HIGH SCHOOL STUDENTS' PROBABILISTIC REASONING WHEN INTERPRETING MEDIA NEWS

Rocío Álvarez-Arroyo[1], José F. Lavela[2], María M. Gea[1], and Carmen Batanero[1]
[1]Universidad de Granada, Spain, [2]IES Marqués de Comares (Lucena), Spain
mmgea@ugr.es

The purpose of this research was to analyse high school students' probabilistic reasoning when interpreting news from the media. To fulfil this goal, 76 high school students were given six questions related to a report on traffic accidents taken from the media news. The questions involved some conditional probabilities and the critical reading of the information presented. Although the computation of the complementary event probability was easy, there were more troubles in dealing with other conditional probability questions. Some students relied too much on the authority principle to interpret the data and did not reach a level of critical reading of the information. Finally, few students were able to identify the missing information needed to apply Bayes' theorem.

INTRODUCTION

There is an increasing interest in the teaching of probability at different educational levels, given that recent research suggests even young children develop ideas of probability (Nikiforidou, 2018). Probability is the only branch of mathematics that deals with uncertainty, leading to a new type of reasoning (Sharma, 2015), and its knowledge is essential for students' lives due to the existence of randomness in many everyday and work situations (Batanero & Borovcnick, 2016). Reacting to media reports or making decisions when facing risk situations are typical situations in which critical reading of information is a fundamental prerequisite (Gigerenzer, 2002), as it is shown in the current study.

To respond to the need of probability literacy for all citizens, the study of probability is included in Spain in Compulsory Secondary Education and is continued in high school with conditional probability, Bayes' theorem, and probability distributions (Ministerio de Educación, Cultura, y Deporte, 2015). However, teaching at these educational levels is based almost exclusively on learning definitions and properties and solving textbook problems (Muñiz-Rodríguez et al., 2016); it is unusual to confront students with real situations based on the news in which they must apply their knowledge and reasoning about probability. To fill this gap, in this paper, we assess the probabilistic reasoning of a group of high school students when interpreting a report on traffic accidents taken from the media, through conditional probability questions and critical reading of the information presented.

FRAMEWORK

Some authors have described characteristics of probabilistic literacy and reasoning. According to Gal (2005), probabilistic literacy implies appropriate use of probability language and competence in computing or estimating probabilities in varied situations. It also involves a basic knowledge of some fundamental ideas, particularly randomness, variability, independence, and predictability/uncertainty. Probabilistic literacy helps people understand probability statements in situations such as interpreting a medical diagnosis or in making a decision under uncertainty. Moreover, a probability literate person appreciates the role of probability in different contexts and can pose critical questions when encountering information about a random situation. Finally, Gal (2005, 2009) suggests that probability literacy includes a number of dispositions, such as a critical stance towards probabilistic information, overcoming mistaken beliefs, and a positive attitude towards learning and using probability.

Probabilistic reasoning arises in solving probability problems and using arguments to prove the truth of a probabilistic assertion (Sánchez & Valdez, 2017). It requires judgments to establish the credibility of evidence (whether the information is sufficient and relevant to the problem under analysis) and its inferential strength (whether such information can be generalized to another population or context) (Schum, 2001). Some components of probabilistic reasoning, according to Borovcnik (2016) are competence to choose the probabilistic model to be applied in each situation, ability to discriminate between conditioning and causality, understanding the asymmetry of conditional probabilities, and correctly interpreting probabilistic statements.

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

PREVIOUS RESEARCH

Most previous research on probabilistic reasoning was related to decision-making under uncertainty and was developed under the heuristics and biases program (Gilovich et al., 2002). As a result, different heuristics that underlie reasoning about sampling, such as those of representativeness or availability, were described.

In a school context, we found a few papers describing probabilistic reasoning of students when confronted with traditional probability tasks. For example, Sánchez and Valdez (2017) analysed the reasoning of 30 high school students in probability and sampling tasks, suggesting their reasoning was supported by their understanding of ideas of variability, independence, and randomness. Sánchez and Carrasco (2018), in turn, analysed the probabilistic reasoning of 34 high school students in relation to the binomial distribution. We found no research dealing with students' probabilistic reasoning in relation to reports taken from the news, although this is a teaching resource widely recommended to improve probabilistic literacy (Borovcnik, 2016; Gal, 2005).

Because the task posed to students in our research involves conditional probabilities, we also drew on research related to this topic. Conditional probability problems have been classified, and variables that affect their resolution have been identified: the problem context, the meaning of implied events and their probabilities, the terminology used, and the specific teaching environment (Huerta, 2009). The wealth of research related to conditional probability has also described numerous reasoning biases (Borovcnik, 2012; Díaz & de la Fuente, 2007). One of them is the fallacy of the transposed conditional (Falk, 1986), which consists of confusing a conditional probability $P(A|B)$ with its transposed conditional $P(B|A)$ without understanding the asymmetry of conditional probabilities (Borovcnik, 2016). Another bias is confusing conditioning and causation (Díaz & de la Fuente, 2007). Finally, some authors (Böcherer-Linder et al., 2018; Hoffrage et al., 2002) have described didactic resources that may help students overcome these biases such as the use of natural frequencies to provide problem data.

METHOD

The sample consisted of 76 students (16–17-year-olds) in the second year of high school (39 in the specialty of Science and 37 in the specialty of Social Sciences) from a Spanish secondary school. The evaluation was carried out as an initial activity to the study of conditional probability, and the students had already previously studied conditional probability in Secondary Education. These students were given the questionnaire displayed in Figure 1, which is based on a news report published in the media about traffic accidents according to data provided by the Directorate General of Traffic (DGT).

On the DGT[i] web page the following figures appear regarding the accident rate on Spanish roads:

- *Toxicological analyses carried out on 751 people killed in 2018 in road accidents (535 drivers, 143 pedestrians and 73 accompanying persons) showed that the consumption of alcohol and other drugs still has a negative influence on road safety.*
- *More than 43% (232) of drivers killed tested positive for alcohol or drugs and this figure has not changed significantly from previous years.*

Another article published in a newspaper[ii] reported that *"23% of those killed in cars in 2018 were not wearing a seatbelt."*

Analyse the situation according to the information provided and answer the following questions:

1. What percentage of the drivers who died had not consumed alcohol or drugs?
2. What is the probability that a deceased person was wearing a seatbelt?
3. Can we know the percentage of drivers who died due to alcohol or drug consumption?

In view of your previous answers, please reply:

4. Is it more likely to die if you have consumed alcohol or drugs, or if you have not consumed any of these substances?
5. Which is more likely: to be killed if you were wearing a seatbelt or if you were driving without a seatbelt?
6. Do you think the manner in which the information is provided is correct, and why? If not, what information would be missing to give a proper picture of the situation?

Figure 1. Questionnaire given to the students

This topic was selected to arouse the students' interest and to make them aware of the possible effects of the consumption of alcohol and other drugs. Let us consider the events D = deceased, + = positive for alcohol or drugs, and S = wearing a seatbelt; then, the problem data are: P(+|D) = $\frac{43}{100}$ = 0.43; P($\overline{S}$|D) = 23/100 = 0.23, where $\overline{S}$ is the complementary event to S. In every question, the student must first understand the statement, identify the data, and then, depending on the specific question, calculate the probability or provide reasoning suggesting critical reading of the report.

After students' responses were collected, a conventional content analysis (Neuendorf, 2017) was carried out, separating responses into categories taken from previous research (Borovcnick, 2012, 2016; Díaz & de la Fuente, 2007; Sánchez & Valdez, 2017). This response classification was reviewed by the study's authors until consensus was reached. The response categories are described below, with examples from Science students coded as Cx and examples from Social Science students coded as Sx.

Responses to questions 1 and 2 have been coded using the following criteria:

- *Correct answer (probability).* Question 1 asks to calculate the conditional probability P(-|D), which is the complement of P(+|D); P(-|D) = 1 – P(+|D) = 1 – 0.43 = 0.57. The second question requests the probability P(S|D), which is the complement of P($\overline{S}$|D) = 0.23, or 0.77. In both questions, after identifying the problem data, the student should apply the probability of the complementary event.
- *Correct answer (percentage).* The student's answer is represented as a percentage (57% in question 1 and 77% in question 2).
- *Repeating the problem data.* The student merely gives the data in the statement without noticing that the probability of the complementary event is requested, which indicates that he/she does not show a comprehensive reading of the question.
- Other answers based on miscalculations or incorrect interpretations of the question.

Questions 3 to 5 test the ability to distinguish a conditional probability from its transposed conditional, as well as to differentiate between conditioning and causation. To obtain the requested probability we should apply Bayes' theorem, but this is not possible, because we are missing some data: the probability of consuming alcohol or drugs and the probability of wearing a seatbelt. In addition, we expect students to understand that a conditional probability does not imply cause and effect. The responses have been coded using the following criteria:

- *Correct answer differentiating conditioning and causality* to rightly argue that the question cannot be answered due to lack of information or missing data to compute the probability, in differentiating between conditioning and causality.
  C69: No, it informs us how many people were under the influence of these substances, but this does not imply that these substances caused the accident (question 3).
- *Correct answer, by applying contextual knowledge* without taking into account the problem data.
  C44: It is more likely to die by drug consumption or other substances, since you lose control of your body (question 4).
- *Confusing conditional probability with its transposed* and mixing up conditioning and causality.
  C41: It is more likely to die wearing a seatbelt because only 23% of people who died did not use a seatbelt (question 5).
- Other answers based on miscalculations or no justify the answer.

Question 6 involves posing critical questions (Gal, 2005) and correctly interpreting probabilistic statements (Borovcnik, 2016) because students should find what sort of information is missing in the news report. The answers have been coded according to the following criteria:

- *Correct answer, pointing to the missing information* such as the number of people wearing a seatbelt who survived.
  C45: I do not think so, because the data given for the total does not take into account accidents that are not caused by alcohol or drug use and not wearing seatbelts. To give a correct picture of the situation, the percentages of casualties in relation to the total number of drunk or drug-using drivers should be given, as well as the number of casualties in relation to the total number of drivers not wearing seatbelts.
- *Correct answer, indicating that more information should be given without specifying what information is needed* such as the answer of C53, who is aware that the information is incomplete.
  C53: It is incorrect, as this information leads us to believe that more people die with their seatbelts or without having consumed substances such as alcohol or drugs.

- *Principle of authority,* claiming that the information is correct on the basis of the authority of the information source (considering it reliable and unquestionable), thus showing a lack of critical stance towards the information, an ability recommended by Gal (2005).
  C76: Yes, as the DGT is a very reliable source, even though alcohol tests sometimes fail.
- *Need to raise public awareness,* arguing the need for public awareness on issues, without reference to problem data. Although they apply contextual knowledge, it is not used to give a correct answer.
  S8: It is not correct, because they have to create and disseminate more awareness campaigns, as this is a huge problem that causes hundreds of deaths every year in Spain.
- *Suggesting irrelevant information* and arguing the need for other information, even though such input would not help to solve the problem posed.
  C44: The procedure is not entirely correct because only 751 deaths have been analysed, i.e., a small part of all the people who have died in road accidents each year. A complete analysis of all persons killed in road accidents is what is missing.

RESULTS

Table 1 displays the percentages of students who gave different answers to the first two questions. Most students gave correct answers for both tasks, with slightly fewer on the second task and mainly for the Social Sciences group. Thus, in general, for both questions there was a good understanding of the question statement and identification of the data; in addition, the probability of the complementary event was remembered well, and its application was recognized. A small group did not correctly identify that the question asked for the probability of the complementary event or made other errors. Many of these errors were related to an improper calculation of the percentage by considering an incorrect total. This especially happened in the Social Sciences students in the second question, where it is worth noting the percentage of nonresponses.

Table 1. Percentages of responses to questions 1 and 2

	Sciences (n=39)		Social Sciences (n=37)		Total (n=76)	
	P1	P2	P1	P2	P1	P2
Correct, probability		7.7	5.4	16.2	2.6	11.8
Correct, percentage	92.3	71.7	83.8	51.4	88.2	61.8
Probability of complementary event	2.6	10.3	2.7	5.4	2.6	7.9
Other errors	5.1	10.3	8.1	18.9	6.6	14.5
Does not reply				8.1		3.9

Table 2 presents the results for questions 3 to 5. In question 3, few students explicitly acknowledged the difference between conditioning and causation, although a few more gave correct answers using their knowledge of the context, mainly the Social Sciences students. This correct application of contextual knowledge, and thus of probabilistic reasoning, increased considerably in questions 4 and 5 for the majority of students in both groups. In all three questions, a high percentage of students reasoned according to the transposed conditional bias (Falk, 1986).

Table 2. Percentages of responses to questions 3, 4, and 5

	Sciences (n=39)			Social Sciences (n=37)			Total (n=76)		
	P3	P4	P5	P3	P4	P5	P3	P4	P5
Distinguish conditioning and causality	2.6			2.7	2.7		2.6	1 .3	
Apply contextual knowledge	7.7	30.8	46.2	16.2	32.4	45.9	11.8	31.6	46.1
Transposed conditional	61.5	61.5	46.2	51.4	62.2	51.4	56.7	61.8	48.7
Other errors	25.6	7.7	7 .7	29.7	2.7	2.7	27.6	5.3	5.3
Does not reply	2.6						1 .3		

Table 3 presents the results for question 6. About 25% of the students (slightly more in Science than in Social Science) gave a correct answer indicating missing data; a significant proportion answered correctly without specifying the data required. These two categories represent almost 50% of the Science students and 30% of the Social Sciences students. It is worth noting the percentage of students who were guided by the principle of authority without showing a real critical stance towards the statistical data provided. A small percentage (around 8%) alluded to the need for greater awareness, which is not relevant from the point of view of probability but is valuable in relation to the formation of the students.

Table 3. Percentages of responses to question 6

	Sciences (*n*=39)	Social Sciences (*n*=37)	Total (*n*=76)
Correct, provides the data	28.2	18.9	23.7
Correct, does not provide data	20.5	13.5	17.1
Suggest other variables	20.5	40.5	30.3
Authority of the data source	7.7	8.1	7.9
Need of public awareness	10.3	8.1	9.2
Does not reply	12.8	10.8	11.8

DISCUSSION AND CONCLUSIONS

Although it is neither possible to generalize to other students because we are working with a non-random sample, nor to have information about their knowledge of probability beyond the questionnaire, the results of our study indicate that some students in the sample had an adequate probabilistic literacy (Gal, 2005), showing their knowledge of probabilistic language and competence in identifying the problem data and computing probabilities in the first two questions. Their probabilistic reasoning was also put into practice, by correctly interpreting probabilistic statements and choosing the probabilistic model to be applied in the situation (Borovcnik, 2016). According to Sánchez and Valdez (2017), this probabilistic reasoning was evidenced when solving a non-routine probability problem and when using arguments to prove the truth of a probabilistic assertion.

The study also points to important shortcomings, with only a few students explicitly recognizing the difference between conditioning and causality, a basic component of probabilistic reasoning (Borovcnik, 2016). Also noteworthy is confusion between a conditioned probability and its transposed, a confusion seen in prior research (e.g., Díaz & de la Fuente, 2007; Falk, 1986). Teachers should consider the use of visualization resources to help overcome such biases (Böcherer-Linder, 2018; Hoffrage et al., 2002). We also remark the percentage of students who only followed the authority of the information source, without adopting a critical stance towards the data (an element of probabilistic culture, according to Gal, 2005) and failing to establish the credibility of the data (Schum, 2001). On the other hand, in some questions students made efficient use of contextual knowledge, which they used to compensate for the lack of data and as a support to discriminate between conditioning and causality.

Students were very highly motivated by the activity and its context, as recommended in the Guidelines for Assessment and Instruction in Statistics Education II (Bargagliotti et al., 2020) framework. We think that this opens a line of research for developing this type of task that could adequately complement traditional problem solving in probability for two reasons: on the one hand, they make students see the usefulness of probability in everyday life while reinforcing probabilistic reasoning (Borovcnick, 2016); and on the other hand, they make them aware of some of their reasoning biases.

ACKNOWLEDGEMENTS

Project PID2019-105601GB-I00/AEI/10.13039/501100011033 and Research Group FQM-126 (Junta de Andalucía).

[i] http://revista.dgt.es/es/noticias/nacional/2019/07JULIO/0718-Informe-alcohol-drogas.shtml#.YDVxonmCFhE (18/07/2019).
[ii] https://www.abc.es/sociedad/abci-23-por-ciento-muertos-accidentes-trafico-2018-no-llevaba-cinturon-seguri dad-201909301335_noticia.html (30/09/2019).

REFERENCES

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II)*. American

Statistical Association; National Council of Teachers of Mathematics. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school*. Sense Publishers. https://doi.org/10.1007/978-94-6300-624-8

Böcherer-Linder, K., Eichler, A., & Vogel, M. (2018). Visualising conditional probabilities—three perspectives on unit squares and tree diagrams. In C. Batanero & E. J. Chernoff (Eds.), *Teaching and learning stochastics* (pp. 73–88). Springer. https://doi.org/10.1007/978-3-319-72871-1_5

Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didáctica de la Matemática, 2,* 5–27. https://doi.org/10.35763/aiem.v1i2.32

Borovcnik, M. (2016). Probabilistic thinking and probability literacy in the context of risk. *Educação Matemática Pesquisa, 18*(3), 1491–1516.

Díaz, C., & de la Fuente, I. (2007). Assessing students' difficulties with conditional probability and Bayesian reasoning. *International Electronic Journal of Mathematics Education, 2*(3), 128–148. https://doi.org/10.29333/iejme/180

Falk, R. (1986). Conditional probabilities: Insights and difficulties. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 292–297). International Statistical Institute. https://www.stat.auckland.ac.nz/~iase/publications/icots2/Falk.pdf

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

Gal, I. (2009). South Africa's mathematical literacy and mathematics curricula: Is probability literacy given a fair chance? *African Journal of Research in Mathematics, Science and Technology Education, 13*(1), 50–61. https://doi.org/10.1080/10288457.2009.10740650

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. Simon and Schuster.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* Cambridge University Press. https://doi.org/10.1017/CBO9780511808098

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. *Cognition, 84*(3), 343–352. https://doi.org/10.1016/S0010-0277(02)00050-1

Huerta, M. P. (2009). On conditional probability problem solving research-structures and contexts. *International Electronic Journal of Mathematics Education, 4*(3), 163–194. https://doi.org/10.29333/iejme/235

Ministerio de Educación, Cultura, y Deporte. (2015). Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. *Boletín Oficial del Estado, 3*, 169–546. https://www.boe.es/eli/es/rd/2014/12/26/1105

Muñiz-Rodríguez, L., Velázquez, P. A., Muñiz, L. J. R., & Valcke, M. (2016). ¿Hay un vacío en la formación inicial del profesorado de matemáticas de Secundaria en España respecto a otros países? *Revista de Educación, 372,* 111–140. https://doi.org/10.4438/1988-592X-RE-2015-372-317

Neuendorf, K. (2017). *The content analysis guidebook* (2nd ed.). Sage. https://doi.org/10.4135/9781071802878

Nikiforidou, Z. (2018) Probabilistic thinking and young children: Theory and pedagogy. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education* (pp. 21–34). Springer. https://doi.org/10.1007/978-981-13-1044-7_2

Sánchez, E., & Carrasco, G. (2018). El razonamiento probabilístico de estudiantes de bachillerato en actividades de distribución binomial. In L. J. Rodríguez-Muñiz, L. Muñiz-Rodríguez, A. Aguilar-González, P. Alonso, F. J. García, & A. Bruno (Eds.), *Investigación en Educación Matemática XXII* (pp. 535–543). SEIEM.

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

Schum, D. A. (2001). *The evidential foundations of probabilistic reasoning*. Northwestern University Press.

Sharma, S. (2015). Teaching probability: A socio-constructivist perspective. *Teaching Statistics, 37*(3), 78–84. https://doi.org/10.1111/test.12075