Students' reasoning about variability in an horizontal modelling process of the stabilized relative frequencies

Ana Serradó Bayés[1], Maria Meletiou-Mavrotheris[2], Efi Paparistodemou[3]

[1]ana.serrado@gm.uca.es, La Salle-Buen Consejo, España
[2]m.mavrotheris@euc.ac.cy, European University of Cyprus
[3]e.paparistodemou@cytanet.com.cy, Cyprus Pedagogical Institute

Abstract

We describe a task that engaged a class of Grade 9 (aged 14-15) students in informal reasoning about the stabilized relative frequency distribution. The task involved a horizontal modelling process of the relative frequency of appearance of each vowel in strings of characters with three cycles: a pseudo-concrete model developed through a statistical process of investigation, a statistical model, and validation. We highlight aspects of students' reasoning about variability when interconnecting this important statistical concept with the related ideas of density, stability, data and distribution.

Keywords: variability, stabilized relative frequency distribution, modelling.

.

1. Introduction

Since its inception, philosophical difficulties have been prevalent in defining probability. Those difficulties are related to the fact that probability is not an inherent property of an event, but it is based on the underlying model chosen (Borovnik & Kapadia, 2014). Difficulties related to the nature of each of the models (classical, subjective, frequentist or logic) have been largely analysed by researchers investigating Probabilistic Thinking (Chernoff & Sriraman, 2014). Acknowledging the difficulties of introducing probability at the school level, teachers tend to be resistant to its teaching. In order to improve this situation, there have recently been various efforts to innovate and reform the teaching of probability by integrating it within the teaching of statistics, considering that Chance and Data developments should be intertwined. Those innovations aspire to link students' real world experiences to sustainable representations of probability.

It is in line with this aspiration that we present, in this paper, a task involving a horizontal modelling process of the relative frequency of appearance of each vowel in strings of characters. The task aims to help students analyse the variability of data and to estimate a confidence interval while informally reasoning about the stability of frequencies. Although the task seems very simple, complex concepts can emerge from it, including variability, stability, and relative frequencies in a modelling process that can eventually lead to the introduction of the Law of Large Numbers. The analysis of the interplay between these notions is introduced through a historical analysis of the Law of Large Numbers, and an aggregate view of the stabilized relative frequency distribution. In the study described in the paper, we sought to investigate whether the task indeed helped students to reason about data variability while analysing the stabilized relative frequency distribution.

En J. M. Contreras, C. Batanero, J. D. Godino, G.R. Cañadas, P. Arteaga, E. Molina, M.M. Gea y M.M. López (Eds.), Didáctica de la Estadística, Probabilidad y Combinatoria, 2 (pp. 415-424). Granada, 2015.

2. The Law of Large Numbers

2.1. A historical view of the Law of Large Numbers

We are exploring in this paper the possibility of introducing stabilized relative frequencies distribution as a mathematical activity, in coherence with the growth of the probabilistic thinking to mathematize the frequentist notion of probability due to the evolution of the ideas of Huygens, Bernoulli and Von Mises.

For Huygens, a prominent Dutch mathematician and scientist who lived in the 17th century, probability was not yet a number (Huygens, 1998). It was a collection of arguments (pros and cons) that could be used to weigh arguments. These arguments were related to decisions made in the context of playing games of chance with equally likely outcomes. From these arguments, Huygens derived the expected value of an event (Shafer, 1996). Huygens' philosophical disquisitions were crucial for Bernoulli when introducing probability as a number instead of an estimation of possibilities. Bernoulli derived a mathematical relation (a kind of convergence) between equal probabilities and the observed frequencies in the repetition of such games, which comprised the very first version of the "Law of Large Numbers".

It was Von Mises' (1964) disquisitions in other contexts, such as the sex of a new born in a population, that were crucial for considering not only individual events, but also the different outcomes of a distribution. Furthermore, this consideration of different events created the need of analysing variation and variability of these outcomes. Von Mises is considered to have played an influential role in the definition of probability. On his works underlie the analysis of the convergence on probability, but also the variability. He analyses the variability due to the density of the different outcomes, expressed by the frequencies, the tendencies and the intervals of confidence that gave sense to the convergence on probability when the number of observation increase.

Von Mises' (1928/51) axiomatic frequentist definition of probability has two main contradictions. In order to prove the "Law of Large Numbers", a definition of the theoretical value of probability, and a definition of the property of convergence of probabilities are needed. For Batanero, Henry and Parzysz (2005) this conception confuses model and reality. Von Mises (1964, pp. 45) understood this difficulty as unavoidable in mathematical science, arguing that: "the transition from observation to theoretical concepts cannot be completely mathematized. It is not a logical conclusion but rather a choice, which, one believes, will stand up in the face of new observations". In the ideas of Von Mises underlie a constructivist perspective about how probabilistic, statistical and mathematic thinking is organized, defined, and proved.

2.2. Tradition and innovation when introducing the Law of Large Numbers at school

Cognitivist researchers, such as Treffers (1987), introduce the idea of progressive mathematizing in terms of two types of mathematical activity- horizontal mathematizing and vertical mathematizing. Treffers describes horizontal mathematizing as "transforming a problem field into a mathematical problem" (p. 247). Horizontal mathematizing might include, but not be limited to, activities such as experimenting, pattern snooping, classifying, conjecturing and organizing. Those horizontal activities can be grounded in, and build vertical mathematizing, which encompasses activities such as reasoning about abstract structures, generalizing and formalizing.

Historically, the works of Huygens, Bernoulli and Von Mises can be thought of as a process of horizontal and vertical mathematizing of the frequentist notion of probability. This growth in probabilistic thinking construction might be transferred to education through the design of learning trajectories that enable students to be involved in horizontal and vertical mathematizing. For some statistics educators, horizontal mathematizing can be understood as the development of mathematical models (Drijvers, 2000). These educators view students' modelling process as going through multiple cycles of developing a mathematical model for a given problematic situation (Lesh & Doerr, 2006). Thus, in designing a learning trajectory to prove the law of large numbers, one should take into consideration the background about the historical evolution of the frequentist notion of probability and its implications in the learning process.

The differences between estimation and probability have been confusing when teaching probability. In the Spanish curricular design and its textbooks, this estimation is taken as the definition of the mathematical value (Chaput, Girard and Henry, 2011; Serradó, Cardeñoso and Azcárate, 2005). In some other countries, this dichotomy between experimental value estimated based on empirical models, and theoretical value based on probability distributions, has been tackled by opting for a modelling perspective, where probability is a theoretical value of the degree of confidence that one can give to a random outcome obtained by an observation of the stabilized relative frequency when the same random experiment is repeated a large number of times under the same conditions (Chaput, Girard, & Henry, 2011).

Some difficulties when proving the "Law of Large Numbers" were anticipated by Bernoulli. He dismissed context, arguing that most phenomena were perceived to be too complex to take context into account, and he derived mathematically a relation (a kind of convergence) between equal probabilities and observed frequencies in the repetition of an experiment (Borovnik & Kapadia, 2014). The frequentist approach does not provide the probability for an event when it is physically impossible to repeat an experiment under the same conditions, or a very large number of times (Batanero, Henry, & Parzysz, 2005). In the school context, students are not usually asked to analyse if it is physically possible to repeat an experiment or not, due to the use of traditional random generators (dice, coins …) or the use of simulation tools constructed based on the theoretical model. We argue that, in order to overcome this obstacle, students should be confronted with situations or contexts in which the data generated is susceptible to be repeated under the same conditions in a statistical environment. In the analysis of context, one should have in mind reasoning about the population, the samples taken and the sampling process. In reasoning about samples and the sampling process, teachers and students should have in mind the difficulties involved in determining the minimum value of required sample size for accepting the estimated sample statistic as representative of the corresponding population parameter (sample representativeness). They should also be able to explain both the similarities, and differences between the estimated and the theoretical value of the probability.

Furthermore, Bernoulli's restriction to only equalprobable events still remains as a prevailing tradition in curricula using random equiprobable generators, where students can anticipate the theoretical value of probability through a classical: Laplacian model. Difficulties can arise again due to equiprobability bias, when distinguishing between the estimation and the theoretical value. In order to improve this situation, it could be useful to introduce students to a context in which the events are not equiprobable, or at least students have an intuition that they are not going to be equiprobable. From a statistical point of view, students should not only analyse an isolated repetition of an event, but should rather analyse the distribution of frequencies, understood as the aggregate properties of the data (Bakker & Gravemeijer, 2004).

The idea of distribution was still hidden in the arguments of Bernoulli, although it formed the basis of his proposed "Law of Large Numbers" (Borovnik & Kapadia, 2014). Von Mises (1928/51) describes the need for the analysis of the outcomes that describe a distribution, but also the variability of these outcomes. Conducted educational studies investigating the conceptual construction of the "Law of Large Numbers" and the concept of probability, inform us that, at the very minimum, an analysis of the following is required: (a) the variability of the results obtained when repeating an experiment, (b) the stability of the frequencies of the observed outcomes, and (c) the relation between the value of the limit of the frequencies, the distribution of possible outcomes, and the theoretical value of the probability (Yánez & Jaimes, 2013).

In sum, we understand these new claims of innovation, when introducing the Law of Large Numbers and the frequentist definition of probability, as: (a) the integration of probability within statistics, with a gradual evolution from informal to formal reasoning and thinking about data and chance to understand the four primary interpretations of probability (Chernoff & Sriraman, 2014), (b) an aggregate view of the data, that distinguishes between data as individual values and distribution as a conceptual entity (Bakker & Gravemeijer, 2004), in which the stabilized relative frequency distribution will help to connect between the real contexts and data to formally define probability distributions (Serradó, 2014), (c) data modelling as a powerful vehicle for illuminating students' learning potential (English, 2012), when involved in horizontal and vertical mathematizing activities, and (d) a growth in the understanding of concepts related to chance and data, in general, and of the Law of Large Numbers, in particular.

Although, the elements of innovation are described independently, we consider them as interconnected, and emerging from the need or evolution of each other. This is the reason why we consider that a learning trajectory might begin with reasoning about the aggregate view of data.

3. Stabilized relative frequency distribution framework (SRFD)

Bakker and Gravemeijer's (2004) theoretical framework allows analysing the relation between data and distribution in an informal way. The two authors examine aspects as centre, spread, density and skewness to structure the relationship between data (individual value), and distribution (conceptual entity). This structure can be read upward, from data to distribution, which is typical for novices in statistics; or, downward, from distribution to data, in which they use probability distributions to model data. From the two perspectives, we are interested in the upward approach, i.e. in how students organize numerical data as a frequency distribution that allows them to take an aggregate view of the data and to describe the distribution's shape, centre and spread (Konold & Kazak, 2008). For an accurate analysis of the cognition process developed when constructing the notion of distribution, we consider the framework developed by Ben Zvi, Gil and Apel (2007). They distinguish between reasoning about variability, distributional reasoning, reasoning about signal and noise, contextual reasoning, and graph comprehension. This framework provides the opportunity to think about the links when reasoning about distributions and variability, but it has omitted the links with sampling reasoning (Dierdrop et al., 2012). Furthermore, if we provide students with the opportunity to think about the variability of a distribution in a statistical modelling process of "growing samples", they might be able to construct the concept of stabilized relative frequency distribution.

In coherence with that aggregate view of data, we propose the Stabilized Relative Frequency Distribution Description Framework (SRFD), in which we describe theoretically the overarching statistical s and features of the distribution: contextual knowledge (population, sample, sample size, variable, interpretation, explanation…), distributional knowledge (shape, skewness, error, reliability, individual cases…), graph comprehension (decoding visual shape, unusual features, smoothing, comparing samples…), variability (spread, density, tendencies, intuitive confidence intervals) and signal and noise (centre, modal clumps).

The variability dimension of the SRFD framework includes both variability as the propensity for something to change, and variation as the description or measurement of this change (Reading and Shaughnessy, 2004). We use Shaughnessy's (2007) artificial categorization about variability in data, among samples, from samples to samples' distribution, from informal to formal inference to analyse the different kinds of reasoning about variability that students do when involved in a task about the stabilized relative frequency distribution.

4. Methodology

To answer the main question as to whether the task can help students to reason about variability when analysing the stabilized relative frequency distribution, we carried out a design-based research study of the task (e.g. Bakker & Gravemeijer, 2004). The task was constructed, refined and validated for developing reasoning on stabilized relative frequency distributions in the context of informal inferences (Serradó, 2014), and developing hypothetical thinking through four cycles of informal stochastical modelling (Serradó, 2014). The main question of the task, or real problem, was: "Can I guess which language my friend is speaking by only counting vowels?". The problem generated three questions, fully described in previous work of the first author (e.g. Serradó, 2014), which allowed students to be involved in an informal modelling process.

We present in this paper two phases of the teaching experiment. The first phase took place in 2013 and lasted for 11 sessions of 60 minutes, while the second cycle was implemented in 2014 and lasted for 8 sessions of 60 minutes. The aim of the first phase was to validate and refine the task and the technological tools developed, and the possibilities that they provide students to conjecture. In the second phase, the validated task has been used to analyse the possibilities that it gave for students to develop an aggregate view of data, as the stabilized relative frequency distribution. In particular, in this paper we analyse students' reasoning about the variability of the data.

Both phases of the study took place in a Grade 9 (ages 14, 15) classroom of Spanish Compulsory Secondary School in a low socioeconomic coast city. A total of 49 and 45 students participated in phases 1 and 2 of the study respectively, grouped in two classrooms A and B,. For each problem and cycle of modelling, the question was first presented to the whole class in order to collectively decide which statistical problem to pose. Next, Students were asked to individually hypothesize about the answer to the posed problem. Then, cooperatively in small-groups of four, students were asked to confront their hypothesis to promote a deliberative dialogue, which guided the successive cycles of investigation. Each group's interpretations were presented to the whole-class in order to draw conclusions about the problem and to structure the components of the SRFD. These conclusions led to new problems, new cycles of modelling, and new methodological cycles of whole class discussion, individual reflection, small-group action, and whole class discussion. We videotaped the whole class discussions, collected individual and small group work samples, and conducted a small number (n=11) of individual interviews. The data was codified by the researchers using the SRFD framework, with the aim of investigating the relationships that students established between the distributional and variability elements.

5. Results and discussion

Previously to the first cycle of investigation, corresponding to the pseudo-modelling process, students were asked to draw their first conjectures, as a hypothesis about the answer of the problem: *"Is letter A the letter that appears most often in Spanish?"* All students (phase 1 and 2) agreed that letter A was the most frequent letter of the Spanish alphabet. We are interpreting this as an indication that students were able to draw hypothesis about the density of the data as a perception of the reality about the structure of the Spanish Language (Serradó, 2014b). We think that students undertaking this particular problem had an initial appreciation of the variability of the data, as a gestalt of the density of the vowel, that allowed them to reason "iconocally" as described by Watson & Kelly (2006).

Then, students were involved in the first cycle of investigation. Each team of students selected the sample - the kind of text and its size - analysed the data and compared the results of their investigation with their initial hypothesis about the relative appearance of letter "a". We can find differences on how the students reasoned about the similarities and differences between their hypothesis and the empirical results. Students considering the process of investigation as a manner to validate their hypothesis noted:

> We were right, that vowel "a" is the one most often used (phase 1, A21).

Meanwhile, other students thought that only one sample is inadequate to validate their hypothesis and argued:

> After the first investigation, we have thought that we should select another text to compare the results, and prove that letter "a" is still the one most used. We have to consider that the text selected has to have the same size of characters. With the results of this sample we have arrived to the conclusion that the letter that appears most often is still "a". However, the frequency of the vowels' appearance varies (Phase 1, B1).

These students are arguing about the variability of the data when comparing samples of the same size. We think that the selection of a second sample and its comparison has enabled them, as additive reasoners, as described by Watson & Kelly (2006), to argue about the variability of the data and to discuss the need of the variability of the sample.

But, for Francisco (phase 1, group A41), the first cycle of investigation does not give him enough information. He describes what he thinks that he has to improve his investigation:

> The precision of the data, because the samples have a lot of variability. The quantity of data analysed. The sample size is not adequately large. I think that the most important thing is that we are not able to analyse all the texts written in Spanish.

We interpret that the student reasons using his contextual knowledge about the samples of texts and the population to look for the reliability of the distribution. But, he is still not able to understand the variability due to the sampling process.

In the second cycle of investigation, students used the affordances of Geogebra to animate the bar graphs of the relative frequencies of appearance of each vowel in relation with the total number of characters of the text, with samples from 1 to 10000 characters (Serradó, 2014a). We can again find differences when decoding the visual shapes of the bar graphs. The analysis of the animation helped students of group A in phase 1 to reason about the relation between sample size, variability and the tendencies of the distribution of vowels. We highlight the description made by students of team A6, phase 1:

> The distribution of frequencies in relation with the sample size of characters has little variation because all the graphs are similar except from the first of 1261.

Using the classification introduced by Shaughnessy (2007), when conceptualizing about variability students tend to focus their attention on particular data values as pointers, as in this case in a very strange value. Furthermore, we consider these students' expression *"all the graphs are similar"*, as an indication that they have begun to move away from seeing data only as individual values that vary, to recognising that entire samples of data can also vary. Students are expressing variability among samples, which came from the analysis of the individual data to samples.Meanwhile, students of phase 1, group B and students of phase 2 looked for differences between variables:

> The differences between [the appearance of] A and E with respect at the number of characters is that the graph increases or decreases in relation with the other vowels" (Phase 1, group B2)

> The relative frequencies of vowels a and e, at the beginning, are more equal, but when the number of characters increases, letter e smoothens, while letter a increases, but finally all the frequencies evolve to the same rhythm. The relative frequency of "a" is approximately 0.12… and it is always around this value" (Phase 2, B3).

When using the expression "it is always around this value", we think that students are describing some kind of convergence to the estimated value. According to Watson and Kelly (2006), students' reasoning is predominantly proportional to conceptualize the variation about the expected value (Shaughnessy, 2007). Students have analysed both the variability of the data corresponding to the vowels, and the variability of the samples to describe the stabilization of the values, but they are still not able to describe the variability of the sampling distribution. If instead of a statistical analysis of students' reasoning, we analyse it from a probabilistic point of view, we can say that students are conjecturing about the expected value to which the samples converge, but still they do not make any reference to the sample.

Jonas (Phase 2, A5) concluded from the analysis of the tendencies that *"when increasing the sample size, the values of the variable vary"*. And, when the teacher interviewed him, asking if he was able to generalize the ideas he proposed, with the aim of provoking a vertical mathematization, he responded: *"Increasing the sample size, increases the values of each variable in relation with the horizontal axis approximately till 7000, at which point they seem to eventually stabilize"*. We consider that the student has attempted to balance sample representativeness with sample variability, with an initial proportional reasoning about the variability of the distribution. When the variation between or among a set of distributions is compared, the spectre of statistical significance arises. Theoretically, probability distributions emerge to help decisions about distributions of data, or about sampling distributions (Baker and Gravemeijer, 2004; Shaughnessy, 2007). In consequence, we think that the task has helped the student to develop his intuitions about the Law of Large Numbers, and it has provided him with background knowledge to define the axiomatic of frequentist definition of probability.

In the third cycle of investigation, when the student is analysing a second animation to compare three samples of texts and validate the model (Serradó, 2014a), he notes the following:

> "[sampling distribution] is varying till they stabilize from a specific sample size onwards. In the first sample, the value is 8000; in the second, 7500 and in the third, 6000" (Jonas, Phase 2, A5).

We think, that the student has been reasoning proportionally looking for the minimum value of sample size, at which there was a convergence of the distribution. The student has only informal intuitions about this convergence, but may be the task has given him the opportunity to reflect about the importance of sample size and sample representativeness. This, in turn, is going to help him in the future to understand the complex nature of formal inference.

6. Conclusions

The task employed in this study, has enabled students to reason statistically about the stabilized relative frequency distribution when solving the problem: "Which vowel is used most often in Spanish?" From a statistical point of view, the task was organized with the aim that the students develop an aggregate view of the data when counting the appearance of vowels on chains of characters. The theoretical framework introduced by Bakker and Gravemeijer (2004) has provided a tool for structuring the relationships between data (individual values), and distribution (conceptual entity). We consider the stabilized relative frequency distribution as a conceptual entity to be analysed prior to the introduction of the Law of Large numbers. This aggregate view has provided the SRFD description framework, including, among others, a dimension that analyses students' descriptions and reasoning about variability.

We have used the artificial categorization introduced by Shaughnessy (2007) to analyse how students reason about the variability in data, among samples, and from samples to distribution. The problem posed and its contextual analysis has provided students with a gestalt of the density of the vowel, reasoning iconically about the variability of the data. In the first cycle of investigation, corresponding to a pseudo-modelling process, students have been reasoning, as additive reasoners, about the variability of the data when comparing samples of the same size. Furthermore, their engagement with the task has prompted the need for them to understand variability due to the sampling process.

In the second cycle of investigation, a modelling process, we have used the affordances of Geogebra to decode the visual shapes of the bar graph. Some students have described contextual aspects such as the sample size and distributional aspects such as the tendencies, and also conceptualized about variability, focusing their attention on particular data vales. Students have begun to move away from seeing data as individual values that vary, to recognising the variability among samples. Students with a proportional reasoning have been able to conceptualize the variation about the expected value, reasoning about the variability of the data corresponding to the vowels as variables and the variability of the samples to describe the stabilization of the values. When trying to generalize the ideas about stability, sample variability and sample representativeness, an initial proportional reasoning about the variability of the distribution has appeared. In the third cycle of investigation, when validating the model through decoding and comparing samples, students were able to reason about the variability of the sampling distribution.

From a probabilistic point of view, we think that the task has helped students to develop their informal reasoning about the Law of Large Numbers. They have been reasoning about the value in which the distribution approximately stabilized, giving them the sense of the need of estimating this value. Furthermore they have been reasoning about the minimum size of the sample that gave sense to the stabilization of the distribution. The modelling process proposed in the task has allowed the students to investigate about density, looking for patterns when analysing the variability and stability of the relative frequencies, and conjecturing about the value estimated in which the distribution stabilized. But, when the students were asked to generalize their conjectures, they were involved in vertical mathematizing of the frequentist notion of probability. The horizontal and vertical mathematizing has allowed linking students' real life experiences to sustainable representations of probability integrated within statistics.

References

Bakker, A., & Gravemeijer, K. P. (2004). Learning to reason about distributions. In D. Ben-Zvi, & J. Garfield, *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147-168). Dordrerch, The Netherlands: Kluwer Academic Publishers.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 16-42). Dordrecht: Kluwer.

Ben Zvi, D., Gil, E., & Apel, N. (2007, August 11-17). What is hidden beyond the data? Helping young students to reason and argue about some wither universe. *Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5).* UK: University of Warwich.

Borovnik, M., & Kapadia, R. (2014). A historical and philosophical perspective on probability. In E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic Thinking. Presenting Plural Perspectives* (Advances in mathematics Education ed., págs. 7-34). Dordrecht: Springer.

Chaput, B., Girard, J.-C., & Henry, M. (2011). Frequentist approach: modelling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school. mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE Study*. (pp. 85-96). Dordrech: Springer.

Chernoff, E. J., & Sriraman, B. (2014). Commentary on probabilistic thinking: presenting plural perspectives. En E. J. Chernoff, & B. Sriraman (Edits.), *Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education Series* (pp. 721-727). Dordrecht: Springer.

English, L. D. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics* (81), 15-30.

Dierdrop, A., Bakker, A., van Mannen, J., & Eijkelhof, H. (2012). Supporting students to develop concepts underlying sampling and to shuttle between contextual and statistical spheres. In D. Ben Zvi, & K. Makar (Eds.), *Teaching and Learning of Statistics. Topic Study Group 12. 12th International Congress on Mathematical Education (ICME-12),* (pp. 249-258). COEX, Seoul, Korea.

Drijvers, P. (2000). Students encountering obstacles using a CAS. *International Journal of Computers for Mathematical Learning* , 189-209.

Huygens, C. (1998). *Du cacul dans les jeux de hazard [The calculus in games of chance]*. In (Euvres coplètes, t. 14, La Haye, 1888-1950. (Original work Ratiocinnis in Aleae ludo, published 1657).

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2* (1).

Lesh, R., & Doerr, L. M. (2006). Symbolizing, communicating, and mathematizing: key components of models and modelling. In P. Cobb, E. Yackel, & K. McClain (Eds.), *Symbolizing and Communicating in Mathematics Classrooms Perspectives on Discourse, Tools, and Instructional Design.* USA: Routledge.

Mises, R. (1928/51). *Wahrscheinlinckeit, Statistik und Wahrheit.* Berlin: Springer.

Mises, R. *Mathematical theory of probability and statistics.* New York: John Wiley.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In J. Garfield, & D. Ben-Zvi (Edits.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201-226). Dordrecht, The Netherlands: Kluwer.

Serradó, A. (2014a). Constructing, refining and validating a task for developing reasoning on stabilized relative frequency distributions in the context of informal inferences. In K. Makar, B. de Sousa, & R. Gould (Ed.), *Sustainability in statistics education Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014).* Voorburg, The Netherlands: International Statistical Institute.

Serradó, A. (2014B, from 21th to 25th of July). Developing hypothetical thinking through four cycles of informal stochastical modelling. In G. Aldon (Ed.), *Pre-proocedings CIEAEM 66* (pp. 104-109). Lyon: Institute Français de l' Éducation.

Serradó, A., Cardeñoso, J. M., & Azcárate, P. (2005). Obstacles in the learning of probabilistic knowledge: influence from the textbooks. *Statistics Education Research Journal, 4* (2), 59-81.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 957-1009). Greenwich, CT: Information Age Publishing, Inc., & NCTM.

Shafer, G. (1996). The significande of Jacob Bernoulli's Ars conjectandi for the philosophy of probability today. *Jounal of Econometrics* (75), 15-32.

Treffers, A. (1987). *Three dimensions. A model of goal and theory description in mathematics education: The Wiskobas project.* Dordrecht, The Netherlands: Kluwer Academic.

Yánez Canal, G., & Jaimes, E. (2013). Efectos de la simulación en la comprensión de la ley de los grandes números. *Revista de Integración , 31* (1), 69-89.

Watson, J. M., & Kelly, B. A. (2006). Expectation versus variation: Students' decision making in a sampling environment. *Canadian Journal of Science, Mathematics and Teaching Education, 6, 145-166.* , 145-166.