

Article

Prospective Mathematics Teachers Understanding of Classical and Frequentist Probability

Carmen Batanero ^{1,*}, Nuria Begué ², Rocío Álvarez-Arroyo ¹ and Silvia M. Valenzuela-Ruiz ¹

¹ Department of Mathematics Education, University of Granada, 18071 Granada, Spain; rocioaarroyo@ugr.es (R.Á.-A.); svalenzuela@ugr.es (S.M.V.-R.)

² Department of Mathematics, Area of Mathematics Education, University of Zaragoza, 50009 Zaragoza, Spain; nbegue@unizar.es

* Correspondence: batanero@ugr.es

Abstract: Strengthening the teaching of probability requires an adequate training of prospective teachers, which should be based on the prior assessment of their knowledge. Consequently, the aim of this study was to analyse how 139 prospective Spanish mathematics teachers relate the classical and frequentist approaches to probability. To achieve this goal, content analysis was used to categorize the prospective teachers' answers to a questionnaire with open-ended tasks in which they had to estimate and justify the composition of an urn, basing their answers on the results of 1000 extractions from the urn. Most of the sample proposed an urn model consistent with the data provided; however, the percentage that adequately justified the construction was lower. Although the majority of the sample correctly calculated the probability of an event in a new extraction and chose the urn giving the highest probability, a large proportion of the sample forgot the previously constructed urn model, using only the frequency data. Difficulties, such as equiprobability bias or not perceiving independence of trials in replacement sampling, were also observed for a small part of the sample. These results should be considered in the organisation of probabilistic training for prospective teachers.

Keywords: probability; relating the classical and frequentist views; teacher's knowledge; assessment



Citation: Batanero, C.; Begué, N.; Álvarez-Arroyo, R.; Valenzuela-Ruiz, S.M. Prospective Mathematics Teachers Understanding of Classical and Frequentist Probability. *Mathematics* **2021**, *9*, 2526. <https://doi.org/10.3390/math9192526>

Academic Editors: Laura Muñoz-Rodríguez and María Magdalena Gea Serrano

Received: 13 September 2021
Accepted: 5 October 2021
Published: 8 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As well as being a relevant part of mathematics, and applicable to other curricular areas, probability is necessary in many fields of science, where it enables us to describe the laws governing random phenomena [1]. Given this relevance, the teaching of probability in Spain currently extends from primary education to high school, with the aim of providing students with basic probabilistic literacy that will enable them to successfully deal with random situations in their daily and professional lives [2].

An essential issue to ensure the success of this teaching is the adequate training of the teachers who are responsible for this content. This preparation should include both the mathematical characteristics of probability and the related pedagogical knowledge [3,4].

Nowadays, the knowledge and education of teachers is one of the broadest and most productive lines of research in mathematics education. Some references can be found in several sources [5–8] or in journals such as the *Journal of Mathematics Teacher Education*. However, an analysis of this literature suggests that these studies have paid less attention to the specific case of statistics and probability than to other mathematics topics. This line of research started to increase following the Joint ICMI/IASE Study, “Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education”, which was organised by the International Commission on Mathematical Instruction (ICMI) in collaboration with the International Association for Statistical Education (IASE), with the purpose of promoting research specifically focused on the education and professional development of teachers to teach statistics [9], but is still scarce.

As with other curricular content, the teachers' knowledge includes content knowledge and various components of pedagogical knowledge, described in different models, such as the Mathematical Knowledge for Teaching (MKT) framework [10], or the Didactic-Mathematical Knowledge and Competences model (DMKC) [11]. Moreover, this research makes clear that pedagogical knowledge is specific to each content knowledge area [8] (in particular, for probabilistic knowledge) and, therefore, that a sound knowledge of probability is the first step for teachers to acquire the different components of the pedagogical knowledge that they will need to teach probability.

The above justifies new research that analyses the probabilistic knowledge of teachers, in particular that of secondary and high school teachers who, in spite of having an excellent mathematical background, have received only formal training in probability. This formal teaching of probability is not in accord with current curricular recommendations in Spain [12], where an empirical approach based on simulations and experiments is recommended to complement the classical definition of probability. Given that the time for preparing teachers to teach probability is limited, it is important to assess their probabilistic knowledge at the beginning of their preparation, as a basis for organising didactic activities directed towards reinforcing both their content and pedagogical knowledge.

Consequently, the aim of this paper was to analyse the way in which prospective Spanish secondary and high school teachers relate the classical and frequentist views of probability, as well as their possible difficulties in establishing this relationship. These perspectives are reflected throughout the Spanish curriculum; specifically, the classical view is included from the first year of compulsory secondary education (CSE), the frequentist approach in the second year, supported by simulation, and both approaches are worked on from the third year until the end of high school [12].

Although we count with some studies, summarised in Section 2.3, of how students establish this link, most of these studies have analysed the evolution of small samples of primary or secondary school students, through teaching experiments based on computer simulation. Consequently, the results of this research are not directly applicable to Spanish prospective secondary and high school teachers, who have previously received a strong formal statistical and mathematical preparation.

To achieve our objective, we analysed the responses to a questionnaire with open-ended tasks of a sample of 139 students who were preparing to become mathematics teachers at these educational levels in Spain, taking part in a master's programme that is compulsory for applying to a teaching position in a public school.

2. Theoretical Background

The paper considers the classical and frequentist views of probability, previous research on the connections between them, and research on teacher education.

2.1. Classical and Frequentist Views of Probability

The concept of probability has been conceived from different points of view throughout history, including the intuitive, classical, frequentist, subjective, propensity, logical, and axiomatic connotations, which still coexist in the applications of statistics, and some of which (intuitive, classical, frequentist, subjective, and axiomatic) are included in school curricula [13–15]. Specifically, in this paper we will consider the classical and frequentist approaches, both considered in secondary education and at high school levels. In the following, we summarise the main features of these two approaches and refer the reader to other sources [13–15] for a deeper description of the different approaches to probability.

The classical definition originated from the resolution of problems related to games of chance, among others those discussed by Pascal and Fermat in their correspondence. A first definition following this conception was provided by de Moivre [16], as follows:

Wherefore, if we constitute a fraction whereof the numerator is the number of chances whereby an event might happen, and the denominator the number of all the chances

whereby it may either happen or fail, that fraction will be a proper definition of the probability of happening (p. 1).

Laplace [17] refined this definition by proposing what is currently taught in schools as “Laplace’s rule”, stating that:

Probability is thus simply a fraction whose numerator is the number of favourable cases and whose denominator is the number of all cases possible [9], p. ix.

This definition has been considered circular since its publication, because it included the term “equiprobable”; moreover, it can only be applied to experiments with a finite number of possibilities, neglecting a large field of applications where these assumptions do not hold. In didactic research, we often find authors who call this view a “theoretical approach” to probability, stating that:

It makes use of an implicit assumption of equal likelihood of all single outcomes of the sample space. It is an a priori approach to probability in that it allows calculations of probabilities before any trial is made [18], p. 41.

The frequentist approach arose from life table studies in the United Kingdom, where collecting and analysing large amounts of data showed the stabilisation over time of each event to relative frequency [15,18]. Bernoulli [19] proved the first version of the law of large numbers (LLN) concerning the difference between the relative frequency of an event and its theoretical probability. The probability that this difference is smaller than a fixed amount can approach 1 as much as desired, when a sufficiently large number of independent repetitions of an experiment are performed [14].

Based on this theorem, Von Mises [20] defined the probability of an event as the limiting value to which its relative frequency tends in a sufficiently large number of independent trials. Although this definition significantly broadens the field of application of probability, it is not free of controversy. For example, with this approach, we cannot obtain the true value of probability, but only an estimate of this; moreover, it is not always possible to perform a large number of independent repetitions of the experiment. The frequentist approach is referred to by some mathematics educators as an “*experimental or empirical approach*” to highlight the fact that the probability is estimated from the relative frequency (e.g., [21]).

As Chaput, Girard, and Henry [22] have pointed out, these two approaches are complementary and both of them require a sound understanding of probability. In the same vein, Steinbring [23] indicated that these two views of probability “*should be related to each other as analogous forms of the same concept without, however, being identified*” (p. 165).

Both the classical and the frequentist definitions of probability are included in the Spanish mathematics curriculum for secondary education [12], the latter being highly recommended because it connects statistics with probability. Prospective teachers must have a good understanding of the characteristics and differences of these two approaches and their relationship to each other. Such an understanding requires knowledge of the LLN [21,24], as well as of the fundamental stochastic ideas of randomness, variability, and independence [25]. This type of relational understanding is the one that we aim to assess in this paper for a sample of prospective Spanish secondary and high school mathematics teachers.

2.2. The Education of Mathematics Teachers

This paper is based on the Mathematical Knowledge for Teaching (MKT) model, which the authors divide into the following components: Common Content Knowledge (CCK), Specialised Content Knowledge (SCK), Knowledge of Content and Teaching (KCT), and Knowledge of Content and Students (KCS) [10]. Hill, Ball, and Schilling [26] further proposed that Horizon Content Knowledge (HCK), and Knowledge of Content and Curriculum (KCC) be included. CCK refers to the knowledge brought into play by an educated person to solve mathematical problems, for which a person with basic knowledge is qualified. SCK describes the teacher’s special knowledge that enables him/her to

plan and develop teaching sequences. HCK refers to the more advanced aspects of the content, which provide insights for the teacher, e.g., knowledge of the history or detection of possible errors with respect to the mathematical ideas underlying the topic.

This paper focuses on prospective high school mathematics teachers' mathematical knowledge of the classical and frequentist approaches to probability. Specifically, we focus on their ability to estimate the composition of an urn model from frequency data on extractions with replacement from an urn, and on their prediction of probability in new experiments using the constructed model. Such knowledge about classical and frequentist probability appears in the Spanish secondary school curriculum [12] and should be taught to the students. However, since we propose some tasks that are not usually found in the textbooks, we evaluate part of the teachers' Common Content Knowledge (CCK) and Horizon Content Knowledge (MHK).

2.3. Relating Different Views of Probability

There are few studies specifically analysing the relationship between the classical and frequentist conceptions of probability, although in some research the authors have described the emergence of intuitive ideas about this link in small samples of school children, while they were working in a computational environment. For example, Pratt [27] studied the meaning that 16 children aged 10 and 11 years assigned to chance when they played in pairs with a computer game, where they had to make sense of the sum of two dice. The author suggested that children constructed new ideas based on the interplay of their previous intuitions and their work with the computer resource. Other examples of case studies in a computer simulation experimental setting can be found in [28], in relation to secondary students' ideas of distribution, and of university students understanding of random processes in [29].

More specifically related to our research, Ireland and Watson [24] conducted research with 27 grade 5–6 students (11–12 years old), using the Tinkerplot software [30], to simulate samples of increasing size from a random experiment consisting of drawing balls from an urn of known composition. Their aim was to explore the understanding of probability, in the transition from the theoretical (classical) to the experimental (frequentist) views, by asking questions, followed by experiments, where the children had to predict the colours of the balls in samples of increasing size. The authors concluded that the most difficult element to understand was the LLN.

Sánchez and Valdez [25] studied the inferences made by a group of 30 high school students in Mexico, using their knowledge of the classical and frequentist conceptions of probability. The students were divided into three groups of 10 students, each of which was given a different questionnaire, with variations of the same task that combined probability comparison and sampling. These tasks provided data from 1000 extractions of black and white balls from two urns with known or unknown composition, asking the students to choose the urn that provided more chance of obtaining a given colour in the next extraction, or to predict the colour of the next extraction, depending on the questionnaire. The proposed tasks required relating the classical and frequentist visions of probability, although the authors' aim was to analyse the students' use of the fundamental stochastic ideas [31] of variability, randomness, and independence. From the analysis of the students' responses, the authors proposed a hierarchy of levels of understanding each of these ideas. These levels are as follows:

- Randomness: (1) making deterministic predictions; (2) deterministic predictions qualified with probabilistic language; (3) recognising that the outcome cannot be predicted with accuracy; (4) although the outcome cannot be predicted with accuracy, recognising the stability of the frequency in the long run.
- Variability: (1) not considered; (2) thinking that differences between specified and observed frequencies are always significant regardless of sample size; (3) considering a difference to be significant in a small sample but not in a large sample; (4) understanding the relationship of variability with sample size.

- Independence: (1) thinking that successive results depend on the previous outcomes; (2) the result depends on whether the sample is representative; (3) using models to determine a possible result; (4) recognising independence.

Sánchez and Valdez [32] analysed the way in which a group of Mexican high school students understood the concept of probability in relation to the LLN. Using interviews, they analysed their responses to physical and computational simulation tasks, using the reasoning levels defined by Jones et al. [33]: subjective, transitional, informal quantitative, and numerical. At the subjective level, probability is not assigned to events or it is done subjectively, without the possibility of using probability to make inferences; at the transitional level, probability is assigned to an event through a priori analysis of the experiment or through the empirical results, without relating to each other; at the informal quantitative level, both approaches are used, but variability is not taken into account; and at the numerical level, both approaches are brought into play, valuing variability appropriately to form inferences.

Regarding work with prospective teachers, we can quote Serrano [34], who analysed the understanding of the frequentist approach to probability in a sample of 130 prospective primary school teachers, using a questionnaire in which he presented problems of generation and recognition of random outcome sequences. In the first type of problem, subjects were asked to write down a sequence of coin-toss outcomes in a way that might appear random to another person, and in the second, they were given several sequences of heads and tails and asked which ones they thought had been randomly generated. The analysis of the responses and the characteristics of the generated sequences showed a high proportion of prospective teachers expecting the convergence of relative frequency to probability in small samples, i.e., reasoning according to the representativeness heuristic [35].

Parraguez et al. [4] analysed the way in which a sample of 60 prospective Spanish primary school teachers related the classical and frequentist views of probability, presenting them with a problem about the sum of two dice. First, the teachers were asked to solve the problem with the classical approach, obtaining correct answers in two thirds of the sample. When asked to estimate the expected value of the frequency over a series of repetitions of the experiment, only 50% of participants were able to provide the estimate. Most participants could suggest that the sample size was too small when given data from 100 repetitions of an experiment where the relative frequency clearly deviated from the expected frequency. The authors also noted biases, such as representativeness and equiprobability [36], for some participants.

Our work complements previous research by focusing on prospective secondary and high school teachers, who are more strongly trained in statistics than primary school teachers. In addition, the way in which the prospective teachers build a model of urns from experimental data of urn extractions is analysed, as well as their use of the model built to make predictions or make decisions about future experiments.

3. Materials and Methods

The sample was composed of 139 students on a master's programme, which is compulsory in Spain in order to become a mathematics teacher in compulsory secondary education (students aged from 12 to 15 years) and high school (students aged from 16 to 17 years). This master's degree is taken after the completion of a discipline-specific bachelor's degree, which is formal in nature and does not provide pedagogical knowledge. This master's degree aims to fill this gap in initial teacher training [37], and provides prospective teachers with didactic, curricular, and school organisation knowledge, as well as experience in teaching practice.

This was a purposeful, controlled sample [38], which included all the participants of the master's degree in two successive academic years at the University of Granada (66 and 73 students in the 2019–2020 and 2020–2021 academic years, respectively). Half of these students had completed a university degree in mathematics and the remainder

had undertaken other scientific subjects (e.g., statistics, physics, chemistry, architecture or engineering).

These prospective teachers were given a questionnaire consisting of two open-ended tasks, as shown in Figure 1.

Task 1. Two urns A and B contain 10 balls each, some of which are black and the others white. We performed 1000 extractions from each urn, each time returning the ball to the urn after it had been extracted. In urn A, we obtained 324 white balls and 676 black balls. In urn B the results were 510 white balls and 490 black balls.

- How many white and black balls do you think there are in urn A? And in urn B? Justify your answer.
- What is the probability of drawing a black ball in the 1001st draw in urn A? What is the probability of drawing a black ball in urn B? Justify your answer.

Task 2. From each of the two urns above, we select 10 balls, obtaining the following results:

Urn A	b	w	w	b	b	b	w	w	b	w
Urn B	w	b	w	b	w	w	w	b	w	w

What urn is preferable to obtain a white ball in the next selection?

Figure 1. Questionnaire given to prospective teachers.

In Task 1, adapted from Sánchez and Valdez [25], the participants were asked to estimate the number of white and black balls in the urn, using the result of 1000 extractions. The probability of drawing a ball of each colour in the urns will be given by $w/10$ and $b/10$, where w is the number of white balls, b the number of black balls, with $w + b = 10$. Since the composition of the urns is unknown, in this first step, the probability of each colour must be estimated from the relative frequency of results in the 1000 extractions. That is, 0.324 and 0.676 in the first urn, and 0.510 and 0.490 in the second; this estimate is based on the fact that the sample size is large and the results are independent (as this is sampling with replacement). Since there are 10 balls (possible cases) in the urns, the expected value of the number of white and black balls (favourable and unfavourable cases) in each urn is given by the product of the relative frequency of each colour determined above, multiplied by 10. That is, 3.24 and 6.76 in the first urn, and 5.1 and 4.9 in the second, but since the number of balls is integer, by rounding to the nearest integer, the best estimate of black and white balls is $(3w,7b)$ in the first urn, and $(5w,5b)$ in the second. Once the most probable composition of the urns has been determined with this procedure, to answer the second part of the task, it would be enough to apply the classical definition to obtain the probabilities 0.7 and 0.3 in the first urn, and 0.5, and 0.5, in the second urn.

In Task 2, the prospective teachers' choice of the urn that is most likely to yield a given ball in further draws is assessed. The prospective teachers were expected to use the urn model they have constructed in the first task and apply the classical view of probability to make their decision. Therefore, the correct answer is that urn B is preferable, which, according to the estimation of its composition in Task 1, has a higher number of favourable cases. As a distractor, the subjects were given the results of 10 new draws from each urn, which are consistent with the variability expected in a short sequence of trials and, in any event, also favour urn B.

Consequently, the questionnaire assessed the knowledge of the classical and frequentist views of probability and their relationship, the estimation of proportion, the expected value over a number of trials, when the proportion is known, and the elementary characteristics of a short series of random outcomes apply. Although these are all elementary ideas of probability, since the task is not common in secondary school textbooks, we consider that the questionnaire assessed aspects of Common Content Knowledge (CCK) and Mathematical Horizon Knowledge (HCK).

4. Results

4.1. Estimating the Number of White and Black Balls in the Urns

In Table 1 the results of the estimation of the number of black and white balls in the two urns in Task 1a are presented. The answers obtained have been classified as follows:

- *Correct.* When 3 white and 7 black balls in the first urn and 5 balls of each colour in the second urn are estimated from the relative frequency, by correctly relating the classical and frequentist approaches to probability.
- *Providing an interval of values.* For one or both urns an interval of values that includes the correct answer is given. For example, “in box A there are between 3 and 4 white balls and between 6 and 7 black balls” (P18). This answer is incorrect, as can be deduced by considering that the mean and standard deviation in a binomial distribution $B(n, p)$ are respectively $\mu = np$; $\sigma = \sqrt{npq}$, where n is the number of trials, p is the probability of success, and q the probability of failure. If urn A is assumed to have 3 white balls, the number of white balls obtained in 1000 draws can be modelled by a binomial distribution $B(0.3, 1000)$, with $\mu = 300$ and $\sigma = 14.5$; and if urn A contains 4 white balls, it is modelled by the binomial distribution $B(0.4, 1000)$, with $\mu = 400$ and $\sigma = 15.5$. By typifying the observed number of white balls in each of the two assumptions, we obtain $Z = 1.65$ for the urn with 3 white balls and $Z = -4.9$ for the urn with 4 white balls. While the first value of the standard normal distribution $N(0, 1)$ is acceptable, the second is too far away from the mean, and would lead us to reject the solution of 4 white balls, as the data are very unlikely with such a composition of the urn.
- Indicating only that there are *fewer white balls than black balls* in urn A, and therefore failing to relate the relative frequency of results to the composition of the urn, i.e., not linking the frequentist estimate of the probability obtained from the 1000 draws to the theoretical probability of obtaining each colour, when defined in the classical sense.
- *Incorrectly estimating* the number of balls of each colour; for example, replying that there are 3 white balls and 4 black balls in urn A, so that the total number of balls is different from 10, or else indicating that the number of balls in each urn cannot be known.

Table 1. Proposed composition of urns in Task 1a.

Number of White and Black Marbles	Percentage
Correct: Urn A: (3w,7b), Urn B: (5w,5b)	71.9
Providing an interval of values for each urn	5.8
Urn A: More black marbles	5
Incorrect values or anything is possible	10.8
No response	6.5

The results of this question are displayed in Table 1, and indicate that the task was straightforward for the prospective teachers, since a high percentage of them correctly related the frequency estimate of the probability, given by the results of the 1000 experiments, to the theoretical value of classical probability, given by the quotient of the number of favourable and possible cases in each urn.

About 6% of the prospective teachers provided a range of values for the composition of black and white balls in the urns. Although these prospective teachers used the idea of variability in estimation, which is fundamental to relating the classical and frequentist approaches to probability [25], they misinterpreted this variability, since, as we have determined, a composition of the urns different to what is expected would be highly improbable. These participants did not correctly consider sampling variability, although they related the classical and frequentist views, and therefore they manifested an informal quantitative reasoning (level 3), not reaching the higher level of understanding of variability in the classification of Sánchez and Valdez [32].

A proportion of the sample stated that any urn composition was possible, thus demonstrating the equiprobability bias [36] by assuming that any urn could have given the results obtained. Or, in other words, they suggested that all possible results were equiprobable with any urn composition, consistent with what was observed in Parraguez et al. [4]. Other participants supplied a composition whose sum did not correspond to the total number of balls in the urns. In conclusion, about 30% of the sample was unable to create a plausible model of the distribution of balls in the urns, given the results obtained in the 1000 extractions. Thus, they failed to connect both approaches to probability, implying a lack of understanding of the LLN [24].

In the following, the justifications given for the composition of the urns are analysed.

4.1.1. Correct Justifications

In correct justifications the participants related the classical and frequentist conceptions of probability using different procedures:

- Estimating the probability by analysing the ratio* between the numbers of black and white balls in the 1000 extractions, and approximating the number of black and white balls in the urn. This is a correct justification, based on proportional reasoning, which is an essential component of probabilistic reasoning [39].

P31: (3,7) and (5,5) but it is not certain. Urn A: Based on the results we could assume a direct relationship between the number of times a colour is drawn and the number of balls inside the urn. A relationship $B/W \rightarrow 7/3$. Urn B: the same as in A. Ratio $B/W \rightarrow 5/5$.
- Estimating the probability value by computing the relative frequency* or percentage of balls of each colour in 1000 outcomes, and then determining the expected number of balls of each colour by multiplying the estimated probability by 10 (number of balls in the urn). Finally, rounding the number of balls to the nearest integer:

P2: $f(w) = \frac{324}{1000} = 0.324 \rightarrow P(w) = \frac{n_w}{10} \rightarrow n_w = 3.24 \approx 3$ white
 $f(b) = \frac{676}{1000} = 0.676 \rightarrow P(b) = \frac{n_b}{10} \rightarrow n_b = 6.76 \approx 7$ black
- Some students also *set up and solved an equation* by equalling the ratio of black and white balls in the results and inside the urn. They were working at the algebraic level [40], since they used the linear function and dealt with equations in which they found the unknown value, while, in the two previous categories, the students worked at the arithmetic level. Finally, the result must be rounded, although some students did not express this step explicitly. For example:

P32: $\frac{324}{1000} = \frac{x}{10} \rightarrow x = 10 \cdot \frac{324}{1000}$
- Convergence of the proportion or of the sample mean to the population proportion or quoting the LLN.* Sometimes, the relationship between the frequentist and classical approach to probability was re-emphasised by recalling the LLN, which states the conditions of convergence of the relative frequency to the theoretical probability (see example P36). A different way of expressing the idea that the relative frequency over a long series of trials (frequentist approach) tends to the theoretical probability (classical approach) is to use the idea of sampling. The composition of balls defines a finite population in each urn and the series of outcomes constitute a sample of 1000 elements, taken with replacement from that population. The proportion of balls of a colour in the urn is a parameter in the population, while the sampling proportion is an unbiased estimator of the population proportion.

P36. According to the law of large numbers, when an experiment is performed a sufficiently large number of times, the probability of an event stabilises.

P25. The sample mean converges to the population mean; the sample mean is an unbiased estimator of the population mean.

4.1.2. Incorrect or Incomplete Justifications

Other participants were inaccurate or relied on the following incorrect reasons:

- They simply answered based on the observed results without being able to estimate the probability. They pointed out that, apparently, the results indicated that there were more or an equal number of balls of one colour than of another. These students did not provide a specific composition for the urns, since they could not relate the relative frequency in the 1000 experiments to an estimate of the number of black and white balls in the urns.

P5: I believe that there are more black balls than white balls in urn A, because when drawing 1.000 balls, 676 were black. In the second urn, the number of black and white balls is more similar, so I think there are approximately the same number of each colour.

- *Incorrect application of classical probability, by interpreting the experiment outcomes as favourable and possible cases.* These prospective teachers failed to understand the difference between event (element of the experiment sample space) and outcome (event that occurred in each trial). Some participants, such as P80, offered a solution that exceeded the total of 10 balls in the urn. In other cases, such as P105, they explained that they did not calculate the probability because they did not know the number of favourable cases.

P80: I think that in box A there are about 324 white balls and about 676 black balls. I believe that in box B there are about 510 white balls and about 490 black balls. I can't know the exact numbers, but it seems reasonable to assume that the values will be similar to these, when assuming that the draws are random. Therefore, the results of the draws and the probability distribution they yield reflect the distribution of balls in the urns.

P105: I don't know how to calculate the number of balls. Not knowing the number of favourable events, I don't know how to calculate the probability. Besides, by putting the ball back in the urn, the number of possible events is not reduced.

- *Equiprobability bias.* This biased reasoning arose when the participant suggested that any composition was possible, because we dealt with a random experiment, so that any outcome had the same probability. These participants explicitly showed the equiprobability bias, described by Lecoutre [36], in which it is assumed that any outcome of a random experiment is equiprobable. Consequently, these participants thought that the given results could be obtained with any composition of the urn, and that an estimation of the number of balls was impossible. They failed to link the two approaches to probability.

P57. Assuming that we obtain 5 white balls and 5 black balls, the sample will always be different, since it is possible to pick up a ball and leave it, and select the same one again. Therefore, it is not feasible to deduce how many white or black balls there are in the urn.

P98: There might be any number of black and white balls. There can be 1 black and 9 white balls or vice versa as each time one is picked up it is replaced. Or there can be 5 and 5 of each colour.

Table 2 lists the justifications, most of which were correct (45.2%). These are supplemented by 22.3% of participants who gave partially correct answers. These prospective teachers related both approaches to probability, and were also aware of the properties that this relationship enables, which have been discussed throughout the section. Therefore, they have reached the top level (4) of understanding variability in the Sánchez and Valdez's model [25].

Table 2. Justifications given to Task 1a.

	Justifications	Percentage
Correct	Ratio between black and white results	13.0
	Frequentist estimation of probability	20.1
	Setting up and solving an equation	7.2
	Convergence of sample proportion/mean or LLN	5.0
Partly correct	Basing on experimental results, correct	22.3
Incorrect	Basing on experimental results, incorrect	12.2
	Not able to estimate probability or misapplying classical probability	5.8
	Equiprobability bias	2.2
	Do not justify or confuse justification	12.2

Some of these students stated some additional correct arguments in addition to the four already mentioned, as follows:

- *The number of trials is high enough.* Some participants added the explicit description of the experiment properties, which demonstrated their high knowledge of the mathematical content. One of these properties is the high number of trials, which sufficiently justifies the estimation of probability from the relative frequency, i.e., the application of the frequentist approach. Thus, the understanding of the LLN, which according to Ireland and Watson [24] is the main obstacle to linking the two approaches to probability, was overcome.
P68: Since there is a large number of attempts, I assumed that the results will tend to approximate to the actual probability of getting a white ball or a black ball in each urn.
- *Independence of results in repeated trials,* a property required to apply the frequentist definition of probability [41]. This requirement is fulfilled in the proposed situation because sampling with replacement was used. In the following example, A47 described the two properties mentioned above; although he did not explicitly refer to the classical and frequentist views of probability, he implicitly related them in his answer. P47: Urn A: 3 white and 7 black; Urn B: 5 white and 5 black. Since there were many extractions and because they are independent (the ball is always returned to its place) we have obtained the probability of obtaining a white ball and a black ball in each urn, because the large number of attempts tends to approach to what really happens.

There were fewer incorrect justifications, usually because of not being capable of estimating the theoretical probability using the relative frequency. Those cases of confusing outcomes with favourable or possible cases or showing the equiprobability bias were infrequent (only 8%). Another part of the sample did not justify their response.

4.2. Assigning Probability in New Experiments

The aim of Task 1b, which explicitly asks for the probability of obtaining a result in the next draw, was to analyse whether prospective teachers used the urn-composition model they obtained in Task 1a. The answers were classified as follows:

- *Correct answer.* The subject uses the urn model that he/she has constructed in Task 1a (3 black and 7 white balls in urn A and 5 of each colour in urn B) to assign the probability of obtaining one black ball in each urn in the next draw. Consequently, in urn A he assigns the probability $7/10$ (or its decimal or percentage expression) to obtain a white ball. Similarly in urn B he assigns a probability $\frac{1}{2}$.
- *Partly correct answer,* giving only the probability of getting black balls in one urn, but taking into account the estimated number of balls of each colour in the urn constructed in Task 1a. Basically the answer is similar to the previous one, although not all calculations are completed.

- *Incorrect answer.* The student uses only the results of the 1000 experiments to re-obtain a frequency estimate of the probability, without taking into account that the extraction will be made from an urn with only 10 balls. The construction of the urn involved a modelling process [22], which started from reality (the results observed in the 1000 drawings) and then, simplified this reality to accept certain hypotheses (the total number of balls in each urn is 10; the relative frequency will be close to, but not exactly equal to, to the theoretical probability). The last step in the modelling process is to work with the mathematical model, in this case, to calculate the probability from the assumed composition in the urn. Students who gave this answer built the model, but were unable to use it to answer the new questions.
- *Suggesting that the requested probability cannot be computed or not answering this question.* Therefore, again, the modelling process was not completed and the two approaches to probability were not connected.

In Table 3 the answers to the probability of getting a black ball in each urn (Task 1b) are presented. We observed a reduction in the correct answers when compared to Task 1a. Thus, some of the prospective teachers who were able to correctly estimate the probability from the relative frequency, and hence provided an adequate composition of each urn, in Task 1b did not use the constructed model to calculate the probability of getting the black ball. About half of those who provided a correct composition of the urns now referred only to the results obtained in the 1000 extractions, a response inconsistent with the constructed model. Finally, 18% indicated that it was not possible to calculate the probability or did not answer.

Table 3. Responses given to Task 1b.

	Justification	Percentage
Correct	Using the urn composition	38.1
Partly correct	Only compute one probability	7.9
Incorrect	Do not take into account the urn composition	36.0
	Suggest it is not possible to compute or do not compute	18.0

Table 4 contains the results of Task 2 on the choice of an urn to obtain a white ball in a new selection. The answers given to this task have been classified according to the urn chosen. An overwhelming majority of participants gave the correct answer, which they have obtained either by using the urn model generated in Task 1, or by employing the probability estimation from the relative frequency. However, 15.8% of them still gave the wrong answer or did not reply.

Table 4. Responses given to Task 2.

	Responses	Percentage
Correct	Urn B	84.2
	Urn A	2.9
Incorrect	Any of them	4.3
	No response	8.6

Table 5 presents the arguments used to choose the urn in Task 2, which have been classified according to the criteria specified below:

- *Correct, basing the argument solely on the estimated composition of the urns and the model built in Task 1, while applying the classical probability approach.* The answer is not affected by the last 10 trials.
P102: Given that the urns are the former, and the probability of drawing a white ball in urns A and B was 0.3 and 0.5 respectively, you would choose urn B, as the probability is higher.

- Correct, using the determined composition of the urns, as well as the 10 new results.* For these participants, the argument was also supported by the model of urns constructed, and the participants giving this answer also alluded to the last 10 results, by comparing them with the greater evidence provided by the 1000 drawings which constitute a larger sample size.

P127: As we saw in the first question, the probability of drawing a white ball in urn A is $1/3$ and in urn B, $1/2$. Therefore, there will be a greater chance of drawing a white ball in urn b, regardless of the results obtained in this last table, since only 10 extractions are taken into account now, and 1000 in the previous questions.
- Partially correct, when using the data from the 1000 draws in task 1, but not the composition of the urns.* Although the answer would be correct if the students had not previously constructed the urn models, in this response the two approaches to probability were not completely linked. Thus, instead of considering the theoretical probability $3/10$ and $7/10$ in urn A, the participant was guided by the relative frequency of outcomes in the experiment, not clearly differentiating between relative frequency and probability, which is a problem described by Chaput et al. [22].

P18: Urn A \rightarrow P (white) = 0.324; Urn B \rightarrow P(white) = 0.51. The previous sample (item 1) is larger, more representative. Therefore, I would choose urn B.
- Partially correct, by using the data from the 1010 extractions, but not the composition of the urns.* As above, the subject used the relative frequency, which is an estimate of the theoretical probability, and not the value of the theoretical probability given by the composition of the urn. The difference is that the relative frequency calculation is adjusted by adding the results of the 10 new trials.

P4: In this case, since we use the same previous urns, we already have 10 more outcomes. I would choose urn B, since $P_A(\text{white}) = 329/1010$, and $P_B(\text{white}) = 517/1010$.
- Incorrect.* Participants who relied solely on the 10 results given in Task 2, without taking into account the estimated composition of the urns in Task 1 or the 1000 results provided in Task 1.

P23: I would choose urn B because out of the 10 draws, 7 balls were white, while in urn A only 5 were white. This leads to the conclusion that the probability of drawing a white ball from urn B is higher.
- Other errors, which indicate biases in the participants' reasoning.* Thus, P29 suggested that a decision cannot be made because we use sampling with replacement, which manifests an equiprobability bias [36]. In another example, P124 showed a *positive recency bias* [35], consisting of assuming that the trend of a short series of outcomes will continue; this participant reasoned at the lowest level of understanding independence in Heitele's model [31].

P29: No matter the urn I select, as the balls are returned to the urn.

P124: Urn B, as we see that the while ball is on a run.

Table 5. Justifications of responses given to Task 2.

	Justifications	Percentage
Correct	Basing on the composition of urns (Task 1)	13.7
	Using the urn composition and the last 10 results	17.3
Partly correct	Using the 1000 results only	6.5
	Using the 1010 results	1.4
Incorrect	Using only the last 10 results	41.7
	Other errors or biases	11.5
	Do not justify	7.9

The results in Table 5 show again that the vast majority of correct answers in this task (84.2%) did not correspond to the correctness of the arguments, as only 31% of participants correctly argued their choice. Moreover, the largest number of students relied only on the

last 10 results, without either taking into account the identified composition of the urns or the frequency information of the 1000 results given in Task 1.

5. Discussion and Implications for Teacher Education

To conclude, most of the prospective teachers taking part in the study showed their competence in Task 1a, when estimating the most feasible composition of both urns from the frequentist data of 1000 extractions, and thus found the theoretical probability of obtaining balls of the two given colours. In other words, they went through the first steps of the modelling process [22], by using the data from reality, simplifying the assumptions about reality, and building a feasible mathematical model. In doing so, they were able to move from the frequentist view of probability to the classical view and vice versa, linking the two approaches together. Nevertheless, about 30% of these participants were unable to estimate the urn composition, because they either misinterpreted the sampling variability or demonstrated the equiprobability bias [36], failing to connect the classical and frequentist views of probability in the first question.

Although one third of the sample provided wrong justifications, the reasons supporting this construction by the remaining participants were mostly correct, which reveals the high mathematical knowledge of these participants, who used both proportional reasoning and ideas of convergence, variability, and independence, which are required to connect both approaches to probability [25].

However, only one third of the sample was consistent with the constructed urn model when assigning the probability of getting a black ball in a new draw in Task 1b. This means that only one third demonstrated an ability to work with the previously constructed mathematical model and to decide when it is preferable to use the classical or frequentist approaches to probability. The remaining participants in the study ignored such a model, instead relying solely on the frequency data to assign a probability. In addition, several reasoning biases were observed, such as equiprobability, or the confusion between favourable and possible cases in an experiment with the experimental outcomes.

In the second task, although a large majority correctly selected the urn that provided the highest probability, only one third of the participants were able to adequately argue for their choice; in this task, most participants did not use the urn model constructed previously, as well as the previous frequency information of 1000 trials, and relied only on the last 10 results.

Consequently, the study adds new information on prospective teachers' probabilistic knowledge, in a topic with almost no previous research. The strong mathematical and probabilistic preparation of the participants, and the problems described, suggest that a formal study of probability alone is not enough to establish a complete link between the classical and frequentist views of probability.

Moreover, although the task used was adapted from Sánchez and Valdez [25], their focus was the analysis of students' understanding of the fundamental ideas of variability, randomness, and independence. We analysed the connection between two approaches of probability and performed a deeper analysis of the participants' responses to our questionnaire. The large sample size served to describe a variety of different correct responses, as well as misconceptions that have been extensively described in the previous paragraphs.

Our results also suggest areas for improvement in the training of prospective teachers in the classical and frequentist perspectives of probability and their articulation. Although the results indicate that the prospective teachers in the sample differentiated between both aspects of probability, they neither used them adequately, nor was the relationship between both approaches to probability complete. Such training should emphasise the transition from the classical to the frequentist conceptions and vice versa, by using tasks such as the one proposed in this paper.

The education of teachers could also be supported by technology, by analysing simulations based on the task presented, or on another, such as, for example, the sampling tasks analysed by Batanero et al. [42], in which the opposite task to those used in this

research was provided: knowing the composition of a population or of a random generator, the teachers were asked to generate possible results from samples obtained with that population. Other useful activities are described by Abrahamson [29] in his experiments with undergraduate students, to develop their reasoning about a binomial situation in the context of sampling. These kinds of tasks are complementary, and a full understanding of the classical and frequentist approaches and their articulation should take these types of tasks into account.

Finally, in agreement with authors such as Chaput et al. [22], Eichler and Vogel [43], and Pfannkuch and Ziedins [44], we believe that it is necessary to consider the teaching and learning of probability from a modelling perspective. In this sense, technology and the available tools help to establish the connections between frequentist and classical approaches to probability; this is a promising field for the exposition of probability modelling.

Author Contributions: Conceptualization: C.B., N.B.; data collection and analysis: N.B.; validation: all authors; writing—original draft: all authors. All authors have read and agreed to the published version of the manuscript.

Funding: Research project PID2019-105601GB-I00/AEI/10.13039/501100011033 (Ministerio de Ciencia e Innovación), and research group FQM-126 (Junta de Andalucía, Spain).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available from N. Begué upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Borovcnik, M. Strengthening the role of probability within statistics curricula. In *Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education*; Batanero, C., Burrill, G., Reading, C., Eds.; Springer: Dordrecht, The Netherlands, 2011; pp. 71–83. [\[CrossRef\]](#)
- Gal, I. Towards ‘probability literacy’ for all citizens. In *Exploring Probability in School: Challenges for Teaching and Learning*; Jones, G., Ed.; Springer: New York, NY, USA, 2005; pp. 43–71.
- Batanero, C.; Díaz, C. Training school teachers to teach probability: Reflections and challenges. *Chil. J. Stat.* **2012**, *3*, 3–13.
- Parraguez, R.; Gea, M.M.; Díaz-Levicoy, D.; Batanero, C. Conectan los futuros profesores las aproximaciones frecuencial y clásica de la probabilidad? *Rev. Digit. Matemática Educ. Internet.* **2017**, *17*. [\[CrossRef\]](#)
- Beswick, K.; Goos, M. Mathematics teacher educator knowledge: What do we know and where to from here? *J. Math. Teach. Educ.* **2018**, *21*, 417–427. [\[CrossRef\]](#)
- Llinares, S. Mathematics teacher’s knowledge, knowledge-based reasoning, and contexts. *J. Math. Teach. Educ.* **2018**, *21*, 1–3. [\[CrossRef\]](#)
- Kaiser, G.; König, J. Competence measurement in (mathematics) teacher education and beyond: Implications for policy. *High. Educ. Policy* **2019**, *32*, 597–615. [\[CrossRef\]](#)
- Scheiner, T.; Montes, M.A.; Godino, J.D.; Carrillo, J.; Pino-Fan, L.R. What makes mathematics teacher knowledge specialized? Offering alternative views. *Int. J. Sci. Math. Educ.* **2019**, *17*, 153–172. [\[CrossRef\]](#)
- Batanero, C.; Burrill, G.; Reading, C. *Teaching Statistics in School Mathematics—Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study: The 18th ICMI Study*; Springer: Dordrecht, The Netherlands, 2011. [\[CrossRef\]](#)
- Ball, D.L.; Thames, M.H.; Phelps, G. Content knowledge for teaching: What makes it special? *J. Teach. Educ.* **2008**, *59*, 389–407. [\[CrossRef\]](#)
- Godino, J.D.; Giacomone, B.; Batanero, C.; Font, V. Enfoque ontosemiótico de los conocimientos y competencias del profesor de matemáticas. *Bolema* **2017**, *31*, 90–113. [\[CrossRef\]](#)
- Ministerio de Educación, Cultura y Deporte. *Real Decreto 1105/2014, de 26 de Diciembre, por el que se Establece el Currículo Básico de la Educación Secundaria Obligatoria y del Bachillerato*; MECED: Madrid, Spain, 2015.
- Batanero, C.; Díaz, C. The meaning and understanding of mathematics. In *Philosophical Dimensions in Mathematics Education*; François, K., Ed.; Springer: Dordrecht, The Netherlands, 2007; pp. 107–127. [\[CrossRef\]](#)
- Batanero, C.; Henry, M.; Parzysz, B. The nature of chance and probability. In *Exploring Probability in School*; Jones, G., Ed.; Springer: New York, NY, USA, 2005; pp. 15–37. [\[CrossRef\]](#)
- Borovcnik, M.; Kapadia, R. A historical and philosophical perspective on probability. In *Probabilistic Thinking*; Chernoff, E., Sriraman, B., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 7–34. [\[CrossRef\]](#)
- de Moivre, A. *The Doctrine of Chances*; Chelsea Publishing: New York, NY, USA, 1967; (original work published in 1718).

17. Laplace, P.S. *Théorie Analytique des Probabilités*; Jacques Gabay: Paris, France, 1995; (original work published in 1814).
18. Borovcnik, M.; Bentz, J.; Kapadia, R. A probability perspective. In *Chance Encounters: Probability in Education*; Kapadia, R., Borovcnik, M., Eds.; Kluwer: Dordrecht, The Netherlands, 1991; pp. 27–71.
19. Bernoulli, J. *Ars Conjectandi- 4ème partie*; IREM: Rouen, France, 1987; (original work published in 1713).
20. von Mises, R. *Probabilidad, Estadística y Verdad*; Espasa-Calpe: Barcelona, Spain, 1952; (original work published in 1928).
21. Prodromou, T. Connecting experimental probability and theoretical probability. *ZDM* **2012**, *44*, 855–868. [[CrossRef](#)]
22. Chaput, M.; Girard, J.-C.; Henry, M. Frequentist approach: Modelling and simulation in statistics and probability teaching. In *Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education*; Batanero, C., Burrill, G., Reading, C., Eds.; Springer: Dordrecht, The Netherlands, 2011; pp. 85–95. [[CrossRef](#)]
23. Steinbring, H. The theoretical nature of probability in the classroom. In *Chance Encounters: Probability in Education*; Kapadia, R., Borovcnik, M., Eds.; Kluwer: Dordrecht, The Netherlands, 1991; pp. 135–168.
24. Ireland, S.; Watson, J. Building a connection between experimental and theoretical aspects of probability. *Int. Electron. J. Math. Educ.* **2009**, *4*, 339–370.
25. Sánchez, E.; Valdez, J.C. Las ideas fundamentales de probabilidad en el razonamiento de estudiantes de bachillerato. *Av. Investig. Educ. Matemática* **2017**, *11*, 127–143. [[CrossRef](#)]
26. Hill, H.C.; Ball, D.L.; Schilling, S.G. Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *J. Res. Math. Educ.* **2008**, *39*, 372–400. [[CrossRef](#)]
27. Pratt, D. Making sense of the total of two dice. *J. Res. Math. Educ.* **2000**, *31*, 602–625. [[CrossRef](#)]
28. Konold, C.; Harradine, A.; Kazak, S. Understanding distributions by modeling them. *Int. J. Comput. Math. Learn.* **2007**, *12*, 217–230. [[CrossRef](#)]
29. Abrahamson, D. Embodied design: Constructing means for constructing meaning. *Educ. Stud. Math.* **2009**, *70*, 27–47. [[CrossRef](#)]
30. Konold, C. Designing a data tool for learners. In *Thinking with Data*; Lovett, M., Shah, P., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2007; pp. 267–291. [[CrossRef](#)]
31. Heitele, D. An epistemological view on fundamental stochastic ideas. *Educ. Stud. Math.* **1975**, *6*, 187–205. [[CrossRef](#)]
32. Sánchez, E.; Valdez, J.C. La cuantificación del azar: Una articulación de las definiciones subjetivas, frecuencial y clásica de la probabilidad. In *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria*; Contreras, J.M., Cañadas, G.R., Gea, M.M., Arteaga, P., Eds.; Departamento de Didáctica de la Matemática de la Universidad de Granada: Granada, Spain, 2013; pp. 39–46.
33. Jones, G.; Langrall, C.; Thornton, C.; Mogill, T. A framework for assessing and nurturing young children's thinking in probability. *Educ. Stud. Math.* **1997**, *32*, 101–125. [[CrossRef](#)]
34. Serrano, L. Significados Institucionales y Personales de Objetos Matemáticos Ligados a la Aproximación Frecuencial de la Enseñanza de la Probabilidad. Ph.D. Thesis, University of Granada, Granada, Spain, 1996.
35. Kahneman, D.; Slovic, P.; Tversky, A. (Eds.) *Judgment under Uncertainty: Heuristic and Biases*; Cambridge University Press: New York, NY, USA, 1982.
36. Lecoutre, M.P. Cognitive models and problem spaces in “purely random” situations. *Educ. Stud. Math.* **1992**, *23*, 557–568. [[CrossRef](#)]
37. Muñoz-Rodríguez, L.; Alonso, P.; Rodríguez-Muñoz, L.J.; Valcke, M. Is there a gap in initial secondary mathematics teacher education in Spain compared to other countries? *Rev. Educ.* **2016**, *372*, 106–132. [[CrossRef](#)]
38. Mendenhall, W.; Scheaffer, R.L.; Lyman, R. *Elementos de Muestreo*; Paraninfo: Madrid, Spain, 2006.
39. Hernández-Solís, L.A.; Batanero, C.; Gea, M.M.; Álvarez-Arroyo, R. Comparing probabilities in urns: A study with primary school students. *Uniciencia* **2021**, *35*, 1–19. [[CrossRef](#)]
40. Godino, J.D.; Aké, L.; Gonzato, M.; Wilhelmi, M.R. Niveles de algebrización de la actividad matemática escolar. Implicaciones para la formación de maestros. *Enseñanza Cienc.* **2014**, *32*, 199–219. [[CrossRef](#)]
41. Borovcnik, M. Mutual influence between different views of probability and statistical inference. *Paradigma* **2021**, *XLII*, 221–256. [[CrossRef](#)]
42. Batanero, C.; Begue, N.; Borovcnik, M.; Gea, M.M. Ways in which high-school students understand the sampling distribution for proportions. *Stat. Educ. Res. J.* **2020**, *19*, 32–52. [[CrossRef](#)]
43. Eichler, A.; Vogel, M. Three approaches for modelling situations with randomness. In *Probabilistic Thinking*; Chernoff, E., Sriraman, B., Eds.; Springer: Dordrecht, The Netherlands, 2014. [[CrossRef](#)]
44. Pfannkuch, M.; Ziedins, I. A modelling perspective on probability. In *Probabilistic Thinking*; Chernoff, E., Sriraman, B., Eds.; Springer: Dordrecht, The Netherlands, 2014. [[CrossRef](#)]