Assessing and developing prospective teachers' understanding of random sequences[1]

Carmen Batanero, Emilse Gómez, Magdalena Gea

Abstract: In this paper we describe a formative activity carried out with a sample of 157 prospective primary school teachers in Spain. Participants first completed a task used in previous research with children and then, discussed their solutions and carried out simulation activities. Results show initial difficulties to discriminate a random and a non-random sequence; difficulties in understanding independence and perceiving the variability linked to randomness. The simulation activities helped many participants to overcome these biases. We conclude with some implications for teacher education.

1 Introduction

Probability is today part of the primary school mathematics curriculum in many countries. Specifically, the Spanish curriculum for primary education (MEC, 2006) includes the following contents in the first cycle (6-7 year-olds)*: "Random nature of some experiences. Difference between possible, impossible and that what is possible but not certain".* In the second cycle (8-9 year olds) the document suggests that children should evaluate the results of random experiences, and understand that there are more and less probable events, and that it is impossible to predict a specific result. In the last cycle (10-11 year olds) children are encouraged to recognize random phenomena in everyday life and estimate the probability for events in simple experiments.

The success of this curriculum depends on the adequate preparation of the teachers. Consequently, it is important assessing teachers' conceptions of probability and finding suitable activities where teachers work with meaningful problems and are confronted to potential probabilistic misconceptions (Batanero & Díaz, 2012). The aim of this paper is to analyze results from one of these activities related to perception of randomness.

2 Previous research

In the extensive research examining perception of randomness (e.g., Bar-Hillel, & Wagenaar, 1991; Falk, 1981; Falk & Konold, 1997; Nickerson, 2002; Engel & Sedlmeier, 2005) two types of task have been used (Falk & Konold, 1997): (a) In *recognition* tasks, people are asked to select the most random series of results that might have been produced by a random device (e.g. flipping a coin); (b) In *generation* tasks, subjects generate sequences to simulate a series of outcomes from a typical random process. One main conclusion of this research is that people do not easily recognise or produce randomness. Systematic biases such as the gambler's fallacy where people believe that, after a long run of a same result, the probability of that event occurring in the following trial is lower have been described. Explanations for these biases include the representativeness heuristics (Tversky & Kahneman, 1982), misperception of independence (Falk, 1981; Falk & Konold, 1997; Engel & Sedlmeier, 2005), or belief that random experiments can be controlled (Fischbein & Gazit, 1984).

These biases have been also observed in children: Green's (1983) research with 11-16 year-olds suggests that, contrary to Piaget's theory, recognition of randomness does not improve with age. Batanero and Serrano's (1999) research with secondary school students (14 and 17 year olds) showed a mixture of correct and incorrect properties associated by students to randomness. On one hand, students perceived local

1. En U. Sproesser, S. Wessolowski, y C. Wörn (Eds.) *Daten, Zufall und der Rest der Welt – Didaktische Perspektiven zur anwendungsbezogenen Mathematik*, (pp. 1-11).. Heidelberg: *Springer Fachmedien Wiesbaden.*

variability, lack of patterns in the lineal or spatial arrangement of outcomes, and unpredictability of the random processes. However some students did not perceived independence of trials or believed in the possibility of controlling a random process.

There is little research exploring teachers' understanding of randomness, and that available suggest that their understanding is poor. Azcárate, Cardeñoso and Porlán (1998) analyzed the responses by 57 prospective primary school teachers to a questionnaire that asked them to classify some events as random or deterministic. Even when many teachers recognised the existence of various possible results and the unpredictability of random phenomena, few of them showed a sound understanding randomness. Chernoff (2009) analyzed the responses from 19 mathematics secondary and primary school prospective teachers to a recognition task, in which short sequences of coin tossing, with the same ratio of heads to tails (3:2) were presented. According to the author, these teachers may hold a non-standard perception of the sample space in the experiment that explained their apparently incorrect responses.

In a previous paper (Arteaga, Batanero, & Ruiz, 2010) we analyzed perception of randomness in a sample of 200 prospective primary school teachers, using a generation task. Our results suggested that many participants may held some misconceptions described by Batanero and Serrano (1999), which match incomplete conceptions of randomness that were common along history. In order to further explore this hypothesis, in this paper we analyze the written responses and justifications given by a new sample of 157 prospective primary school teachers to a classical recognition of randomness task that was previously used with children. Below we first describe the method, then present our results and, finally, describe a teaching activity oriented to confront these prospective teachers with the misconceptions revealed in their responses to the task.

3 Method

A total of 157 prospective primary school teachers took part in the sample. The data were taken as part of a practical activity in a mathematics education course. These participants had studied probability in secondary school, and in the previous academic year in a mathematical course that included randomness and probability (classical and frequentist approach). Three sessions (1 hour long each, with an interval of a week between sessions) were spent in the activity. In the first session the prospective teachers were given a questionnaire with some questions taken from Green's (1983) research. These tasks were selected because of the availability of children's responses that would be used to discuss with the prospective teachers the similarities and differences between their conceptions and those shown by the children. In the second session the prospective teachers debated their answers and justifications to the tasks; the aim was to reveal potential conflicts and misconceptions. Also, simulations activities were organised by the teacher educator in order to confront participants with their own misconceptions and help them overcome the same (as suggested by Batanero et al., 2005). In the third session, participants were given responses taken from Spanish children to the same tasks, and worked in pairs to evaluate these responses. In this way the whole practice helped to assess and develop the probabilistic and pedagogical knowledge of these prospective teachers.

The task analyzed (Figure 1) was adapted from Green (1983) and is similar to other tasks used in previous research on perception of randomness. A possible strategy to solve the task is counting the number of heads in each sequence and comparing them with the expected value. The theoretical number of heads in 150 trials can be modelled by the Binomial distribution B (150, 0.5); so that the expected number of heads is 75.

When comparing this theoretical number with the observed frequency of heads in the sequence of Clara (72) and Luisa (67), we observe no match in either case; however, some variation would be expected in a random process. We could use the Chi-squared test to decide whether the differences are reasonable or not. When applying this test to Clara sequence we obtain $\chi^2_{obs}=0.24$, $p=0.6.$, *1 d.f.* and for Luisa $\chi^2_{obs}=1.71$, $p=0.19$; therefore no result is statistically significant. We could repeat the procedure counting the sequences by pairs (i.e., as successive tossing of two coins) and obtain the results presented in Table 1.

Now, the Chi-square test is statistically significant for Clara's sequence ($\chi^2_{obs}=9.84$, *p=0.02, 3 d.f.*), but not for Luisa's results ($\chi^2_{obs}=4.89$, *p=0.18)*. Consequently, we reject the hypothesis that Clara's sequence is random.

Clara and Luisa were each told to toss a coin 150 times. One did it properly. The other just made it up. They put 0 for Heads and 1 for Tails.

Clara: 01011001100101011011010001110001101101010110010001
010100111001101011001011001011001001011101100110 11
010100101100101011000100110101100111011101011000 11

Luisa: 10011101111010011100100111001000111011111101010101
11100000010001010010000010001100010100000000011001
00000001111100001101010010010011111101001100011000

Question 1. Did Clara or Luisa make it up?

Question 2. How can you tell?

Figure 1. Task given to participants

Table 1. Observed and expected frequencies for possible events in throwing two coins

	HH	HT	TH	TT
Clara	12	30	18	15
Luisa	25	21	12	17
Theoretical	19	19	19	19

Our participants had not enough statistical knowledge to use the Chi-square test; however, they were able to make an informal inferential procedure similar to the reasoning of many children in Green's research. They may count the frequency of heads and tails in both sequences, and base their response on the difference between the expected and observed values. Other participants could base their answer on the analysis of longest run (3 in Clara's sequence and 9 in Luisa's). According to Schilling (1990), the expected value of length of the longest run in n trials of flipping a coin tends to $\log_2 n - 2/3$; in this case $\log_2 (150) - 2/3=6.56$, then the expected length of the longest run is near to 7, so that Luisa's result is closer to the theoretical value that than of Clara's.

4 Results and discussion

4.1 Identification of random sequences

The analysis of responses to Question 1 in the assessment performed during the first session (Who made it up?) provided the following results: 42 prospective teachers (26.8%) replied that Clara cheated (correct response), 89 suggested that Luisa cheated (56.7%), 18 did not know (11.4%) and 8 did not answer the question (5.1%). Few participants showed a correct intuition to recognise random sequences and results were worst than those observed in previous research (the correct answer was given by 34% of children in Green's research). Results were better in shorter sequences, since in Batanero and Serrano (1999) 54% and 59% of 14 and 17 year olds students correctly identified random sequences of length 40 (40% and 64% correctly identified non-random sequences).

4.2 Arguments used to support randomness

We also studied the reasons given by the participants to select the random sequence in Question 2 (How can you tell?). First, we classified these arguments according to whether they were based on the frequency of heads, the length of runs, the pattern in the sequence or unpredictability. Afterwards, each of the above categories was classified according to the concepts and ideas involved in the responses. With this procedure we got the following categories:

A1. *Frequencies are too different from the theoretical values:* This argument is consistent with the frequentist view of probability, where we expect that relative frequency will approach the theoretical probability (Batanero, Green, & Serrano, 1998). Participants giving this reason showed a comprehension of convergence, and moreover they performed an informal inference process, where they rejected or accepted the hypothesis of randomness in each sequence, by comparing the empirical data with a mathematical model (expected number of heads). Those giving this argument to support that Clara cheated were also able to recognize the variability inherent to a random process. On the contrary, when the argument was used to defend that Luisa cheated, the person showed a wrong conception of randomness, and did not recognised the variability of random processes.

A2. *Frequencies are too similar to the theoretical value:* Here again, the participant performed an informal inferential process as described in the above category. Those providing his reasoning to support that Clara cheated recognised random variability, as in the previous category; however, those participants suggested that Luisa cheated, because her frequencies should be closer, or even match the theoretical value hold a wrong conception of randomness.

A3. *Long runs:* Many participants rejected Luisa's sequence as random because of the existence of long runs; which shows a poor understanding of independence in repeated trials (Batanero & Serrano, 1999). Only a prospective teacher observed the lack of long runs, as a reason to reject Clara's sequence as random; in this case his perception of independence was good.

A4. *Short runs*: Some prospective teachers found that runs in Clara's sequence were too short to be random, showing a correct perception of independence.

A5. *Pattern or order in the sequence:* Some participants suggested that both symbols should alternate often (they expected a pattern of alternations in the sequence) since heads and tails were equally likely; this argument was used to decide that Luisa cheated. Underlying this reasoning is the outcome approach described by Konold (1989); the reasoning also involves a poor understanding of the frequentist meaning of probability. Other participants viewed the regularity in the alternation pattern as an indicator of lack of randomness (Clara cheated). These participants associated randomness with no model or pattern, a view close to von Mises' (1952/1928) modelling of randomness; this author conceived a sequence as random if the sequence contains no predictable patterns.

A6. *The sequence does not follow a pattern*: This argument is the opposite of A5 and was used by one participant to support that Luisa cheated.

A7. *Unpredictability:* Some participants incorrectly generalized the unpredictability of a random process (which applies to isolated outcomes) and assumed the impossibility to predict the frequencies of different outcomes in a series of trials. Again, here the reasoning is close to the outcome approach.

A8. *Other arguments:* less frequent explanations were: equally likely outcomes (showing a conception of randomness close to the classical view of probability, according to Batanero, & Serrano, 1999), personal beliefs or very confuse explanations.

In Table 2 we present the frequencies (and percentages) of the different reasoning used to support what girl cheated. The arguments supporting that Clara cheated related mainly to the existence of a pattern (50%) or too short runs (28.6%); both arguments indicate correct conceptions of randomness. Those participants deciding that Luisa cheated were primarily based on long run (58%), misunderstanding independence; another high percentage expected that the observed frequencies should be closer to the theoretical value (28.6%), and then were unaware of the variability inside randomness.

Table 2. Frequencies and percentages of reasons on part b of the task (n=148)

	Did Clara or Luisa make it up?							
	Clara		Luisa		Don't know		Total	
How can you tell?	Freq.	%	Freq.	%	Freq.	%	Freq.	%
A1. Frequencies are very different	1	2.4	19	21.6			6	4.1
A2. Frequencies are very close	3	7.1	5	5.7			22	14.9
A3. Runs are too long	1	2.4	51	58.0			52	35.1
A4. Runs are too short	12	28.6					12	8.1
A5. There is a pattern	21	50.0	7	8.0			28	18.9
A6. There is no pattern			1	1.1			1	0.7
A7. Unpredictability					15	83.3	8	5.4
A8. Other arguments	4	9.5	5	5.7	3	16.7	19	12.8

The most frequent arguments were those based on the length of runs, followed by the existence of an apparent pattern and, with smaller frequency, by differences between observed and expected frequencies; 59% of all the prospective teachers gave wrong arguments to support that Luisa cheated, 27% of them provided correct reasons to support that Clara cheated; the remaining participants were unable to decide what sequence was non random or to provide a sound argument. These percentages were very close to those obtained in previous research with children (22% correct answer in Green's research). Prospective teachers had higher argumentative capacity than children; since no response rate in children was 14%.

4.3 Debates of solutions and simulations

After the teachers completed the assessment questionnaire, a second session was spent in a debate, where the teachers presented their solutions to the task and their justification, with the aim of revealing potential conflicts in understanding randomness and related misconceptions. After each participant supported his of her view, with the arguments analyzed in the previous section, the teacher educator suggested to perform a statistical analysis of Luisa's and Clara's sequences to examine if they were likely or not, when accepting the hypothesis of equally likelihood for tails and heads in the coin.

Participants worked in groups to count the frequencies of tails and heads in both sequences (Clara; 72-78) (Luisa; 67-83); results were compared with the theoretical frequencies (75-75). As the difference were not conclusive, the group continued counting the frequencies of possible results in flipping two coins (analysing the 75 pairs of outcomes in the sequences) and obtaining the results presented in Table 1, where now Clara's empirical results differ more than Luisa's from theoretically expected. The longest run in both sequences was also analyzed and a discussion was held as regards the teachers' expectations about the length of the longest run in a long sequence of outcomes.

To investigate this point and to confirm the likelihood of results in Table 1, the discussion was supported with some simulation activities of coin- flipping using an Applet from the experiments provided by Statistics Online Computational Resource (http://www.socr.ucla.edu/SOCR.html). Specifically participants worked with the Binomial coin experiment that serves to simulate the throwing of n coins with a probability p of head; where both n and p can vary and investigated the longest run and the distribution of possible pairs in the compound experiment (flipping two coins), increasing the length n of the sequence, until $n=150$. The different representations in the Applet (outcomes in a series of n trials; frequencies and graphical representation of the binomial distribution) helped participants evaluate the likelihood of frequencies of outcomes and pairs of outcomes as well as the likelihood of short or medium-sized runs in Luisa's and Clara's sequences. By the end of the session most participants understood the experiment and why Luisa's sequence was closer to what we expect in randomness than Clara's sequence. They accepted that long runs are expected in the experiment; that unpredictability is applicable only to individual outcomes, but we can use the probability to predict more or less likely results. Moreover, randomness cannot be identified to absence of models, since they could recognise the binomial model in the experiment.

All this new knowledge was put in practice by the participants in the third session when they were asked to evaluate school children's responses to some probability tasks that involve the randomness concept.

5 Implications for teacher education

Our results coincide with other previous studies (Falk, 1981; Falk & Konold, 1997; Nickerson, 2002), which confirm adults' difficulty to perceive randomness; biases such as the outcome approach, and misconceptions about equiprobability or independence were common in the first session. This is not surprising, since according to Bar-Hillel and Wagenaar (1991) randomness refuses a simple definition and can be applied only through the analysis of sequences of outcomes (and not from the analysis of a single outcome). Moreover, even when expressions like "random number" or "random experiment" frequently appears in school textbooks these books do not usually include a precise definition of the concept (Batanero, Green, & Serrano, 1998).

Prospective teachers in our study initially showed a mixture of correct and wrong intuitions and beliefs about randomness, part of which may interfere in their teaching of probability, since according to Ball, Lubienski, and Mewborn (2001), daily tasks of teachers, such as assessing students or organizing teaching, depends on their mathematical knowledge. Our results also suggest that teacher educator should help prospective teachers to build a sound understanding of randomness, starting from the prospective teachers' correct intuitions.

In agreement with Arteaga, Batanero and Ruiz (2010), some prospective teachers' views described in this study were close to primitive conceptions of randomness, accepted as correct in different historical periods. Teacher educators may start from these correct intuitions to help prospective teachers to complement their understanding of randomness:

- The view of randomness as equally likely results should be restricted to the limited situations where it is applicable.
- The frequentist view, where convergence from observed to theoretical frequencies is expected, should be complemented with an adequate perception of variability and independence on successive trials.
- Recognition of unpredictability of isolated outcomes should be complemented with assuming the possibility of predicting the distribution of results in a long sequence of trials and with understanding of the difference between probability (theoretical value) and frequency (estimation).
- The identification of randomness with lack of pattern should be changed in favour of recognizing the multiplicity of underlying patterns in a random sequence of outcomes.

The task analyzed in this paper, together with the discussion with prospective teachers, of the possible correct and incorrect answers and reasoning biases in their future students, complemented with simulations, may help increase the prospective teachers' mathematical and pedagogical content knowledge about randomness.

Acknowledgment: Project EDU2010-14947, grant FPI-BES-2011-044684, (MCINN-FEDER) and group FQM126 (Junta de Andalucía), Spain; National University, Colombia.

References

Arteaga, P., Batanero, C., & Ruiz, B. (2010). Pre-service primary school teachers' perception of randomness. In M. Pinto & T. Kawasaki (Eds.), *Proceedings of the XXXIV Conference of the International Group for the Psychol-*

ogy of Mathematics Education (v. 2, pp. 183-190). Belo Horizonte, Brazil.

Azcárate, P., Cardeñoso, J. M., & Porlán, R. (1998). Concepciones de futuros profesores de primaria sobre la noción de aleatoriedad (Primary school prospective teachers' conceptions of randomness). *Enseñanza de las Ciencias*, *16*(1), 85-97.

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.

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in applied mathematics, 12*(4), 428-454

Batanero, C., Green, D.R., & Serrano, L. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematical Education in Science and Technology, 29*(1), 113-123.

Batanero, C., Biehler, R., Engel, J., Maxara, C., & Vogel, M. (2005). Simulation as a tool to bridge teachers' probabilistic and pedagogical knowledge. Paper presented at the *ICMI Study 15. Professional development of mathematics teachers.* Aguas de Lindoia, Brazil, 2005.

Batanero, C., & Díaz, C. (2012). Training teachers to teach probability: Reflections and challenges. *Chilean Journal of Statistics 3*(1), 3-13.

Batanero, C., Green, D.R., & Serrano, L. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematical Education in Science and Technology, 29*(1), 113-123.

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education,* 30(5), 558-567.

Chernoff, E. (2009). Subjective probabilities derived from the perceived randomness of sequences of outcomes. Doctoral dissertation. Simon Fraser University.

Engel, J. & Sedlmeier, P. (2005). On middle-school students' comprehension of randomness and chance variability in data. *Zentralblatt fur Didaltik der Mathematik, 37*(3), 168-177.

Falk, R. (1981). The perception of randomness. In C. Laborde (Ed*.), Proceedings of the Fifth International Conference for the Psychology of Mathematics Education* (pp. 222-229)*.*University of Grenoble.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104*, 301-318.

Fischbein, E. & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, *15*(1), 1-24.

Green, D. R. (1983). A Survey of probabilistic concepts in 3000 pupils aged 11-16 years. In D. R. Grey et al. (Eds.), *Proceedings of the First International Conference on Teaching Statistics* (v. 2, pp. 766-783). Universidad de Sheffield: Teaching Statistics Trust.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*, 59-98.

MEC (2006). *Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación Primaria.* España: Ministerio de Educación y Cultura.

Mises, R. von (1952). *Probabilidad, estadística y verdad [Probability, statistics and truth]*. Madrid: Espasa Calpe (Original work published 1928).

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*, 330-357.

Schilling, M. F. (1990). The longest run of heads. *The College Mathematics Journal*, *21*(3), 196-207.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 117-128). New York: Cambridge University Press.