PROSPECTIVE TEACHERS' PROBABILISTIC REASONING WHEN SOLVING A SAMPLING TASK

Carmen Batanero[1], Nuria Begué[2], and Juan J. Ortiz[1]
[1]University of Granada, Spain and [2]University of Zaragoza, Spain
batanero@ugr.es

The purpose of this report is to describe the probabilistic reasoning of prospective teachers when working with a sampling situation. We analyze written responses of 139 prospective secondary school mathematics teachers to a task where they had to estimate an urn's composition given a sample of 1000 drawings. Afterwards they had to estimate the probability of the next extraction. Although most teachers built an urn model consistent with the sample selected, few of them used the model to predict the next selection. The study highlights the need to improve the probabilistic reasoning of prospective secondary school mathematics teachers, who, having received formal statistical training, lack an education in the psychological and didactical components of teaching probability.

INTRODUCTION

There is growing interest in the education of mathematics teachers, as is reflected in journals such as the *Journal of Mathematics Teacher Education,* although research in the specific case of statistics and probability is still scarce. This line of research increased after 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) (Batanero et al., 2011), although most of it has focused on prospective or in service primary education teachers.

We focus on prospective secondary and high school mathematics teachers, who have solid mathematical backgrounds as graduates in mathematics, statistics, science, or engineering. However, they only receive formal training in probability, which is contrary to the current curricular recommendations in Spain (Ministerio de Educación, Cultura y Deporte [MECD], 2015), where an empirical approach based on simulations and experiments is recommended in the teaching of probability and sampling.

BACKGROUND

We assume that probabilistic reasoning is based on probabilistic literacy, including understanding of randomness, variability, independence, and predictability/uncertainty according to Gal (2005), as well as capacity to estimate probabilities in random situations. Besides, we consider the following components of probabilistic reasoning described by Borovcnik (2016): understanding and relating the different meanings of probability and equilibrating the psychological and formal components of probability.

Teacher's Mathematical Knowledge

We base our work on the Mathematical Knowledge for Teaching model, which consists of Common Content Knowledge (CCK), Horizon Content Knowledge (HCK), Specialized Content Knowledge (SCK), Knowledge of Content and Teaching, Knowledge of Content and Students, and Knowledge of Content and Curriculum (Hill et al., 2008). CCK is the knowledge brought into play by an educated person to solve mathematical problems, for which a person with basic knowledge is qualified. SCK describes a teacher's special knowledge that enables him/her to plan and develop teaching sequences. HCK refers to the more advanced aspects of the content, e.g., solving tasks that are not commonly found in textbooks.

This paper focuses on prospective high school mathematics teachers' mathematical knowledge of sampling. Specifically, we focus on their ability to estimate the composition of an urn after they are given a sample of extractions with replacement from the urn, and on their prediction of probability in new experiments using the constructed model. Such knowledge is part of relating sampling and estimation, as well as classical and frequentist probability, which appears in the Spanish secondary school curriculum (MECD, 2015) and should be taught to the students. Because tasks such as this are not usually found in textbooks, we evaluate not only part of the teachers' CCK, but also their HCK.

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

Understanding Sampling

Understanding sampling requires linking two apparently opposed ideas: sample representativeness and sample variability (Chance et al., 2004). Sample representativeness implies that a random sample of adequate size from a population will approximately reproduce population characteristics, whereas sample variability implies that different samples will differ. Understanding also requires discriminating among three types of distributions: the theoretical population distribution (e.g., the distribution of black and white balls in the random generator urn); the distribution of sample data collected from the population (e.g., the results of 1000 drawings) and the sampling distribution of statistics from different samples (Batanero et al., 2020). Research has dealt with this issue, asking students to generate samples from a given population (e.g., see a survey in Begué, et al., 2019).

In Batanero et al. (2020) we proposed the opposite task to those used in this research to 235 high school students. That is, knowing the composition of a random generator, participants were asked to generate possible results from samples obtained from that population. The analysis of students' responses indicated a good understanding of the relationship between the theoretical proportion in the population and the sample proportion. However, sampling variability was overestimated in bigger samples. We also observed various types of biased thinking in students: specifically, the equiprobability (Lecoutre, 1992) and recency (Kahneman & Tversky, 1972) biases. The former considers results of any random phenomenon as equally likely, whereas the latter describes a tendency in giving priority to past sample results over information about the population.

In this paper we use the complementary activity, which consists of estimating the composition of a population when the results from a sample are given and for which there is scarce research. In particular, Sánchez and Valdez (2017) studied the inferences made by three groups of 10 high school students in Mexico in variations of the same task. The task 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 greater chance of obtaining a given colour in the next extraction, or to predict the colour of the next extraction, depending on the group of students. From the analysis of students' responses, the authors proposed a hierarchy of levels of understanding randomness, variability, and independence, as follows:

- *Randomness*: (L1) making deterministic predictions; (L2) making deterministic predictions qualified with probabilistic language; (L3) recognising that the outcome cannot be predicted with accuracy; (L4) recognising the stability of the frequency in the long run even though the outcome cannot be predicted with accuracy.
- *Variability*: (L1) not considered; (L2) thinking that differences between specified and observed frequencies are always significant regardless of sample size; (L3) considering a difference to be significant in a small sample but not in a large sample; (L4) understanding the relationship of variability with sample size.
- *Independence*: (L1) thinking that successive results depend on previous outcomes; (L2) thinking the result depends on whether the sample is representative; (L3) using models to determine a possible result; (L4) recognizing independence.

In other research, Sánchez and Valdez (2013) analyzed the way in which a group of Mexican high school students understood the Law of Large numbers and related the frequentist and classical views of probability. Using interviews, the researchers studied students' responses to physical and computational simulation tasks, describing the subjective, transitional, informal quantitative, and numerical levels of reasoning. 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. In this paper we use a variation of the tasks used by Sánchez and Valdez (2013, 2017) to analyze prospective teachers' reasoning in the context of sampling; more specifically, to evaluate if the prospective teachers are able to build an urn model consistent with results from a sample and whether they use the model to compute the probability of a given result in the next trial. The variation consists in requesting that participants justify their responses.

METHOD

A sample of 139 students preparing to become mathematics teachers in secondary education (ages 12–17) took part in the study as a part of a workshop directed to improve their knowledge of probability education. 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 students were following a masters' course with pedagogy, didactic, and curricular contents that complemented their previous university studies. These prospective teachers were given the task reproduced in Figure 1, adapted from Sánchez and Valdez (2017), to be solved individually. After their written responses were collected, we performed a content analysis (Krippendorff, 2013). A priori categories were developed from Sánchez and Valdez (2017) and were adapted and expanded when analyzing the responses. To assess the reliability of the procedure, another researcher reviewed all coding, and disagreements were discussed until consensus was obtained.

Two urns A and B contain 10 balls each, some of which are black and the others white. We make 1000 extractions from each urn, each time returning the ball to the urn after it has 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.
Question 1. How many white and black balls do you think there are in urn A? And in urn B? Justify your answer.
Question 2. 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?

Figure 1. Task given to prospective teachers

In Question 1, participants were asked to estimate the number of white (w) and black (b) balls in each urn, using the results given from a sample of 1000 extractions from each urn. Because the sample size and the results are independent, the probability of each colour must be estimated from the relative frequency. For the first urn, the values are 0.324 (w) and 0.676 (b), and the values are 0.510 (w) and 0.490 (b) for the second urn. The expected value of the number is given by the product of the relative frequency of each colour and the number of balls in the urns, or 3.24 and 6.76 for the first urn and 5.1 and 4.9 for the second. Because the number of balls is a natural number, rounding provides the best estimation for numbers of black and white balls. After the most probable composition of the urns has been determined, participants should apply the classical definition to answer the second question. Thus, the probabilities in the first urn are 0.7 (b) and 0.3 (w), and 0.5 (b and w) in the second urn.

RESULTS

Estimating the Urns' Composition

Responses to Question 1 about the estimation of the number of black and white balls in the two urns have been classified as follows:

- *Correct response*, suggesting there are 3 white and 7 black balls in the first urn and 5 balls of each colour in the second urn. This response involves relating the sample to the population and being able to estimate the parameter (proportion of black balls) in the population, when the statistics summary (proportion of black balls in the sample) is known.
- *Providing an interval of values*, such as, for example, "there are between 3 and 4 white balls in box A" (P18). This answer is incorrect because 325 is a very unlikely value in the assumption that the number of white balls is 4, as can be seen with the value $Z = -4.9$ (highly significant) in the Z test to check the assumption of obtaining 325 balls or less for the urn with 4 white balls.
- Just suggesting that there are *fewer white balls than black balls* in urn A, thus not providing an exact number of balls and therefore failing to relate the results in the sample to the population from which the sample was drawn.
- Replying that the number of balls in each urn cannot be known because it is a random situation, and anything is possible.

Results are provided in Table 1. Most prospective teachers solved the tasks successfully and were able to relate the urn composition with the results of sampling, using their CCK of sampling. However, we observed errors in an important proportion of participants, suggesting lack of CCK:

5.8% of them misinterpreted variability in sampling by suggesting an interval including a highly unlikely urn composition. Further, 10.8% of them stated that any urn composition was possible, thus demonstrating the equiprobability bias (Lecoutre, 1992); 5% replied that there were more black marbles; and another 6.5% provided no response.

Table 1. Proposed composition of urns in Question 1

Number of white and black balls	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 balls	5
Anything is possible	10.8
No response	6.5

In the following paragraphs, the justifications given for the composition of the urns are analysed, using the prospective teachers' HCK of sampling. These justifications were classified as:

- *Correct justification*: The participant justified that the requested probability was estimated 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.
- *Correct justification*: The requested probability was estimated by computing the relative frequency or percentage of balls of each colour in a sample of 1000 trials, 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.
- *Correct justification*: 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. Finally, they rounded the results.
- *Correct justification*: The participant based their justification on the convergence of the proportion or of the sample mean to the population proportion or by quoting the law of large numbers.
- *Incorrect response*: The participant simply based their justification on the observed results without being able to estimate the probability. These students did not provide a specific composition for the urns because they could not relate the relative frequency in the sample of 1000 experiments to an estimate of the number of black and white balls in the urns.
- *Incorrect application of classical probability*: The participant interpreted the experiment outcomes as favorable and possible cases.
- *Equiprobability bias*: This biased reasoning arose when the participant suggested that any composition was possible. Because we dealt with a random experiment, any outcome had the same probability.

Table 2. Justifications given for Question 1

	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
	Misapplying classical probability	5.8
	Equiprobability bias	2.2
	Do not justify or confuse justification	12.2

Table 2 lists the justifications given by the participants to Question 1, most of which were correct (45.2%). These are supplemented by 22.3% of participants who gave partially correct answers. These prospective teachers related estimation and sampling, as well as classical and frequentist probability, and were also aware of the properties that this relationship enables. Therefore, they reached the top level (4) of understanding variability in Sánchez and Valdez's (2017) hierarchy. There were fewer incorrect justifications, usually because of not being capable of estimating the population proportion using the sample proportion. 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.

Finding the Probability of a Black Ball in the Next Trial

In Question 2 students were asked for the probability of obtaining a black ball in the next draw, in order to analyse whether prospective teachers used the urn-composition model they obtained in Question 1. The answers were classified as follows:

- *Correct answer*: The participant used the urn model that he/she has constructed in Question 1. Consequently, in the urn A he or she assigned the probability 7/10 and in urn B the probability ½.
- *Partly correct answer:* The participant gives the probability of getting black balls in one urn without considering the number estimated in response to Question 1.
- *Incorrect answer*: The student uses only the results of the 1000 experiments to re-obtain a frequentist estimate of the probability, without taking into account that the extraction will be made from an urn with only 10 balls. Students who gave this answer built the model but were unable to use it to answer the new questions.
- *The participant replies that the requested probability cannot be computed or does not answer.*

In Table 3 the answers to the probability of getting a black ball in each urn (Question 2) are presented. We observe a reduction in the correct answers when compared to Question 1. Thus, some of the prospective teachers who provided an adequate composition of each urn in Question 1 did not use the constructed model to calculate the probability of getting the black ball and then failed in their HCK. About half of those who provided a correct composition of the urns now referred only to the sample of 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 Question 2

	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

IMPLICATION FOR TEACHER EDUCATION

The results showed that most participants built an urn model adequate for the data provided, thus relating the results of sampling to the urn composition and showing understanding of randomness, independence, and predictability/uncertainty, which are fundamental to understand sampling according to Sánchez and Valdez (2017). However, the remaining participants provided an improbable composition of the urn, misinterpreting sampling variability, while others showed the equiprobability bias (Lecoutre, 1992) and failed to connect the results of sampling with the random generator producing the sample, which is part of CCK. This indirectly also suggests that these prospective teachers were unable to link the classical and frequentist meanings of probability. Thus, they did not control the subjective and formal components of the situation, which is a component of probabilistic reasoning according to Borovcnik (2016).

In the second part of the task, few prospective teachers used the urn model constructed to estimate probability in the next drawing, and others believed that this probability could not be estimated. This result suggests that the prospective teachers were not confident in their responses to Question 1 and also highlights the difficulties of these participants with modelling in probability which is part of their HCK. The construction of the urn involved a modelling process (Chaput et al., 2011), 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. All these results were discussed with the prospective teachers, who were given the chance to identify their faulty reasoning and main probabilistic biases. The participants found the activity to be helpful to recognize the challenging task to teach probability to young students. The study implication is the need to improve the probabilistic reasoning of prospective secondary school mathematics teachers, who, having received a formal statistical training, lack an education in the psychological and didactical components of teaching probability. The value of this research is providing information about components to be reinforced in the education of prospective teachers. The paper also contributes to the analysis of probabilistic reasoning and its components.

ACKNOWLEDGEMENT*:* Research project PID2019-105601GB-I00/AEI/10.13039/501100011033 and research group FQM-126 (Junta de Andalucía, Spain).

REFERENCES

Batanero, C., Begué, N., Borovcnik, M., & Gea, M. M. (2020). Ways in which high-school students understand the sampling distribution for proportions. *Statistics Education Research Journal, 19*(3), 32–52. https://doi.org/10.52041/serj.v19i3.55

Batanero, C., Burrill, G., & Reading, C. (Eds.), (2011). *Teaching statistics in school mathematics-challenges for teaching and teacher education: A joint ICMI/IASE study.* Springer. https://doi.org/10.1007/978-94-007-1131-0_10

Begué, N., Batanero, C., Ruiz, K., & Gea, M. M. (2019). Understanding sampling: A summary of the research. *Boletín de Estadística e Investigación Operativa–BEIO, 35*(1), 49-78.

Borovcnik, M. (2016). Probabilistic thinking and probability literacy in the context of risk. *Educação Matemática Pesquisa, 18*(3), 1491-1516. https://revistas.pucsp.br/index.php/emp/article/view/31495/21953

Chance, B., del Mas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. B. Garfield (Eds), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295–323). Springer. https://doi.org/10.1007/1-4020-2278-6

Chaput, M., Girard, J.-C., & Henry, M. (2011). Frequentist approach: Modelling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education* (pp. 85–95). Springer. https://doi.org/10.1007/978-94-007-1131-0_12.

Gal, I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring probability in school* (pp. 39-63). Springer. https://doi.org/10.1007/0-387-24530-8_3

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*, 372–400. https://doi.org/10.5951/jresematheduc.39.4.0372.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Sage.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics, 23*, 557–568. https://doi.org/10.1007/BF00540060

Ministerio de Educación, Cultura y Deporte (MECD). (2015). *Real decreto 1105/2014, de 26 de diciembre, por el que se Establece el Currículo Básico de la Educación Secundaria Obligatoria y del Bachillerato*. https://www.boe.es/eli/es/rd/2014/12/26/1105/con

Sánchez, E., & Valdez, J. C. (2013). La cuantificación del azar: Una articulación de las definiciones subjetivas, frecuencial y clásica de la probabilidad. In J. M. Contreras, G. R. Cañadas, M. M. Gea, y P. Arteaga (Eds.), *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp.39-46). Departamento de Didáctica de la Matemática de la Universidad de Granada. http://www.estadis.net/3/documentos/ACTAS/1%20Ponencia%204.pdf

Sánchez, E., & Valdez, J. C. (2017). Las ideas fundamentales de probabilidad en el razonamiento de estudiantes de bachillerato. *Avances de Investigación en Educación Matemática, 11*, 127–143. https://doi.org/10.35763/aiem.v1i11.180