13th International Congress on Mathematical Education
Hamburg, 24-31 July 2016

PROBABILISTIC REASONING OF HIGH SCHOOL STUDENTS ON SAMPLE SPACE AND PROBABILITY OF COMPOUND EVENTS

Pedro Landín-Vargas and Jesús Salinas-Herrera

Colegio de Ciencias y Humanidades, UNAM, Mexico

A probabilistic reasoning hierarchy, for the concepts of sample space and probability of a compound event, is proposed as a tool to describe the performance of high school students when they solve two text-book problems, before and after a teaching and learning period. As expected, the responses to the posttest indicate an improvement in the levels of reasoning with respect to the pretest. In the pretest, the most used procedures to count the elements of the sample space and calculate the probabilities of compound events were list and table making. In the posttest, they were list making and tree diagrams. Such diagrams seem a useful tool to improve the probabilistic reasoning level.

INTRODUCTION

A documented fact in literature is the difficulty to build sample spaces and calculate probabilities of compound events (e.g. Polaki, 2005; Jones, Langrall & Mooney, 2007; and Chernoff & Zazkis, 2011). In addition, such topics are present in the high school education curriculum in several countries (Jones et al., 2007) and authors as Heitele (1975) consider the sample space and the measure of probability as fundamental concepts. For these reasons, we propose the following research questions:

- How is the high school students' reasoning to find the *sample space* of a compound random experiment and to calculate the *probability of a compound event*, using the classical approach before and after a teaching and learning process?
- Which difficulties do the students face when solving such problems?

CONCEPTUAL FRAMEWORK

The SOLO (Structure of Observed Learning Outcomes) Taxonomy by Biggs, J. & Collis, K. (1991) states five modes of representation of levels of abstraction that constitute the basis of development and the age at which they typically emerge. The five modes are: sensorimotor (from birth), ikonic (from 18 months), concrete-symbolic (from 6 years and on), formal (from 14 years) and postformal (from 20 years and on). Every mode includes learning cycles constituted by unistructural, multistructural and relational levels, which are defined according to the structural complexity of the responses to a task. The Probabilistic Reasoning Framework (Jones, Langrall, Thornton & Mogill, 1997) distinguishes, in increasing order, the subjective, transitional, informal quantitative and numeric levels. Finally, the Probabilistic Thinking Framework (Mooney, Langrall & Hertel, 2014, p. 495) considers levels of growth in probabilistic reasoning (which "result from maturational or interactionist effects in both structured and unstructured learning environments") for a set of concepts and seeks to summarize research results of the past 20 years for elementary and middle school.

Below are presented some of the concepts developed by the Onto-Semiotic Approach (OSA) to mathematics education. These concepts will be useful to analyze the students' responses to the problems presented.

The OSA states an ontology of mathematical objects which considers mathematics as: a) a socially shared problem solving activity; b) a symbolic language; and c) a logically organized conceptual system. According to OSA, a *mathematical practice* is any action or manifestation that a subject carries out to solve mathematical problems and communicate the solution to other people so as to validate and/or generalize the solution to other contexts and problems (Godino, Batanero & Font, 2007, p. 129).

When carrying out a mathematical practice, a subject involves a set of primary mathematical objects: *problem-situations*, *language, concepts, properties, procedures,* and *arguments*. The situations (problems, exercises, etc.) are tasks that induce mathematical activity while the language is constituted by terms, expressions, notations or graphs. The procedures are operations, algorithms or techniques carried out to solve a mathematical task; the concepts are given through definitions or descriptions and properties in the form of statements or propositions. Finally, an argument (deductive, inductive, etc.) is used to explain or validate propositions and procedures (Godino, 2002, pp. 245-246).

METHODOLOGY

Participants

Twenty-eight third-year high school students (ages 17-18), from a Mexican public school, participated in a probability course.

Instruments

Pretest and posttest questionnaires, with eight items each one, were used. By the time the pretest took place, the group had already taken a course focused on descriptive statistics topics as well as probability elements of one-stage experiments. The posttest was applied to half of a class taking their second four-month course on probability distribution (binomial and normal) and sample distributions. Probability problems in two or three dimensions which involve the concepts of sample space and measure of probability (classical approach) were included in the questionnaires.

Below, two of the common problems in the questionnaires are shown and their responses are analyzed. Problem 1 involves a random experiment in two stages while problem 2 includes a random experiment composed of three stages.

Problem 1**.** A couple plans to have two children. Suppose it is equally likely that a boy or a girl be born and that the gender of any child not influence the gender of another.

a) Write down all arrangements of the genders of two children.
b) Supposing that the listed results in a) are equally likely, what is the probability that two girls be born?
c) What is the probability of having a child of each gender exactly?

(Prepared by the authors based on Moore, 2007, p. 268).

Problem 2. We toss a $1 coin, a $2 coin, and a $5 coin in the air.

a) Determine the number of possible results of tossing the three coins in the air, based on getting heads or tails for each coin.

b) What is the probability of getting no tails when tossing the three coins?

c) What is the probability of getting one tail exactly when tossing the three coins?

d) What is the probability of getting two tails exactly when tossing the three coins?

e) What is the probability of getting three tails exactly when tossing the three coins?

(Pastor, 1998, p. 76)

Procedure

We analyzed, classified and compared the pretest and posttest responses based on a probabilistic reasoning hierarchy. In addition, the mathematical objects used to solve each problem were analyzed with some of the OSA categories.

Hierarchy

Table 1 shows a description of the probabilistic reasoning hierarchy that includes the execution elements as well as the components to construct the sample space and the probability of a compound event. In order to construct it, we considered elements of the reasoning hierarchy by Sánchez and Landín (2014), the framework of probabilistic reasoning (Mooney et al., 2014), and the responses provided by the students in the sample. We analyzed the responses as the basis to validate and adjust the hierarchy we propose here. The names of the levels of reasoning "reflected a continuum from subjective to numerical reasoning" (Jones et al., 1997, p. 101).

Probabilistic Reasoning Level	*Characteristics of the responses*
Subjective	The sample space of the experiment is omitted. The responses are idiosyncratic or influenced by cognitive biases as equiprobability[1]. The responses have features from knowledge components without any coherence or with a great number of mistakes.
Transitional	The responses show incomplete sample spaces with or without the aid of tree diagrams for experiments of two or three stages as well as a tendency to go back to subjective probabilistic reasoning. The students use Laplace's definition of probability and the rule of product of probability with mistakes or missing calculations.
Informal quantitative	The responses show systematic sample spaces, supported by tree diagrams or lists of results. The subjects count and use proportions, probabilities or possibilities when judging probabilistic situations. The classic definition of probability is used in the correct responses. A random variable is identified

[1] The probability bias came to be known through Lecoutre's work (1992, p. 557) and is characterized by a reasoning of the following type: "the results to compare are equiprobable because it's a matter of chance; thus, random events are thought to be equiprobable *by nature*."

	although the calculation of probability is incorrect. The addition rule or the product rule of probabilities is used to obtain partial and correct responses.
Numerical	Sample spaces for experiments of two or more stages are systematically generated with or without a list of results. Sample spaces and probabilities are connected. Probabilities for complex situations, including non-equiprobable ones, are determined. The addition rule, the product rule of probabilities or a random variable is identified and used in the correct responses.

Table 1: Probabilistic reasoning hierarchy for the concepts of sample space and probability of a compound event.

RESULTS AND DISCUSSION

The following are examples of the responses for problem 1 in the posttest; the responses are classified by probabilistic reasoning level.

Baby 1 -	$\frac{1}{4}$	Girl	Probability two girls be born
	$\frac{1}{4}$	Boy	$\frac{2}{4}=\frac{1}{2}=0.5$ posibilities of two girls be born
Baby 2-	$\frac{1}{4}$	Girl	Probability of having a child of each gender exactly
	$\frac{1}{4}$	Boy	$\frac{1}{4}=0.25$ of girl and $\frac{1}{4}=0.25$ of boy

Figure 1: Response classified in the subjective level.

The response in Figure 1 was classified in the subjective reasoning level because it does not show a sample space of compound events (boy-girl, girl-girl, etc.). It only considers simple events while the probabilities requested are incorrect.

a) Girl-girl
Girl-boy
Boy-boy
b) P(a)=$\frac{1}{3}$
c) P(a)=$\frac{1}{3}$

Figure 2: Response classified in the transitional level.

The response in Figure 2 was one of the most frequent. We classified it as transitional since the sample set ("any set of all possible outcomes, where the elements of this set do not need to be equiprobable"; Chernoff & Zazkis, 2011, p. 18) includes three compound events and does not correspond to the sample space. In consequence, the probabilities calculated through the definition by Laplace are not correct. This student might have committed an equiprobability bias if he or she considers the event Girl-Boy equivalent to the event "exactly one child of each gender", which can occur in two ways: Boy-Girl or Girl-Boy.

(boy-boy) (boy-girl)

(girl-girl) (girl-boy)

P(two girls be born)= $1\ de\ 4 = \frac{1}{4} = 25\%$

P(of having a child of each gender exactly) $= (niño - niña)\ (niña - niño) = \frac{2}{4} = 50\%$

Figure 3: Response classified in the informal quantitative level.

The response in Figure 3 was classified in the informal quantitative reasoning level because the student wrote down a complete sample space, used Laplace's definition of probability and correctly calculated the probabilities of the events "two girls" and "exactly a child of each gender."

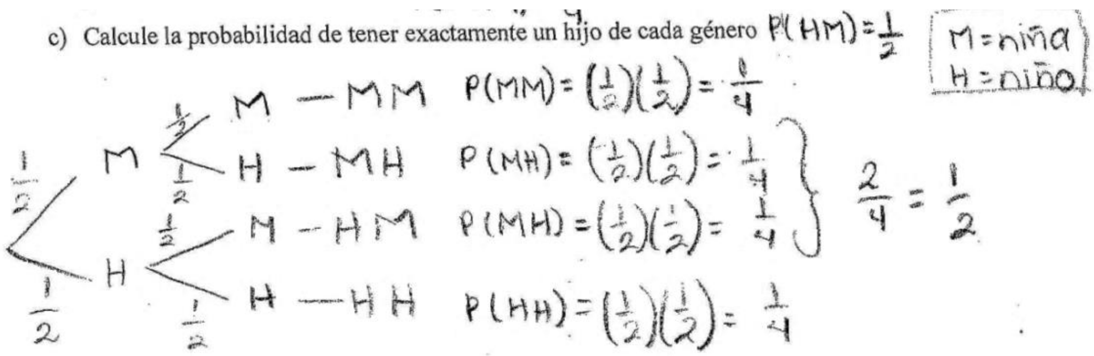

Figure 4: Response classified in the numerical level.

The response in Figure 4 was classified in the numerical reasoning level because, based on a tree diagram, the student wrote down a complete sample space and calculated the probabilities needed for the events using the product rule and the addition rule of probabilities.

Summary of responses to problem 1

From the data in Table 2, we have that:

i. In the pretest, 71% of the responses are classified in the subjective or transitional level. Only 29% of the responses are classified in the informal quantitative or numerical level.
ii. In the posttest, 43% of the responses were classified in the subjective or transitional level while 54% of the responses were classified in the informal quantitative or numerical level.

	Time of application	
Probabilistic reasoning level	Pretest	Posttest
Subjective	8	1
Transitional	12	11
Informal quantitative	8	8
Numerical	0	7
No response	0	1
TOTAL	28	28

Table 2: Responses to problem 1 by probabilistic reasoning level and time of application.

Mathematical objects brought into play when solving problem 1

The Table 3 shows that the majority of the students who answered the pretest and the posttest used the concept of sample space. In both instruments, nearly a third of them used Laplace's concept of probability and obtained the correct value of probability. In the posttest, another third used the addition rule or the product rule of probability to answer b) and c). With respect to the procedures employed, the enumeration strategy (Batanero, Godino & Navarro-Pelayo, 1996, p. 75) stands out in the pretest through lists or tables to obtain the sample space and calculate the probability. In the posttest, we can highlight the use of the tree diagram to enumerate the elements of the sample space, the calculation of probabilities with Laplace's definition, and the use of the addition rule or the product rule of probability. The most used representations by the students were: lists and tables in the pretest, and lists and diagrams in the posttest.

Mathematical objet	*Problem 1*		*Problem 2*	
Definitions or properties	Pretest	Posttest	Pretest	Posttest
Sample space	21	26	12	17
Laplace's concept of probability	10	8	11	7
Adittion rule or product rule of probability.		9		9
Function of binomial probability		2		
Strategies or procedures				
Enumeration of sample espace elements	21	26	12	17
a) With tree diagram	1	13		12
b) Without tree diagram	20	13	12	5
Arithmetic operations and mistaken probabilities	15	11	18	9
Use of Laplace's concept of probability	10	8	11	7
a) With tree diagram		4		3
b) Without tree diagram	10	4	11	4
Use of adittion rule or product rule of probability.		9		9
a) With tree diagram		8		8
b) Without tree diagram		1		1
Use of function of binomial probability		2		
Language or representations				
Verbal expressions	6	2		
Lists o tables	27	14	12	5
Graphs	1	12		12
Algebraic notation		2		6

Table 3: Mathematical objects brought into play when solving problems 1, 2

Summary of responses to problem 2

From the data in Table 4, we have that:

i. In the pretest, 68% of the responses are classified in the subjective or transitional level. Only 18% of the responses are classified in the informal quantitative or numerical level.
ii. In the posttest, 43% of the responses are classified in the subjective or transitional level and 46% of the responses are classified in the informal quantitative or numerical level.

	Time of application	
Probabilistic reasoning level	Pretest	Posttest
Subjective	10	6
Transitional	9	6
Informal quantitative	5	6
Numerical	0	7
No answer	4	3
TOTAL	28	28

Table 4: Responses to problem 2 by probabilistic reasoning level and time of application.

As expected, in the responses to problems 1 and 2 of the posttest, we see an improvement in the probabilistic reasoning levels with respect to the responses to the same problems of the pretest. The problem 2 was more difficult for the students than the problem 1 since the level of reasoning in the responses to problem 1 was highter than the level in the responses to problem 2 and there were several students who did not answer the problem 2 in either application.

Mathematical objects brought into play when solving problem 2

The Table 3 shows that nearly half of the students who answered the pretest used the concept of sample space. In the posttest, around 60% of the students used such concept. In both tests, nearly a third of the students used Laplace's definition of probability and obtained the values of probability requested. In the posttest, another third of the students used the addition rule or the product rule of probability to correctly calculate the probabilities. With respect to the procedures shown: in the pretest, 43% of the students employed enumeration through lists or tables to construct the sample space and calculate the probabilities; in the posttest, 43% used the tree diagram to enumerate the elements in the sample space or calculate probabilities (29% through the addition rule or the product rule of probabilities and 11% with Laplace's definition). The most used representations by the students were lists and tables in the pretest, and tree diagrams and lists in the posttest. Six of the students (21%) represented the random variable through algebraic notation.

CONCLUSION

We propose a hierarchy of probabilistic reasoning for the concepts of sample space and probability of a compound event as a tool to describe the performance of high school students when they solve problems that involve such concepts.

The tree diagram seems to be an instrument that, in the adequate contexts, makes it easy for students to represent the stages of a compound experiment and to understand the concepts of sample space, probability of a compound event, addition rule and product rule of probability.

One of the difficulties observed among the students is that the three-stage problem was more difficult for them than the two-stage problem. Furthermore, few of the students were able to identify and represent the random variable through algebraic notation.

References

Batanero, C., Godino, J.D., & Navarro-Pelayo, V. (1994). *Razonamiento combinatorio.* Madrid: Síntesis.

Biggs, J. B. & Collis, K. F. (1991). Multimodal learning and the quality of intelligence behavior. In H. A. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57-76). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chernoff, E. J. & Zazkis, R. (2011). From personal to conventional probabilities: from sample set to sample space. *Educational Studies in Mathematics, 77*(1), 15-33.

Godino, J.D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques, 22*(2-3), 237-284.

Godino, J. D., Batanero, C. & Font, F. (2007). The ontosemiotic approach to research in mathematics education. *ZDM. The International Journal on Mathematics Education, 39* (1-2), 127-135.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics, 6*(2)*,* 187-205.

Jones, G. A., Langrall, C. W. & Mooney, E. S. (2007). Research in probability. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909-955). Charlotte, NC, USA: Information Age-NCTM.

Jones, G. A., Langrall, C. W., Thornton, C., A. & Mogill, T. (1997). A framework for assessing and nurturing young children´s thinking in probability. *Educational Studies in Mathematics, 32*(2), 101-125.

Lecoutre, M. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics 23*(6), 557-568.

Mooney, E., Langrall, C. & Hertel, J. (2014). A Practitional Perspective on Probabilistic Thinking Models and Frameworks. In E. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 495-507). Springer.

Moore, D. (2007). *The Basic Practice of Statistics.* New York: W. H. Freeman and Company.

Pastor, G. (1998). *Estadística Básica.* México: Trillas-SEP-Conalep.

Polaki, M. (2005). Dealing with compound events. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 191-214). New York: Springer.

Sánchez, E. & Landín, P. (2014). Levels of Probabilistic Reasoning of High School Students About Binomial Problems. In E. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 581-597). Springer.