

DEVELOPING A QUESTIONNAIRE TO ASSESS THE PROBABILITY CONTENT KNOWLEDGE OF PROSPECTIVE PRIMARY SCHOOL TEACHERS

EMILSE GÓMEZ-TORRES
Universidad Nacional de Colombia
egomez@unal.edu.co

CARMEN BATANERO
Universidad de Granada, Spain
batanero@ugr.es

CARMEN DÍAZ
Universidad de Huelva, Spain
carmen.diaz@dpsi.uhu.es

JOSÉ MIGUEL CONTRERAS
Universidad de Granada, Spain
jmcontreras@ugr.es

ABSTRACT

In this paper we describe the development of a questionnaire designed to assess the probability content knowledge of prospective primary school teachers. Three components of mathematical knowledge for teaching and three different meanings of probability (classical, frequentist and subjective) are considered. The questionnaire content is based on curricular guidelines and primary school textbooks in Spain. The items were selected and adapted, after expert judgment, from previous research. The responses of 157 prospective primary school teachers were used to analyze the psychometric properties of the questionnaire and to provide information about various aspects of participants' probability content knowledge.

Keywords: *Statistics education research; Probability content knowledge; Meanings of probability; Prospective primary school teachers; Assessment*

1. INTRODUCTION

The topic of probability is part of the primary school mathematics curriculum in many countries, including Spain (Ministerio de Educación, Cultura y Deporte [MECD], 2014). Given that this topic was introduced only recently at the primary school level in Spain, many of today's prospective primary school teachers did not themselves study probability in primary school and may have forgotten the relevant ideas that they did learn in secondary school. Accordingly, courses and workshops designed for prospective teachers should take into account prior assessments of their knowledge obtained through adequate (valid and reliable) instruments.

The aim of our research is to contribute to this need by developing a valid and reliable

questionnaire that can be used to assess some components of probability content knowledge required for teaching. The questionnaire has been designed to take into account different meanings of probability, as well as the probability concepts that are referred to in curricular guidelines for primary school mathematics (e.g., National Council of Teachers of Mathematics [NCTM], 2000; MECD, 2014) and in primary school textbooks.

In this paper, we describe the process that was used to design the questionnaire, as well as the process used to analyse its psychometric properties using the responses from a sample of 157 prospective primary school teachers in Spain. We also analyze the performance of the participants with respect to various aspects of probability content knowledge for teaching and different meanings of probability. The questionnaire is included in Appendix 1.

2. BACKGROUND

Although research concerning teachers' knowledge for teaching mathematics is abundant, research related specifically to teachers' understanding of probability is scarce (Callingham, & Watson, 2011). Prior research suggests the existence of difficulties or misconceptions in probability, especially among primary school teachers (both prospective and in-service), and has mainly focused on teachers' knowledge of specific concepts. Our research complements these studies by introducing a comprehensive questionnaire that can be used to assess several components of teachers' probabilistic knowledge for teaching in primary school. The assessment results may also serve to organize actions directed at improving teachers' probabilistic knowledge.

2.1. PROBABILISTIC KNOWLEDGE FOR TEACHING

After Shulman's (1987) seminal work, many authors have analysed the nature of knowledge needed by teachers to achieve effective teaching outcomes. Ball and her colleagues (Ball, Lubienski, & Mewborn, 2001; Ball, Thames, & Phelps, 2008) developed the notion of "mathematical knowledge for teaching" (MKT), distinguishing different categories of such knowledge. Three of these are relevant to our study:

- *Common content knowledge* (CCK), according to Ball, Thames and Phelps (2008, p. 399), is "the mathematical knowledge and skill used in settings other than teaching." In our interpretation, this type of knowledge includes basic skills and general knowledge about the topics that are to be taught to students.
- *Specialized content knowledge* (SCK) is "the mathematical knowledge that allows teachers to engage in particular teaching tasks" (Hill, Ball, & Schilling, 2008, pp. 377-378). The authors suggest that this type of knowledge supports the teacher in representing mathematical knowledge, providing explanations, and understanding students' solutions to problems.
- *Knowledge in the mathematical horizon* is described in this paper as *advanced content knowledge* (ACK). "Horizon knowledge is an awareness of how mathematical topics are related over the span of mathematics included in the curriculum. First grade teachers, for example, may need to know how the mathematics they teach is related to the mathematics students will learn in third grade to be able to set the mathematical foundation for what will come later" (Ball et al., 2008, p. 403). We interpret this as knowledge that is broader than what the teacher explicitly teaches; it involves, for example, knowledge of the content that students will encounter in future courses.

The questionnaire developed as part of this research assesses knowledge about

probability related to these three categories of content knowledge, according to our interpretation of these categories. It also incorporates three different meanings of probability that appear in Spanish curricular guidelines and textbooks.

The questionnaire was administered to participants in a mathematics education course, over two 1-hour sessions (participants responded to six items in each session).

2.2. MEANINGS OF PROBABILITY IN THE PRIMARY SCHOOL CURRICULA

According to Batanero (2013), philosophical controversies surrounding the nature of probability still influence school curricula. In a previous analysis of school textbooks (Gómez & Contreras, 2014), we identified the following meanings of probability (Batanero, Henry, & Parzys, 2005), all of which are taken into account in our questionnaire:

- *Classical meaning.* Probability is conceived of as a ratio of the number of favourable outcomes to the total number of possible outcomes. This approach was popular in teaching until the 1980s. The main application is in computing probabilities for single events in games of chance that are familiar to children. When dealing with compound events, however, children need to apply combinatorial reasoning, which is a difficult task for many children; moreover, it is difficult to apply this conception of probability outside the context of games of chance.
- *Frequentist meaning.* Probability is defined as the limit of the relative frequencies of an event when an experiment is repeated a large number of times. This approach is common in teaching today, given the availability of computers and simulators that can reproduce random experiments quickly and thus easily illustrate the effect of sample size and demonstrate stochastic convergence (Batanero, Henry, & Parzys, 2005). A problem with this approach is that some random situations are not readily reproducible through computer simulation.
- *Subjective meaning.* According to this definition, probability is a personal degree of belief and can be updated via Bayes' theorem when new information is obtained. Godino et al (1987) suggest that it is possible to introduce the subjective meaning of probability at primary school in an intuitive way: The approach involves assigning qualitative probabilities to events or locating these events on a probability scale, and later revising these probabilities after new experiences or through the use of conditional probabilities.

3. METHOD

3.1. STEPS IN DEVELOPING THE QUESTIONNAIRE

In developing a questionnaire to assess the probability content knowledge of prospective primary school teachers, we took into account the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (1999) methodological suggestions. Consequently, we followed four steps:

1. *Semantic definition:* As is the case with many educational or psychological traits variables, “probability content knowledge for teaching” is not directly observable. We must therefore infer this knowledge from empirical indicators (in this case, the subjects' responses to the items in the questionnaire; McDonald, 2013). When developing a questionnaire, the first step is to carefully define and describe the aspects of knowledge

that are to be assessed. This specification should be grounded in adequate theoretical background. In our study, it was based on prior analysis of Spanish curricular guidelines and textbooks at the primary school level (Gómez & Contreras, 2014).

2. *Components of teacher knowledge:* We considered three types of content knowledge related to mathematical knowledge for teaching (Ball et al., 2008). To assess *common content knowledge*, we included items that evaluate knowledge of probability concepts included in the primary school curriculum (identified in Gómez & Contreras, 2014). *Advanced content knowledge* is assessed by additional items, whose solutions require a more advanced knowledge of probability (in comparison to the probability content considered in primary school). This advanced knowledge is expected to have been acquired by prospective teachers during their middle and secondary school education. Some items also assess certain common probability reasoning biases in order to make the prospective teachers aware that these biases may be held by their future students. For both common and advanced content knowledge, we fixed a list of the concepts we wanted to evaluate in the questionnaire (see Table 1, column 1). Finally, to assess *specialized content knowledge*, we asked the participants to describe how they would justify their solutions of eight of the items to their students.

Table 1. Components of teacher's knowledge and probability content assessed in questionnaire

	Meaning of probability	Item	Included in primary school	
			Curricular guidelines	Textbooks
<i>Common content knowledge</i>				
Listing (counting) favourable / possible outcomes	Classical	1	x	x
Comparing probabilities	Classical	2	x	x
Joint probability (product rule): Independent experiments	Classical	3		x
Joint probability (product rule): Dependent experiments	Subjective	4		x
Estimating probability	Frequentist	5	x	x
Fair game	Classical	6a		x
<i>Advanced content knowledge</i>				
Expectation	Classical	6b		
Conditional probability	Subjective	7		
Equiprobability bias	Classical	8		
Sampling	Frequentist	9		
Perception of randomness	Frequentist/ Subjective	10		
Representativeness heuristic	Frequentist /Subjective	11a		
Binomial distribution	Frequentist/ Subjective	11b		
Outcome approach: Prediction	Frequentist	12a		
Outcome approach: Validity of prediction	Frequentist /Subjective	12b,c		
<i>Specialized content knowledge</i>		Arguments in items 2, 3, 4, 6a, 10, 11a, 11b, 12b		

3. *Item selection:* To identify suitable items for the questionnaire that assessed the intended probability content, we surveyed prior research on children's and adolescents'

probabilistic reasoning. A preliminary pool of items was created by translating and adapting items from different researchers. Through an iterative process, each item was analysed several times and compared to the intended content with the help of several colleagues and researchers who volunteered to provide feedback regarding the suitability of potential items. During each revision, some items were discarded, and others were adapted to align more closely with the intended content. As we wanted to limit the length of the final questionnaire (so that participants could complete it within a reasonable amount of time), we finally selected 12 items that assessed the intended content (see Table 1).

4. *Item format and wording*: Once the items were fixed, their format and wording were revised. We decided to include both multiple choice and open-ended questions, as we were interested not only in the accuracy of the responses, but also in the strategies and arguments used to arrive at the solutions. The final questionnaire is included in Appendix 1. The items were translated or adapted from Azcárate (1995), Batanero, Garfield, and Serrano (1996), Chernoff (2011), Díaz and Batanero (2009), Falk and Wilkening (1998), Fischbein and Gazit (1984), Green (1982, 1983), and Shaughnessy and Ciancetta (2002).

As Table 1 shows, the questionnaire assesses three components of teachers' knowledge. It incorporates the main probability concepts identified in our previous analysis of the primary school curriculum (Gómez & Contreras, 2014), and is also designed to evaluate some common probabilistic reasoning biases. Furthermore, we included items that take into account three different meanings of probability (classical, subjective, and frequentist). Although some of the problems may be solved by applying different meanings of probability, in our analysis we consider the main meaning we expect the students to apply (see Table 1, column 2). For example, a correct solution of Item 12a, designed to evaluate use of the outcome approach, requires the application of the frequentist meaning of probability; however, some students may also use the subjective meaning of probability in their responses.

3.2. CONTENT VALIDITY

The questionnaire was designed such that it assesses a wide range of content (three categories of teachers' knowledge; three meanings of probability; basic probability content in primary school; some potential reasoning biases). The construct of probability knowledge for teaching was defined through a careful description of its dimensions and facets, which contributes to the content validity of the questionnaire (content validity refers to the extent to which a measure represents all facets of a given construct; McDonald, 2013). Additional evidence of content validity is obtained by comparing the assessed content (summarized in Table 1) with the probability content in the primary school curricula (identified in Gómez & Contreras, 2014). Furthermore, several expert researchers in probability education contributed to the item selection process by analysing to what degree each item contributed to measuring the underlying construct. This external support helped to reinforce the content validity of the questionnaire.

3.3. SAMPLE AND ANALYSES

The pilot sample consisted of 157 prospective primary school teachers enrolled in a mathematics education course during their second year of undergraduate studies at the Faculty of Education in the University of Granada, Spain. All of them had studied simple

and conditional probability during the previous academic year. These students should have also studied probability in middle and secondary school, where the curriculum includes the concepts of sample space, probability distributions, expected value of random variables, conditional and compound probability, total probability and Bayes' theorem, and binomial and normal distributions.

In the following section, we summarize the two main psychometric properties of the questionnaire: the difficulty and the discrimination of the items. Then, we analyse global performance on the questionnaire as related to the three components of teachers' knowledge, and, finally, discuss the reliability and generalizability of the questionnaire.

4. ITEM ANALYSIS

In Table 2, we present the difficulty (as measured by the proportion of correct answers) and discrimination of the items, along with information related to the content assessed by each item, the implicit meaning of probability, and the relevant component of teacher knowledge.

The table shows a wide range of item difficulty (ranging from 11% to 94% of correct responses). To evaluate the items' discrimination (the extent to which the items discriminate between students with high and low levels of knowledge), the sample was divided into three groups (lower, middle, and upper scores) according to their relative performance on the entire questionnaire. We compared the proportion of participants in the upper and lower groups (each having 50 participants) that provided correct responses to each item using a test of difference in proportions in independent samples. The high discriminant power of most items on the questionnaire is evident. In the next section, we analyse the items' difficulty in relation to the three different categories of teacher knowledge (common, advanced, and specialized) and the main meaning of probability implicit in each item.

Table 2. Item difficulty index by component of knowledge and probability meaning

Teacher knowledge	Meaning of Probability	Item	Item difficulty	Discrimination ¹
Common Knowledge	Classical	1a. Enumeration of sample space (three elements)	0.94	0.10*
		1b. Enumeration of sample space (four elements)	0.59	0.36**
		1b. Strategy	0.21	0.36*
	Subjective	2. Comparing probabilities	0.76	0.24**
		3. Joint probability. Independent experiments	0.57	0.48**
		4a. Dependent compound experiment: Most likely result	0.50	0.56**
		4b. Dependent compound experiment: Least likely result	0.55	0.52**
	Frequentist	4c. Consistency between response to 4a and 4b	0.39	0.50**
		5a. Estimating probability	0.54	0.36**
		5b. Sampling variability	0.45	0.28**
Advanced Knowledge	Classical	6a. Fair game	0.90	0.26**
	Classical	6b. Expectation	0.50	0.70**
		6b. Strategy	0.39	0.72**
	Subjective	7. Conditional probability	0.23	0.20*
		7. Strategy	0.15	0.20**
	Classical	8 ² . Independent compound experiment: Equiprobability bias	0.27	0.08
	Frequentist	9a. Sampling: Total estimation	0.20	0.40**

		9a. Strategy	0.15	0.30**
		9b. Re-sampling: Predicting a second sample	0.54	0.64**
		9b. Strategy	0.35	0.50**
	Frequentist/ Subjective	10. Perception of randomness	0.27	0.16
		11a. Representativeness heuristics: Insensibility to sample size	0.20	0.46**
		11b. Binomial distribution	0.39	0.24*
	Frequentist	12a. Outcome approach: Predicting an average	0.87	0.22**
	Frequentist/ Subjective	12b. Outcome approach: Validity of forecast with one contradictory observation	0.85	0.32**
		12c. Outcome approach: Validity of forecast with two contradictory observation	0.28	-0.08
Specialized Knowledge	Classical	2. Comparison of probabilities (Justification)	0.74	0.22**
		3. Joint probability: Independent experiments (Justification)	0.25	0.28**
	Subjective	4. Joint probability: Dependent experiments (Justification)	0.21	0.30**
		6a. Fair game (Justification)	0.89	0.72**
	Frequentist/ Subjective	10. Perception of randomness (Justification)	0.23	0.20*
		11a. Representativeness heuristic (Justification)	0.17	0.40**
		11b. Representativeness heuristic: Binomial distribution (Justification)	0.13	0.16*
		12b. Outcome approach: Validity of forecast (Justification)	0.11	0.06

¹Difference in proportion of correct responses between the groups with upper and lower scores on the entire questionnaire; (*) $p < 0.05$ (significant difference in proportions); (**) $p < 0.01$ (highly significant difference in proportions).

²The answers to questions 8a and 8b were analysed conjointly.

4.1. COMMON CONTENT KNOWLEDGE

Classical meaning The degree of common content knowledge related to the classical meaning of probability was acceptable on many items in our sample, which agrees with results observed in prior research involving Spanish primary school prospective teachers (e.g., Azcárate, 1995; Mohamed, 2012). Enumerating a sample space with a small number of elements (Items 1a and 1b) and comparing probabilities of events involving drawing balls from urns (Item 2) were easy tasks for the majority of participants. More than half of the participants correctly computed the joint probability in independent experiments (Item 3), and 90% of them correctly identified an unfair game (Item 6a).

However, some difficulties with the classical meaning of probability were also apparent in the results. Although 60% of the prospective teachers gave the correct number of permutations of four elements, few of them used systematic enumeration or the permutations formula (21% in total); the remaining participants used non-systematic enumeration strategies to obtain the solution. Although most solutions to Item 2 were correct, only a few participants were able to propose urns with proportional composition (having the same probability but different absolute frequencies of colours); though two different correct responses were possible for this item (identical composition of urns and proportional composition), very few participants proposed the second solution.

Frequentist meaning We only included one item (Item 5), adapted from Green (1983), to evaluate common content knowledge related to the frequentist meaning of probability. In his study, Green (1983) only asked the students to provide an estimation of the relative frequencies of pins landing “up” and “down” in the next sample of 100 drawing pins. Our

task is more difficult, because we asked for four different estimates in the next four samples. With this change, we can analyse both the frequencies of the up and down results (Item 5a) and the variability of the estimates in the four samples (Item 5b).

We computed the average number of pins landing “up” using the four values provided by each participant and considered a correct response to be one where the distance from the expected value in the binomial distribution was smaller than two standard deviations: that is, when the average value of the participant’s estimates was not far from the theoretical value of 68 pins landing “up” provided in the problem statement. Most participants demonstrated a correct perception of the expected number of pins landing up (Item 5a); moreover, almost 50% of participants provided values with an acceptable variability, as compared with what is theoretically expected in samples of 100 elements (item 5b).

However, some difficulties were also apparent in the responses to this item. For example, some of the participants provided equiprobable estimates (close to 50% of pins landing up), a result that may suggest misunderstanding of the frequentist meaning of probability or provide evidence for the equiprobability bias (Lecoutre, 1992). Other students tried to compensate for the teacher’s result (and, consequently, provided values close to 32 pins landing “up”) or repeated the same value four times. Nevertheless, the participants in our study performed better on this item than the Spanish primary school prospective teachers in Mohamed (2012), who used the original prompt from Green (1983).

Subjective meaning About half of the participants correctly solved a joint probability problem in the context of dependent experiments, where it was necessary to first compute a conditional probability (Items 4a and 4b). They provided correct responses to the first two questions in this item. The participants in this study outperformed those in previous research, which posed the same question to 10-15 year-old children (Green, 1982; Cañizares, 1997). For example, Green found 60% of children answering that all of the possible results were equiprobable, as compared to only 8% of our prospective teachers.

However, when comparing responses in parts a) and b) of this item, we observed little consistency (only 39% of the participants provided consistent responses). These results (Item 4c) suggest the need to reinforce the subjective meaning of probability in the preparation of prospective teachers.

4.2. ADVANCED CONTENT KNOWLEDGE

Classical meaning In agreement with previous research involving prospective teachers (Azcárate, 1995; Mohamed, 2012), only half of the participants in this sample were able to establish the price that a player should pay to transform an unfair game into a fair one (Item 6b). Those who failed to do so either showed little understanding of the ideas of expectation or were unable to properly compute the expected price for each of the players. There was a large difference between the percentage of correct responses to the first part of the item (Item 6a, with a correct response given by 90% of participants), where many participants correctly computed the probabilities of winning for each player and concluded that the game was unfair, and the percentage of correct responses to the second part of the item (Item 6b). Only 39% of the strategies used to establish the price for each player were correct (involving proportional reasoning or balancing the expectations for both players). This percentage is higher than that the percentage reported in Mohamed (2012), where only 11% of prospective teachers used correct strategies on a similar item.

Moreover, only 27% of prospective teachers correctly solved the compound probability problem in Item 8, and many of them reasoned according to the equiprobability bias (Lecoutre, 1992), considering all of the results as being equally likely. To conclude, these

two items suggest that advanced content knowledge related to the classical meaning of probability was weak in this sample of prospective teachers and needs to be improved if probability is to be taught effectively in primary school.

Frequentist meaning Our results also suggest poor advanced content knowledge related to the frequentist meaning of probability among the participating prospective teachers. Only 20% of participants were able to estimate the size of a population in a capture-recapture sampling process (Item 9a), with results close to those of 11-year-old children (23% correct responses) reported in a study by Fischbein and Gazit (1984) and worse than the results (reported in the same study) of 12-year-old children (44% correct responses). Only 15% of participants correctly used proportional reasoning to solve this part of the item; others used additive reasoning, or guessed the population size at random. When participants were asked to predict the proportion in a second sample (Item 9b), results improved, with 54% of participants providing a correct response. This suggests that for these prospective teachers, it was easier to predict results in a replication of sampling (Item 9b) than to use the sample information to estimate the population size (Item 9a). Even so, the strategies used to solve this second part of the problem remained intuitive, with only 35% of participants correctly using proportional reasoning. Others used incorrect proportions or additive strategies, and some simply tried to guess the result randomly.

Results were also poor on an item related to the perception of randomness (Item 10, which presents a comparative likelihood task; adapted from Chernoff, 2011). Only 27% of participants chose the correct response when they were asked to distinguish between a random and a non-random sequence of outcomes. On the contrary, on Item 12a (related to the outcome approach [Konold, 1989]), where the prospective teachers were asked to predict the average number of rainy days using a frequentist estimate for the probability of rain, 87% of participants provided a satisfactory response.

Subjective meaning Once again, performance on items related to the advanced meaning of subjective probability was weak. Only 23 % of the participants provided a correct solution to Item 7, where they needed to compute a conditional probability in a compound dependent experiment. Many participants computed other probabilities; for example, some confused the joint and the conditional probability, or computed the simple probability that the number on one dice was 6. The multiple difficulties in understanding a statement related to conditional probability, described by Contreras (2011), were apparent in the responses to this item. Most participants tried to solve this problem using enumeration (instead of developing the definition of conditional probability), and their strategies failed because of incomplete enumeration or because the order of elements in the sample space was not taken in to account. Only 15% of participants used correct strategies in responding to this item.

With respect to subjective heuristics, only 20% of participants correctly solved Item 11a, and the majority reasoned according to the representativeness heuristic (Tversky & Kahneman, 1974). Results improved slightly (Item 11b) when participants were asked to compare different outcomes in a binomial distribution (39% correct responses), but were again poor (28% correct responses) when they were asked to explain two contradictory results on an item designed to evaluate the outcome approach (Item 12c).

To conclude, though the prospective primary teachers in this study demonstrated a reasonable degree of common probability content knowledge, our results suggest that they possess scarce advanced knowledge related to topics that should have been acquired during middle or secondary school. This suggests in turn that such knowledge needs to be improved during primary teacher education.

4.3. SPECIALIZED CONTENT KNOWLEDGE

We used the arguments provided in response to some items on the questionnaire to evaluate the prospective teachers' specialized content knowledge, as the task of explaining problem solutions to students is given as an example of specialized content knowledge by Hill et al. (2008). Below, we describe the results and offer an analysis of these arguments.

Classical meaning The prospective teachers showed an acceptable level of specialized content knowledge related to this meaning of probability, as they were generally able to explain their solutions clearly and correctly. Most participants were able to explain why two probabilities were equal or unequal in response to Item 2 and why a change in the composition of the urns affects the probability of drawing each colour. Justifying why a given game is fair or unfair was also very easy for the majority of the participants; the percentage of correct justifications was almost identical to the percentage of correct responses. The correct arguments were based on probability, the proportion of favourable cases, or enumeration of cases. Even so, some prospective teachers tended to make decisions based on the difference between the number of favourable and unfavourable outcomes, without integrating the concepts of probability and proportionality.

Lastly, it was more difficult for participants to provide a satisfactory explanation of their solutions to a compound probability problem involving independent experiments (Item 3). Only 25% of the arguments were correct, even though the percentage of participants that provided correct responses to the item was much higher (57%). Errors in the arguments included misinterpreting Laplace's rule, incorrect enumeration of the sample space in the compound experiment and reasoning according to the equiprobability bias.

Frequentist meaning Responses to items of the questionnaire involving the frequentist meaning of probability suggested that specialized content knowledge related to this meaning was weak among the participants. The arguments used to determine whether a sequence was random (Item 10) varied with the sequence selected as being random. Those who believed that Clara was cheating (the correct response) used arguments related to the existence of patterns or very short runs and generally demonstrated a good perception of randomness (23% of the arguments). Participants who argued that Luisa cheated based their explanation on the existence of long runs, disregarding the independence of trials; other subjects argued that the observed frequencies of heads and tails should be closer to 50%, evidently assuming little variability in a random process.

Most explanations of responses given to Items 11a and 11b evidenced use of the representativeness heuristic and an insensitivity to sample size, with very few correct arguments given. A few participants also exhibited the equiprobability bias in their arguments. Correct arguments (representing only 17% and 11% of the responses, respectively) discussed sample variability and the effects of changing sample size.

Subjective meaning Responses to items involving the subjective meaning of probability indicated that specialized content knowledge was also poor with respect to this conception. Responses to Item 4 were typically justified by referring to the maze characteristics, such as the number of paths, the distance to origin, and the number of junctions. Few participants based their arguments on probabilistic reasoning, for example by drawing a tree diagram or by calculating probabilities.

The percentage of correct explanations for Item 12b was also very low (11%). In responses to this item, arithmetic arguments prevailed, evidencing the extreme expectation of a deterministic replication of the forecast prediction. Correct justifications (the forecast is still valid) involved probabilistic reasoning and correct interpretations of a frequentist

estimate of probability.

These results were poorer than those reported in studies involving Spanish prospective secondary school teachers (Contreras, 2011) or future psychologists (Díaz & Batanero, 2009) responding to similar items. This result was expected, given that subjective probability and conditional probability are complex topics and that participants in our sample typically have less extensive mathematical backgrounds than psychology students or prospective secondary school teachers. A relevant percentage of incorrect justifications were linked to use of the representativeness heuristic, equiprobability bias, misinterpretation of the statement, or changing of the question, even though many of the participating prospective teachers could correctly compute the probabilities involved and solve parts of the problems correctly.

5. GLOBAL ANALYSES

We assigned a numeric score to each correct response and to each correct argument for all of the items in order to compute an individual quantitative measure for every student on each component and on the entire questionnaire.

5.1. GLOBAL KNOWLEDGE

The global score on the questionnaire theoretically varies between 0 (all responses and all arguments incorrect) and 34 (all responses and all arguments correct). In this study, however, the maximum value was not achieved (See Figure 1), with scores varying between 2 to 28. The observed average score (14.32) was lower than the theoretical average score (17). A quarter of participants received scores lower than 11.

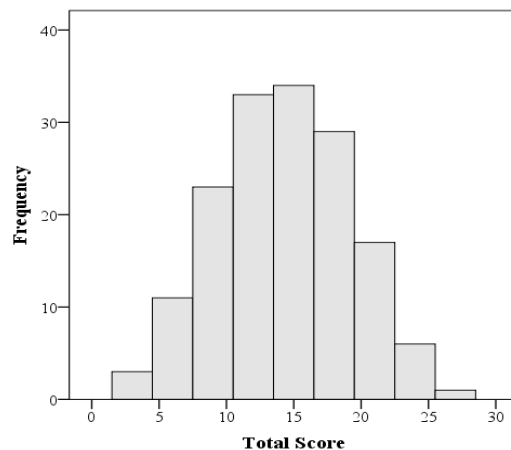


Figure 1. Total scores on the questionnaire within the sample of prospective teachers

5.2. COMPONENTS OF PROBABILITY KNOWLEDGE

In Figure 2, we compare the participants' scores in the areas of common, advanced and specialized probability content knowledge. Each of these scores was obtained by adding the scores on all items measuring the given component (see Table 1), then dividing by the maximum possible score in the component in order to obtain an index that ranges between 0 and 1, to facilitate comparisons (see Figure 2). The theoretical average value for each component of content knowledge assessed is 0.5.

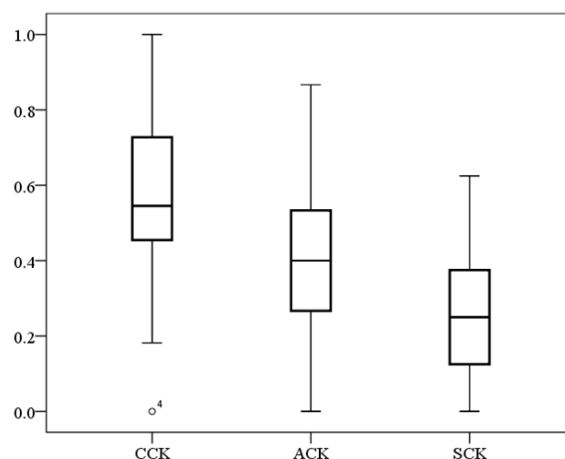


Figure 2. Component of probability content knowledge indexes in the sample

In agreement with previous research, the participants in the present study obtained better scores on items assessing common probability content knowledge, in comparison with the other content knowledge components. The median (0.55) was higher than the theoretical value (0.5), and only 25% of participants obtained a score lower than 0.36. In other words, most participants showed a reasonable degree of common content knowledge for teaching probability in primary school, only one participant (labeled with 4) had zero correct answers in this component (as a future teacher, this is worrisome). Although there was a high proportion of correct responses to the items assessing this component of content knowledge, there were some exceptions, with difficulties related to enumeration (Item 1b), selecting the least likely and most likely result in a dependent compound experiment (Item 4c), and sampling variability in a frequentist estimate of probability (Item 5). Errors in responses to items assessing common probability content knowledge were mostly due to the use of arithmetic strategies, errors in proportional reasoning, or comparing favourable cases instead of comparing probabilities. We finally note the great variability in scores: For example, one participant did not give a single correct response to the items related to this component, and a few participants provided correct responses to all of the items.

The advanced probability content knowledge of the participants was weaker according to the results, with a median value of 0.40 and only 25% of prospective teachers giving 53% or more correct responses to items assessing this component of content knowledge. Again, there was a great variability in the results, with half of the participants giving more than 40% correct responses, but none succeeding in correctly answering all 15 of the items related to this component. These results are a cause of concern, considering that all of these items are similar to problems that the participants should have encountered during their secondary education and in the first year of their teacher education. Unlike the responses to items assessing common content knowledge, most correct responses to these items were obtained through valid probabilistic strategies, which are required in more advanced tasks. Furthermore, a relevant percentage of participants solved the tasks using only proportional reasoning, though some difficulties with inverse proportionality were evident in responses to Items 6b and 9a.

Specialized content knowledge, which was assessed by evaluating the arguments used to justify solutions to certain items, was very low among the participants and also very variable. Only in response to a few of the relevant items were the majority of the participants able to provide valid explanations of their solutions and strategies. The analysis of the items' difficulty (Table 2) shows that only two SCK items received a correct argument from more than half of the participants (item 2, 74% and item 6, 89%). Correct

arguments were more likely to be provided when the task was successfully solved; for example, most of the participants who correctly computed the joint probability on Item 6 also used the appropriate probabilistic reasoning to argue why a game was unfair.

In general, the participants used the same types of arguments in response to most of the items, regardless of the content of the problem; low competence in justifying reasoning was evident. It is possible that this difficulty is not exclusive to the topic of probability and may extend to other areas of mathematics. Given that the ability to explain a mathematical solution or strategy is an aspect of teachers' specialized content knowledge, the results suggest the need to strengthen this competency among prospective teachers.

In this part of the questionnaire (i.e., on the items assessing specialized content knowledge), a significant proportion of the participants exhibited non-probabilistic reasoning—in particular, in response to the questions involving the frequentist meaning of probability (10, 11a, 11b, 12b), where the proportion of valid arguments was less than 23% (Table 2). Some responses also evidenced the equiprobability bias (Lecoutre, 1992), use of the representativeness heuristic (Tversky & Kahneman, 1974), or use of the outcome approach (Konold, 1989).

5.3. MEANINGS OF PROBABILITY KNOWLEDGE

In Figure 3, we compare the participants' scores on questions involving the three different meanings of probability. Each of these scores is determined by computing the average score on the items involving the given meaning of probability (see Table 2), which results in an index ranging between 0 and 1. Each score has a theoretical average value of 0.5. Because we have discussed the results of specific items in the previous section, here we only discuss the global results as related to each meaning of probability.

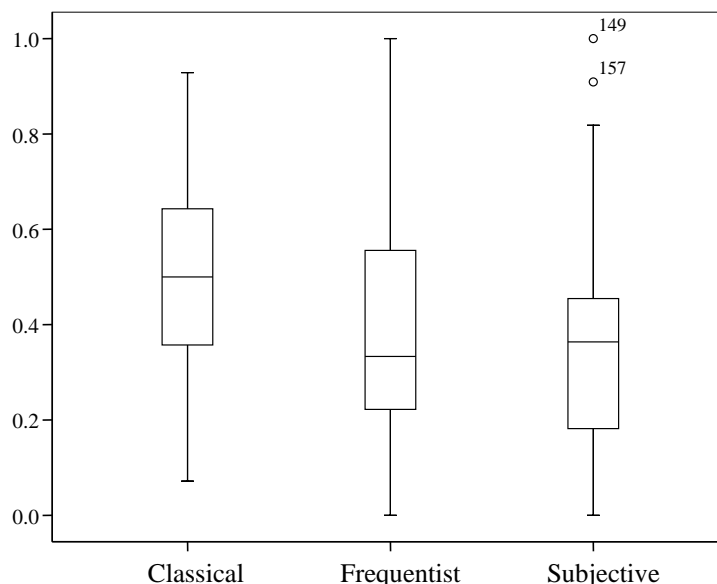


Figure 3. Knowledge indexes for different meanings of probability in the sample

It can be seen in Figure 2 that an average value of 0.5 was only obtained on items involving the classical meaning of probability, with the lowest median value obtained on items involving the frequentist meaning. This is a matter of concern, as the current curricular guidelines in Spain for primary education suggest introducing the frequentist

conception of probability at this level. A potential explanation for this result is that the frequentist approach to probability is a recent addition to curricula in Spain, and that consequently, the prospective teachers in this study did not have much prior experience with this approach.

The boxplots in Figure 3 suggest a high variability in the prospective teachers' knowledge, with some giving correct responses to all items corresponding to the frequentist and subjective meanings, and others giving incorrect responses to almost all items in all three categories. Two participants (149 and 157, labelled in the figure) stand out positively for their high scores in subjective meaning, which is the more complex of these three.

5.4. DISCRIMINATION, RELIABILITY, AND GENERALIZABILITY

The scores on the questionnaire were also used to estimate a reliability coefficient. For this questionnaire, Cronbach's $\alpha = 0.768$, which is a reasonable value considering the multidimensional character of the questionnaire (assessing different components of knowledge). Generalizability coefficients, which are based on an analysis of different sources of variance that could affect the observed scores, were also computed for the questionnaire. A generalizability coefficient $G_1 = 0.768$ was obtained when the variation among the questionnaire items was the only source of variance considered for this computation. Both α and G_1 , which may have values between 0 and 1, describe the variability of the measure due to variation in the questionnaire items. A second generalizability coefficient, which describes the variability that may be attributed to the participants in the sample, was computed and found to have a value of $G_p = 0.984$, suggesting a high generalizability of our results to different participants taken from the same population—that is, to other Spanish prospective primary school teachers who share the main sociodemographic and educational characteristics of our sample.

5.5. DISCRIMINATION, RELIABILITY, AND GENERALIZABILITY

The scores on the questionnaire were also used to estimate a reliability coefficient. For this questionnaire, Cronbach's $\alpha = 0.768$, which is a reasonable value considering the multidimensional character of the questionnaire (assessing different components of knowledge). Generalizability coefficients, which are based on an analysis of different sources of variance that could affect the observed scores, were also computed for the questionnaire. A generalizability coefficient $G_1 = 0.768$ was obtained when the variation among the questionnaire items was the only source of variance considered for this computation. Both α and G_1 , which may have values between 0 and 1, describe the variability of the measure due to variation in the questionnaire items. A second generalizability coefficient, which describes the variability that may be attributed to the participants in the sample, was computed and found to have a value of $G_p = 0.984$, suggesting a high generalizability of our results to different participants taken from the same population—that is, to other Spanish prospective primary school teachers who share the main sociodemographic and educational characteristics of our sample.

6. CONCLUSIONS

The results of this study partly support other previous research, in the sense that they suggest that prospective primary teachers' probability content knowledge is insufficient. A high proportion of the participants demonstrated poor combinatorial reasoning, made errors in computing conditional probability and in interpreting frequentist probabilities, and

evidenced use of common heuristics and biases.

A unique aspect of the present study is that the questionnaire used to analyse the prospective teachers' probability content knowledge can also be used to obtain information about participants' knowledge of the different meanings of probability embedded in the primary school curriculum. In our sample, participants' understanding of the classical meaning was stronger than their understanding of the frequentist and subjective meanings. Although the prospective teachers demonstrated a reasonable degree of common content knowledge, there is still need for improvement, and the study shows that both advanced and specialized content knowledge need to be strengthened among prospective primary school teachers. In particular, specialized content knowledge related to probability was very low among the participants, although appropriate arguments were more likely to be given when the task was successfully solved, in agreement with Mohr (2008) and Callingham and Watson's (2011) results. These results provide information about particular areas of knowledge that need to be reinforced in primary teacher education programs in Spain.

The present study also suggests the need to strengthen prospective teachers' skills in justifying solutions and strategies, not only to improve their skills in teaching probability, but also in teaching other content areas in primary education.

It must be noted that these results should be interpreted with caution, as the sample in the study was limited; moreover, the questionnaire used in this study should be complemented with other assessment instruments and tasks related to the different components of pedagogical content knowledge. These limitations suggest the need for continued research that contributes to an improved understanding of prospective teachers' knowledge and their formative needs.

ACKNOWLEDGEMENTS

Project EDU2013-41141-P (MEC) and group FQM126 (Junta de Andalucía) provided funding for this research.

REFERENCES

- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Azcárate, P. (1995). *El conocimiento profesional de los profesores sobre las nociones de aleatoriedad y probabilidad* (Teachers' professional knowledge of randomness and probability). Unpublished Ph.D. University of Cádiz, Spain.
- Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433-456). Washington, DC: American Educational Research Association.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. What makes it special? *Journal of Teacher Education*, 59(5), 389-407.
- Batanero, C. (2013). Teaching and learning probability. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 491- 496). Heidelberg: Springer.
- Batanero, C., Garfield, J. B., & Serrano, L. (1996). Heuristics and biases in secondary school students' reasoning about probability. In L. Puig & A. Gutiérrez (Eds.), *Proceedings of the Twentieth Conferencere of the International Group for the Psychology of Mathematics Education* (Vol 2, pp. 51-59). Valencia, Spain: IGPME.

- Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 15-37). New York: Springer.
- Callingham, R., & Watson, J. (2011). Measuring levels of statistical pedagogical content knowledge. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI and IASE study* (pp. 283-293). New York: Springer.
- Cañizares, M. J. (1997). *Influencia del razonamiento proporcional y combinatorio y de creencias subjetivas en las intuiciones probabilísticas primarias* (Influence of proportional and combinatorial reasoning and of subjective beliefs in primary probabilistic intuitions). Unpublished Ph.D. University of Granada, Spain.
- Chernoff, E. J. (2011). Investigating relative likelihood comparisons of multinomial, contextual sequences. In M. Pytlak, T. Rowland, & E. Swoboda (Eds), *Proceedings of the Seventh Conference of the European Society for Research in Mathematics Education* (591-600). Reszlow: ERME.
- Contreras, J. M. (2011). *Evaluación de conocimientos y recursos didácticos en la formación de profesores sobre probabilidad condicional* (Assessing knowledge and didactic tools for training teachers in conditional probability). Unpublished Ph.D. University of Granada, Spain.
- Díaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education*, 4(3), 131-162.
- Falk, R., & Wilkening, F. (1998). Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology*, 34(6), 1340-1357.
- Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, 15(1), 1-24.
- Godino, J. D., Batanero, C., & Cañizares, M. J. (1987). *Azar y probabilidad. Fundamentos didácticos y propuestas curriculares* (Chance and probability. Didactic foundations and curricular proposals). Madrid: Síntesis.
- Gómez, E., & Contreras, J. M. (2014). Meanings of probability in Spanish curriculum for primary school. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics*. Flagstaff, Arizona, USA: International Statistical Institute. [Online: iase-web.org]
- Green, D. R. (1982). *Probability concepts in school pupils aged 11-16 years*. Unpublished Ph.D. Loughborough University of Technology. [Online: dspace.lboro.ac.uk]
- Green, D. R. (1983). From thumbtacks to inference. *School Science and Mathematics*, 83(7), 541-551.
- Hill, H. C., Ball, D. L., & Schilling, S. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372-400.
- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98.
- Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, 23, 557-568.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. Sussex, UK: Psychology Press.
- Ministerio de Educación, Cultura y Deporte (MECD). (2014). *Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la educación primaria* (Royal Decree establishing the basic curricula for primary education). Madrid: Author.

- Mohamed, N. (2012). *Evaluación del conocimiento de los futuros profesores de educación primaria sobre probabilidad* (Assessing prospective primary school teachers on probability). Unpublished Ph.D. University of Granada.
- Mohr, M. J. (2008) Mathematics knowledge for teaching: The case of preservice teachers. In G. Kulm (Ed.), *Teacher knowledge and practice in middle grades mathematics* (pp. 19-43). Rotterdam: Sense Publishers.
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Shaughnessy, J. M., & Ciancetta, M. (2002). Students' understanding of variability in a probability environment. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on the Teaching of Statistics*, Cape Town, South Africa [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

EMILSE GÓMEZ TORRES

Departamento de Estadística, Facultad de Ciencias,
Universidad Nacional de Colombia, Carrera 30, N. 45-02
Bogotá, Colombia

APPENDIX 1. QUESTIONNAIRE

Solve the following tasks, explaining your solution in writing when required to do so:

Item 1. (Adapted from Green, 1982)

Three boys take turns playing a video game. They have to line up in a row and wait for their turn. Suppose the boys are called Andres, Benito and Carlos (A, B, C, for short). We want to write down all of the possible orders in which they could play this game: for example, one possible order is BCA.

- Write down all of the different orders. How many different ways are there altogether?
- If four boys (A, B, C, D) want to play, how many different ways are there?

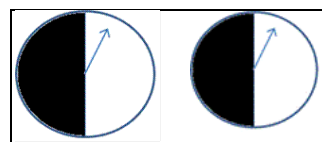
Item 2. (Adapted from Falk & Wilkening, 1998)

Pablo puts 5 white balls and 7 black balls into an urn. Miguel puts 3 white balls and 5 black balls into another urn. How many black or white balls should be moved from one urn to another if we want both children to have the same chance of drawing a black ball? Why?

Item 3. (Adapted from Shaughnessy & Ciancetta, 2002)

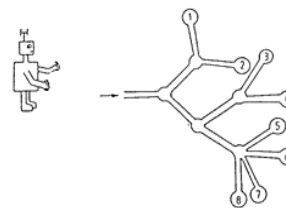
These two fair spinners are part of a carnival game. A player wins a prize only when *both* arrows land on black after each spinner has been spun once.

- What is the probability of winning the game?
- Why?





Item 4. (Adapted from Green, 1982)

A robot is put into a maze, which it begins to explore. At each junction, the robot is as likely to follow any one path as any other (however, it will not go back the way it came). There are eight traps at the end of each of the eight paths (see the picture).



- In which trap (or traps) is the robot most likely to finish up? Why?
- In which traps or traps is the robot least likely to finish up? Why?

Item 5. (Adapted from Green, 1982, 1983)

A packet of 100 drawing pins is emptied out onto a table by a teacher. Some drawing pins landed “up”  and some landed “down” . The results were as follows: 68 landed up and 32 landed down. The teacher then asked four students to repeat the experiment. Each student emptied a packet of 100 drawing pins and got some landing up and some landing down. In the following table, write possible results for each student:

Daniel	Martin	Diana	Maria
up:	up:	up:	up:
down:	down:	down:	down:

Item 6. (Adapted from Azcárate, 1995)

Miguel and Luis play a game that involves rolling two fair dice (each numbered from 1 to 6). They roll both dice and multiply their numbers.

- Miguel receives 1 euro if the product is an even number
 - Luis receives 1 euro if the product is an odd number
- Is this game fair? Why?
 - If Miguel receives 1 euro every time the product of both dice is even, how many euros should Luis receive every time the product is odd if the game is to be fair?

Item 7. (Adapted from Díaz & Batanero, 2009)

Two dice (one red and one blue die) are rolled, and the product of the two resulting numbers is 12. What is the probability that one of the two numbers is a six? (We take the order of the numbers into account.)

Item 8. (Batanero, Garfield, & Serrano, 1996)

When three dice are rolled simultaneously:

- Which of the following results is most likely?
 - a 5, a 2, and a 3 _____
 - two 5s and a 3 _____
 - three 5s _____
 - The chances of obtaining each of these results are the same _____
- Is one of these results less likely than the others? Which one?

Item 9. (Adapted from Fischbein & Gazit, 1984)

On a farm, there is a fishing pool. The owner wants to know how many fish there are in the pool. He took out 200 fish and marked each of them with a coloured sign. He released the marked fish back into the pool and let them get mixed in with the others. On the second day, he took out 250 fish in a random fashion and found that, among them, 25 were marked.

- a. What is the approximate number of fish in the pool?
- b. If owner randomly takes 100 more fish, approximately how many will be marked?

Item 10. (Green, 1982)

A teacher asked Clara and Luisa to each toss a coin 150 times and to record whether the coin landed on heads or tails on each toss. For each “Heads,” a 1 is recorded, and for each “Tails,” a 0 is recorded. Here are the two sets of results:

Clara: 01011001100101011011010001110001101101010110010001
 01010011100110101100101100101100100101110110011011
 01010010110010101100010011010110011101110101100011

Luisa: 10011101111010011100100111001000111011111101010101
 11100000010001010010000010001100010100000000011001
 00000001111100001101010010010011111101001100011000

One girl followed the instructions, tossing the coin on each turn; the other girl cheated and just made the sequence up.

- a. Which girl cheated?
- b. How can you tell?

Item 11. (Adapted from Batanero, Garfield, & Serrano, 1996; this is a version of the hospital problem introduced in Tversky & Kahneman, 1974)

In a certain town hospital, a record of the number of boys and girls born in the hospital is kept.

- a. Which of these cases is more likely:
 - There will be 8 or more boys among the next 10 babies born at the hospital ____
 - There will be 80 or more boys among the next 100 babies born at the hospital ____
 - Both results are equally likely ____

Explain your answer:

- b. Which of these cases is more likely among the next 10 babies born at the hospital:
 - There will be 7 or more boys ____
 - There will be 3 or less boys ____
 - The number of boys will be between 4 and 6 ____
 - These three results are equally likely ____

Explain your answer:

Item 12. (Adapted from Batanero, Garfield, & Serrano, 1996; the problem is inspired by a situation described by Konold, 1989)

A weather forecaster says that this year, there is a 70% chance of rain in Santiago de Compostela.

- a. If this forecaster is right, how many rainy days would you expect this year in Santiago de Compostela?
- b. Suppose that the forecaster said there was an 80% chance of rain this week and that it did not rain on Monday. What would you conclude about the statement that there was a 80% chance of rain?
- c. If the prediction was 80% chance of rain, but it did not rain on Monday or Tuesday, what would you conclude?