieved without having the same number of balls per box.	1 (26; 19.55)	
			PT67. In the case of Luis and Alba, I believe their solutions are correct because they based their reasoning on the equivalence of fractions, 1/3 = 5/15.	0 (4; 3.01)	
AL2	Equal probability means an equal number of balls.	Luis stated that if the probability of success is the same, the same number of balls of each colour in both boxes is considered adequate. In this category, the PTs consider Luis's answers (globally) correct.	*PT3*. Luis' answer is correct because he has taken into account the equiprobability factor, taking into account the statement that both boxes have the same probability and therefore he has concluded that both have the same number of balls of the same colour (15 black and 5 white).	0 (16; 12.03)	16 (12.03)
AL3	The iconic strategy employed by Luis is considered appropriate.	The iconic strategy followed by Luis is considered appropriate without any justification.	*PT36*. The answer given by Luis is correct. He uses the drawings to indicate that for every white ball, there are three black balls, giving him 5 white and 15 black balls.	1 (1; 0.01)	16 (12.03)
			PT95. Luis' answer is correct. With drawings he adds 1 white ball plus 3 black balls repeatedly until he had 20 balls in total. Once represented by pictures, he must count only the total of black and white balls in box B.	0 (15; 11.28)	
AL4	The iconic strategy employed by Luis is considered inappropriate.	The strategy followed by Luis is assessed as incorrect because he relied on a graphical representation or because of the absence of mathematical calculations.	*PT58*. Luis' answer is not correct, as he simply draws a representative picture and for every white ball he draws 3 black balls.	0 (6; 4.51)	6 (4.51)
IC	Inconclusive	The PT's response does not align with any pre-established category and does not provide a clear interpretation of their assessment of the student's response.	*PT53*. His [Luis's] answer is correct, the justification makes sense, and he has provided reasons because he has based his answer on box A, that for every white ball, there are 3 black balls.	0 (13; 9.77)	13 (9.77)

*The relationships between different relevance category values are presented, along with the count and percentage of PTs' responses classified in each category.

Table 6 shows greater diversity in how the PTs evaluated Alba's solution. In category AA1, the PTs adequately identified the relationship between favourable and unfavourable cases. However, they often failed to recognise arithmetic errors or did so ambiguously, leading most evaluations to be classified as partially relevant. In categories AA4 (misuse of fractions), AA5 (absence of probabilistic reasoning), and AA6 (misuse of ratio 1:3 for 1:4), which had similar frequencies of responses, a high proportion of not relevant responses was observed. These were due to an incorrect interpretation of fraction use in category AA4, an improper consideration of probability in AA5, or a misunderstanding of the ratio in AA6. In all cases, these evaluations were superficial and failed to address the meaning of the mathematical elements involved in the solution. The few relevant responses reflected on the use of proportionality (AA1) and operations with fractions (AA4), identifying the error or interpreting the potential meaning as ratios.

In the case of the solution provided by Carla, more than three-quarters of the PTs (see Table 4) recognised that both the student's response and justification were incorrect. As shown in Table 7, the frequencies in the categories of PTs' responses in which errors were identified in their solutions (i.e., AC2, AC3, AC4, and AC5) were similar. When the assessment was as globally correct, either the description was inconclusive or it was based on the use of Laplace's rule when posing the proportional equation (AC1). In any case, their evaluations did not allow assessment of the degree of correctness of the solution or did not recognise any errors; thus, they were considered not relevant. In categories AC2, AC3, and AC4, PTs attempted to identify specific errors in the appropriate strategy, such as inappropriate composition (AC2), confusion between favourable and possible cases (AC3), and misuse of ratio 1:3 for 1:4 (AC4). Thus, there was a greater presence of at least partially relevant responses, close to the institutional didactic-mathematical practices of reference, in these categories. It is noteworthy that the highest percentage of non-response (7.52%) was associated with Carla, which could be due to the greater complexity of the procedure used by students to solve the task. In this regard, the majority of responses categorised as not relevant (15.04%) were inconclusive (IC) assessments of Carla's response.

When comparing the percentages of responses with a correct evaluation (Table 4) to the percentages of inconclusive responses (Tables 5, 6, and 7), we observed that a higher percentage of correct evaluations did not always correspond to a lower percentage of inconclusive responses. In Luis's case, the lowest percentage of inconclusive responses (9.77%) was observed, but it also had the lowest percentage of correct evaluations (35.34%). Carla's case had the highest percentage of correct evaluations (78.19%), although 15.04% were inconclusive. Similarly, Alba had the highest percentage of inconclusive responses (24.81%) despite 45.86% of the PTs correctly evaluating both the response and the justification. This indicates that although the PTs more frequently recognised the correctness of Alba's response or identified the conceptual and procedural errors in Carla's solution, their interpretations did not clearly articulate which mathematical elements presented in the students' practices formed the basis of their evaluations.

Table 6. Synthesis of categories, examples, and relevance classification in PTs' evaluations of Alba's solution (n = 129 responses; 4 non-responses)

Code	Category	Description	Example(s)	Relevance[*]	Count (percent)
AA1	Adequate proportional strategy related to probability	Alba's strategy is considered adequate based on the use of ratios or the equivalence of fractions, and its connection to probability is explicitly mentioned.	*PT14*. Alba's answer is correct because she uses the ratio 1/3 (for every white ball there are 3 black balls). She correctly calculates that there are 5 times as many balls in box B and multiplies these 5 times as many in the ratio (1×5)/(3×5)=5/15 having 5 white balls and 15 black balls. Alba knows that there does not have to be the same number of balls of each type in the box; she knows that having the same ratio gives the same probability regardless of the number of balls in the box.	2 (2; 1.50)	18 (13.53)
			PT97. [Alba]Correctly establishes the proportionality relationship of box A and adequately translates it to box B. This idea is supported by equiprobability.	1 (11; 8.27)	
			PT108. [Alba]solves the exercise correctly because, based on her justification, the probability of drawing a white ball is the same in both boxes. Use equivalent fractions.	0 (5; 3.76)	
AA2	Adequate strategy not related to probability	Alba's strategy is based on the use of ratios or the equivalence of fractions, but its connection with probability is not explicitly mentioned.	*PT92*. Alba performs the exercise well using equivalent fractions.	1 (2; 1.50)	5 (3.76)
			PT46. Alba represents the situation correctly using fractions, and performs the operations correctly, so I can't find any errors.	0 (3; 2.26)	
AA3	Use of probabilistic concepts and procedures	The PT does not analyse the use of proportionality in Alba's response but instead focuses on the use of procedures and terms specific to probability.	*PT88*. Alba not only provides a correct answer, but also an adequate justification, demonstrating her understanding of the concept of probability by clearly differentiating between favourable and possible cases.	0 (5; 3.76)	5 (3.76)
AA4	Misuse of fractions	The PT indicates that Alba made a mistake while operating with fractions or that what she represented with fractions was not what she intended.	*PT93*. [Alba's] answer is correct. Note that it does not multiply 1/3×5 but rather (1×5)/(3×5) that is, for the same ratio in the numerator and denominator, although she solves the operation correctly.	2 (2; 1.50)	17 (12.78)
			PT9. Alba gives a correct solution, but she makes a mistake; just like Luis, she considers P(A)=1/3. What happens is that when she multiplies 1/3x5, she does it wrong and puts 5/15 when, in fact, it is 5/3, but it happens that she gets the solution right. However, the development is wrong.	1 (6; 4.51)	
			PT115. The [Alba's] answer is correct, but the justification is not, as the fraction for [the composition of] box A was not obtained correctly, and the calculations are not well structured.	0 (9; 6.77)	

Code	Category	Description	Example(s)	Relevance*	Count (percent)
AA5	Absence of probabilistic reasoning	The error made by Alba is considered to come from not taking into account equiprobability, not calculating or not comparing probabilities, or wrongly considering the number of favourable cases by indicating that the probability of drawing a white ball is 1/3 instead of 1/4.	*PT11*. Alba approaches the problem using equivalent fractions and determines the composition of box B based on an assumption for box A. She understands both concepts but not their relationship to probability.	1 (1; 0.75)	19 (14.29)
			PT26. Alba did not consider that if we consider a total of 4 balls, the probability of obtaining a white ball is 1/4 and not 1/3, as she indicated. Therefore, the solution to the remaining exercise is incorrect.	0 (18; 13.53)	
AA6	Misuse of ratio 1:3 for 1:4 (not as probabilities)	Alba's strategy is considered incorrect because of the misuse of ratio 1:3 for 1:4, but there is no reference to probability.	*PT73*. Her [Alba's] answer and reasoning are incorrect because the ratio in box A with 4 balls would not be 1/3 but 1/4 and 3/4 for white and black, respectively.	0 (20; 15.04)	20 (15.04)
AA7	Incorrectly assuming there are 4 balls in A	The PT states that Alba made the mistake of considering that there are 4 balls in box A without supporting this assumption.	*PT129*. Alba assumes, not takes it for granted, that there are only 4 balls in box A, but does not talk about favourable cases among possible cases to establish this equivalence.	1 (1; 0.75)	12 (9.02)
			PT91. [Alba's] fault comes when she admits that there are only 4 balls in box A, for which there is no information.	0 (11; 8.27)	
IC	Inconclusive		*PT6*. Both the answer and the justification are correct. It is evident that you have correctly acquired the knowledge of probability.	0 (33; 24.81)	33 (24.81)

*The relationships between different relevance category values are presented, along with the count and percentage of PTs' responses classified in each category.

Table 7. Synthesis of categories, examples, and relevance classification in PTs' evaluations of Carla's solution (n = 123 responses; 10 non-responses)

Code	Category	Description	Example(s)	Relevance*	Count (percent)
AC1	Posing and solving proportional equations	The PT recognises that Carla knew and used Laplace's rule adequately to arrive at the equation, but she made mistakes during the equation-solving process.	*PT13*. Although at the beginning [Carla] demonstrates her knowledge of probability when she uses Laplace's rule, we see that she makes a mistake. She states the equation well, but the probability is 0.25 and not 0.3.	1 (2; 1.50)	11 (8.27)
			PT11. Carla adopted a more mathematical than visual approach, calculating probability through an equation, from the classical perspective of Laplace's rule.	0 (9; 6.77)	
AC2	Composition of box B does not respect the ratio or the probability of box A	Carla's answer was considered incorrect because the composition obtained from box B did not respect the ratio of white to black balls in A or did not comply with the probability of drawing a white ball being the same as in A. There was no reflection on the origin of this disparity.	*PT50*. According to Carla, box B would have 6 white balls and 14 black balls, so we see that there was already a higher probability of drawing them than in box A.	1 (20; 15.04)	22 (16.54)
			PT63. The answer is incorrect because she states that box B has 6 white balls and 14 black balls when there are 5 white balls and 15 black balls.	0 (2; 1.50)	
AC3	Confusion between favourable and possible cases	Carla was considered to have difficulties in determining or differentiating between possible and favourable cases in the sample spaces of both boxes.	*PT121*. [Carla] does not correctly relate the number of favourable cases to the number of possible cases. Thus, she incorrectly poses the equation, and when solving it, she does not obtain the correct number of balls that should be in box B, 15 black and 5 white balls.	2 (3; 2.26)	24 (18.05)
			PT39. [Carla] gives a wrong answer by not properly identifying the relationship between favourable and possible cases.	1 (16; 12.03)	
			PT88. Carla provides an incorrect response as she is unable to identify the possible and favourable outcomes, leading to flawed reasoning and justification.	0 (5; 3.76)	
AC4	Misuse of ratio 1:3 for 1:4	The error in Carla's response was attributed to considering the ratio of favourable to unfavourable cases (1 to 3) rather than favourable to possible cases (1 to 4).	*PT107*. Her answer is incorrect. Carla's general idea is correct; however, she makes a mistake in establishing the equivalences, considering the probability of drawing white as 1/3 when it should be 1/4. This error leads to incorrect results throughout.	2 (3; 2.26)	26 (19.55)
			PT14. Carla's answer is incorrect because when she started to make fractions to compare, she did not take into account that it is not 1/3, it is 1/4 because in the denominator, she has to put the total of balls in box A if she then compares it with a fraction that has the total of box B in the denominator. She may have got confused when trying to use the ratio of 1/3 (for one white ball there are 3 black balls).	1 (20; 15.04)	
			PT73. Her answer and reasoning are incorrect since the ratio in box A with 4 balls would not be 1/3 but ¼.	0 (3; 2.26)	

Code	Category	Description	Example(s)	Relevance*	Count (percent)
AC5	Error in calculating the probability of drawing a white ball in A	Carla's approach was considered adequate and that if she had properly determined the probability of drawing a white ball in A, she would have arrived at the correct solution. The focus was on miscalculating a specific probability.	*PT15*. [Carla] She performs the exercise correctly; however, when she gives the probability, she indicates that it is 0.3 when it should be 0.25 because [for] each white ball, there are 3 black balls in the box. If we add all these balls together, then there would be 4, which makes [that the probability is] 1/4 of drawing a white ball. Ignoring the numbers, the approach is correct; by doing the proportion correctly, the exercise is valid.	1 (13; 9.77)	20 (15.04)
			PT66. Her mistake was that she did not know how to calculate the probability correctly, since there is 1 white for every 3 blacks, the probability is 25% and not 30% as she obtained.	0 (7; 5.26)	
IC	Inconclusive		*PT131*. Carla arrives at the solution by performing operations to calculate the probability and deduces the number of white and black balls in box B. In her approach, if we take out one white and one black ball until the white balls run out, box A would have 8 black balls left and box B would have 10 black balls.	0 (20; 15.04)	20 (15.04)

*The relationships between different relevance category values are presented, along with the count and percentage of PTs' responses classified in each category.

4.2. ERROR MANAGEMENT PROPOSALS

After analysing the incorrect answers and justifications, the PTs were asked to describe how they would explain and resolve the errors identified in the learners' answers. Content analysis of their reports showed that the PTs use similar approaches with different learners, which leads to classifying their error management approaches into general categories.

Table 8. Synthesis of categories, examples, and relevance classification in PTs' error management proposal

Code	Categories	Description	Example(s)
EMP1	Reviewing probabilistic notions	The PTs proposed recalling basic probabilistic concepts, either in a generic way or directed at one of the students (Luis, Alba or Carla), particularly focusing on the definition of probability and equiprobability.	*PT38:* I would explain to them the term equiprobability of events, that is, that there exists the same probability of the events happening.
EMP2	Determining the sample space	The PTs aimed to help students identify the sample space so that they could perceive the influence on probability determination. The PTs emphasised the determination of the favourable, unfavourable, and possible cases.	*PT39*: In Luis' case, although he determines the number of balls of each colour in box B, he does not take into account all the existing possibilities in box A. To solve this problem, I would ask him to determine the sample space, which would allow him to realise that in box A there may be several options.
EMP3	Explaining Laplace's rule	It is considered necessary to explain the meaning and use of Laplace's rule in the case of Luis and Alba (due to lack of knowledge) and in the case of Carla (due to confusion or absent-mindedness).	*PT10:* To Luis and Alba, although they obtain the result they are looking for, they have not taken into account Laplace's rule, so I would explain its meaning, and I am sure they would have followed the steps to obtain the result properly. *PT13*: With Carla, I would go over the meaning of "n" [possible cases] in Laplace's Rule because I think it was just an oversight.
EMP4	Using iconic or graphic representations and manipulative materials	PTs proposed using iconic visual representations and/or manipulatives (such as boxes, balls, etc.) to help students understand the experiment, the structure of the task, and the probabilistic concepts involved.	*PT114:* Luis does not understand equiprobability. This concept should be worked on with figures with which this property can be clearly seen, accompanied by graphical [representations] and manipulable material
EMP5	Reviewing operations and equivalences with fractions	PTs considered that the errors come from the difficulties associated with operations or the equivalence of fractions; therefore, they recommended recalling or reinforcing these notions.	*PT79*: As both [Carla and Alba] do not understand the relationship between fractions and probability, previous activities will be conducted in which fractional numbers will be addressed. We begin with fractions [...], then with operations with these [...] and, finally, exercises in which they can understand the relationship between fractional numbers and probability.

Code	Categories	Description	Example(s)
EMP6	Encouraging reflection on process and justification	The PTs chose to ask students to reflect on the process followed or the justification provided as a way to discover errors made by themselves.	*PT81*: I would first ask her [Carla] if she understood what the exercise was asking her to do, and I would also ask her to explain how she did it to see if she could realise the error herself.
EMP7	Re-reading the statement	It is considered that the error made by the student is due to a lack of understanding or interpretation of the statement; therefore, it is recommended that a new reflective reading be conducted.	*PT119*: In this case, I would tell the student [Alba] to review the exercise and read the statement again because her justification is correct, even though her answer is incorrect. This may have occurred because she had read the statement too quickly.
EMP8	Using examples	It is proposed to use examples and activities related to the content of the task beforehand to consolidate the mathematical knowledge required to tackle the task successfully.	*PT88*: For Alba to understand her mistake, as a teacher, I would review the idea of probability and then perform small, simple activities that would progressively increase complexity. Once she had internalised the concept of probability, I would ask her to repeat the exercise again.
EMP9	Clarifying the relationship between the number of balls in both boxes and probability	This category is specific to Luis's answer. PTs relied on the distribution or equivalence of fractions to help Luis understand that it was not necessary to assume that both boxes must have the same number of balls to obtain the same probability of drawing a white ball.	*PT26:* We must point out to Luis that for there to be the same probability of obtaining a white ball in both boxes, it is not necessary to be composed of the same balls. For example, we can have 4 balls (1 white and 3 black) in box A and 20 balls (5 white and 15 black) in box B, and in both [boxes], we have the same probability of obtaining a white ball.
NC	Inconclusive		*PT89*: I would only tell her to be careful when entering and indicating the data, and to check it properly.
NR	No reply		---

Figure 2 provides an overview of the distribution of error management proposals according to the categories described in Table 8. Of the 133 participants, only 10 did not propose any way of handling errors. A higher number of indications to solve the error made by Carla can be observed, 115 in total, whereas, in the cases of Luis and Alba, these numbers were 89 and 81, respectively. If we compare these numbers with what is displayed in Table 4, we observe that the quantity of responses in which PTs identified an error in this respect does not coincide with these values: 83 in the case of Luis, 68 in the case of Alba and 104 in the case of Carla. Thus, even when the PTs had assessed the response as correct, they proposed actions to manage difficulties.

The category of "inconclusive (NC)" was the most frequent, representing 21% of the total number of coded responses (60 out of 285). This indicates that the PTs, when explaining and solving the errors encountered, expressed general ideas that did not fit in with the mathematical practices of the students. This difficulty was mostly found in deciding how to deal with Carla's error, for which half of the proposals were inconclusive. In many cases, the participants insisted again on the mistake she had made, and in others, their proposal did not take the mistake into account. For example, PTs suggested Carla use another strategy (without specifying which one) on the understanding that "clearing the unknown" could lead to more mistakes. Some recommended that Carla pay more attention: "I would only tell her to be careful when entering and indicating the data, and to check it properly" (PT89).

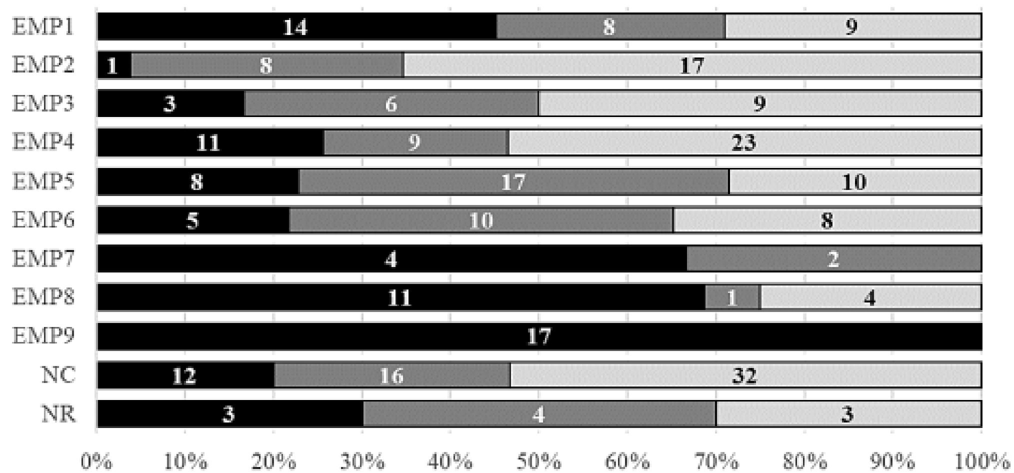


Figure 2. Distribution of PTs' proposals for error management according to each student (Luis: black; Alba: dark grey; Carla: light grey)

The next most frequent category, with 15% of responses, concerned the use of visual representations or manipulative materials (EMP4). Some PTs recommended that Alba or Carla use the representation Luis used. For example,

> I would encourage Carla to see that the exercise is not as complicated as she has presented in her procedure. From my point of view, it is easier to explain to her the approach developed by Luis; it is visual and without difficulty [...] Through drawings, the relationship between boxes is directly observed. I would make her [Carla] know that it is not necessary to always present favourable and total cases, but that doing it in box "A" would be enough. (PT33)

In these cases, PTs proposed strategies that they considered would avoid the error, not necessarily help the student to understand it, and overcome the lack of knowledge that caused it. A frequent response to help Luis was seen when PTs proposed the use of boxes and balls whose composition was the same as proposed in the task, without proposing other compositions that included equivalent urns and might lead students to reflect on them.

Twelve per cent of PTs' proposals were based on the need to review operations with fractions and the equivalence relation (EMP5), understanding in the case of Luis that this would allow him to work with probability. For example, PT12 indicated, "We would review equivalent fractions so that [Luis] would bear in mind that they are useful when looking for the same probability." In the case of Alba and Carla, PTs aimed to avoid operational errors with fractions.

Finally, 11% of the proposals were associated with category EMP1, review probabilistic notions. Responses in this category sometimes referred to all three students. When responses were not generic, such as in the case of Luis, answers focused almost exclusively on the idea of equiprobability. In the case of Carla and Alba, the notions of favourable cases, possible cases, and sample space were considered.

Analysing the specific actions for each student, the most frequent category in the case of Luis (17 out of 89 proposals), aimed to help him understand that it was not necessary for there to be the same number of balls in both boxes to guarantee equal probability (EMP9). In the case of Alba, the largest number of proposals (17 out of 81) were associated with the review of operations, such as the equivalence of fractions (EMP5), as well as with a reflection on the process followed (10 out of 81). In the case of Carla, although the most frequent category related to inconclusive answers (32 out of 115), it is worth noting that 20% of the PTs considered it appropriate to resort to the use of graphical representations or manipulative materials to manage the error made by the student.

4.3. OVERALL ASSESSMENT OF THE DEGREE OF RELEVANCE

In this section, we present the results of the statistical analyses conducted to evaluate the differences and relationships in the relevance scores assigned by participants to items (a) and (b). The analyses were guided by the criteria outlined in Section 3.3 to classify responses based on their degree of relevance: not relevant (0 points), partially relevant (1 point), and relevant (2 points). These criteria were applied to analyse the relevance of the responses provided by the PTs to both tasks. As previously indicated, not all participants responded to both questions. Figure 3 shows a predominance of not relevant responses, both in interpreting the students' practices and in the error management proposal. The exception was the interpretation of Carla's response, in which more than half of PTs' responses were partially relevant.

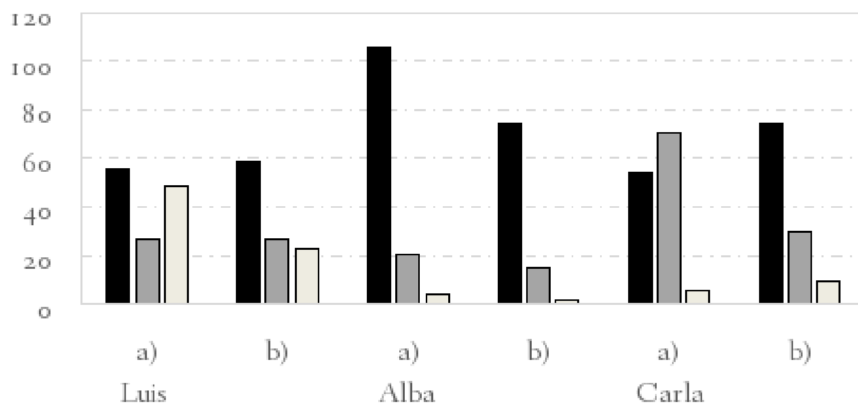


Figure 3. Distribution of PTs' proposals according to the correctness degree of responses to items (a) and (b) by the PTs (Not relevant: black; Partially relevant: dark grey; Relevant: light grey)

Table 9 presents an overview of the data, including medians and interquartile ranges. Responses to item (a) for Luis and Carla display an interquartile range of 2, indicating greater variability compared to item (b), where the interquartile range decreases to 1. In contrast, responses for Alba exhibit an interquartile range of 0 in both items, reflecting a strong concentration in the lowest category. Overall, the median scores for Luis and Carla indicate greater variability and higher performance in item (a) compared to item (b), whereas Alba's responses remain consistently low across both items.

Table 9. Descriptive statistics of PTs' responses to items (a) and (b)

Student	Item	N(*)	Median	Interquartile Range
Luis	a)	108	1	2
	b)	108	.5	1
Alba	a)	90	0	0
	b)	90	0	0
Carla	a)	114	1	2
	b)	114	0	1

Note: (*) Null values (missing responses) were removed to ensure comparability between items (a) and (b).

The results of the nonparametric Wilcoxon signed-rank test show significant differences in the PTs' scores between items (a) and (b) for the cases of Luis and Carla. For Luis, the PTs achieved higher scores on item (a) than on item (b), with a Wilcoxon statistic of $W = 164$ and a p-value $< .001$, which is statistically significant. The effect size, calculated as $r = Z/\sqrt{n^*}$ where n^* represents the number of pairs with non-zero differences (i.e., zero-differences are excluded), is $r = -.698$ ($Z = -4.985$, $n^* = 51$), indicating that this difference is large. These results suggest that the PTs perform considerably better in interpreting Luis's responses than in managing errors. For Carla, the values $W = 494.5$ with a p-value $< .001$ also indicate a significant difference between items (a) and (b). The effect size in this case is moderate to high, $r = -.463$ ($Z = -3.620$, $n^* = 61$), suggesting a relevant but less pronounced difference than in Luis's case between response interpretation and error management. In contrast, for Alba, there is no significant difference between items (a) and (b), with a p-value of .879. The effect size is very small, $r = -.030$ ($Z = -0.153$, $n^* = 26$), which reinforces the conclusion that there is no substantial difference in the PTs' performance between interpreting Alba's response and the actions proposed to manage errors.

The 95% confidence intervals, calculated using a Monte Carlo simulation with 10,000 samples, provide additional support for these findings. The parameter evaluated represents the differences in relevance scores assigned to participants for items (a) and (b) for each case (Luis, Alba, and Carla). This parameter reflects the PTs' ability to analyse student responses and propose error management strategies, offering valuable insights into their performance across the two tasks. For Luis and Carla, both ends of the confidence intervals are negative (−5.259, −4.711)[1] and (−3.869, −3.371)[1], respectively, indicating a consistent and significant difference between PTs' performance on the two items. Specifically, participants performed better on interpreting responses (item a) than managing errors (item b). In contrast, for Alba, the confidence interval spans both positive and negative values (−0.539, 0.233), including zero, which suggests that the differences between items (a) and (b) are not significant and are very close to zero, suggesting no statistically significant difference in performance between the two tasks.

Additionally, correlation analysis using Kendall's Tau-*b* coefficient reveals a significant positive correlation for the relationship between interpreting responses and managing errors for Luis (coefficient of .534, $p < .001$) and Carla (coefficient of .240, $p = .007$). These results indicate consistency in the relevance achieved by the PTs in interpreting these students' responses and making action decisions to manage errors. For Alba, no significant correlation is detected between responses to items (a) and (b) (coefficient of .028, $p = .787$), indicating that there is no consistent pattern between interpretation and management proposals in her response.

5. CONCLUSIONS

The mathematics teacher's competence to recognise aspects that are relevant in teaching and learning situations and interpret them professionally is essential for quality teaching (Godino et al., 2017; Ivars et al., 2018). Adopting a professional perspective to analyse students' mathematical thinking involves (a) describing the solution strategies that students use by discerning the mathematical details in their responses, (b) recognising the relationships between the identified elements and the characteristics of students' mathematical thinking, and (c) analysing and using this information to decide how to act on students' understanding or lack thereof (Buforn et al., 2020).

The aim of this paper was to report on the knowledge and skills of a group of primary school PTs in interpreting students' solutions to a task that seeks to determine the composition of an urn such that the probability of success is the same as that in which the ratio of favourable cases to unfavourable cases is known. This task is substantially different from those employed in previous research on sample space understanding (Hernández-Solís et al., 2021; Supply et al., 2020) and involves one of the most complex capabilities of probabilistic reasoning, "creating probabilities" (Supply et al., 2020). Therefore, the findings offer new insights into the didactic-mathematical knowledge of future

[1] Although individual differences are limited to a range of –2 to 2, the simulation estimates the sampling distribution of the median difference. As a result, the confidence intervals reflect the uncertainty in this overall estimate and can extend beyond the bounds of individual differences.

teachers, specifically in the cognitive and interactional aspects related to the proportional nature of probability calculation and the understanding of sample spaces (Batanero et al., 2015; Ortiz & Mohamed, 2014; Vásquez & Alsina, 2015a, 2015b, 2017).

The results show that interpreting students' responses, identifying errors, and proposing actions to help students who provide inadequate solutions to understand and overcome the difficulties that generated them was a complex task for the PTs. Specifically, PTs demonstrated varied success across the three cases analysed (Luis, Alba, and Carla). Although many PTs correctly identified the errors in Carla's response, difficulties were evident when justifying their evaluations, particularly for the responses from Luis and Alba. The quantitative analysis, particularly the effect size results, provided insights beyond statistical significance, emphasising the magnitude of the differences observed in PTs' performance across tasks. The results suggest that PTs performed more effectively when interpreting students' responses than when proposing strategies to manage errors, with larger differences noted for Luis and Carla than Alba. This lower score for the error management proposal aligns with Mohamed (2012). Although the PTs identified deficiencies in students' responses, the PTs did not use this information to decide how to address the deficiencies, or the proposed solution was not always related to the PTs' prior evaluation of the student's responses. These results also revealed gaps in didactical-mathematical knowledge about probabilistic reasoning that should be addressed. For example, as Vásquez and Alsina (2017) observed, when analysing students' responses, incorrect strategies were considered valid. Thus, some PTs considered Luis's answer adequate (and for the same reason, Alba's and Carla's answers inadequate), accepting that equal probability requires an equal composition of the urns and disregarding the proportional nature of probability calculations (Begolli et al., 2021). In the case of Alba, who had the highest percentage of inconclusive ratings, PTs did not correctly interpret the use of the 1:3 ratio or considered it incorrect. For some PTs, probability ratios represented ratios of favourable cases to possible cases. A biased or incomplete understanding of proportional reasoning and the meaning of ratio and proportion in probabilistic contexts may prevent PTs from adopting multiple or relativistic perspectives needed to go beyond a single correct answer (Buforn et al., 2020). The assessment of Carla's response was somewhat better, perhaps because it was closer to the expected strategy. On the other hand, it was also significant that most of the PTs recognised Luis's argument, which linked equal probability of events with the same number of balls, as inappropriate. This demonstrated an adequate understanding of the concept of sample space and its relationship with the proportional nature of probability calculation. These results showed an improvement over those obtained by Hernández-Solís and colleagues (2024) with students and by Batanero and colleagues (2015) with PTs.

The difficulties of the PTs in recognising the relationship between the mathematical practices (operational and discursive) of the students and the characteristics of students' thinking were subsequently reflected in how PTs proposed managing the errors or lack of knowledge encountered. In particular, more than 20% of the proposals for action were not relevant. The strategies highlighted include reinforcing basic probabilistic concepts (in Luis's case, it is believed that his strategy stems from a lack of understanding of equiprobability; in Alba's and Carla's cases, it is assumed that their use of the ratio 1/3 is due to a mistaken conception of the sample space), reviewing operations with fractions and their equivalence, and using iconic representations and manipulatives. However, these are not always appropriate or are used in a limited way (conceptual or procedural). As Vásquez and Alsina (2015a) observed, although PTs considered concrete materials to be a good strategy for teaching probability, they were not clear about which concrete materials were most suitable or how to use them to assist students in their learning. The tendency to focus interventions on direct correction of the task (expert "step-by-step" solution by the teacher on the blackboard so that pupils can check how to arrive at the correct result) without allowing pupils to understand their errors other than by confronting what was expected suggests room for improvement in cognitive and interactional aspects of didactic-mathematical knowledge.

However, compared with the results of Burgos and colleagues (2022) in the context of a task of comparing probabilities in urns, we observed that, on this occasion, despite a more complex situation than the one addressed in the previous work, the PTs obtained better scores both in the interpretation of the answers and in error management, with the improvement in the latter being more evident. In both aspects, better results were also obtained than those obtained by Burgos and colleagues (2023) when analysing students' responses for a task in which the prize had to be decided for the game to be fair. This suggests that success in dealing with students' responses when solving problems does not exclusively depend on the underlying mathematical content.

The qualitative approach adopted in our research does not aim to generalise but rather to contribute or reflect on plausible generalisations from the specific case studied to other communities with certain similarities (Montes, 2021). In this case, the community to which the findings could be "plausibly generalisable" would be defined by PTs who receive similar training during their primary education degree program studies. Of course, gaining certainty about this "reasonable possibility" requires new cycles of research.

Teachers' abilities to recognise and interpret students' strategies and errors are crucial for effective educational interventions. Our results highlight the need to reinforce different skills and knowledge related to probabilistic reasoning—materials that may be suitable for its teaching, how it is learned, and the difficulties that students may encounter—to enable PTs to incorporate them into decision-making in the classroom. The instrument type used in this research could be adapted to design teaching materials focused on the development of PTs' competences to analyse students' written work. This instrument should be extended to include multiple solutions (several strategies) and the analysis of the knowledge involved, as in the works of Batanero and colleagues (2015) or Ortiz and Mohamed (2014), with the intention of identifying whether the shortcomings observed have their origins in the epistemic dimension. Given that many of the descriptions were inconclusive, it would be advisable to complement the intervention with interviews and subsequent sharing sessions in which the PTs explain the reasons behind their assessment of the students' responses or their proposed actions, which we were unable to do given that the data collection was part of the final assessment of the subject.

This study has certain limitations that should be considered when interpreting the results. On the one hand, the characteristics of the participants (who were in the early stages of their training and might have had limited experience in the area of probability) may influence the generalisation of the results to other contexts or groups of teachers with more advanced training. On the other hand, although the scoring system and criteria were based on prior analysis, and the coding, categorisation, and evaluation process was conducted and agreed upon by different researchers, we acknowledge the inherent subjectivity in the qualitative data analysis. These limitations suggest the need for further studies that include more diverse samples and complementary methodological approaches. This would not only verify the generalisability of the findings but also enable a deeper study of PTs' didactic-mathematical competences in probabilistic reasoning across various contexts, providing a more robust and comprehensive understanding of their training and needs. Moreover, it would be beneficial to investigate competence in interpreting students' responses and error management among in-service teachers, with the intention of identifying other aspects that should be integrated into teacher education programmes, extending the study to secondary school teachers.

ACKNOWLEDGEMENTS

The research was developed within the framework of the project PID2022-139748NB-100 funded by MICIU/AEI/10.13039/501100011033 and FEDER, EU; Groups FQM-126 and HUM-886 (Junta de Andalucía, Spain).

REFERENCES

Australian Curriculum, Assessment and Reporting Authority. (2014). *Foundation to year 10 curriculum: Statistics and probability* (ACMSPO24).

Batanero, C., & Álvarez-Arroyo, R. (2024). Teaching and learning of probability. *ZDM - Mathematics Education*, *56*(1), 5–17. https://doi.org/10.1007/s11858-023-01511-5

Batanero, C., Gómez, E., Contreras, J. M., & Díaz, C. (2015). Mathematical knowledge of primary school teachers in training for the teaching of probability: An exploratory study. *Práxis Educativa*, *10*(1), 11–34. https://doi.org/10.5212/PraxEduc.v.10i1.0001

Batanero, C., Hernández-Solís, L. A., & Gea, M. M. (2023). Analysing Costa Rican and Spanish students' proportional reasoning and comparison of probabilities. *Statistics Education Research Journal*, *22*(3), Article 7. https://doi.org/10.52041/serj.v22i3.659

Begolli, K. N., Dai, T., McGinn, K. M., & Booth, J. L. (2021). Could probability be out of proportion? Self-explanation and example-based practice help students with lower proportional reasoning skills learn probability. *Instructional Science, 49*, 441–473. https://doi.org/10.1007/s11251-021-09550-9

Bryant, P., & Nunes, T. (2012). *Children's understanding of probability: A literature review*. The Nuffield Foundation.

Buforn, A., Llinares, S., Fernández, C., Coles, A., & Brown, L. (2020). Pre-service teachers' knowledge of the unitizing process in recognizing students' reasoning to propose teaching decisions. *International Journal of Mathematics Education in Science and Technology, 53*(2), 425–443. https://doi.org/10.1080/0020739X.2020.1777333

Burgos, M., López-Martín, M. M., Aguayo, C. G., & Albanese, V. (2022). Cognitive analysis of probability comparison tasks by prospective primary school teachers. *Uniciencia, 36*(1), 1–24. https://doi.org/10.15359/ru.36-1.38

Burgos, M., López-Martín, M. M., Albanese, V., & Aguayo-Arriagada, C. G. (2023). Analysis of primary school students' answers to fair game tasks: An experience with preservice teachers. *BEIO, Bulletin of Statistics and Operations Research, 39*(3), 48–69.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*. Routledge. https://doi.org/10.4324/9780203224342

Dröse, J., Griese, B., & Wessel, L. (2022). Prospective teachers' diagnostic judgments on students' understanding of conditional probabilities. In J. Hodgen, E. Geraniou, G. Bolondi, & F. Ferretti (Eds.), *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)* (pp. 3080–3088). Free University of Bozen-Bolzano and European Research in Mathematics Education.

Franco, J., & Alsina, Á. (2022). El conocimiento del profesorado de Educación Primaria para enseñar estadística y probabilidad: Una revisión sistemática. [Primary school teachers' knowledge to teach statistics and probability: A systematic review]. *Aula Abierta, 51*(1), 7–16. https://doi.org/10.17811/rifie.51.1.2022.7-16

Godino, J. D., Giacomone, B., Batanero, C., & Font, V. (2017). Enfoque Ontosemiótcio de los conocimientos y competencias del profesor de matemáticas [Ontosemiotic approach to mathematics teacher's knowledge and competences]. *Bolema: Boletim de Educação Matemática, 31*(57), 90–113. https://doi.org/10.1590/1980-4415v31n57a05

Hernández-Solís, L. A., Batanero, C., & Gea, M. M. (2024). Relación entre la construcción de espacios muestrales y el razonamiento proporcional de estudiantes costarricenses [Relating the construction of sample spaces and proportional reasoning in Costa Rican students]. *Revista de Educación Estadística, 3*, 1–28. https://doi.org/10.29035/redes.3.1.1

Hernández-Solís, L. A., Batanero, C., Gea, M. M., & Álvarez-Arroyo, R. (2021). Construcción de espacios muestrales asociados a distintos tipos de sucesos: Un estudio exploratorio con estudiantes de Educación Primaria [Construction of sample spaces associated with different types of events: An exploratory study with primary education students]. *Educación Matemática, 33*(1), 181–207. https://doi.org/10.24844/em3301.07

Ivars, P., Fernández, C., Llinares, S., & Choy, B. H. (2018). Enhancing noticing: Using a hypothetical learning trajectory to improve pre-service primary teachers' professional discourse. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(11), Article em1599. https://doi.org/10.29333/ejmste/93421

Jakobsen, A., Ribeiro, C. M., & Mellone, M. (2014). Norwegian prospective teachers' MKT when interpreting pupils' productions on a fraction task. *Nordic Studies in Mathematics Education, 19*(3–4), 135–150.

Leavy, P. (2022). *Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches*. Guilford Publications.

Ministerio de Educación y Formación Profesional (2022). Real Decreto 157/2022, de 1 de marzo, por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria [Royal Decree 157/2022, of March 1, Establishing the organization and minimum curricular standards for primary education]. *Boletín Oficial del Estado, 52*(I), 24386–24504.

Mohamed, N. (2012). *Evaluación del conocimiento de los futuros profesores de primaria sobre probabilidad* [Assessment of prospective primary school teachers' knowledge of probability] [Unpublished doctoral dissertation]. University of Granada.

Montes, M. (2021). Introducción al Seminario: Investigación cualitativa en educación matemática [Introduction to the seminar: Qualitative research in mathematics education]. In P. D. Diago, D. F. Yáñez, M. T. González-Astudillo, & D. Carrillo (Eds.), *Investigación en Educación Matemática XXIV* (pp. 53–63). Servicios Educativos Integrados al Estado de México.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics.*

Ortiz, J. J., & Mohamed, N. (2014). Prospective teachers' knowledge of sample space. *Quadrante, 23*(2), 5–22.

Sánchez, E., & Valdez, J. (2017). Las ideas fundamentales de probabilidad en el razonamiento de estudiantes de bachillerato [Fundamental ideas of probability in the reasoning of high school students]. *Avances de Investigación en Educación Matemática*, *11*, 127–143. https://doi.org/10.35763/aiem.v1i11.180

Simpson, A., & Haltiwanger, L. (2017). This is the first time I've done this: Exploring secondary prospective mathematics teachers' noticing of students' mathematical thinking. *Journal of Mathematics Teacher Education, 20*(4), 335–355. https://doi.org/10.1007/s10857-016-9352-0

Supply, A. S., Van Dooren, W., & Onghena, P. (2020). Can we count on early numerical abilities for early probabilistic reasoning abilities? *Mathematical Thinking and Learning, 24*(1), 19–37. https://doi.org/10.1080/10986065.2020.1805551

Supply, A. S., Vanluydt, E., Van Dooren, W., & Onghena, P. (2023). Out of proportion or out of context? Comparing 8- to 9-year-olds' proportional reasoning abilities across fair-sharing, mixtures, and probability contexts. *Educational Studies in Mathematics, 113*(3), 371–388. https://doi.org/10.1007/s10649-023-10212-5

Van Dooren, W. (2014). Probabilistic thinking: Analyses from a psychological perspective. In E. Chernoff & B. Sriraman, (Eds.), *Probabilistic thinking: Advances in mathematics education.* Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7155-0_7

Vásquez, C., & Alsina, Á. (2015a). Conocimiento didáctico-matemático del profesorado de Educación Primaria sobre probabilidad: Diseño, construcción y validación de un instrumento de evaluación. [Didactic-Mathematical knowledge of primary education teachers about probability: Design, construction and validation of an assessment instrument]. *Bolema: Boletim de Educação Matemática, 29*(52), 681–703. https://doi.org/10.1590/1980-4415v29n52a13

Vásquez, C., & Alsina, Á. (2015b). El conocimiento del profesorado para enseñar probabilidad: Un análisis global desde el modelo del Conocimiento Didáctico-matemático [Teachers' knowledge to teach probability: A global analysis from the didactic-mathematical knowledge model].

Avances de Investigación en Educación Matemática, *7*, 27–48. https://doi.org/10.35763/aiem.v1i7.104

Vásquez, C., & Alsina, Á. (2017). Aproximación al conocimiento común del contenido para enseñar probabilidad desde el modelo del Conocimiento Didáctico-matemático [Approaching common content knowledge to teach probability from the didactic-mathematical knowledge model]. *Educación Matemática*, *29*(3), 79–108. https://doi.org/10.24844/EM2903.03

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “$p < 0.05$.” *The American Statistician*, *73*(s1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

MARÍA DEL MAR LÓPEZ-MARTÍN
University of Almería
Almeria, Andalucía, ES