Secondary education students’ understanding of sampling
Gilda Lisbôa Guimarães, Tâmara Marques da Silva Gomes

To cite this version:
Gilda Lisbôa Guimarães, Tâmara Marques da Silva Gomes. Secondary education students’ understanding of sampling. CERME 9 - Ninth Congress of the European Society for Research in Mathematics Education, Charles University in Prague, Faculty of Education; ERME, Feb 2015, Prague, Czech Republic. pp.672-677. hal-01287070

HAL Id: hal-01287070
https://hal.science/hal-01287070
Submitted on 11 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Secondary education students’ understanding of sampling

Gilda Lisbôa Guimarães and Tâmara Marques da Silva Gomes

University Federal of Pernambuco, Recife, Brazil, gilda.lguimaraes@gmail.com, tamara_msg@hotmail.com

This study aims to identify what students in the 5th and 9th grades from Secondary Education (10 to 14) understand of sampling. Semi-structured, person-to-person audiotaped interviews were carried out with 40 students (20 from each grade), to whom questions were asked concerning some different skills related to sampling. It was noted that, besides the differences between the years of schooling, there was no significant difference in association with the students’ performance. This study shows that, albeit all difficulties, students in the 5th grade are able to answer the sampling aspects.

Keywords: Sampling, sample, population, statistics education.

STATISTICS EDUCATION

Statistics data have been present in various different social situations, thus becoming increasingly essential to understand them so that we can consciously exert our role in society. Due to this fact, statistical literacy has been important to Basic Education, since it enables students to develop their critical, investigative, and reflexive attitude. This posture is greatly necessary for the current society characterized by data dissemination (Campos, Jacobini, Wodewotzki, & Ferreira, 2011).

In the Brazilian curriculum, the National Curriculum Parameters (NCP) have outscored that, in the early years (9 to 10 years old), statistical content should be looked upon in the classroom to enable students “to analyze relevant data regarding knowledge and establish the highest number of relation among them by using mathematical knowledge to interpret them and critically assess them”. (Brasil, 1997, p. 48). By the 3rd cycle (11 and 12 years old), abilities will need to be developed by students to build sample space and show success possibility of an event by using reasoning (Brasil, 1998, p. 75).

After analyzing Probability and Statistics teaching in the Mathematical Curriculum for Secondary Education, Lopes (1998) took into account the importance of these themes for students’ formation as they provide for interdisciplinary approach, experiments, and exploration of random events, which lead to a change of the deterministic model characterized in mathematics.

Among the several statistical concepts that must be worked with students, sampling should not be overlooked. It is part of the investigative cycle of a research and helps to elaborate teaching systematic situations.

For data collection, it is essential to figure out who and how many individuals should be investigated, and how to approach them and put the questions to them in such an efficient way as to gather the most possible representative data. At the moment of data analysis and interpretation in statistical research, researchers must take into consideration the way the data are selected, the methods are applied, the variables analyzed, so that they can understand what is being investigated in context and then compare them to other situations. Sampling is noted implicitly and explicitly through all of these steps. Therefore, an approach using these steps along with mathematical teaching may contribute to important reflections upon the use and the importance of this concept for statistical activities.

It’s important to show that this article doesn’t provide a discussion of the various types of samples, but investigates the importance of working the concept of sampling from the early years of schooling so that students understand the relationship between the variability of the sample and the generalizability of the data for the entire population.
SAMPLING

The sampling concept is one of several statistical concepts to be developed with students, as part of the investigative cycle research. According to Stevenson (1981), sampling is the part of the group being examined, and the entire group is called population or universe. To him, “the purpose of sampling is to make general assumptions about the whole group, without examining every and each of its elements” (Idem, p. 158).

Some authors as Rubin, Bruce and Tenny (1990), Garfield (2003) and Innabi (2006) have mentioned difficulties students and lay people alike face when tackling statistics to understand the basic concepts of sampling. Besides, when pondering on the rigor of the samples, these people may be neglecting the quintessential factors of their representativeness, namely, size and variability.

Ben-Zvi, Makar, Bakker and Aridor (2011) noticed that by using a sequence of activities with 11-year-old students, in which an increase of the size of the samples had been identified, they got stimulated to think about the relations population-sample. The analysis of the individuals’ inferential reasoning showed a development in points of view on what may be concluded from a small sample (contradictory and full of uncertain inferences) and with large ones, which favor the use of informal inferences.

The same way, during a didactic sequence applied to 12-year-old children, Gil and Ben-Zvi (2010) noticed that the students’ first explanations, after being questioned about the rigor of the information based on the sample, reflect a conflict between random sample perception and reliability in their inferences. However, along with the activities and discussions, Gil and Ben-Zvi undertook the development of conceptions about sample and sampling, as students tried to infer from a random sample on population. Ideas at random and random sampling were partially understood and used by students at that age group. They were able to understand the implications in sampling representativeness and their variability. However, they were not able to understand their relations.

Pfannkuch (2008) also noticed that, when carrying out a study aiming to identify conceptual growth on students from New Zealand about sampling variability, 14-year-old students started to build some relations about sample variability upon studying samples of different sizes. It was concluded that the comprehension of sampling concepts became even more complex due to limited knowledge of other concepts, such as distribution and variability. Nevertheless, by involving students in contexts composed of discussions over sampling, students were able to show development regarding ideas of sampling variability and to relate population and sample.

The study of Pfannkuch (2008) ratifies the idea that the aspects approach linked to sample and population concepts, such as representativeness, sampling variability, inference, and distribution, when carried out with interrelation, facilitates understanding and building of sampling.

Given the importance and the needs to develop skills related to statistical reasoning, a research realized by Gomes (2013) was carried out to verify what students in the 5th and 9th grades of schooling understood of sampling. This article show some questions about this study emphasizing the definition of the sampling concept, an example of sampling, selection, size, purpose and representativeness of the sample, and population definition.

METHODOLOGY

Our aim is what students of 5th and 9th grades understand about sampling? Forty students in the 5th grade (10 years old) and 9th grade (14 years old) of Secondary Education took part in this study, 20 of each group. These groups belong to the final grade of each level of the Brazilian curriculum. Semi-structured, person-to-person audiotaped interviews composed of 10 questions on sampling were carried out as follows by one of the research. Based on the answers given by the students, categories were created observing and valuing the diversity of responses on each question: not respond, incorrect, partially correct and correct. A comparison of these grades allows observing the teaching of these concepts in school has provided a meaningful learning.

RESULTS

Regarding sample concept, students were encouraged to answer two questions: “What do you think sample means?” and “To find out which candidate running for
mayor of Recife has a greater chance to win the elections, researchers interviewed a sample of one thousand voters. What does this sample mean in this case?" Most of the students did not answer or answered incorrectly these two questions about sample definition (33 individuals). It was noticed that the word “sample” was mostly associated with the verb to show (they are homographic in Portuguese). This association shows the influence of regional linguistic characteristics in the formulation of concepts, provided that this word is many times applied as synonym of ‘to show’ by people that live in Pernambuco, as shown in the following example:

Student 1: Amostrar alguma coisa, um objeto (To show something, an object.)

Student 2: Quando a pessoa pode mostrar alguma coisa. (When someone can show something.)

A great number of appropriate answers for the first question from students in the 9th grade were observed. The difference in the years of schooling is really meaningful ($X^2 = 4.329$, gl 1, $p<0.037$). It was also common among the appropriate answers examples to be given in lieu of defining the concept of sample, as seen in the following example:

Student 37: Uma amostra de sangue que as pessoas tiram. (A sample of blood people collect.)

Student 40: Eu acho que é uma pequena parte de tudo. (I guess it is a small part of everything.)

In the second question, the students who gave appropriate definition for sample were those who defined sample as a part of all voters in Recife.

Student 34: Eu acho que essa tá significando uma parte dos eleitores pra saber o que eles preferem. (I guess it means a part of the voters to know what they prefer.)

At this question, although the 9th graders had shown a greater number of appropriate answers, unlike the previous question, there was no significant difference in the years of schooling ($X^2 1.558$, gl 1, $p < 0.212$). These data imply that the same concept has different ways of tackling its definition, that is, by approaching directly or by a context.

When students were once again asked to give an example of sample, most of them (30 individuals) did not exemplify nor presented incorrect examples. A great deal of the given examples did not show any association with statistics. In fact, they were linked to the concept of sample defined by the students at the first question.

Student 01: Eu quero lhe amostrar (mostrar) meu caderno, ai eu amostr o à senhora. (I want to show you my notebook, then I show to you.)

Student 14: Eu pegar esse lápis e lhe amostrar (mostrar). (I take this pencil and show you (to show).)

Students in the 9th grade gave significantly better answers than students in the 5th grade ($X^2 4.800$, gl 1, $p < 0.028$). In studies carried out by Rubin, Bruce and Tenny (1990), it was shown that several answers from the students had been grounded on their own personal experiences. Because of that, we believed this question would have been quite easy for them to respond. We found daily situations in which sample was used, such as: free sample of products, or else blood or saliva sample for DNA test. Yet, it was clear the participants did not relate these examples to the applied question.

Nevertheless, the appropriate answers presented by the students in the 5th grade show that it is possible for the students at this age to learn this concept. Those who defined this concept partially or totally correct proved to understand sample as part of a whole.

Student 10: Por exemplo, um suco, uma amostra de suco. (For instance, some juice, a sample of juice.)

Student 32: Aqueles vidrinhos de perfumes ou aqueles papeizinhos com perfume quando a gente passa e entregam pra gente. (Those little bottles of perfume or those pieces of paper sprayed with perfume given to us when we pass by the stores.)

Concerning the sample selection, two questions were asked chiefly to understand whether students were able to list important criteria in order to select a sample. During the first activity, students were encouraged to come up with a strategy to select a sample, as being the most possible representative of the studied
population. Thus, the following situations were introduced: “A researcher wanted to know the students’ favorite snack in the public schools of Recife. As he could not interview all students, he decided to interview only two hundred of them. How should he choose these students in order to have a better scope on their preference?”

The performance of most participants, once again, was below the expected. The number of appropriate answers was higher among participants in the 9th grade. However, there was again no significant difference in the years of schooling ($X^2 = 0.476$, gl 1, $p < 0.490$). By suggesting the way of sample selection, these students showed similar answers to those found in Rubin, Bruce and Tenny’s (1990) studies, with students in High School (14 to 17 years old). The participants presented variable models in association with sample selection so that it could represent the expected population, as long as their answers were grounded on personal intuition.

Answers with at least one aspect referring to probability sample were taken as correct; in other words, those that may be used with the entire population. In this kind of answer, it was common for the participants to suggest that the sample be chosen by drawing the lots, characteristic of random sampling, as described below:

Student 28: Ele dividia pelas escolas que tem no Recife e sorteava. (He divided by schools in Recife and drew the lots.)

The second question about the sample selection showed a hypothetical context: “Five friends wanted to know approximately how many books the people who live in their neighborhood read a year. As the district had about 10,000 residents, they couldn’t to interview every one”. Participants were asked to indicate the most appropriate sample among five options: 100 residents who frequent de local library (large and biased); 100 residents of the district (large and unbiased); 10 residents who frequent de local library (small and biased); 10 residents of the district (small and unbiased); men, women, boys and girls (without information about size and sample selection criterion). Thirty-one students gave inappropriate answers, supporting their answers without taking into account whether the sample had characteristics of the studied population or a bias of selection. Once again, there was no statistical difference between grades analyzed ($X^2 = 1.290$, gl 1, $p < 0.256$).

In the same question, answers were analyzed taking into consideration, this turn, size in association with sample selection. Although there has been a difference between grades, it was not shown to be significant ($X^2 = 3.750$, gl 3, $p < 0.053$). The results achieved revealed a more appropriate performance of the participants, since twenty-three participants had their answers classified as partially correct, and only one as correct.

Answers were considered partially correct when students took into account the larger number of residents, disassociating population with the number of individuals of the sample as well as one of the criteria for their representativeness.

Student 1: Porque a opção um tem mais moradores. (Because option one has more residents.)

Student 13: [...] A opção 2. Porque tem mais pessoas. [...] (Option 2. Because it has more people.)

The single participant with a correct answer was a student in the 9th grade, who emphasized the size of the sample as one of the important factors for its representativeness. Likewise, although a smaller sample could also represent the population of the mentioned example, a larger number of individuals would be more appropriate.

Student 34: [...] porque o três pegou só 10. Vai dar, mas vai dar muito pouco para saber do bairro todo. [...] because the number three had only 10. It will do, but it will not be sufficient to learn about the entire neighborhood.

Regarding the purpose of using samples, participants had to answer the following question: “To find out which candidate running for mayor of Recife has a greater chance to win the elections, researchers interviewed a sample of one thousand voters. Why do you think they used a sample and not all voters in Recife”?

It should be pointed out that, despite the fact that a large number of participants could not produce a definition, nor provide a correct exemplification, when led to a situation of research by using samples
and asked to explain why using them, almost half of the students (16 individuals) answered appropriately. Among these participants, 8 (eight) were students of the 5th grade, who explained their answers based on the convenience of using samples, as well as declared to have skills to broaden their knowledge on the purpose of using samples, as answered the student 34:

Student 34: Se tiver 2 milhões de habitantes no Recife e ele pegou mil já dá pra ter pelo menos uma ideia. (If there are 2 million inhabitants in Recife and he got a thousand, it is enough, at least, to have an idea.)

There was no significant difference between the levels of schooling ($X^2 = 0.000, gl 1, p < 1.000$). Results achieved by this question highlighted the importance of activities with different contexts, involving sample concept to enable learning.

Still, another aspect verified has to do with sample representativeness. It is important to remember that it had been assessed indirectly in other questions; however, being representativeness the main purpose of sampling, it had already been used directly in specific questions. A situation was presented to the students, in which a research would be carried out in a school assigned beforehand. They were asked whether to select individuals for a research in order to represent the school, it would be more appropriate to draw students of all years of schooling or rather only a class would be enough. More than half of participants responded incorrectly (22 individuals). When answers were classified as appropriate or inappropriate, it was clearly proved that the difference in the performance of 5th and 9th graders was not significant ($X^2 = 2.558, gl 1, p < 0.110$). Some students chose the sample without justifying the reason for the option; others presented justifications unrelated to the sample representativeness; and there were those who inappropriately used aspects associated with sampling, as Student 25 explains:

Student 25: Melhor (só de uma turma). (It is better (only one classroom).)

Researcher: Por que melhor? (Why is it better?)

Student 25: Porque vai ser mais rápido para entrevistar eles. (Because it is going to be quicker to interview them.)

Better posed answers indicate that, if encouraged, students of different ages would be able to develop the skills needed to list eligible criteria of a sample. For justifications partially correct or correct, the relevance of sample variability was implicit.

Student 27: Pegar de uma sala só é pior, professora. Não é melhor escolher da escola toda que tem mais variedade?! Da mesma sala vai mostrar só daquela. (Collecting from only one classroom is worse, teacher. Isn’t it better to choose from the entire school that has more variability?! From the same classroom it will show only from that one.)

This comes to show that, although students could elaborate a more formal and complete answer, they noted that, to validate a sample, it must have some specific features of the population.

To investigate the understanding of population, two questions were asked. The first question raised the idea of population as a group of people. Thus, it was included in the previous situation about the candidate running for mayor of Recife, the following question: “What would the analyzed population be?” More than half of participants (25 individuals) responded appropriately; however, more students in the 9th grade responded in an appropriate and significant fashion ($X^2 = 5.227, gl 1, p < 0.022$). The correct answers narrowed the population as to target variable, in other words, those who could vote.

Student 29: Os eleitores. Do Recife. (The voters. From Recife.)

The second question took population as a group of objects: “If a research intending to find out how long computers of a specific brand last were carried out; what type of population would be analyzed in this research?” In this situation, most of them responded incorrectly (35 individuals) without any significant statistical difference within the grades studied ($X^2 = 5.227, gl 1, p < 0.022$). Five students, three in the 5th grade, answered correctly, proving to understand that population not always are people, but the whole being researched, as it is clear in the answer of student 06:

Student 25: Os computadores da marca. (Computers of a brand.)
This understanding shows that even more complex definitions could be more simply worked in the early years of schooling.

Thus, it is important that these students’ performance differences be discussed during activities that involve the same definition. As Ben-Zvi, Makar, Bakker and Aridor (2011) state that a variety of situations encourage students to think about the population-sample relations.

CONSIDERATIONS

Investigating different skills with the same students enabled us to identify what skills are more easily understood by them. Similarly, when we verify partial understanding of these different skills, we have some suggestions to put into use during the teaching-learning process, starting with the situations understood by the students in order to build more structured and complex knowledge.

This research showed that despite the great difficulty of the students to understand concepts related to sampling, students from the 5th grade already show to understand concepts associated with sampling in some situations. The identification of this students’ aptitude make us to ratify that it is necessary to rethink of what schools can and should offer to the students. The learning ability of these contents from the early years, as shown by the individuals of this research, outscors the idea that studies on sampling should not be limited to the last years of schooling; rather, it must seek strategies to improve these skills in order to find positive changes since the initials levels of education.

REFERENCES


