THE MEANING OF RANDOMNESS FOR SECONDARY SCHOOL STUDENTS

Carmen Batanero and Luis Serrano
University of Granada, Spain
Journal for Research in Mathematics Education, 30(5), 558-567, 1999
(*Extended version)*

SUMMARY

Subjective perception of randomness has been researched by psychologists using a variety of production and judgement tasks, resulting in a number of different descriptions for the biases that characterise people's performances. These research findings, especially those concerned with children's and adolescents' understanding of randomness, are highly pertinent to didactic practices, as new mathematics curricula for compulsory teaching levels are being proposed that incorporate increased study of random phenomena. In this article, we first analyse the complexity of the meaning of randomness from the mathematical point of view and outline philosophical controversies associated with this. Secondly, we complement previous research by comparing the meaning of randomness for 277 secondary students in two age groups (14 and 17 year-olds), through the identification of the mathematical properties they associate to random and deterministic sequences and to two-dimensional distributions. Some implications for teaching and future research are then suggested.

New mathematics curricula are being proposed in different countries that place more emphasis on the study of randomness and probability and that suggest the performing of simulation experiments, based on random sequences. For example, in the Spanish curriculum for compulsory secondary-level teaching (MEC, 1992), we find the following topic*: "Random phenomena and the associated terminology"*. Reference is made to *"using the vocabulary adapted to describe and quantify random situations"* and to *"building frequency tables to represent the random phenomena's behavior"*. Also, within the list of algorithms and skills to develop, *"obtaining random numbers with various devices"* and the *"detection of common mistakes in the interpretation of chance"* are suggested. In the general strategies we find *"recognizing random phenomena in everyday life and in scientific knowledge", "formulating and testing conjectures on the behavior of simple random phenomena"* and *"planning and performing simple experiments to study the behavior of random phenomena"*. This curriculum is not an exception, since similar terms or expressions are found in recent curricula adopted in other countries, for example those in France, England and Wales, and the United States.

However, epistemological analysis of the concept, as well as psychological research results have shown that the meaning of randomness is not easy for statistically naive people. There is an apparent contradiction in people's understanding of random processes and sequences, which is related to the psychological problems associated with the concept, namely that randomness implies that *"anything possible might occur"*. Subjectively, however, many people believe that only the outcomes without visible patterns are "permissible" examples of randomness (Hawkins et al, 1991, p.104). Despite the relevance of the topic in probability and statistics, little research has been carried out within mathematics education and most psychological research has merely concentrated on the accuracy of people's judgement of randomness.

Before introducing a new topic in the curriculum, a fundamental issue is to determine students' preconceptions, because they inevitably construct their own knowledge by combining their present experiences with their existing conceptions. Shaughnessy (1996) also suggest the need of documenting the changes (or lack of change) of students' thinking over time and after instruction to identify the problematic teaching areas in which we need to revise our curriculum or instructional practices (Shaughnessy, 1997).

The experimental study reported here was intended to examine possible differences in secondary students' conceptions of randomness, before and after instruction in probability, which occurs in the Spanish curriculum between the ages of 14 and 17. At the same time we wanted to explore the similarities between our student's conceptions and various historical conceptions of randomness.

To achieve these aims, the written responses of 277 secondary students in two different groups (14 and 17 year-olds) to some test items taken from Green (1989, 1991) concerning the perception of randomness in sequences and two dimensional distributions are analyzed from different points of view. Our results show that students' subjective meaning of randomness is close to some interpretations that randomness has received throughout history. Because of the fact that Green's studies were carried out with 11-16 years old students, we also provide a complement to this research by extending its results to 17 year-old students.

In the following sections all these aspects will be described, starting off with a brief analysis of the mathematical meaning of randomness and some philosophical controversies relating thereto, as well as a summary of earlier research on which our study is based.

MATHEMATICAL MEANING OF RANDOMNESS

Randomness has received various interpretations at different periods in history. Even today, there are still serious difficulties involved in defining randomness (Fine, 1973; Zabell, 1992; Toohey, 1995).

Randomness and causality

An early meaning of randomness was that it was the opposite of something that had some known causes. 'Chance' was then assumed to be the cause of random phenomena. This meaning was given to randomness throughout the period extending from antiquity to the beginning of the Middle Ages: *"Nothing happens by chance, but rather everything occurs for a reason and a necessity"* (Leucippus, Vth Century BC, quoted by Bennet, 1993). If we adopt this meaning, randomness is due to our ignorance and has a subjective nature.

Poincaré (1936), however, found this meaning to be unsatisfactory because, if we accept it, certain phenomena with unknown laws (such as death) would be considered to be deterministic. Among the phenomena for which the laws are unknown, Poincaré chose to differentiate between random phenomena, for which probability calculus would give us some information, and those non-random phenomena, for which there is no possibility of prediction until we discover their laws. As Ayer (1974) stated, a phenomenon is only considered random if it behaves in accordance with probability calculus, and this definition will still hold even when we have found the rules for the phenomenon.

Randomness and probability

With the first theoretical developments of probability, randomness was related to equi-probability (for example in the *Liber of Ludo Aleae* by Cardano), because this development was closely linked to games of chance, where the number of possibilities is finite and the principle of equal probabilities for the elementary events of the sample space in a simple experiment is reasonable.

Nowadays, we sometimes find randomness explained in terms of probability, although such an explanation would depend on the underlying understanding of probability. If we adopt a *classic* interpretation, we say that an object (or an event) is a random member of a given class, if there is the same probability for selecting this than for any other member of its class. This definition of randomness may be sufficient for random games based on dice, coins, etc., but Kyburg (1974) suggests that it imposes severe, non-natural restrictions on its applications. We can, for example, only consider that an object is a random member of a class if this class is finite. If the class is infinite then the probability associated to each member is always null, and therefore still identical, even when the selection method is biased. Furthermore, this particular explanation precludes any consideration of randomness applied to elementary events that are not equi-probable.

When we transfer the applications of probability to the physical or natural world, for example studying the blood type of a new-born baby or any other hereditary characteristic, we cannot rely on the equi-probability principle. Here, we may consider an object as a random member of a class if we can select it using a method providing a given ‘a priori’ relative frequency to each member of this class in the long run. Thus, we use the *frequentist* basis of probability, which is most appropriate when we have data from enough cases. However, we are left with the theoretical problem of deciding how many experiments it is necessary to consider in order to be sure that we have sufficiently proven the random nature of the object.

Within either of these two frameworks, randomness is an ‘objective’ property assigned to the event or element of a class. Kyburg (1974) criticizes this view and proposes an interpretation of randomness composed of the following four terms:

- the object that is supposed to be a random member of a class;
- the set of which the object is a random member (population or collective);
- the property with respect to which the object is a random member of the given class;
- the knowledge of the person giving the judgement of randomness.

Whether an object is considered to be a random member of a class or not, depends, under this interpretation, on our knowledge. This view, supported by the subjective conception of probability, is more appropriate when we have some information affecting our judgement about the randomness of an event.

Formalization of randomness

At the end of the XIXth Century, Edgeworth, Galton, Pearson, Fisher and their collaborators started to base statistical analysis in applied research on random samples of data. Before this period, rather than being oriented towards inference, statistics was often confined to the descriptive study of complete populations. When samples were used, they were not selected at random, because there was not an awareness of the need for this if valid and precise inferences about the population were to be achieved. Once theoretical developments began to show the importance of random sampling, then interest in finding models for processes providing long sequences of random digits was born.

With the advent of tables of pseudo-random numbers, concern about how to ensure the ‘quality’ of those numbers began to emerge. The possibility of obtaining pseudo-random digits with deterministic algorithms also suggested the need for examining the sequence produced, regardless of the process by which it had been generated. It was discussions about such things that led to the formalization of the concept of randomness (Fine, 1973).

Von Mises based his study of this topic on the intuitive idea that a sequence is considered to be random if we are convinced of the impossibility of finding a method that lets us win in a game of chance where winning depends on forecasting that sequence. He defined the randomness of a sequence by proposing the invariance of the relative frequency for each possible event in all the possible sub-sequences of the given sequence.

This definition of randomness is the basis for statistical tests that are used for checking random number tables before presenting them to the scientific community. However, since in all statistical tests there is the possibility of error, we can never be totally certain that a given sequence, in spite of having passed all the tests, does not have some unnoticed pattern within it. Thus, we cannot be absolutely sure about the randomness of a particular finite sequence. We only take a decision about its randomness with reference to the outcomes of test techniques and instruments. This explains why a computer-generated random sequence (which is not random in an absolute sense) can still be random in a relative sense (Harten & Steinbring, 1983).

Another attempt to define the randomness of a sequence was based on its computational complexity. Kolmogorov’s interpretation of randomness reflected the difficulty of describing it (or storing it in a computer) using a code that allows us to reconstruct it afterwards. In this approach, a sequence would be random if it cannot be codified in a more parsimonious way, and the absence of patterns is its essential characteristic. The minimum number of signs necessary to code a particular sequence provides a scale for measuring its complexity, so this definition allows for a hierarchy in the degrees of randomness for different sequences. It is important to remark that in both theoretical approaches perfect randomness would only apply to sequences of infinite outcomes and therefore, randomness would only be a theoretical concept.

SUBJECTIVE PERCEPTION OF RANDOMNESS

There has been a considerable amount of research into adults’ subjective perception of randomness (e.g., Wagenaar, 1972; Falk, 1981; Bar-Hillel and Wagenaar 1991). These psychologists have used a variety of stimulus tasks, which have been classified in a recent review by Falk and Konold (1977) into two main types. In the first type (*generation tasks*), subjects generate random sequences under standard instructions to simulate a series of outcomes from a typical random process, such as tossing a coin. The second type of research approach uses *recognition* tasks. People are asked to select the most random of several sequences of results that might have been produced by a random device or to decide whether some given sequences were produced by a random mechanism. Similar types of research have also been performed using two-dimensional random distributions, which essentially consist of random distributions of points on a squared grill.

In these investigations, systematic biases have consistently been found. One such bias is known as the *gambler's fallacy*, by which people believe that, after a long run of a same result in a random process, the probability of that event occurring in the following trial is lower. Related to this is the tendency of people in sequence generation tasks to include too many alternations of different results (such as heads and tails in the flipping of a coin) in comparison to what would theoretically be expected in a random process. Similarly, in perception tasks people tend to reject sequences with long runs of the same result (such as a long sequence of heads) and consider sequences with an excess of alternation of different results to be random. Comparable results are found in people’s performance with two-dimensional tasks, in which clusters of points seem to prevent a distribution from being perceived as random. In addition to various psychological mechanisms such as *local representativeness* (Kahneman *et al*, 1982) that may be in operation and may explain these biases, some authors (e.g. Falk, 1981; and Falk and Konold, in press) believe that individual consistency in people's performance with diverse tasks suggests underlying misconceptions about randomness.

Konold *et al* (1993) used a different type of task consisting of having people judge the randomness of different natural phenomena. The results suggested that subjects’ conceptions about random experiments could be classified into the following categories that we believe parallel the historical interpretations of randomness outlined earlier:

- Subjects for whom an experiment is random, only if the possible results are equally probable. If the probabilities of the events involved are very different, they would not be considered to be random - such as raining on a day for which the possibility of it raining has been forecast as 80%.
- Randomness as opposed to causality, or as a special type of cause.
- Randomness as uncertainty; existence of multiple possibilities in the same conditions.
- Randomness as a model to represent some phenomenon, dependent upon our information about it.

Research into children's conception of randomness

From a didactic point of view, a crucial question is whether these biases and misconceptions are spontaneously acquired or whether they are a consequence of poor instruction in probability. Below, we outline a number of key research studies looking at children’s and adolescents’ conceptions of randomness, and their performance when faced with tasks requiring the generation or recognition of sequences of random results.

According to Piaget and Inhelder, chance is due to the interference of a series of independent causes, and the ‘non presence’ of all the possible outcomes when there are only a few repetitions of an experiment. Each isolated case is indeterminate or unpredictable, but the set of possibilities may be found using combinatorial reasoning, thereby making the outcome predictable. This notion of probability is based on the ratio between the number of possible ways for a particular case to occur and the number of all possible outcomes. This would suggest that chance and probability cannot be totally understood until combinatorial and proportional reasoning are developed and, for Piaget, this does not happen until a child reaches the formal operations stage (12-14 years).

Piaget and Inhelder (1951) investigated children's understanding of patterns in two-dimensional random distributions. They designed a piece of apparatus to simulate rain drops falling on paving stones. The desire for regularity appeared to dominate the young children's predictions. When they were asked where the following rain drop would fall, children at stage 1 (6 to 9 years) allocated the rain drops in approximately equal numbers on each pavement square, thereby producing a uniform distribution. With older children, proportional reasoning begins to develop, and Piaget and Inhelder report that such children tolerate more irregularity in the distribution. They believed that children understood the law of the large numbers, which explains the global regularity and the particular variability of each experiment simultaneously.

Green's (1983) findings, however, contradict those of Piaget and Inhelder. His investigations with 2930 children aged 11-16, using paper and pencil versions of piagetian tasks, showed that the percentage of children recognizing random or semi-random distributions actually decreased with age. A second study with 1600 pupils aged 7 to 11 and 225 pupils aged 13 to 14 is described in Green (1989, 1991). Using a slightly different task, Green found that general reasoning ability was a significant factor influencing children's responses, and concluded that more detailed investigations of children's concepts of randomness were needed.

In his second study, Green gave the children generation and recognition tasks related to a random sequence of heads and tails representing the results of flipping a fair coin. The study demonstrated that children were able to describe what was meant by equi-probable. However, they did not appear to understand the independence of the trials, and tended to produce series in which runs of the same result were too short compared to those that we would expect in a random process. In both studies, Green found that the perception of random sequences does not improve with age. Children did base their decisions on the following properties of the sequences: results pattern, number of runs of the same result, frequencies of results, and unpredictability of random events. However, these properties were not always correctly associated to randomness or determinism.

Toohey (1995) repeated some of Green's studies with 75 11-15 year-old students and concluded that some children have only a local perspective of randomness, while that of other children is entirely global. The local perspective of randomness emphasized the spatial arrangement of the outcomes within each square, while the global perspective concentrates on the frequency distribution of outcomes.

Other authors, such as Fischbein and Gazit (1984) and Fischbein *et al* (1991), have also documented children's difficulties in differentiating random and deterministic aspects, and their beliefs in the possibility of controlling random experiments. In contrast to the Piagetian view, these authors have inclined towards stating that even very young children display important intuitions and precursor concepts of randomness. They argue that it is not didactically sound to delay exploiting and building on these subjective intuitions until the formal operations stage is reached.

AIMS AND METHODOLOGY OF THE EXPERIMENTAL RESEARCH

In the following we present the results from a wider study (Serrano, 1996) in which 277 secondary school students were asked to answer a written class questionnaire including generation and recognition tasks, which were taken from Green (1991). About half of the students (n=147) were in their third year of secondary school (14 years-old) and had not studied probability. The rest of the students (n=130) were in their last year of secondary education (17-18 years-old, pre-University level). This second group had studied probability with a formal, mathematical approach for about a month when they were 15 years old and for another month during their previous school year. Five different schools in the town of Melilla (Spain) were used. The test took between 45 and 60 minutes to complete.

Our research was intended to compare secondary students' conception of randomness before and after instruction in probability, which occurs in the Spanish curriculum between the ages of 14 and 17. As Shaughnessy (1996) points out, there is still a need to document the growth of students' understanding of probability concepts over time. We also wanted to refine Green's research methodology, in particular to permit further analysis of children's reasoning when judging the randomness of two-dimensional distributions, and to compare students' conceptions with some of the interpretations that randomness has received along history. In this paper we present the responses to recognition rather than generation tasks. In Konold and Falk's opinion(in press), these reflect subjective conceptions of randomness more directly, because people may be able to perceive randomness, even if they are not able to generate an example of it.

A total of eight items will be analyzed, four of which refer to random sequences and the reaming to random two-dimensional distributions. For each type of item (random sequences and two-dimensional distributions) we first compare the percentages of students considering the situation to be random, using the Chi-squared test of independence between the response and group of students. Then, the students' arguments supporting or rejecting randomness are classified and compared by items, age group and type of judgement (randomness/ determinism), using again the Chi-squared test. Synthesis of the general tendency in these subjective meanings of randomness is finally achieved through correspondence analysis. This is a multivariate data analysis technique, which us applicable to qualitative data and serves to identify factors in a multiway contingency table. In our case, the 2216 responses (277 students x8 items) were classified by item, type of response (random or fake) and argument provided (7 different categories). The type of item (random sequence or two dimensional distribution) and group of student were used as supplementary variables, which does not intervene in the analysis, but are plotted on the graphs to help in improving the interpretations of factors.

Recognition of properties of random sequences

Firstly, we analyze the responses to items 1 to 4, in which we studied the students' capacity for discriminating random models in Bernoulli sequences. Two different variables were changed in the different sequences:

a) The proportion of heads: In items 1 to 3 this proportion is very close to the theoretical probability (P(H)=.53 in item 1, P(H)=.48 in items 2 and 3), while in item 4 the proportion of heads in only .30.

b)The length of runs, and, consequently the proportion of alternations (change in the type of outcome, from head to tail or from tail to head). This proportion is the quotient between the number of alternations and the sequence length and is very close to the theoretical value .5 in items 1 (P(A)=.54), 2 and 3 (P(A)=.51). On the contrary P(A)=.74 in item 2 that have too many alternations (too short runs) to be considered to be random.

Therefore, from a normative point of view, we would consider that the correct response to items 1 and 3 is that the child was playing correctly and for items 2 and 4 that the child was cheating. However, as in a test of randomness there is always a small probability that the sequence is not random, in spite of having passed the randomness tests, we are not just interested in whether the students responses coincide or not with the normative answer, but in the properties of sequences that associate to randomness.

Items 1 to 4. Some children were each told to toss a coin 40 times. Some did it properly. Others just made it up. They put H for Heads and T for tails.

MarÌa: T T T H T H H T T T H T H H H H H T H T H T T H T H H T T T T H H H T H H T H H

Daniel: H T H T T H H T H T H H T T H T T H H T T H T H H T T H T H T H T H T H T T H T

MartÌn: H T T T H T T H H H T H T T T T T H T H T H H T H T T H H H H T T T H T T H H H

Diana: H T T T H T T H T H T T T H T T T T H H T T T H T T H T T H T T T T H T T T H T

Item 1: Did MarÌa make it up? How can you tell?

Item 2: Did Daniel make it up? How can you tell?

Item 3: Did MartÌn make it up? How can you tell?

Item 4: Did Diana make it up? How can you tell?

The students' answers to these four items are presented in Table 1. Most students considered all the sequences to be random, except that in item 4, in which the frequency of heads (12) is quite different from the theoretical frequency expected in a random sequence.

Fifty-four 14 year-old students and thirty 17 year-old students gave a positive response to item 2, but most of the students considered the sequence to be random, in spite that the number of alternations (and consequently the length of runs) has been biased. This result confirms Green’s findings (1991) that suggested that the students have more difficulty in recognizing runs properties than frequency properties. The consistent percentages of students considering the sequences to be random in items 1, 2 and 3 indicates that the similarity between observed and expected frequencies may be more important than the length of the runs in deciding whether a sequence is random.

In all items, the Chi-squared test of independence between response and students group yielded a significant result (p-values were equal to .0011, .0045, .0028, and .0192 in items 1 to 4). A smaller proportion of 17-year-old students considered that the sequences were made up, and a higher proportion did not reach a conclusion. This also coincided with Green’s findings, and suggest that the older students are more cautious in their judgment, because they better appreciate the variability of random phenomena.

Table 1: Percentages of students' responses to items 1-4 and levels of significance in the

Chi-squared test

Sequence	Response	Age=14	Age=17
Maria: P(H)=.53	Real	60	58
P(A)=.54	Fake	34	27
	No decision	6	15
Daniel: P(H)=.48	Real	58	63
P(A)=.74	Fake	37	23
	No decision	5	14
Martin: P(H)=.48	Real	53	56
P(A)=.51	Fake	42	29
	No decision	5	15
Diana: P(H)=.30	Real	36	37
P(A)=.51	Fake	56	48
	No decision	8	15

We asked the students to justify their answers. The reasons that they gave were classified according to the scheme followed by Green (1991), which is described below. In some cases students might be reasoning according to the representativeness heuristic described by Kahneman *et al* (1982), where people tend to estimate the likelihood for an event based on how well it represent some aspects of the parent population.

a) *There is a regular pattern in the sequence:* This reasoning refers to the order in which heads and tails appear in the sequence, and to the regularity of this pattern: *"He did not cheat, because there is a very regular sequence of heads and tails, which is most probable to happen", "He might have made it up because the results are very uniform; heads and tails almost alternate".*

b) *There is an irregular pattern in the sequence:* For example, *"He did not make it up, because the sequence does not follow any order".*

In the last example or argument (b) the student associates the lack of pattern to randomness, which agree with the *complexity approach* to randomness described before, where the absence of patterns is an essential characteristic.

c) *The frequencies of the different results are quite similar:* For example, *"Because the numbers of heads and tails are very balanced", "The proportion of heads and tails is very similar", "The results are about 50-50".*

d) *The frequencies of the different results are quite different:* This is the opposite of the previous reasoning.

These two reasons, (c) and (d), are both based on some sort of comparison between the observed frequencies and the theoretical probability distribution. Here, the *frequentist approach* to probability and randomness, where an object is considered as a random member of a class if we can select it using a method providing a given “a priori” relative frequency in the long run, might underlay.

e) *There are long runs:* Here students show misconceptions concerning the idea of independence between successive trials*: "Because there are many consecutive tails".* Again, the students might be relying on the representativeness heuristic here.

f) *There are no runs:* This is the opposite to the previous reason, for example, *"There are too few sequences of heads".*

g) *The probability of different events must be the same:* This reason is related to (d), but here there is an allusion to probability. A few subjects' responses do fall into this category, stating for example, *"There must be equal probability for heads and tails".* In this case students relate randomness to equiprobability, using a *classical approach* to this concept, where an object is a random member of a class if there is the same probability for this object and for any other member of its class.

h) *It is unpredictable; it is random:* Students giving this response might say, for example, *"That is luck", "Though correctly flipped it can give those results, because it is a game", "When flipping a coin, there is no certain result, here randomness plays a very important role".*

Here the "outcome approach" described by Konold, where people interpret questions about probability in a non probabilistic way and rely on the unpredictability of random events to reject taking a decision. Belief in an underlying causal mechanism, which parallel the earlier meaning of randomness, might also be implicit in some of these categories of student explanations. This point is difficult to decide with our data, and would require further research using interviews and a more varied types of tasks.

As shown in Table 2, the highest frequencies appear in responses based on unpredictability or luck, somewhat in contrast to Green's findings (1991). We are also able to detect variations in the more common responses to the different sequences: regular pattern in item 1; irregular pattern in items 2 and 4; difference of frequencies in item 2, 3 and 4; long runs in items 2 and 3. This variation suggests that the students were able to discriminate between the features of the different sequences. Green (1991) did not look at variation by items in this way.

Table 2: Percentages of students' arguments in Items 1-4 according age

	MarÌa		MartÌn		Daniel		Diana	
Argument	14	17	14	17	14	17	14	17
Regular pattern	6	3	7	4	44	18	10	2
Irregular pattern	13	5	18	10	9	1	8	5
Similar frequencies	4	12	0	10	5	23	3	2
Different frequencies	9	8	9	7	0	4	13	27
Runs	21	13	24	18	4	5	22	16
Impredictibility	35	33	32	28	31	29	31	28
No argument	12	26	10	23	7	20	13	20

In each item we carried out the chi squared test of independence between the argument and group of students (p-values were equal to .001, .002, .0001 and .02). Younger students made more reference to the *pattern* of a sequence, whereas 17 year-old students used *frequency* arguments more often. As it is shown in Table 3, the responses that they gave also differed depending on whether the sequence was considered to be random or not (p. values in the Chi-squared test of independence between argument and type of response were equal to .0001 in the four items). The main argument given to support randomness was unpredictability. Irregular pattern, long runs, and non-coincidence of the outcome frequencies were associated with lack of randomness.

Table 3: Percentage of students' arguments in items 1 to 4 according to whether the sequence is considered to be random or not

	MarÌa		MartÌn		Daniel		Diana	
Argument	Real	Fake	Real	Fake	Real	Fake	Real	Fake
Regular pattern	5	6	1	13	12	81	13	3
Irregular pattern	14	2	23	5	8	1	7	7
Similar frequencies	10	6	7	2	19	6	5	1
Different frequencies	5	18	4	18	3	1	5	35
Runs	2	53	4	54	4	6	4	33
Impredictibility	52	12	48	5	43	5	51	18
No argument	12	3	13	3	11	0	15	3

RECOGNITION OF PROPERTIES IN TWO-DIMENSIONAL ARRAYS

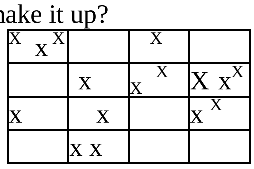
We included four further items in which students were requested to indicate their perception about whether the points in a two-dimensional array were distributed randomly or not. Our aim was to analyze the properties attributed by students to random two-dimensional distributions and the differences between the two age-groups. In order to make the stochastic model clear for the students the following introductory activity was set:

Paul plays a game using 16 counters numbered 1, 2, 3, 4,...16. Paul puts all the counters in a tin. He shakes the tin a lot. Rachel shuts her eyes and picks out a counter. It is number 7. Paul puts a cross in box 7. The 7 is put back in the tin and someone else picks out a counter.

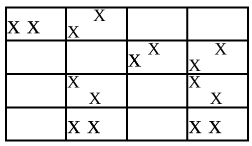
1	2	3	4
5	6	7x	8
9	10	11	12
13	14	15	16

After reading out this text, children were told to play this game until they understood the rules. They were then given two items exploring the children's perception of what might happen in the experiment if continued over 16 and 30 selection (generation tasks). The last four questions (the recognition tasks) were as follows:

Items 5 to 8: Some children were told to play the counters game by themselves using 16 real counters. Did some cheat and make it up?



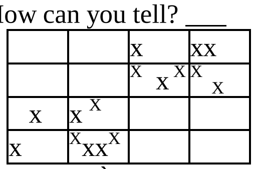
Jaime



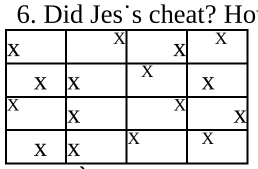
Jes˙s

5. Did Jaime cheat? How can you tell? ___

6. Did Jes˙s cheat? How can you tell? ___



MarÌa



LuÌs

7. Did MarÌa cheat? How can you tell? ___

8. Did LuÌs cheat? H--------------------

Table 4: Percentages of students' responses to items 5-8

Array	Response	Age=14	Age=17
Jaime: χ^2=1	Real	85	85
	Fake	14	12
	No decision	1	3
Jes˙s: χ^2=11	Real	57	58

	Fake	42	38
	No decision	1	4
MarÌa: χ^2=2.3	Real	36	38
(Diagonal)	Fake	63	58
	No decision	1	5
LuÌs: χ^2=24.3	Real	25	22
	Fake	74	75
	No decision	1	3

In a theoretical random distribution of the number of counters, about 6 empty squares, 6 squares with only one counter, 3 with two counters and 1 with 3 o more counters will be expected. The chi-squared test of goodness of fit between the theoretical and the observed distribution in each of the items yielded a value of 1 for item 5 (Jaime), 11 for item 6 (Jes˙s), 2.3 for item 7 (MarÌa), and 24.3 for item 8 (Luis). In addition, Maria's pattern has too many adjacent empty squares, all of them on the main diagonal. Consequently, only Jaime's distribution would be considered random from a normative point of view.

There was a greater spread in the number of students considering the distributions to be random now (Table 4). The percentage of students ranged from 85.1% and 85.4% in item 5, to 25.2% and 22.3 % in item 8, where exactly one point is distributed in each square. There were no significant differences depending on the age-group in the proportions considering the distributions to be random in the Chi-squared test. This is not surprising as, in Spain, teaching does not include this type of activity. The percentages were also similar to those obtained by Green (1991) in 13-14 year old students. Toohey's study (1995) did not include such comparison data. In neither Green (1989, 1981) nor Toohey (1995) were the students asked for reasons to justify their answers, as they were in the present study. We have grouped the students' responses into categories that can be compared with those used in items 1 to 4:

a) *There is a regular pattern in the distribution of points:* Similarly to the previous items, the subjects noted the presence of some regularity in the spatial arrangement of results presented and argued that the points distribution followed a regular pattern. For example, *"It is all too correct, all in its place, well ordered", "It is too lucky to have all the squares in the top right and bottom left corners with dots and nothing in the others".*

b)*The sequence follows an irregular pattern:* This is the opposite reason to the previous one. When the student associates the lack of pattern to randomness, the *complexity approach* to randomness might be implicit.

c) *The frequencies of the different results are similar:* This characterizes the reasoning of students who find the frequencies of different results to be too similar. It was the favorite argument to justify item 5, referring to Luis. The following statements are variations on the same theme. *"It is very difficult not to have some square repeated", "It is very difficult that all the squares should appear the same number of times".*

d) *The frequencies of the different results are different:* In contrast, subjects responding in this category gave more importance to the differences between the results presented: *"The results are quite different, from an X to 4X and therefore reasonable", "I believe that we should get more even results, not so different the number of times that squares are repeated".*

We can observe the *frequentist approach* to probability implicit in these arguments c) and d).

e) *There is a cell with too many points:* This argument is frequently used to justify items 6 and 8: *"Some squares have too many crosses"*. This would be equivalent to an argument that the runs are too long in items 1-4, because it implies similar misconceptions about the idea of independence.

The following arguments were hardly used; although we have included them in order to maintain consistency with the classification of arguments in items 1 to 4.

f) *There are no cells with several points.* This is the reversal of the previous reason.

g) *It must be equal possibilities in the number of points by square.*

h) *Unpredictability of the results in random experiments.* Students giving this type of response seems to reason by the 'outcome approach' or by belief in an underlying causal mechanism. For example, *"Such is each person's luck, we cannot guess the result"*.

Table 5:Percentages of students' arguments in Items 5-8 and level of significance in the Chi-squared test

	Jaime		Jes˙s		MarÌa		LuÌs	
Argument	14	17	14	17	14	17	14	17
Regular pattern	1	1	52	45	7	45	38	6
Irregular pattern	36	21	10	5	24	5	0	0
Similar frequencies	1	0	3	5	0	5	41	65
Different frequencies	9	11	0	0	5	0	1	1
Clusters	8	2	0	0	26	0	1	0
Unpredictability	30	44	24	35	29	35	10	22
No argument	15	21	11	10	9	10	9	6

The students' responses are presented in Table 5. The more frequent reasons were: too regular pattern, irregular pattern, similar frequencies, different frequencies, cells with too many points, and unpredictability. The most frequent response did change according to the item, which suggests that the students are able to distinguish the characteristics presented in the distributions. Regular pattern is more frequently associated with items 7 and 8, irregular pattern with items 5 and 6, equality of frequencies with item 8, and cells with too many points with items 5 and 6.

We found significant p-values in the Chi squared test of independence between student group and arguments, except in item 6 (p-values were equal to .004, .001, and .001 in items 5, 7, and 8). The older students seemed to favor frequency-based reasoning and unpredictability, whilst the 14 year-olds tended to comment on the regularity or irregularity of the pattern.

Table 6: Percentages of students' arguments according to whether the sequence is considered to be random or not

	Jaime		Jes˙s		MarÌa		LuÌs	
Argument	Real	Fake	Real	Fake	Real	Fake	Real	Fake
Regular pattern	3	1	9	74	2	17	14	27
Irregular pattern	31	20	12	6	28	6	0	0

Similar frequencies	0	0	0	1	0	0	25	63
Different frequencies	11	5	4	2	4	5	3	0
Clusters	1	33	0	2	6	47	0	0
Unpredictability	39	25	62	10	46	20	40	8
No argument	17	14	14	5	14	5	18	2

We also found significant p-values in the Chi-squared test of independence between argument and whether the student considered the distribution to be random or not (p-values were equal to .0001, .009,.004, and .0001). In Table 6 we can see that the reasons given when the student did not believe that the distribution was random were; regular pattern, the equality of frequencies of the different results, and the existence of cells with too many points. The irregularity of pattern, and unpredictability, were more often used to support a belief in the randomness of the distribution.

Factors affecting the perception of randomness

One research issue was whether the students employed the same arguments to reject or to confirm randomness, and whether these arguments changed according to the task variables in the items. In the previous section we observed that particular responses appeared to be associated with different items. We have also found variations between the responses employed to justify saying that a sequence was random (the child was not cheating) or otherwise. Since the classification of responses was made to allow comparison in the eight situations proposed, we have completed our results with a correspondence analysis, using BMDP software, with the aim of identifying which task variables were associated with each argument.

The results of the analysis support our hypothesis that students recognized the characteristics that we changed in the items and, furthermore, that they used them to decide if the situation was random or not. These characteristics may be summarized in terms of the three factors identified below which, between them, explain 87.5% of the total variance:

1. Students attribute a local variability to random phenomena. Because of this, they expect frequent alternation of results and a lack of pattern in the order or in the spatial arrangement of results.

2. Students expect a global regularity given by the similarity between the frequencies of different results and the theoretical probabilities.

3. However, some discrepancy (though not much) between the theoretical and the observed distribution is expected. These discrepancies are easily observed in the linear sequences, but are not so obvious in the items concerned with distributions of points in arrays.

DISCUSSION OF FINDINGS

Our results show a mixture of correct and incorrect properties associated by students to randomness. On one hand. they perceived the local variability, lack of patterns in the lineal or spatial arrangement of outcomes, and unpredictability of the random processes underlying the tasks we gave them. In many cases, our students carried out a statistical analysis of the frequencies for the different events in the random sequences and compared these frequencies with an underlying equiprobability model. As regards to two dimensional distributions, they rejected the item with only one counter by square as well as that with too many adjacent squares. We should take into account all these correct intuitions when organizing the teaching of probability, because when learning something new students construct their own meaning by connecting the new information to what they already believe to be true.

We also must consider the students' mistaken conceptions a review our ways of teaching in the light of changes after instruction: There was a too strong emphasis on the unpredictability and luck to justify randomness and this tendency seemed to increase in the older students when analyzing the two dimensional distributions, a type tasks that are not considered in the Spanish curriculum. The existence of runs of 5 and even 4 identical results in the sequences (squares with cluster of 3-4 markers in two dimensional distributions) was mistakenly associated by the students to lack of randomness, though after instruction a smaller proportion of students provided this type of arguments. Because understanding the existence of runs and clusters implies understanding independence, our results suggest that independence is not an intuitive idea and that students continue having difficulties with the idea of independence after instruction.

The students' arguments and responses also suggest underlying conceptions that parallel some of the meanings that randomness has received along history. In particular, students relate randomness to luck and unknown causes, and to probability in its classical and frequentist approach. Some of the features of the complexity approach (lack of pattern) and selection algorithms (unpredictability) are also shown in their responses. All these results are very close to that by Green, and therefore, his results could be extended to our group of 17 year-olds students. In addition, we have extended the classification of students' arguments which Green carried out for linear sequences to two-dimensional distributions and have achieved a general synthesis using correspondence analysis.

In summary, our experimental results as well as our previous analysis reveal the complexity of the meaning of randomness for which different properties need to be understood. It may in fact be preferable to consider the term *randomness* as a 'label' with which we associate many concepts, such as experiment, event, sample space, probability, etc (Konold et al., 1993). In this sense, the word 'randomness' refers us to a collection of mathematical concepts and procedures that we can apply in many situations. We should think more about an orientation we take toward the phenomenon that we qualify as 'random' rather than a quality thereof. We apply a mathematical model to the situation, because it is useful to describe it and to understand it. But we do not believe that the situation will be 'identical' to the model. Deciding when probability is more convenient or adapted to the situation than other mathematical models is part of the work of modeling that we should encourage among our students.

REFERENCES

Ayer, A. J. (1974). El Azar [Chance]. In Kline, M. (Ed.), *Matemáticas en el mundo moderno* [Mathematics in modern world] (pp. 172-171). Barcelona: Blume.

Ayton, P., Hunt, A. J., & Wrigth, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making, 2,* 221-238.

Batanero, C., & Serrano, L. (1995). La aleatoriedad, sus significados e implicaciones educativas. *Uno: Revista de didáctica de las matematicas,* (5), 15-28.

Batanero, C., Green, D. R., & Serrano, L. R. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematical Education in Science and Technology,* 29(1), 113-23.

Batanero, C., Serrano, L., & Garfield, J. (1996). Heuristics and biases in secondary students' reasoning about probability. In Proceedings of the XX PME Conference (Vol. 2, pp. 43-50).

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

Bennett, D. J. (1993). *The development of the mathematical concept of randomness; educational implications.* Ph. D. New York University. (DAI n. 9317657).

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. (1994). Random means hard to digest. *Focus on Learning Problem Solving in Mathematics, 16,* 2-12.

Falk, R., & Konold, C. (In press). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review.*

Fine, T. L. (1973). *Theories of Probability: An examination of foundations.* New York: Academic Press.

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

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics, 22,* 523-549.

Green, D. R. (1983). A Survey of probabilistic concepts in 3000 students aged 11-16 years. In D.R. Grey et al. (Eds.), *Proceedings of the FirstInternational Conference on Teaching Statistics* (v.2, pp. 766-783). University of Sheffield.

Green, D.R. (1989). Schools students' understanding of randomness. In R. Morris (Ed.), *Studies in Mathematics Education. v.7: The Teaching of Statistics* (pp. 27-39). Paris: Unesco.

Green, D.R. (1991). *A longitudinal study of pupil's probability concepts.* Technical Report ME90/91. Loughborough University of Technology.

Hawkins, A., Jolliffe, F., & Glickman, L. (1991). *Teaching Statistical Concepts.* London: Longman.

Harten, G., & Steinbring, H. (1983). Randomness and stochastic independence. On the relationship between intuitive and mathematical definition. In R. W. Scholz (Ed.), *Decision making under uncertainty* (pp. 363- 373). Amsterdan: Reidel.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: heuristics and biases.* Cambridge: Cambridge University Press.

Konold, C. (1991). Understanding students' beliefs about probability. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 139-156).

Dordrecht: Kluwer.

Konold, C., Lohmeier, J., Pollatsek, A., & Well, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education,* 24, 392-414.

Kyburg, H. E. (1974). *The logical foundations of statistical Inference.* Boston: Reidel

M.E.C. (1992). *Matem·tica. Secundaria Obligatoria.* [Secondary school mathematics]. Madrid: Ministerio de Educaci¤n y Ciencia.

Piaget, J., & Inhelder, B. (1951). *La genése de l'idée de hasard chez l'enfant.* [The origin of the idea of chance in children]. Paris: Presses Universitaires de France.

Poincar*é*, H. (1936). El Azar. [Chance].*Journal of the American Statistical Association,* v. 31, pp. 10-30. In J. Newman (Ed): *Sigma. El mundo de las Matemáticas* [World of mathematics] (v. 3, pp. 68-82). Barcelona: Grijalbo.

Shaughnessy, J. M. (1996). Emerging research issues in the teaching and learning of probability and statistics. Paper presented at the *Eigth International Congress on Mathematics Education (ICME 8),* Sevilla, July, 1996.

Serrano, L. (1996). *Significados institucionales y personales de objetos probabilÌsticos ligados a la aproximación frecuencial de la enseñanza de la probabilidad* [Institutional and personal meanings of mathematical objects linked to the frequentist approach to teaching probability]. Unpublished Ph. D. University of Granada.

Toohey, P.G. (1996). *Adolescent perceptions of the concept of randomness.* Unpublished Master Thesis. University of Adelaide.

Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin,* 77, 65-72.

Zabell. S. L. (1992). The quest for randomness and statistical applications. In F. Gordon and S. Gordon (Eds.), *Statistics for the XXI Century* (pp. 139-150). The Mathematical Association of America.