THEORETICAL DOGMATISM AND EMPIRICAL COMMITMENT IN THE INFORMAL PROBABILISTIC REASONING OF HIGH SCHOOL STUDENTS

Ernesto Sánchez[1], Miguel Mercado[2] & Jaime García[1]

[1]Departamento de Matemática Educativa, Cinvestav, IPN

[2]Colegio de Ciencias y Humanidades, UNAM, México

esanchez0155@gmail.com, miguelmm@unam.mx, jaime_garcia1980@hotmail.com

This paper explores student responses to two binomial tasks, one of prediction and another of distribution, in an effort to understand how they express variability in their predictions before and after simulation activities. To collect data, a four-step study was conducted with two student groups: one that had not taken a probability course and another that had taken one. The first and fourth steps consisted of applying a questionnaire related to a binomial situation b (x, 2, ½). In the second and third steps, students undertook manipulative and virtual simulations, respectively. SOLO taxonomy was used to analyze their progress in reasoning through 2 questions. Moreover, by counting their cross frequency responses to the two questions analyzed, the authors were able to ascertain the difficulties students face in integrating variability into their thoughts, despite their experience with simulation. Two patterns of student responses are salient from the set of answers.

INTRODUCTION

A general problem in the learning of statistics is to understand the connection between the model and the data; in the process of understanding statistics, learners should be faced with many instances of this problem. In the arena of learning and teaching probability a particular piece of the problem consists of students learning the relationships between the classical and frequentist approaches of probability; this is not a minor task when considering the many subtleties the connection contains. A classical probability follows from the symmetry of the random device assuming fairness, while through frequentist approach a probability is estimated by processing data observed through repetitions of an experiment in which an underlying probability is assumed to exist. Unlike what happens in the study of natural phenomena in which building a theoretical model can be arduous, in gambling situations (coins, dice, spinners) the classical approach allows for a model that can be considered theoretical (i.e. a probability distribution) to be proposed easily. Also, as game situations are repeatable under similar conditions it is possible to generate data sets and apply the frequentist approach to estimate the probabilities, which are considered empirical. These characteristics coupled with the availability of software for performing simulations make game situations appropriate for the study of the relationship between chance and data, and more generally models (random) and 'reality'. How do high school students connect them? As variability is inevitable, in data from game situations it is crucial for students to develop a sense of variability to then be considered in operating the model. This paper reveals some ways in which students deal with the variability of a simple binomial situation before and after simulation activities.

INFORMAL PROBABILISTIC REASONING

Probabilistic reasoning overlaps mathematical reasoning, although the former is not contained in the latter. The aspects that distinguish them can be characterized with the help of the big ideas of probability proposed by Gal (2005) in his analysis of probabilistic literacy. Notions of randomness, variation, independence and the pair of complementary ideas, predictability and uncertainty, are the underpinnings of the system of probabilistic statements, and they are not part of any other mathematical system. Gal (2005) points out that: “Some aspects of these big ideas can be represented by mathematical symbols or statistical terms, but their essence cannot be fully captured by technical notations”. Said clarification ties in to our interest in focusing solely on informal reasoning, since our study subjects are students who are or should be beginning to think probabilistically. The authors of this paper understand informal reasoning as a process in which the student builds a model of the situation, articulating several of its elements and obtaining consequences with the help of common sense and previous knowledge (Perkins, 1985). Informal probabilistic reasoning is informal reasoning that involves some of the big ideas of probability.

This research is particularly interested in observing how students reason with or ignore variability when faced with a situation of prediction/uncertainty with the underlying distribution of $b(x, 2, ½)$. Variability of a sequence of experiment outcomes is given by the difference between the probability and the relative frequency of a target event. Student variability reasoning is elicited by asking them about their expectations of the frequency of results for a run of 1,000 random numbers from the distribution. Are students able to perceive the structure and variability of a simple binomial situation?

In response to Jones who warned “…there is a void in the research associated with the frequentist approach to probability…" (Jones, 2005, p. 368) there is already some literature related to the problem of articulation between the frequentist and classical approaches of probability, most of which employ educational software. We mention three sources where more references to this issue can be found. Sthol, Rider, & Tarr (2004) explored the ways in which six grade students make inferences from data to determine if a distribution is uniform or not. Ireland & Watson (2009) documented the eighth grade students’ understanding of relationships of several components involved in transition from experimental (concrete) to theoretical probabilities (abstract), among them variability and sample size. Konold et al. (2011) point out some pitfalls of the common teaching practice of introducing students to both “theoretical” and “experimental” probability. They show how naive it is to believe that the teaching of both definitions lead students to establish their connection. In this paper we explore the same connection but with older students and a fixed sample size.

METHOD

A four-step study was undertaken with two groups of high school students. The first had 37 students who had not previously taken a statistics and probability course, and the other had 66 who had already taken a first probability course. In the first step of the study, a probabilistic situation with several questions (in this paper only 2 of such questions are reported) was administered. In the second, students were guided to simulate the situation with manipulatives, to observe the results, and they were then asked to respond to the same questions posed in the first step based on their observations. The third step involved asking students to simulate the situation and to observe what happened, but this time on the computer using the Fathom software. Finally, in the fourth step students were asked to answer the same questions that were administered in step one. The situation is the following:

The Smith family consists of mother Ana, son Billy and father Charles. Every night the three gather to watch TV, but they are never able to agree on what TV program to watch. Ana likes to watch movies; Billy prefers cartoons; and Charles likes the news. As there is only one TV at home, it would be easier for them to take turns. However, Billy proposes something more fun: a test of luck. He proposes throwing two coins with the following results: Ana wins if the coins land with two heads; Billy wins if one coin lands head and the other tail; and Charles wins if both are tails.

After receiving the above information, the students were asked 13 questions, two of which are analyzed and reported here (they are re-enumerated), namely:

Question 1. What do you think will happen if the control is thrown a thousand times?

Number of times Ana wins: _____

Number of times Billy wins: _____

Number of times Charles wins: _____

Question 2. Assign the probability to each value of the variable, that is, complete the following:

Probability of A (X=0) =

Probability of B (X=1) =

Probability of C (X=2) =

The second question is common in teaching as it requires performing the procedures of the classic definition of probability or applying the frequency approach when data are available. This is contrary to the first question, which is a prediction/uncertainty question with a high degree of sophistication, as it requires considering the probability distribution and a sense of variability. Assuming that students know the probability distribution, they probably state that 250, 500, 250 is the most likely, yet they would be more surprised by this outcome than by results like, for instance, 257, 510, 233. The first event is judged by rules of chance, while the second is probably determined by representativeness (Kahneman and Tversky, 1982); conflicts of this type are often solved in conversations using the expression "the event X or something like that" instead of simply "the event X". In fact, a mathematical expression to specify "…something like this" would provide an event made up of "intervals" around each expected frequency with the probability that it occurs; for example: $E = \{(a, b, c) \mid a \in N, b \in N, c \in N, 200 \leq a \leq 300, 450 \leq b \leq 550, 200 \leq c \leq 300\}$ where the probability of E is a little more than 80% (calculated by simulation using Fathom). This solution is the mathematical way of considering the variability of the situation. It is difficult for students to model variability in this way. Although the language to do so is not out of their reach, they are not yet equipped with the concepts that would allow them to express variability in this way. However, responses that approximate expected frequencies can be interpreted as attempts to represent variability in the absence of concepts that do so more formally.

Procedure of Analysis. To analyze data, first the responses to pre- and post-questionnaires are classified in SOLO categories for each question. Second, the frequencies of answers classified as Unistructural or better for both questions are arranged in a double entry table. The way we use the SOLO taxonomy is for identifying and isolating some relevant aspects (for better or worse) of the solution to the task; these aspects or components are constructed by considering both theoretical elements and features observed in the responses. Combining those aspects, the categories Pre-, Uni-, Multi-structural, and Relational levels are defined in a similar way (but not exactly) as the characterization given by Biggs & Collis (1991).

In general, students responded to question 1 by giving three integer values, each representing the frequency of values 0, 1 and 2. Such responses are normatively incorrect because the probability that the actual frequencies match three given numbers is very small, hence it is unwise to expect it to occur. However, some features of the responses can be highlighted, in particular those related to the notions of distribution and variability as they can help define quality levels of responses. The aspects that we propose are the following:

1) *Coherence*: The sum of the three frequencies is 1000. 2) *Distribution pattern*: The frequency of value X = 1 is greater than the frequency of values X = 0 and X = 2. 3) *Variability*: The frequencies of 0, 1, and 2 fall within the ranges 250 ± 50, 500 ± 50 and 250 ± 50, respectively, but also considering the following: *Lack of variability*, we add condition 4 in order to differentiate responses where the variability is not reflected from those in which it is, namely: 4) The frequencies of 0 and 1 are equal. Of course, this includes responses which assign exactly 250, 500 and 250, as the frequencies of 0, 1 and 2, respectively.

If an answer is incoherent or it does not satisfy conditions of distribution pattern or of variability it is Prestructural. If it satisfies the condition of distribution pattern but not of variability, or of variability but not of distribution pattern, or alternatively if it satisfies 4 it is Unistructural. If the responses satisfy distribution pattern, and neither 3 nor 4 then it is Multistructural. In this case the variability is too wide. Finally, if an answer satisfies distribution pattern and variability, but does not satisfy 4 it is Relational.

For question 2, students also assign three numbers. In this case, the correct answer is (¼, ½, ¼). For succinctness the SOLO categories will be described directly. Responses are classified as *Prestructural*: 1) if they refer to strange elements of the situation or they are incomprehensible; 2) when the answer consists of three numbers which sum is different from 1 or 100% without being proportional to (1, 2, 1). The answers that meet one of the following three conditions are classified as *Unistructural*: 1) They assign values which sum is different from 1 or 100 (generally integers), but they are proportional to (1, 2, 1); 2) values are given in the response which sum is 100 or 1000, but they are not proportional to (1, 2, 1); 3) The answers are given with three decimal or fractional numbers which sum is 1 but not proportional to (1, 2, 1). The responses that are classified in the *Multistructural* level meet any of the following conditions: 1) They are given in decimal numbers, fractions or percentages which sum is 1 or 100% with more probability assigned to X = 1 than to those assigned to X = 0 and X = 2; but with values that are far from the theoretical probabilities (¼, ½, ¼). 2) Values are given which sum to 100 (usually integers but not necessarily) and greater probability is assigned to X = 1, but the "%" sign is omitted. 3) They express the probabilities but indicating that it is just an approximation: "(*about* ¼, *about* ½, *about* ¼)". The responses classified in the *Relational* level are probability distributions, some of which are theoretical distributions and others are obtained via the frequentist approach.

The SOLO classification was used to evaluate the quality of responses independently of one another. However, the joint consideration of responses to questions 1 and 2 can be valuable for understanding how students address variability. To this end, we made contingency tables dividing the responses to each question into two complementary classes. In question 1, one class contains responses in which the expected frequencies are given (250, 500, 250) and the other contains the remaining responses. With respect to question 2, responses in which the probabilities were obtained by the classical approach are distinguished from those obtained by frequentist approach.

Table 3. Categories

	Expected frequencies	Not-expected frequencies
Classical approach	Theoretical dogmatism	Connecting theoretical probabilities and data
Frequentist approach	Blurred relationship between theory and data	Empirical commitment

We describe and interpret each category as follows.

The responses classified as *theoretical dogmatism* start assigning theoretical probabilities and the prediction is calculated simply by multiplying the probability of each event by the number of repetitions of the experiment (in this case 1000). Students whose responses are in this category have not learned anything from experiments and therefore avoid or ignore the variability of the situation in their responses. Some of those students probably already knew the classical approach to probability, but their thinking is that a prediction problem is only a mathematical or theoretical issue without consideration for the results of real situations.

The students whose responses are classified as *empirical commitment* neglect the theoretical distribution and they propose a distribution based on the frequentist approach. For them there is no underlying distribution to the situation since what they proposed is only a description of what they experienced during simulation or from their previous experience. Even though they give 3 numbers that are different from expected frequencies in their answers to question 1, it cannot be stated that they consider variability since they do not have the theoretical distribution as a reference for assessing differences.

Better quality responses would involve the students *connecting theory and data*, i.e. when they accept a theoretical distribution underlying the situation, but realize that in practice results usually vary from theoretical expected values, so they accept that randomness cannot be eliminated. When they do so, it is an indication that they have properly perceived variability and begin to express it. This may allow them to understand the idea of the law of large numbers in approximating the frequencies of the theoretical probabilities. However, it is necessary that they translate the expected frequencies to relative frequencies and compare them with the corresponding probabilities.

Finally, responses in the "Expected frequencies – Not theoretical distribution" cell suggest a *blurred relationship between theory and data*. Said responses are odd as it is difficult to imagine an argument that leads to the expected frequencies, without knowing the theoretical probability distribution.

RESULTS

Tables 2 and 3 below show, the frequencies in which responses to questions 1 and 2 are classified for pre- and post- questionnaires. Table 2 refers to students that had not taken any statistics or probability course, while Table 3 refers to students who had taken a course.

Table 2. Frequencies of responses according to SOLO categories of a group without a probability course. P = Prestructural; U = Unistructural M = Multistructural, R = Relational.

Question 1							Question 2					
	P	**U**	**M**	**R**	**Total**			**P**	**U**	**M**	**R**	**Total**
Pre	51%	34%	9%	6%	100		**Pre**	34%	21%	13%	32%	100
Post	17%	37%	20%	26%	100		**Post**	11%	19%	27%	43%	100

It is worth noting that in both questions and both groups (Tables 2 and 3), the frequencies of responses in the post–questionnaire classified in the two lower SOLO levels (two upper levels) are lesser (greater) than the corresponding frequencies in pre-questionnaire. This means that in general the quality of post questionnaire responses is greater than the quality of pre-questionnaire responses.

Table 3. Frequencies of responses according to SOLO categories of a group with a probability course.

Question 1							Question 2					
	P	**U**	**M**	**R**				**P**	**U**	**M**	**R**	
Pre	52%	45%	3%	0	100		**Pre**	48%	15%	0	37%	100
Post	27%	50%	0	23%	100		**Post**	4%	16%	12%	68%	100

We can say that after the activities, students achieved a better understanding of the problems. For example, the total of frequencies they proposed is 1,000, and the total of probabilities is 1 or a multiple of 100. They learned to associate 0, 1 and 2 with the events "Ana wins", "Billy wins" and "Charles wins". They also learned that 1 is more likely than 0 and 2, and that the probability of the latter two is approximately equal. Yet it seems very difficult for them to articulate the knowledge of the distribution and their intuitive sense of variability.

Table 4. Cross frequencies of responses to questions of Diagnostic questionnaire (considering only Unistructural and better responses)

	Question 1						
		Expected frequencies		Not expected frequencies		Total	
		Pre	Post	Pre	Post	Pre	Post
Question 2	Theoretical Distribution	14%	14%	14%	11%	28%	25%
	No theoretical Distribution	7%	7%	65%	68%	72%	75%
	Total	21%	21%	78%	22	100%	100%

Table 4 refers to data from students that had not taken a course in probability. The results of the pre-questionnaire are noteworthy. The relative frequency of the responses classified as “empirical commitment” (see Table 1) is 65% (9 of 14); as “theoretical dogmatism” is 14% (2 of 14); those as “connecting theoretical probabilities and data” is 14% (2 of 14) and as “blurred relationship between theory and data” is 7% (1 of 14). Second, notice the results of the post-questionnaire, the corresponding frequencies are almost proportional to the above data since the total in this case is 28:

68% (19 of 28), 14% (4 of 28), 11% (3 of 28) and 7% (2 of 28%). What is relevant about this data is that in both questionnaires the majority of responses are classified as “empirical commitment”.

Table 5 refers to data from students that had taken a probability course. Noting that the results of the pre-questionnaire, 79% (19 of 24) are classified as “theoretical dogmatism”, 0% as “empirical commitment”, 8.5% (2 of 24) as “connecting theoretical probabilities and data” and 12.5% (3 of 24) as “blurred relationship between theory and data”. With respect to post-questionnaire data, the corresponding frequencies are 60% (26 of 43), 19% (8 of 43), 16.5% (7 of 43) and 4.5% respectively. It is worth noting that the majority of responses in both questionnaires are classified as “theoretical dogmatism”; in addition to the fact that in the pre-questionnaire no students gave a response that was classified as “empirical commitment”, but after the activities 19% (8 of 43) were placed in this category.

Table 5. Cross frequencies of responses to questions in Diagnostic questionnaire

		Question 1					
Question 2		Expected frequencies		Not expected frequencies		Total	
		Pre	Post	Pre	Post	Pre	Post
	Theoretical Distribution	79%	60%	8.5%	16.5%	87.5%	76.5%
	No theoretical Distribution	12.5%	4.5%	0	19%	12.5%	23.5%
	Total	91.5%	64.5%	8.5%	35.5%	100%	100%

DISCUSSION

The foregoing analysis is primarily descriptive because the concepts employed were designed to fit the situation herein described. However, some general dimension of categories is hidden in other situations that involve uncertainty. To take part in *theoretical dogmatism* is to distort or to skew theoretical propositions or accept illusory consequences of some of them as true results, generally, because of students’ inability to address the uncertainty, randomness or variability that is present in the situation. This resolution behavior shows some similarity to responses to other problems or situations in which statistical misconceptions also occur. For example: “If the probability of rolling 5 with a die is 1/6 then exactly one 5 will be rolled when rolling a die six times”, or “If a sample is random then it is like a miniature replica of the population”; this is not unusual because of the conception of mathematics as a set of definitions and procedures unrelated to reality. The *empirical commitment* is to believe only in what can be observed, limiting the possibilities of inference and theorizing. Faced with the difficulties of conceptualizing frequencies and variability of the results, students are limited to describing them. There are other situations where a similar behavior is found. For example, when the concept of *event* as is understood in probability is identified with *outcome* because it can be seen. Also, the empirical commitment recalls the *approach to outcome* (Konold, 1991) when an answer to a probability question is assessed as correct or incorrect depending on whether it is coincident or not with the results. The development of probabilistic reasoning must integrate model and data regarding variability, this implies conceptualization and not just description; concepts cannot be seen in reality. It is naive to think that just performing simulation activities and observing their results allows students to abstract an underlying distribution. In teaching probability, classical and frequentist approaches should be treated under the more general framework of integrating theory and data; as well as taking the pitfalls of neglecting variability for the sake of simplicity into account in chance situations.

REFERENCES

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

Gal, I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemma. In G. A. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 39-63). New York: Springer.

Hoffman, T., Maxara, C., Meyfarth, T., Prömmel, A. (2014). Using the software FATHOM for learning and teaching statistics in Germany – A review on the research activities of Rolf Biehler's working group over the past ten years. In T. Wassong et al. (eds.), *Mit Werkzeugen Mathematik und Stochastik Lernen*. Heidelberg: Springer.

Jones, G.A. (2005). Reflections. In G.A. Jones (Ed.). *Exploring Probability in School. Challenges for Teaching and Learning* (pp. 367-372). New York: Springer.

Kahneman, D., Tversky, A. (1982). Variants of uncertainty (pp. 509-520). In D. Kahneman, P. Slovic, A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press.

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W. Hortosn, N.J., Kazak, S. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning, 13*(1 & 2), pp. 68-86.

Perkins, D.N. (1985). Reasoning as imagination. *Interchange 16*(1), 14-26.

Sánchez, E., Landín, P.R. (2014). Levels of probabilistic reasoning of high school students about binomial problems. In E.J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking. Presenting Plural Perspectives* (581-597). New York: Springer.

Shaughnessy J. M., Research in probability and statistics: Reflections and directions. In Grouws D. A., (Editor), 1992, *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). New York: Macmillan Publishing Company.

Sthol, Rider, & Tarr (2004). *Making connections between empirical and theoretical probability: Students' generation and analysis of data in a technological environment.*

http://www.probexplorer.com/Articles/LeeRiderTarrConnectE&T.pdf

Ireland, S., Watson, J. (2009). Building a connection between experimental and theoretical aspects of probability. *International Electronic Journal of Mathematics Education*, 4(3). 339-370.