WHICH PROBABILITY DO WE HAVE TO MEET …? A CASE STUDY ABOUT STATISTICAL AND CLASSICAL APPROACH TO PROBABILITY IN STUDENTS' BEHAVIOUR

Rossella Garuti, Aurelia Orlandoni and Roberto Ricci
National Agency for Educational System Support, Bologna, Italy
ricci@irreer.it

This case study report describes a training activity for in-service teachers working on a project related to statistical literacy. Collected data from the solutions to a problem given to the students are taken into account. The problem was chosen to give teachers the opportunity to reflect on their methods for teaching statistics and probability. Results about student understanding of statistical and classical probability and about the teacher methodologies to present these key concepts are described.

INTRODUCTION

Statistical and probabilistic reasoning has recently assumed a new role within the context of secondary level mathematics. Indeed, educational science recognises the need for secondary school mathematics to be more closely related to citizens' daily life (Ottaviani, 2005). Surely citizens should be capable of thinking quantitatively, for instance reading a newspaper intelligently, making decisions based on quantitative information (Moreno, 2002). This new way of understanding a citizen's mathematical literacy is a relevant facet of international skill assessment projects, such as the Programme for International Student Assessment (PISA). In particular, in Italy statistical and probabilistic concepts have been included at every level of schooling. This makes it very important for students and teachers to have access to tools that aid their learning-teaching experience in order to handle concrete problems connected with real life and with the scientific disciplines included in their educational curricula to stimulate the scientific interests of the students.

The principal aim of the present report is to deal with some aspects related to students' (grades 7 to 10, ages ranging from 12 to 15) behaviour when given a task about the connection among different approaches to probability (classical, frequentist, and subjective approach). Furthermore, teachers' reflections on their students' behaviour throw a light on the need to connect statistics and probability. The following section provides a rationale for analysing students' behaviour when they are asked to complete a task where different approaches to probability could conflict. According to Batanero, Henry, and Parzysz (2005), "epistemological analysis plays a fundamental role for mathematics educators because (…) it can help us understand students' difficulties in learning mathematics."

Three distinct interpretations of probability are possible and therefore can emerge in the teaching and learning of statistics and probability. The first one is based on the classical view of probability where outcomes of a sample space are assigned known probabilities. In contrast the frequentist approach is based on the frequency view of probability. A third probability view is the subjective view, where probability reflects an idea of the likelihood of an event. To throw a light on these aspects is crucial to further learning of statistics, in particular when students are asked to make inferences and to avoid misconceptions arising from the lack of connections between statistics and probability.

CONTEXT

About 50 teachers, especially of mathematics and also of natural sciences, history and Italian, are involved in the project "Bring up to Uncertainty." The project is supported by funding from the Italian region Emilia Romagna, and it is addressed to teachers of the grades 7 to 10. These grades are crucial because they represent the transition phase from the lower to the upper secondary school. In order to provide proper tools consistent with the way of teaching statistics advocated by the project, the Emilia Romagna department of the National Agency for Educational System Support, in association with the Faculty of Statistical Science of the

 In C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.*

University of Bologna, has added to Fardiconto, a web environment devoted to mathematics, a new project called Statisticamente, a way of reasoning about real life problems.

Despite the introduction of statistics and probability at every school level in the curricula starting from the nineteen eighties, in Italy the teaching of these topics is still not integrated in the curriculum. According to Borovcnik and Peard (1996), in Italy data-analysis is also not adequately addressed in classrooms.

METHOD

Involved teachers were asked to present to their students the problem given in Figure 1 taken from Nucleo di Ricerca in Didattica della Matematica (NRD) Project of Genoa University, independent of the scheduling of the year's normal teaching activity. The main ideas are to observe students' reactions when given a problem that is aimed to enlighten the relationship between a classical and a frequentist approach to probability and to share with the teachers reflections about the students' behaviour.

The boxing problem.

The serious accidents that in the past happened in the boxing world awaken a vivid debate about the opportunity to forbid such a violent sport. Following are the results of some opinion polls:

- SAMPLE 1: In a technical high school of Genoa 808 students (75% males) are interviewed. 51.7% are against a boxing ban; 48.3% agree with the prohibition of such a sport.
- SAMPLE 2: A Naples newspaper promotes an opinion poll in which the reading public can express by phone what they think about the boxing issue. 1120 persons responded by phone to the newspaper; 790 of them are for the prohibition of boxing.
- SAMPLE 3: A journalist of a well-known newspaper has interviewed 338 individuals of different ages all around the country, half females and half males. 237 persons are for the boxing ban, and 107 are against it.

a. Fill in the following table by using both the absolute data and relative data (percentages):

		IN FAVOUR		AGAINST	
	Total	Absolute Data	Relative Data (%)	Absolute Data	Relative Data (%)
SAMPLE 1					
SAMPLE 2					
SAMPLE 3					

b. In your opinion, which sample gives more reliable information about the belief of Italian people relating to the boxing ban? Why? (Explain your answer).

c. If you are interviewing a randomly chosen Italian person, can you evaluate the probability that he or she is for the boxing ban? Can you argue for your answer?

Figure 1. The boxing problem

In the *a priori* analysis two crucial aspects of the problem arise:

- The first one concerns the expression "more reliable information" (point *b*) and the meaning the students would have given to it. Not many classes had explicitly worked on this topic; therefore we are particularly interested in finding out the leading factors in students' thinking: i.e., we are interested in enunciating the reasoning steps that students follow: sample dimension? Or the characteristics of the various samples taken into account? Or the everyday life common sense attributed by the students to the term "reliable"?
- The second matter concerns the question about probability asked in point *c*. In general, in Italy, in Grades 7 and 10, probability begins with game examples like coins and dice,

where the event space is always built on constituents that are assigned equal probabilities. Seldom are examples of events that are not equally likely provided to the students (Consogno, Gazzolo & Boero, 2006). Furthermore we can observe that the link between statistics and probability is, in general, ignored, so that the students do not have an opportunity to face situations in which the probability of an event depends on a statistical evaluation. The boxing problem, therefore, can be a good tool to investigate this situation.

RESULTS

The boxing problem was presented to 608 students who were distributed according to Table 1.

Table 1. Students' distribution

Grade	7	8	9	10	Total
Students	34	345	211	18	608

Students' responses: Relative and absolute values

About 30% of students showed consistent difficulties in filling in the table in part *a*. In particular, about 17% of the students made mistakes by transforming absolute data into percentages or vice versa. The missing answers were about 17% of the first type of transformation, while the second one was about 10%. Typical mistakes arose from a lack of understanding the meaning of the results; as an example, consider the value of "number of persons" expressed by a non-integer number that is the result of a percentage (48.3% of 808 = 390.26). Moreover, many students failed to check the sum of the percentages, which rarely is equal to 100.

An interesting aspect concerns the different strategies used to calculate the percentage. While Grade 9 and 10 students used the ratio 231/338 = 0.6834, obtaining 68.34%, students in Grade 8 generally proceed through the proportion $x:100=231:338$ and produced "meaningless" results. In some cases we also found wrong results such as 146%. It seems that the students' attention was focused on the proportion procedure losing the meaning of the result. On the other hand those who used the ratio-strategy showed a greater control of the meaning.

From the mathematical point of view the two strategies are equivalent, but, according to Arzarello, Bazzini, and Chiappini (2000), different procedures can be interpreted according to the distinction between "Sinn" (sense) and "Bedeutung" (reference, denotation also meaning) of an expression: "(…). In mathematics some expressions have senses that are different but which have the same denotation. For example, the expressions $4x+2$ and $2(2x+1)$ mean a different rule (sense) but denote the same function (Arzarello et al., 2000). In our problem the meaning is the same, the percentage, but the senses are different: in one case there is a proportion to be solved, in the other a ratio to be calculated. This could explain, at least in part, the fact that those who used the proportion had a lesser control on the result, losing the meaning, while the students who used the ratio (part in relation to the whole) maintained a greater connection with the meaning of percentage.

Students' responses: Sample choice

As expected from the a priori analysis of the problem, the students attributed various meanings to the expression "reliable sample." In some cases they were led by their impressions about the question under consideration. "I think that sample is the most reliable, because I don't want boxing to be forbidden" (Grade 8). Other answers were more complex though not exhaustive. " I think that number 2 is the most reliable sample because many people have been interviewed, and many of them are in favour of boxing, and the opinion poll has been performed in Naples where everyday fights take place" (Grade 9). The size of sample 2 as well as the typology of the opinion poll were the reason for the choice by many students: sample 2 gives the most trusted indication because it has taken the greatest number of people into consideration. Moreover the 1120 persons have spontaneously called, showing more attachment

to their choice" (Grade 9). The expectation about the possible results played a role in the choice of the sample: "I think that sample 1 is the most reliable because the percentages are quite the same, around 50%, while in samples, 2 and 3 the results differ a lot " (Grade 10). Those who accurately analysed the characteristics of the three samples in general produced an answer of this kind: "Sample 3 is the most reliable because it includes various ages, in comparison with sample 1, and most of all it is based on the opinion of people from many areas of Italy" (Grade 8).

Students' responses: Which probability do we have to meet a…?

We classified students' behaviour according to the epistemological analysis of the probability interpretation and to the main answer typologies.

1. The statistical student–"The most reliable sample is number 3, because the opinion poll includes all regions of Italy, different ages and involves half men and half women. Sample 1 includes only students and sample 2 only people from Naples. The probability to meet a citizen favourable of boxing prohibition is 68.3%, because out of 338 persons taken at random, 238 are in favour and 107 against. I have calculated the percentage and the probability is more or less than one" (Grade 9). The students who answered in this way showed the ability to analyse the three samples introduced, grasping their characteristics and relating the probability question to the statistical data.
2. The mathematical student–"I think that sample 2 is the most reliable because it includes 1120 persons and gives absolute values and not percentages. The probability to meet a citizen in favour is ½ because the other option is certain to be against, and you don't know what a citizen could answer" (Grade 8). Another student writes, "Sample 2 is the most reliable because it includes the greatest number of people, so there are more opinions than in the other sample. More opinions depict better the opinion of all Italians. Since there are only two answers, in favour or against, the probability to be in favour is 1/2, or 50%" (Grade 9). In this case students showed the ability to grasp only the different characteristics of the samples and did not take into account the samples chosen in order to answer the question about probability. They ignored the link between statistics and probability, or they had a very superficial understanding of it, and most of all they did not have a critical knowledge of the classic definition of probability, taking for granted that the events would be equally likely.
3. The sympathetic student–"The most reliable sample is 3 because it includes few people against boxing, which is, in my opinion, a very exciting sport. The probability to meet a citizen in favour of prohibition is very high, since the majority of Italians are not interested in boxing" (Grade 8). These students did not take into consideration the data and based their answer on their feelings and impressions; in fact they seemed to answer questions like: "What do you think about forbidding boxing?" Or " What do you think of the opinions reported in the text?" by mixing up the data and personal feelings. These kinds of answers seemed to give some elements of the subjective approach to probability. This is very interesting since it reflects common beliefs of people about statistical investigation on various topics.

Table 2 summarizes the distribution of the students who reacted to the question (70% out of total).

Table 2. Response Distribution, *n*= 608

Student type	%
Statistical	49.6
Mathematical	29.3
Sympathetic	21.1
Total	100

Teachers' reflections

The analysis reported in the previous section was the focus of a discussion with teachers involved in the project "Bring up to Uncertainty." The first step of the discussion was a teacher debate about the ideas and foresights of the student behaviour when facing the problem and of the student difficulties when solving the boxing problem. A large number of teachers thought that students could have calculation difficulties and problems in arguing their choices, particularly the selection of the most reliable sample. Teachers were quite confident of the student capability to calculate percentages and of the possibility students would giving a correct answer to the probability questions. Another teacher discussion followed underlining the following aspects as the most relevant.

- Percentage meaning: Teachers were disappointed with student difficulties when calculating percentages and with the mistakes that students made by filling the table with absolute data. However, after this first phase, the teachers' reflection about the meaning of percentages considering the results was particularly fruitful. Lower secondary school teachers (Grade 8) explained the student tendency to use a proportion in order to calculate the percentages as a consequence of the emphasis that this topic receives in the mathematics curriculum. The most interesting aspect of this discussion phase was the teachers' reflection about the different meaning of percentages (proportion and ratio) and about the consequences when checking the results that follow from these two different approaches.
- Sample choice: Teachers were positively surprised by the large variety of student answers. This provided a meaningful opportunity to expand in further research projects a discussion with students to reach an explicit and shared definition of representative sample.
- Probability: Students' behaviour analysis about the probability questions shows evidently the lack of a didactic aimed to build a fruitful connection between statistics and probability and, on the other hand, a clear tendency to face separately the two topics. In general, a statistical approach to probability is quite uncommon, and the classical definition is the most used in the everyday didactic praxes. Furthermore, it is quite evident that a deficit exists even in the approach to classical probability, which is often presented in a simplistic way, and the relevance of assumption of equally likely possible results of an event is seldom enlightened (Gazzolo & Massi, 2003). Even some teachers said that at first sight they had used the classic approach to the probability without reflecting in depth about the nature of the problem. Although the number of teachers that declared this problem was not large, it showed, in some cases, a lack of knowledge of probability and statistics. This deficit may cause difficulties when approaching the statistical definition of probability. Several teachers wondered at the large number of "sympathetic students", who sometimes gave imaginative answers. These answers showed that students had difficulty in perceiving the connection with the statistical meaning of the problem to solve. Students often reacted to the question on the basis of their personal impressions, and they did not take into account the information that was given by the problem. By discussing in depth this latter aspect, the teachers reflected on the opportunity to improve the statistical literacy of their students, so that their students, future citizens, will be capable of thinking quantitatively.

DISCUSSION

Statistics and probability are present in all curricula of the Italian school system, but, at the moment, the didactic of these two crucial fields is not sufficiently developed. We think that even with very young students it would be opportune to introduce the notion of "probability of an event" as the ratio between the number of favourable outcomes and the number of all possible equally likely outcomes. In the early history of probability, this construction meant the separation between the magic view of random events and their evaluation in terms of objective measures of probability (Hacking, 1975). However, it is also crucial to connect the introduction of probability with the statistical aspects that are involved in everyday life. Students are quite

often not able to appropriately take statistical information into account to evaluate probabilities. On the other hand, it is quite remarkable that teachers lack specific knowledge in this field. In particular, it is extremely important that teachers are conscious of the three approaches to probability in order to understand and overcome the students' difficulties.

Moreover, according to Estrada, Batanero, and Fortuny (2005), it is " (…) useful for teachers' training to consider the appropriate formative experiences that will foster the prospective teachers' capacity for ongoing statistical learning, help them to reflect on the nature of statistics and help them to value statistical knowledge and literacy in improving the education of all the citizens."

REFERENCES

Arzarello, F., Bazzini, L., & Chiappini, G. (2000). A model for analysing algebraic processes of thinking. In R. Sutherland, T. Rojano, A. Bell & R. Lins (Eds.), *Perspective on school algebra* (pp. 61-82). Dordrecht: Kluwer.

Batanero, C., Henry, M., & Parzysz (2005). The nature of chance and probability. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 20-42). New York: Springer.

Borovcnik, M., & Peard, R. (1996). Probability. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 239-288). Dordrecht: Kluwer.

Consogno, V., Gazzolo, T., & Boero, P. (2006). Developing probability thinking in primary school: A case study on the constructive role of natural language in classroom discussions. *Proceedings of International Group for the Psychology of Mathematics* Education (Vol. 2, pp. 353-360). Prague, Czech Republic: PME

Estrada, A., Batanero, C., Fortuny, J. M., & Diaz, C. (2005). A structural study of future teachers' attitudes towards statistics. In A. Mariotti (Ed.), *Proceedings of the Third Conference of the European Society for research in mathematics education*. [CD-ROM]. Bellaria Italy. CERME.

Gazzolo, T., & Massi, L. (2003). Alle radici del pensiero probabilistico. (To the root of statistical thinking). Online: didmat.dima.unige.it/miur/miur_dima/B/probabil/pres.html.

Hacking, I. (1975). *The emergence of probability*. Cambridge, MA: Cambridge University Press

Moreno, J. L. (2002). Toward a statistically literate citizenry: What statistics everyone should know. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. [CD-ROM]. Cape Town: International Association for Statistical Education.

Ottaviani, M.G. (2005). Research into statistics education as a discipline. *International Statistical Review, 73*(2), 207-209.