HOW DO TERTIARY STUDENTS ASSESS SAMPLING ON SOCIAL CONTEXTS AND MEDIA REPORTS? EXPLORING STUDENTS' STATISTICAL LITERACY WITH RELATION TO SAMPLING

Eleazar Silvestre[1], Manuel Urrea[1], and Miguel Medina[2]
[1]Department of Mathematics, University of Sonora, Mexico
[2]Campus 2 "100 Meters," Seniors College, Mexico
eleazar.silvestre@unison.mx

Sampling is a foundational concept for a proper development of students' statistical literacy. This paper focuses on the evaluation of Mexican tertiary students' abilities to evaluate statistical conclusions or claims in relation to sampling found in a social and media report context. The SOLO taxonomy is used to evaluate the richness of students' responses in relation to sampling for two scenarios. General results reveal students have trouble correctly identifying or calculating sample size for a given study and a tendency to accept claims based on inappropriate sampling, even if the selection method has an evident source of bias. Implications for teaching in relation to sampling and statistical literacy are discussed.

BACKGROUND AND RESEARCH QUESTION

Necessary components for statistical literacy include the understanding of why data are needed, how data can be produced, and how statistical conclusions or inferences are reached (Gal, 2002). Early instruction on sampling attempts to develop the foundations for this kind of knowledge and the associated abilities by addressing the concept of representativeness, which is grounded in embracing random sampling methods and acknowledging the representative power of a sample based on a sufficiently large sample size. However, even if this teaching is already quite restrictive when referring to key ideas of sampling in relation to inference (Harradine et al., 2011), the specialized research literature reveals that these notions are not trivial to students and that applying this knowledge outside the classroom is especially complicated. This is a big concern for reaching a main goal of statistical literacy: being able to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena found in diverse contexts (Gal, 2002).

When referring to conceptions of sampling, Watson and Moritz (2000) show that elementary students' (9–15 years old) concept of sampling gradually incorporates the part-whole relationship to their definition of what a sample is, a preference for large sample sizes, and a selection method that avoids bias. Nevertheless, most of Watson and Moritz's subjects often neglected the presence of bias in sampling selection methods when analyzing claims based on sample data and were prone to believe in the law of small numbers. Furthermore, Watson (2004) shows that student's concept of sampling improves over the schooling years, but only a small fraction of students reaches the most sophisticated category of her hierarchy. Only these students (approx. 13–18 years old) "suggest selection based on a random process or distribution by geography" and "identify biased samples in newspaper articles reporting on results of surveys" (Watson, 2004, p. 282).

In a more recent study, Ruiz and Contreras (2021) provide evidence that Chilean secondary students (13–18 years old) share some of the difficulties of Watson's (2004) subjects; about 40% of their participants answered an item that required distinguishing between representative and non-representative samples based on their selection method correctly; when suggesting a sampling method, the same percentage provided an answer but were unable to justify their proposal. This research suggests students are prone to admit biased methods of sampling and only focus, at the most, in sample size as a main factor for representativeness.

Wroughton et al.'s (2013) research points to the influence of college students' attitudes and beliefs about the context of the situation in which the sampling process occurs. Their results suggest that, when students' opinions about a topic are strong, their process of validating claims arising from statistical studies focuses on contextual aspects of the situation instead of the underlying statistical issues, thus permitting a belief and confirmation bias to be triggered. According to the nature of their methodology, the authors also warn this pattern may appear more strongly outside the school setting and call for more research that could reveal its extent and possible causes.

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

There are no available studies like those described in the case of Mexico, but in Mexico, the mathematics curricula include probabilistic and non-probabilistic sampling methods in high school under the disciplinary axis "From information management to stochastic thought" (Secretaría de Educación Pública, 2017). Still, many educational systems do not require the subject of statistics to be mandatory. Gómez-Blancarte et al.'s (2021) research suggests that high school teachers teach statistics according to a unified vision of statistical literacy, statistical reasoning, and statistical thinking, with most of them stating that basic statistical knowledge referring to sampling (i.e., "understand the different ways to select samples") and critical questioning (i.e., "question whether the sample in a message or statistical study leads to a valid inference about the study population") are considered in their teaching practices in a "much more than what is indispensable" manner (Gómez-Blancarte et al., 2021, p.11). The authors notice that the emphasis of teaching may be placed more strongly into the procedural aspects of statistical concepts rather than applying this knowledge in genuine statistical projects or when evaluating statistical information found in everyday contexts and situations.

Considering this scenario, the current paper responds to the need of generating evidence in the Mexican context related to students' abilities in assessing the validity of claims based on sample data found in everyday contexts. Our motive is that given that the studies mentioned above exhibit difficulties for students leaving the high school level, we wonder if those difficulties are still present at the tertiary level. Thus, we are driven by the research question: how do tertiary students assess sampling in social contexts and media reports? We present preliminary results of an exploratory study in which Mexican tertiary students evaluate the validity of a claim based on an inappropriate use of sampling in two scenarios: one that (incorrectly) refers to a different target population and another one that presents conclusions based on a biased sampling method.

METHOD

Participants and Instruments

An online survey with five sampling problems was administered to 59 tertiary students enrolled at the University of Sonora, Mexico. Students came from the areas of socials, engineering, and medicine. At the time of the study, students were beginning a tertiary course on inferential statistics; we assume their statistical background was mainly comprised of descriptive statistics (representations, central tendency, and variability measures) and probability notions (introduction to classical and frequentist approach). We report analysis of responses on two problems that mirror Watson's (2004) problems about statistical literacy in relation to sampling. The problems were posed to students as follows.

- *Problem 1.* Xochitl claims that a study shows that one out of ten Mexicans suffered from discrimination issues during the last year. In the study from the collective 'Hermosillo, ¿Cómo vamos?'—who administered 1,586 random surveys to Sonoran adults distributed in the localities of Hermosillo City, Miguel Aleman, La Victoria, and El Saucito—there was evidence that the first cause for discrimination was religion, followed by features such as skin color or place of origin. *1(a) Would you make any criticisms of Xochitl's claims? 1(b) Does the information referred to by Xochitl make you feel insecure if you are practicing a non-conventional religion in some other city inside Mexico?*
- *Problem 2.* About 96% of callers to a radio station located in Mexico City expressed that marijuana consumption should be decriminalized in all of Mexico. The phone-in listener poll, which closed yesterday, showed that 9,924 out of the 10,000-plus callers favored decriminalization. Many callers expressed they did not smoke marijuana but still believed in decriminalizing its use and only 389 considered that possession should remain illegal. *2(a) What was the sample size of the phone-in listener poll? 2(b) Is the sample reported here a reliable way to determine public support for the decriminalization of marijuana? Why or why not?*

Both problems demanded that participants analyze and question the acceptability of given conclusions based on sampling. In Problem 1, students questioned statistical claims from a social context and in Problem 2, claims from a media report context. Problem 1 solicited participants to critique claims expressed by a university student who is interpreting information found in a local newspaper. The problem required participants to recognize that it is not valid to extrapolate statistical conclusions to non-sampled populations. Thus, in 1(b), students should reject the study's information for making decisions outside the Sonoran state. Problem 2 asked participants to analyze the conclusions made by a radio station based on a phone-in listener poll. Problem 2(a) required

participants to identify the sample size of the poll and, in Problem 2(b), to reject the sample as a reliable source to determine public support for marijuana decriminalization in all of Mexico given that the respondents were not randomly selected among the country.

Data Collection and Analysis

Students completed the online survey as part of a homework assignment during the initial part of their Probability and Statistics course (August 2021) and at the end of another math-related course (December 2021). The researcher-professor and first author of this paper told students they should consider the tasks as if they were responding to a survey on the street, in other words, not worrying about providing correct answers but providing honest opinions and reactions to the problems. Some students were interviewed (including videotape) during class time to gain insights about specifics aspects of their reasoning when completing the survey.

For this report, the SOLO Taxonomy (Biggs & Collis, 1991) is used to analyze students' written responses. The SOLO model is useful to account for students' reasoning by analyzing the structure and complexity of responses to a given problem. The level of an answer (pre-structural, unistructural, multistructural, or relational) depends on the number of relevant components evidenced in solving the problem and the quality of connections among them. For our analysis, reference components were the identification of a sample population context, usage of random selection, and acknowledgment of the power of sample size as main aspects for representativeness. We also interpret results and students' performance by contrasting them with related studies.

RESULTS

Problem 1: Questioning Claims in Social Contexts.

The distribution of student responses to Problem 1 is shown in Table 1. Almost 50% of participants offered no criticism of Xochitl's claim, and a majority of them expressed a mix of personal opinions about the context of the survey that in some way or another acknowledged the reliability of the sample. Student 4 (S4) exemplifies this type of response:

- 1(a): I wouldn't make any criticism, I think she's right about this, people get discriminated because of their religion, [skin] color or origin, which is horrible.
- 1(b): Yes, but it wouldn't matter … I'd still practice what I believe in.

Table 1. Responses according to SOLO levels in Problem 1

SOLO Level	Number of responses	Relative frequency	Properties
Pre-structural	29	0.49	Offers no criticism of sampling and considers the sample a reliable source for decision making.
Unistructural	14	0.24	Identifies inconsistency in target population or points at insufficient sample size but acknowledges reliability.
Multistructural	10	0.17	Identifies inconsistency in target population or points at insufficient sample size and discards it in decision making.
Relational	6	0.10	Identifies inconsistency in target population and points at insufficient sample size; discards it in decision making.

About a quarter of responses were coded as unistructural. These participants identified an inconsistency in the target population expressed in Xochitl's claim or pointed to an insufficient sample size but still accepted the reliability of the sample when expressing personal opinions about what to do with the study's information in relation to their beliefs or decision making 1(b). An example of this type of responses is given by S18.

- 1(a): I don't think that the population of 4 listed cities represents all Mexico's adults.
- 1(b): Most likely yes, given that most of [Mexico's] population is catholic.

A little under 20% of responses were coded as multistructural. Almost all these participants argued that data collection did not include the rest of Mexico's population and thus inferences could only be restricted to the state of Sonora, and just one asserted that sample size was not sufficiently large enough. All these students considered the sample a non-reliable source when answering 1(b); S51 exemplifies this kind of responses:

- 1(a): I don't think she can affirm a claim about [all] Mexicans if the study is only based in the state of Sonora.
- 1(b): No, because in fact it [the study] only talks about residents of the state of Sonora, does not offer results about any other state.

At the relational level, achieved by the remaining 10% of responses, participants identified an inconsistency in the target population and pointed to an insufficient sample size, which made them conclude the avoidance of using the study's conclusion when elaborating a response in question 1(b). S26's response exemplifies this type of reasoning:

- 1(a): Yes … the study was not made in all of Mexico but only in certain neighborhoods of Hermosillo.
- 1(b): No because the sample is very small and comes from a specific city.

*Problem 2: Questioning Claims in Media Repo*rts

The distribution of responses for Problem 2 is shown on Table 2. Almost 15% of responses were coded at the pre-structural level given that participants provided an incorrect sample size, offered no criticism to the statistical issues regarding sampling, and considered the sample a reliable source to find out public support for decriminalization of marijuana. For Problem 2(a), several students noted the number of voters favoring decriminalization, the percentage of voters in the sample that favored decriminalization, or a personal and qualitative appreciation about the sample size of the poll (e.g., "it's an appropriate sample size"). S49's response exemplifies this kind of reasoning:

- 2(a): 9,924 out of 10,000 plus. [It represents] approximately 90%.
- 2(b): Yes because most of the population participated and justified their choice, given that its' just their opinion and not a test the results can vary depending on the person.

Table 2. Responses according to SOLO levels in Problem 2

SOLO Level	Number of responses	Relative frequency	Properties
Pre-structural	8	0.14	Doesn't identify sample size and considers the sample a reliable source.
Unistructural	32	0.54	-Calculates or approximates sample size but considers the sample a reliable source (n = 19). -Doesn't identify sample size but identifies bias in sampling (n = 13).
Multistructural	18	0.31	-Calculates or approximates sample size and identifies one source of bias (n = 15). -Doesn't identify sample size but identifies two sources of bias (n = 3).
Relational	1	0.02	Calculates sample size and identifies two sources of bias.

Over half of the participants responded at the unistructural level. In 19 out of these 32 responses, participants correctly computed the sample size of the poll or provided an approximation but offered no criticism on the methodological aspects of sampling; accordingly, responses to 2(b) reflected that these participants considered the sample a reliable source. The rest of the participants failed to identify sample size but succeeded in identifying a source of bias in sampling, whether a lack of random selection among Mexico's population or an insufficient sample size. Accordingly, these participants rejected the sample as a reliable source. S13's response exemplifies this last type of response.

- 2(a): The poll was made using a sample size of 10,000 people…
- 2(b): I don't think it's a reliable source, it's likely that most of this radio audience is young or open minded…it is necessary to include those people that are not [part of] the specific audience of the radio station.

At the multistructural level, achieved by 31% of participants, all students rejected the sample as a reliable source. Most students correctly computed the sample size of the poll or provided an approximation and identified one source of bias in sampling; three participants failed at identifying the sample size but pointed out these two sources of bias. S44'response illustrates most of this type of response.

- 2(a): It'd be more than 10,000 people[,] as shown in the text.
- 2(b): No, I don't think [marijuana] benefits that much and it's better to make the poll randomly to get different kinds of people.

Finally, only one participant provided a response coded at the relational level. S20 correctly computed the sample size of the poll and rejected the reliability of the sample, arguing the use of an insufficient sample size and the lack of systematic control in applying a random sampling. He also suggested a sampling method that reflected a concern for large sample sizes to assure reliability of the study.

- 2(a): Well, if we add the people who agree with decriminalization with the ones against, we get a sample of 9,924 + 389 = 10,313.
- 2(b): Not really, I think it's a very small sample. We don't know who phoned in and, as we all know, Mexico is a country with lots of states, 32 in total, so, in my view, I would take at least 10,000 samples [participants] per state.

DISCUSSION AND CONCLUDING REMARKS

Students' performance on both problems shows a preponderant percentage of pre-structural and unistructural responses. For the social context scenario, more than half of participants made no criticism to the sampling aspects behind the claim and focused on providing opinions about the context of the situation (discrimination issues). For the media report scenario, more than half of the participants could only identify (or approximate) the sample size or identify one source of bias, most of the time referring to an insufficient sample size.

Despite a better performance in the media report scenario (54% Unistructural and 31% Multistructural), about half of participants accepted one way or another the reliability of the sample in both scenarios. When making a criticism to the claim or reliability of the sample, a significant portion of students focused on sample size and argued it was not large enough. For the media report scenario, about 40% of the total mentions of sampling bias were pointed at sample size (Problem 2(b)). This result is consistent with Reyes and Contreras' study (2021) because students tend to focus on the use of a large sample size as a main criterion to consider a sample as representative and a reliable source for action or decision making.

Differences in students' performance on the problems is consistent with Watson and Moritz's (2000) results; considering the number of respondents that identified the "obvious" sources of bias in both scenarios, the media report scenario turned out to be "easier" for the identification of bias in sampling, with approximately 60% of total mentions pointing to a lack of random sampling as the main reason for bias (Problem 2(b)). This result may strengthen Wroughton et al.'s (2013) conjecture that the more neutral the opinions of students about the context of the situation, the more likely they will succeed at identifying sample bias; in the case of our study, the social context scenario (discrimination topic) could have evoked more personal feelings and attitudes than the media report scenario (marijuana decriminalization topic). Although we did not previously measure the strength of our participants' opinions about the topic of each scenario, we did observe more answers with personal opinions about the context in the social context problem (Problem 1); additionally, we consider our experimentation settings may have provoked students feeling more relaxed about providing a "correct answer" and thus be closer to obtaining more genuine insights of their ways to assess the validity of the statistical claims in each scenario. For these reasons, as Wroughton et al. (2013) suggest, students' performance could be better when surveyed in the regular classroom settings.

A particular result of our study points to the notorious difficulty for students to identify (and correctly calculate) the sample size of a study under scrutiny, given the 40% of incorrect answers in

the media report scenario. Whether a lack of arithmetic skills or confusion between the concepts of sample and population, this pattern points to difficulties in basic skills related to the domain of descriptive statistics, one of the necessary components of Gal's (2002) framework for statistical literacy. This particular result contrasts with the fact that statistics and probability curricula in Mexican educational levels prior to the university level favor topics of descriptive statistics over probability and statistical inference (Inzunza, 2020).

In relation to teaching and considering that our conclusions are restricted to the sample of participants of the study, our students' general performance also contrasts with Gómez-Blancarte et al.'s (2021) result that suggest teachers do encourage the development of statistical literacy skills in the classroom, including those referring to evaluating basic aspects of sampling. On the one hand, perhaps the fact that students tend to focus on sample size more than the method for sampling selection (i.e., probabilistic vs non-probabilistic sampling) may reveal that teachers could be focusing only on the power of a large sample size as the determining factor when looking for representativeness. On the other hand, the fact that only half of students were able to reject claims based on inappropriate use of sampling may point more generally to a lack of attention to this type of tasks within classroom practice, therefore letting the contextual dimension override essential statistical reasoning required for statistical literacy in relation to sampling when confronting everyday situations.

As a general conclusion, our study adds to the body of research that reveals considerable difficulties of young adult students when assessing basic but fundamental aspects of sampling in everyday contexts, as well as to those that call for more research and attention into what the teaching practices about statistical literacy in the classroom settings might really be.

REFERENCES

Biggs, J., & Collis, K. (1991). Multimodal learning and the quality of intelligent behavior. In H. Rowe (Ed.), *Intelligence, reconceptualization and measurement* (pp. 57–76). Laurence Erlbaum.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*, 1–51. https://doi.org/10.2307/1403713

Gómez-Blancarte, A. L., Rocha, R., & Chávez, R. D. (2021). A survey of the teaching of statistical literacy, reasoning and thinking: Teachers' classroom practice in Mexican high school education. *Statistics Education Research Journal*, *20*(2). Article 13. https://doi.org/10.52041/serj.v20i2.397

Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. 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. 235–246). Springer. https://doi.org/10.1007/978-94-007-1131-0_24

Inzunza, S. (2020). Fundamental statistical ideas in primary, secondary and high school Mexican curriculum: reflections from the international perspective. In A. I. Sacristán, J. C. Cortés-Zavala, & P. M. Ruiz-Arias (Eds.), *Mathematics education across cultures: Proceedings of the 42nd Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1308–1316). Cinvestav/AMIUTEM/PME-NA. https:/doi.org/10.51272/pmena.42.2020

Ruiz, K., & Contreras, J. M. (2021). Understanding sampling by Chilean secondary school students. *Statistics Education Research Journal*, *20*(2). Article 11. https://doi.org/10.52041/serj.v20i2.383

Secretaría de Educación Pública. (2017). *Planes de Estudio de Referencia del Marco Curricular Común de la Educación Media Superior* [Reference study plans of the common curriculum framework for higher middle education]. https://www.gob.mx/sep/documentos/planes-de-estudio-educacion-media-superior

Watson, J. (2004). Developing reasoning about samples. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277–294). Kluwer. https://doi.org/10.1007/1-4020-2278-6_12

Watson, J., & Moritz, J. (2000). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, *19*(1), 109–136. https://doi.org/10.1016/S0732-3123(00)00039-0

Wroughton, J. R., McGowan, H. M., Weiss, L. V., & Cope, T. M. (2013). Exploring the role of context in students' understanding of sampling. *Statistics Education Research Journal*, *12*(2), 32–58. https://doi.org/10.52041/serj.v12i2.303