DEVELOPING DATA SENSE: MAKING INFERENCES FROM VARIABILITY IN TSUNAMIS AT PRIMARY SCHOOL

SOLEDAD ESTRELLA
Pontifical Catholic University of Valparaiso
soledad.estrella@pucv.cl

ANDREA VERGARA
Catholic University of Maule
avergarag@ucm.cl

ORLANDO GONZÁLEZ
Assumption University
ogonzalez@au.edu

ABSTRACT

In order to study the manifestation of data sense and identify ways of thinking about variability in authentically realistic problems in a group of Chilean fifth-grade students, a lesson plan was designed and implemented, within the framework of statistical literacy and using the "lesson study" modality, in which students were urged to make inferences based on the analysis of data corresponding to the tsunami that struck the Chilean coast in 2010. This article focuses on the qualitative study of the data representations produced by two groups of students during the implementation of the lesson plan. The analysis of the behavior of the tsunami carried out by the students led them to work simultaneously with nominal qualitative, ordinal qualitative, discrete quantitative, and continuous quantitative variables; to create new variables; to construct representations of data (multiple bar graphs and frequency tables); and to make inferences based on the data. We conclude that the use of an authentic context and the construction of their own representations promoted data sense in students and facilitated the development of their statistical thinking, through which they were able to recognize, describe, and explain the variability of the phenomenon.

Keywords: Statistics education research; Data sense; Statistical variability; Data representations; Lesson study

1. INTRODUCTION

1.1. REGARDING LEARNING AND TEACHING STATISTICAL VARIABILITY

The primary-school curriculum in Chile, as part of the education reform process implemented in 2012, establishes the necessity of promoting the teaching of statistics based on students' own experiences and interests, highlighting the importance of solving problems in context (Langrall et al., 2011; Ministerio de Educación de Chile [MINEDUC], 2018). This approach is part of what is known as statistical literacy (Estrella & Olfos, 2012), which promotes the use of real data, the understanding of how the data are generated, and the making of statistical judgments and inferences (Pfannkuch & Ben-Zvi, 2011), as well as the development of data sense based on solving problems in context (Estrella, 2018). However, in the current curricular guidelines for mathematics in Grades 1 to 6, neither data sense nor the making of inferences appears explicitly (MINEDUC, 2018).

For students to be able to authentically involve themselves in statistical analysis of situations characterized by uncertainty, it is necessary to make visible the omnipresent role of variability in random phenomena. Variability is a complex notion, difficult to conceptualize and define, even in the area of formal statistics, but considering understanding of variability is indispensable for students to develop adequate statistical thinking (Shaughnessy, 2007). It will allow them to identify, characterize, and quantify random processes, in order to predict and reduce variability (Snee, 1993). As such,

reflecting on the variability of data requires having developed data sense, as this is built on a set of abilities and knowledge that favor the use and creation of strategies for investigating statistical problems (Estrella, 2018).

Given the above, our research question is "How do primary-school students reason about variability and how do they manifest their data sense?" To respond to this inquiry, the productions of fifth-grade students (11 years old) were analyzed as they reflected on the tsunami experienced in Chile in 2010. The task proposed in the lesson design asked students to extract, organize, and communicate the behavior of the data from an infographic so that they could create their own strategies to represent aspects of the phenomenon under study. The task was based on real data, under the belief that communicating information based on data contributes to improve decision making for future events.

In considering variability, evoking other concepts such as variation, dispersion, or deviation seems inevitable. In agreement with Shaughnessy (2007), we distinguish variability from variation, as the first corresponds to the attribute of an observable entity, while the second is the measurement of this characteristic. Variability of data is a global characteristic of random phenomena whose measurement can be made considering different statistics such as range, variance, and standard deviation. As such, variability can be expressed through measurements, but it cannot be reduced to a measurement, as it is a perceptible and influential quality throughout all stages of the statistics research cycle (Pfannkuch & Wild, 2004) that must be conceptually perceived as well as measured.

Variability as a teaching object is situated as a concept rightly belonging to the discipline of statistics. The importance of including the statistical practice of modeling variability in situations of uncertainty at the school level has been investigated thoroughly in the last ten years (e.g., Lehrer & English, 2018). A large part of the drive to investigate variability at the school level arises from the difficulties in understanding it at different educational levels and the constant tensions that its perception offers in relation to other statistical ideas, such as distribution, centrality, sampling, and representativity (Garfield et al., 2015; Chan & Ismail, 2013).

Recent studies have addressed teaching and learning variability considering diverse perspectives. Some of these studies were focused on the use of technology for the dynamic exploration of data in order to promote a more qualitative understanding of statistical variability (Ekol & Sinclair, 2016; Inzunsa-Cazares, 2016). Other studies addressed the state of knowledge for teaching variability of both pre- and in-service teachers (Isoda et al., 2018; Lovett & Lee, 2018). In general, these studies highlighted the existence of different ways of considering variability, as well as the limited available research on how to design a lesson plan to determine the relevance of these particular ways of understanding the concept and how to promote them at school level.

1.2. CHILE, A COUNTRY OF NATURAL DISASTERS: TEACHING STATISTICS FOR DECISION MAKING

We understand that learning in general is strongly influenced by sociocultural factors, which determine contexts and activities that interconnect community, family, and the individual (Phan, 2012). One of the most difficult aspects to consider in relation to educational statistics, both in teaching proposals and in research design, corresponds to the sociocultural dimension. From this perspective, we extract the importance of considering those sociocultural activities, linked to uncertainty, that can contribute to improving teaching processes connected to statistics literacy.

The ways of thinking statistically about random phenomena require not only theoretical knowledge but are also determined by a vision of the real world and a broad knowledge about contexts and their meanings (Gal, 2005). Consequently, the challenge of including statistical problems based on real data demands consideration of situations that end up being contextual and incidental on the shared problems of the community in which the subjects are immersed, as they are the ones affected by the decisions made.

With regard to the inclusion of contextual situations for teaching variability, Torok and Watson (2000) recommended to include the analysis of situations that manifest the real world's own variability, as these allow for considering many sources of variability influencing the results. Similarly, other studies have proposed involving students in learning situations based on the analysis of natural phenomena, which are related to their own experiences. Shaughnessy and Pfannkuch (2002), for example, presented real data from a series of waiting times for geyser eruptions, with instructions to

analyze the data using and creating representations. Meanwhile, Lehrer and colleagues (2007) analyzed how primary-school students model plant growth based on their own perceptual experiences of the phenomenon and using drawings to complement the measurements of change.

In the kind of activities described, the variability of the data is not completely random, but still the signal of the variability is much stronger than the signal of the centrality. Experiences of these characteristics, which aim to bring students closer to understanding and modeling variability in order to give answers to situations of real local interest, are rare in the literature reviewed regarding primary education (e.g., Smith et al., 2019; English & Watson, 2018). At this stage of education, experiences trying to introduce children to professional statistics practices predominate instead (Lehrer & English, 2018).

In this scenario, and considering the complexity that comes with promoting learning about variability and developing data sense from an approach of authentic statistical practices, we propose analyzing the early forms of thinking about variability in data in Chilean fifth-grade students, who reflected on tsunamis as natural phenomena that form a regular part of their sociocultural surroundings. The choice of the tsunami was due to a problem that arose at that time, which deeply affected the country's seismic culture, as many poor governmental decisions were made regarding the appropriate times to emit evacuation warnings in the affected areas, and also poor personal decisions were made in thinking that the tsunami was a single wave, which caused the loss of many lives.

In the present study, the data used in the lesson corresponded to the tsunami that struck the shores of cities of southern and central Chile on February 27, 2010. The design and implementation of this classroom situation was carried out in the framework of a lesson study, which led to a process of empirical investigation of exploratory data analysis based on real-life problem solving (Estrella et al., 2014). The analysis of problems related to natural disasters is relevant to the understanding of data as evidence for making better decisions. This study takes advantage of the significance of this context to reveal ways of thinking about variability using a conceptual framework that coordinates an adaptation of the analytical categorization established by Dierdorp et al. (2017), with the conceptual construct of data sense proposed by Estrella (2018).

2. CONCEPTUAL FRAMEWORK

2.1. STATISTICAL VARIABILITY

Dierdorp et al. (2017) developed a framework for analyzing variability in order to determine if the linear regression tasks proposed for twelfth-grade students favored processes of reasoning about variability. To do so, they based their work on the ways in which statisticians consider variability, as the activities were inspired by authentic professional practices. The ways of thinking established by the authors were expressed as adverbs and correspond to five actions regarding variability: considering; recognizing and describing; measuring; explaining; and using research strategies. We consider these actions not to be mutually exclusive or hierarchical, given that, for example, a student could want to explain the sources of variability without having tried to measure it. The adaptation of these ways of thinking about variability, given the goal of this study, is presented in Table 1.

2.2. DATA SENSE

Developing data sense is necessary in an information society, and even more so as big data takes on greater relevance in society. McIntosh et al. (1997) conceived number sense as the ability and propensity to use numbers and operations flexibly. In the early nineties, Garfield (1993) commented that developing data sense should occur especially in the process of solving statistical problems; while Friel et al. (1997) considered that an essential part of data sense lies in the understanding that a statistical investigation is a process.

In Chile, beginning in 2012, Estrella and colleagues have proposed teaching and learning statistics from the first years of formal education and have addressed the idea of data sense. So, Estrella (2017, 2018) focused on early statistics and positioned her research in observing and discovering fundamental ideas and common procedures in statistics processes from the beginning of primary school. She positions data at the center of statistics and, as such, seeks to define the meaning of one of the

fundamental ideas of the discipline, studying its development in learning school statistics, concluding that this implies facing a problem with a certain number sense in context, which gives meaning and integrates uncertainty (Estrella, 2018; Estrella et al., 2020).

In this way, data sense involves a set of knowledge (e.g., representing data in multiple ways depending on context), intuitions (e.g., expressing oneself regarding data without making any calculations), and abilities (e.g., making comparisons between data in different data representations) that a subject develops regarding data and their behavior. This set of knowledge, intuitions, and abilities translates into the capacity to solve problems in contexts based on real data, to understand how these real data are generated, to make statistical judgments and informal inferences (Makar & Rubin, 2009) and, lastly, as mentioned by Gal (2004), to pose questions that activate a critical posture. Consequently, data sense can be developed through teaching that promotes the use of reasoning and strategies belonging to statistics but considers students' experiences and interests. Based on the understanding that data sense is not determined by a definition, Estrella (2018) proposed ostensible actions in performance and characterized data sense through ten components (see Table 1).

2.3. CONCEPTUAL INTEGRATION PROPOSAL: PERCEIVING VARIABILITY THROUGH DATA SENSE

The ways of thinking about variability suggested by Dierdorp et al. (2017) can be connected to data sense, as proposed by Estrella (2018), given that data sense exhaustively establishes a wide range of actions that emerge when facing situations of a statistical nature. The ten actions delimited by data sense have been qualitatively reorganized, based on their relevance, as indicators that describe the different ways of thinking about variability with greater precision (see Table 1). Additionally, an example has been incorporated for each way of thinking, to highlight the specific learning situation that is the object of analysis of this study. However, despite all this, we consider that this conceptual proposal is feasible to be adapted and utilized in other instances that aim to understand ways of thinking about variability.

Table 1. Conceptual proposal for analyzing variability through data sense

Way of thinking about variability	Indicators of data sense	Example
Recognizing variability (REC)	Recognizing the necessity of data, looking for connections between new information and previous contextual knowledge. Being able to move appropriately between real-world situations and the statistical world of data.	Noting one of the variables involved, such as the height of the waves, the time of the waves' arrival, city, or number of waves per city. Referring to the variability of the tsunami's behavior.
Describing variability (DES)	Being able to think about or express in a sensible manner the behavior of the data in a statistics problem, without making a precise calculation. Being able to understand numbers in context as references for measuring variability in the real world.	Expressing that in each city the height of the waves varies. Numerically describing the quantity of waves, the time of arrival of a wave, or the height of the waves.
Measuring variability (MEA)	Approximating or estimating from the evidence of the data's behavior. Making comparisons of numerical magnitude data among different data representations.	Estimating the range of the heights of the waves or of the time of their arrival to shore. Comparing the number of waves with the height of the waves in each city.

Explaining and representing variability (E&R)	Being able to represent data in multiple ways depending on the context and the purpose of the data representation, moving among different representations to obtain a greater understanding of the behavior of the data and make predictions beyond the given data.	Representing in a table the height of the waves in each city, pointing out that a tsunami does not consist of one big wave, but rather of various waves in different moments.
	Being able to use numbers and quantitative methods to communicate, process, and interpret information provided by data and contextual knowledge.	Explaining the visualization of data using a graph, considering one or more variables, in order to use this resource to interpret or give meaning to variability.
Using investigative strategies to manage variability (INV).	Recognizing investigation as a statistical process that includes a problem, a plan, data, data analysis, and conclusions, using data as evidence.	Manifesting that it is important to engage in methodical investigation in order to make better decisions with regard to evacuation processes in natural disasters.
	Becoming aware that an investigative process is being carried out through a real experience with data and the understanding of statistical concepts, using them to make everyday decisions.	Proposing, with evidence based on the data, a way of adequately warning people living near the coast about the irregularity in the time between waves, or that a tsunami is more than one wave.

3. METHODOLOGY

This is a qualitative study that seeks to determine and characterize the ways of thinking about variability manifested by students, using indicators provided by data sense. To do so, the written productions of students' work groups are analyzed.

3.1. PARTICIPANTS & DATA

A class of 30 fifth-grade students in a school in the region of Valparaiso, Chile, participated in a problem-solving based statistics lesson within a lesson study framework. The students reflected on the experience first individually and later worked in pairs. The written productions used in the study were taken from their work in pairs.

3.2. STATISTICS LESSON DESIGNED WITHIN A LESSON STUDY FRAMEWORK

The lesson was designed jointly by four teachers and two researchers in a lesson study group that promotes the incorporation of research processes as part of the continuing professional development of classroom teachers (Estrella et al., 2018; Isoda & Olfos, 2009). The lesson was prepared considering the specific statistical reasoning to be promoted, the curricular content for the grade level, and the lesson management structure. The teachers, in collaboration with the researchers, specified the learning goals and anticipated possible difficulties the students could have, together with other critical aspects of the lesson goal and the respective teacher actions. With this background, the lesson plan was designed (see Appendix) and underwent two successive implementations in different classes of the same level, which permitted observing, discussing, and improving the design of the lesson.

3.3. LESSON PROBLEM AND DATA COLLECTION

The lesson began with the teacher inviting the students to reflect on tsunamis, based on simple questions and watching a video on the topic. The students talked about their experiences of the event, which gave way to a whole-class discussion of the relationship between tsunamis and earthquakes. The teacher completed this first discussion by presenting the infographic that appeared in a nationwide

periodical about the tsunami that occurred after the magnitude 8.8 earthquake at 3:34 AM, February 27, 2010 (Figure 1).
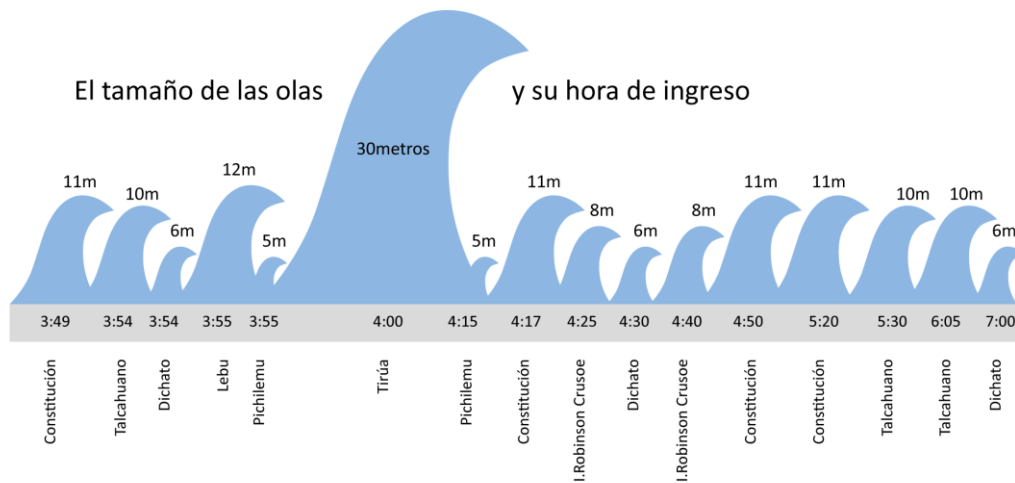

Figure 1. Infographic given to students on the worksheet
(Source: National periodical "El Mercurio", digital version, 03/04/12)

Each student received a worksheet with the infographic shown in Figure 1. The teacher encouraged students to observe the infographic and reflect individually on the data contained in it, asking them to organize the data in a simpler manner. Once the teacher confirmed that each student had understood the problem, he asked them to work in pairs, reiterating the challenge, "Extract and arrange the data to communicate information to towns that were or could be affected by tsunamis and to help to avoid another tragedy."

The students worked with pencil and paper, and their productions were collected at the end of the lesson. These productions constitute the medium for data collection for this study. From the twelve collected productions, two were chosen given the diversity of the representations created. Through these student data representations, two cases are presented and analyzed using our conceptual integration proposal: the ways of thinking about variability using data sense (see Table 1). Fictional names were given to the students in the working pairs.

3.4. THE CASE OF ANITA AND LUISA

Under the title "Size of the waves and time of arrival at towns after the earthquake," Anita and Luisa represented three variables in a multiple bar graph (one to four bars per category) based on the data presented in the infographic (Figure 1). Although the pair committed an error in classifying one piece of data regarding the towns, assigning to *Dichato* a wave that corresponds to *Constitución*, the analysis and organization of the data were not affected.

The x-axis presents the categories of the nominal variable (seven towns affected by the tsunami) and on the y-axis, the continuous variable, the height of the waves. On the y-axis, correctly graded by twos and beginning with an explicit zero, the bars were designed proportionally to this grading, according to the data provided in the infographic (which indicated the height as the size of the image). Also, using different colors, they designated a new ordinal variable (the order of arrival of the waves to the coast) associated with time. In this way, the bar graph showed for each town the height of the first, second, third, and fourth wave, and their respective times of arrival (see Figure 2).
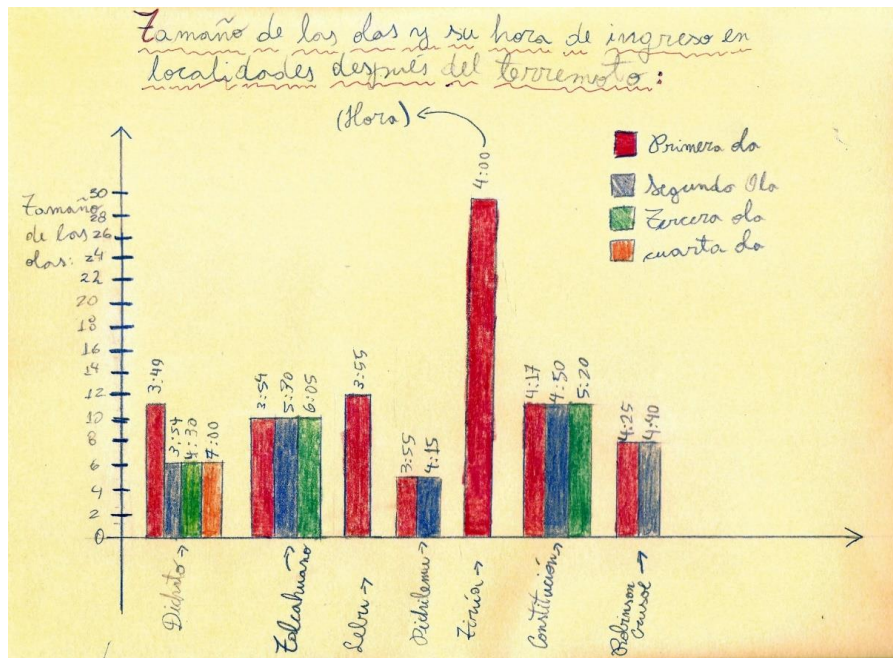
Figure 2. Bar graph with three variables made by Anita and Luisa

According to the representation made by Anita and Luisa, a multiple bar graph, and considering the conceptual proposal for analysis, we observed that the students presented three explicit forms of thinking about variability: recognizing variability (REC), describing variability (DES), and explaining and representing variability (E&R). There is also evidence that the students implicitly measured variability (MEA), because writing down the time of arrival of each wave over the bar was a reference to the time elapsed between one wave and another, as well as between the first and last waves. Similarly, the fluctuation in height of the bars was a reference to the behavior of the height of the waves in each town, in that each group of bars represented a comparison of numerical data (the height of the waves).

To organize the analysis of Case 1, each explicit form of thinking about variability is accompanied by the indicators of data sense and a description of the behavior that justifies each assigned category (see Table 2). The data representation elements used by Anita and Luisa manifest data sense, which allowed them to implicitly coordinate relationships among variables and visually describe the variability of one variable with respect to another.

Table 2. Analysis of ways of thinking about variability using data sense in a bar graph (case of Anita and Luisa)

Way of thinking about variability	Indicators of data sense	Justification
REC	Recognizing the necessity of data, looking for connections between new information and previous contextual knowledge.	Anita and Luisa recognized the need to make reference to the variability of the data using different types of variables, such as the qualitative nominal variable associated with the towns, the qualitative ordinal variable associated with the order of arrival of the waves to the coast, and the quantitative continuous variable associated with the height of the waves in meters. The determination of these variables showed evidence of knowledge of the problem's context, because all the information is integrated through the explanation of three convenient variables.

DES	Being able to understand *numbers in context* as references for measuring variability in the real world.	Anita and Luisa described the variability of the data principally through the contextual relationships of the three variables. This relationship allowed them to describe the variation both in height and in time of arrival to the coast as a function of the town, indicating numerically the different heights in meters and the different times as references for visualizing the phenomenon's variability.
E&R	Being able to represent *data* in multiple ways depending on the context and the purpose of the data representation, moving among different representations to obtain a greater understanding of the behavior of the data and making predictions beyond the given data.	Anita and Luisa represented in the same graph the height of the waves according to their order of arrival at the coast and the number of waves in each town. This can be observed both by reading the numerical and textual expressions in the bar graph and by the different colors and heights of the bars.

3.5. THE CASE OF ROBERTO AND JUAN

Roberto and Juan made an eight-by-four data table, with each of the headings of the four columns designating a specific variable in relation to the central elements of the phenomenon, with the data presented in the image in Figure 1. The cells of the body of the table summarized the values associated with the four types of variables created: two continuous (time of arrival and height of the waves in meters), one nominal (names of the seven towns), and one quantitative discrete variable denominated "how many times it repeats", which expresses the frequency, that is, the number of waves per town. The creation of this last variable was completely novel in relation to the information contained in the infographic. Each row of the table corresponded to a town, describing and connecting the variables involved (see Figure 3).

Hora de ingreso	Tamaño de la ola	Localidad	Cuantas veces se repite
3:54, 4:30, 7:00	6m, 6m, 6m	Dichato	3
3:54, 5:30, 6:05	10m, 10m, 10m	Talcahuano	3
3:55	12 m	Lebu	1
3:55, 4:15	5m, 5m	Pichilemu	2
4:00	30 m	Tirúa	1
3:49, 4:17, 4:50, 5:20	11m, 11m, 11m, 11m	Constitución	4
4:25, 4:40	8m, 8m	Isla Robinson Crusoe	2

Figure 3. Data table with four variables, created by Roberto and Juan

According to the tabular representation created by Roberto and Juan, and considering the forms of thinking about variability and the indicators of data sense, it can be seen that the students in Case 2 explicitly manifested the forms of thinking about variability REC, DES, and E&R. However, additionally, we identified evidence of measuring variability (MEA), because the last column expressed an estimation (or transnumeration) of the behavior of the variables (of columns 1 and 2), which

facilitated comparing data of numerical magnitude (e.g., looking at the data in column 3 in relation to column 4 allows one to say that *Talcahuano* was hit by one less wave than *Dichato*). To organize the analysis of the case, we will use a table similar to the previous case (see Table 3).

Table 3. Analysis of ways of thinking about variability using data sense in a frequency table (case of Roberto and Juan)

Way of thinking about variability	Indicators of data sense	Justification
REC	Being able to move appropriately between the real world of quantities and the *statistical world of data*.	Roberto and Juan, in addition to identifying three variables, relatively explicitly named in the infographic (Figure 1), added a new variable, expressed as an absolute frequency, that allowed them to quantify the number of waves per town. The numerical expression of this variable allowed the students to summarize information, facilitating movement from real-world events to statistical data.
DES	Being able to understand *numbers in context* as references for measuring variability in the real world.	Roberto and Juan described the variability of the data based on the relationship between two continuous variables identified ("time of arrival" and "size of the wave") and the categorical variable. The categorical variable allowed them to situate the correspondence between the continuous quantitative variables because each town contextualized the events. The continuous variable "time of arrival" is expressed through lists of numbers, which provided an ordinal description of variability. Similarly, the variable "how many times it repeats" provided a cardinal description of variability with respect to the number of waves affecting each town.
E&R	Being able to use numbers and quantitative methods to communicate, process, and interpret information provided by *data* and *contextual knowledge.*	Roberto and Juan achieved a tabular representation of four variables in a very creative way, as two of the variables manifested data made up of groups of data. Additionally, the use of a table promoted the joint visualization of the lists made up by the quantitative variables and facilitated the use of quantitative methods with them. The quantification of the number of waves per town provided an absolute frequency that numerically expressed a little-known characteristic of tsunamis, that is, that they are not made up of one large wave, but rather occur in a succession of waves.

4. DISCUSSION

Based on the experience described here, we maintain that the statistical representations created by the students manifested an aggregate perspective of the data in the sense of Konold et al. (2015), because they used quantitative descriptors that allowed them to summarize and highlight general characteristics of the phenomenon, which in the raw data were not easily seen. Similarly, the use of representations shows how the students involved themselves in transnumeration (i.e., the creation of data representations and change from one data representation to another to facilitate understanding). This occurred when, for example, Anita and Luisa reorganized the data in the infographic to achieve what Watson (2016) called data reduction and hence created a three-variable bar graph. In this way, transnumeration and the aggregate perspective are unequivocal signs of the presence of data sense in the reasoning analyzed.

Given that the lesson design did not include instructions regarding what data to represent and how to represent it, the coherence of the representations created by the students is outstanding. We consider that both representations presented indicators of informal inference (Makar & Rubin, 2009), given that they made affirmations beyond the data (Lehrer & English, 2018), used the data as evidence (Pfannkuch et al., 2015), considered the aggregate (Konold et al., 2015), and achieved integration of the knowledge the students possessed about the context (Langrall et al., 2011). We agree with Lehrer and English (2018) that the creation of representations help children to make informal inferences shaped by variability. Extracting data from the infographic to diagram, explain, and communicate the effects of the tsunami on the Chilean coast, in an integrated and creative manner, demonstrated what we interpreted as a form of inferential reasoning based on the use of representations, supported by numerical arguments rather than verbal arguments.

However, one of the limitations of the study lies precisely on not having considered verbal arguments as a data source for analysis, as these could have allowed for a more complete interpretation of how the students manifested data sense. As such, a possible suggestion to achieve more satisfactory results aims at carrying out the implementation of the lesson, broadening the data collection system through clinical interviews, and strengthening the analysis of cases, so as to categorize the performance of all the students in the class exhaustively and, as such, understand the capacities associated with the development of data sense beyond the elaboration of external representations.

In making an external representation based on data, which involved specific components such as variables, frequencies, and graphic elements, the students utilized knowledge that associated logical, numerical, and geometrical elements (Estrella et al., 2017). The coordination of this knowledge, especially at an early age, required inferential reasoning that allowed the students to make affirmations and give explanations in the face of uncertainty. Currently, statistics education places informal inference at the center of the school curriculum, which leads to rethinking the ways in which inferential reasoning is understood and identified and its repercussions for statistics teaching (Garfield et al., 2015).

5. CONCLUSIONS AND PROJECTIONS

This study investigated how primary school students facing an authentically realistic problem reason about variability and how they manifested data sense. A lesson created within the lesson study framework allowed fifth-grade students to analyze the behavior of data from a tsunami and work simultaneously with qualitative and quantitative variables, creating new variables, making representations (a multiple bar graph and a frequency table), and making inferences based on data. We concluded that the use of an authentic context and the construction of their own representations promoted data sense in students and facilitated the development of their statistical thinking, through which they were able to recognize, describe, and explain the variability of the phenomenon.

In the two cases analyzed, differences were found in the ways of representing variability. Case 2 presented a tabular reorganization of the data of the variables (expressed in columns) that facilitated the comparison of values of the same variable, but not the values of different variables. In comparison, Case 1 presented a bar graph that displayed the height of the waves (with the order of arrival and time of arrival) that favored visual comparison among bars from the same town as well as among bars from different towns. The ways of thinking identified about variability can be described using data sense and communicated using representations. Data sense is observed in the creativity shown in both representations analyzed, given that they constructed new variables, one ordinal variable (the order in which the waves arrive to the coast) and one discrete variable (frequency). Working with several variables and using axes and scales, rows and columns, allowed the students to operate with the primordial elements of representations. These representations reflected a particular form of informal thinking because they integrated meanings that helped in decision making. In this way, the students debunked the myths that a tsunami is only one wave or that there is a temporal regularity to the arrival of waves to the coast, and then disseminated this information to the public to prevent future disasters.

We consider the use of research questions (Arnold, 2008)—questions posed to examine a data set— framed in realistic contexts, related to the students' everyday life, but also culturally and scientifically important, fundamental to the development of data sense. Given the importance of creating opportunities for developing, from the early years of school, critical attitudes and knowledge in future citizens, we propose addressing the abilities included in data sense, that is, those that allow to solve

problems framed in contexts based on real data, understanding how this data are generated, making judgments and statistical inferences based on the variability present in the data, and exhibiting a critical position, as elements that promote statistical literacy in general and the development of informal inference in particular.

ACKNOWLEDGEMENTS

REFERENCES

Arnold, P. (2008). Developing new statistical content knowledge with secondary school mathematics teachers. *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study, 18,* 1–6.

Chan, S., & Ismail, Z. (2013). Assessing misconceptions in reasoning about variability among high school students. *Procedia-Social and Behavioral Sciences, 93*, 1478–1483.

Dierdorp, A., Bakker, A., Ben-Zvi, D., & Makar, K. (2017). Secondary students' considerations of variability in measurement activities based on authentic practices. *Statistics Education Research Journal*, *16*(2), 397–418.

Ekol, G., & Sinclair, N. (2016). Undergraduate students' conceptions of variability in a dynamic computer-based environment. In D. Ben-Zvi & K. Makar (Eds.), *The teaching and learning of statistics* (pp. 193–203). Springer.

English, L., & Watson, J. (2018). Modelling with authentic data in sixth grade. *ZDM, 50*(1-2), 103–115.

Estrella, S. (2017). Enseñar estadística para alfabetizar estadísticamente y desarrollar el razonamiento estadístico [Teaching statistics to statistically literate and develop statistical reasoning]. In A. Salcedo (Ed.), *Alternativas Pedagógicas para la Educación Matemática del Siglo XXI* (pp. 173–194). Centro de Investigaciones Educativas, Universidad Central de Venezuela.

Estrella, S. (2018). Data representations in early statistics: Data sense, meta-representational competence and transnumeration. In A. Leavy, M. Meletiou, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education*: *Supporting early statistical and probabilistic thinking* (pp. 239–256). Springer.

Estrella, S., Mena, A., & Olfos, R. (2018). Lesson study in Chile: A very promising but still uncertain path. In M. Quaresma, C. Winsløw, S. Clivaz, J. Pedro da Ponte, A. Ní Shúilleabháin, A. Takahashi (Eds.), *Mathematics lesson study around the world*: *Theoretical and methodological issues* (pp. 105–122). Springer.

Estrella, S., & Olfos, R. (2012). La taxonomía de comprensión gráfica de Curcio a través del gráfico de Minard: Una clase en séptimo grado [Curcio's taxonomy of graphical comprehension through Minard´s graph: A class in the seventh grade]. *Revista Educación Matemática, 24*(2), 119–129.

Estrella, S., Olfos, R., & Morales, S. (2014). What can we learn from natural disasters to prevent loss of life in the future? In J. W. Lott & C. J. Lott (Eds.), *Lessons learned from across the world: PreK–8* (pp. 66–71). National Council of Mathematics Teachers.

Estrella, S., Olfos, R., Morales, S., & Vidal-Szabó, P. (2017). Argumentaciones de estudiantes de primaria sobre representaciones externas de datos: componentes lógicas, numéricas y geométricas. [Arguments of primary school students on external representations of data: logical, numerical and geometric components] *RELIME, Revista Latinoamericana de Investigación en Matemática Educativa, 20*(3), 345–370.

Estrella, S., Zakaryan, D., Olfos, R., & Espinoza, G. (2020). How teachers learn to maintain the cognitive demand of tasks through lesson study. *Journal of Mathematics Teacher Education*, *23*(3), 293–310.

Friel, S., Bright, G., Frierson, D., & Kader, G. (1997). A framework for assessing knowledge and learning in statistics (K–8). In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 55–63). IOS Press.

Gal, I. (2004). Adults' statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Kluwer.

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

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, *1*(1). http://ww2.amstat.org/publications/jse/v1n1/garfield.html

Garfield, J., Le, L., Zieffler, A., & Ben-Zvi, D. (2015). Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking. *Educational Studies in Mathematics*, *88*(3), 327–342.

Inzunsa-Cazares, S. (2016). Razonamiento de estudiantes universitarios sobre variabilidad e intervalos de confianza en un contexto inferencial informal. [University students´ reasoning on variability and confidence intervals in an informal inferential context] In M. B. Wood, E. E. Turner, M. Civil, & J. A. Eli (Eds.), *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 969-983)*.* The University of Arizona.

Isoda, M., & Olfos, R. (2009). *El enfoque de resolución de problemas en la enseñanza de la matemática a partir del estudio de clases.* [The problem-solving approach in the teaching of mathematics from the classroom] Ediciones Universitarias de Valparaíso, Pontificia Universidad Católica de Valparaíso.

Isoda, M., Chitmun, S., & Gonzalez, O. (2018). Japanese and Thai senior high school mathematics teachers' knowledge of variability. *Statistics Education Research Journal, 17*(2), 196–215. https://doi.org/10.52041/serj.v17i2.166

Konold, C., Higgins, T., Russell, S., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

Langrall, C., Nisbet, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning, 13*(1–2), 47–67.

Lehrer, R., & English, L. (2018). Introducing children to modeling variability. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 229-260). Springer.

Lehrer, R., Kim, M., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modeling variability. *International Journal of Computers for Mathematical Learning*, *12*(3), 195–216.

Lovett, J. N., & Lee, H. S. (2018). Preservice secondary mathematics teachers' statistical knowledge: A snapshot of strengths and weaknesses. *Journal of Statistics Education*, *26*(3), 214–222.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.

McIntosh, A., Reys, B., Reys, R., Bana, J., & Farrell, B. (1997). *Number sense in school mathematics: Student performance in four countries.* Mathematics, Science & Technology Education Centre, Edith Cowan University.

Ministerio de Educación de Chile. (2018). Matemática. In *Bases Curriculares Primero a Sexto Básico* (pp. 214–261). MINEDUC.

Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 323–333). Springer.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17–46). Springer.

Pfannkuch, M., Arnold, P., & Wild, C. (2015). What I see is not quite the way it really is: Students' emergent reasoning about sampling variability. *Educational Studies in Mathematics, 88*(3), 343–360.

Phan, H. (2012). A sociocultural perspective of learning: Developing a new theoretical tenet. Paper presented at the *Joint Australian Association for Research in Education and Asia-Pacific*

Educational Research Association Conference (AARE-APERA 2012) World Education Research Association (WERA) Focal Meeting, Sydney, New South Wales, December 2–6.

Shaughnessy, M. (2007). Research on statistics' reasoning and learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 957–1009). Information Age Publishers.

Shaughnessy, M., & Pfannkuch, M. (2002). How faithful is old faithful? Statistical thinking: A story of variation and prediction. *Mathematics Teacher*, *95*(4), 252–259.

Smith, C., Fitzallen, N., Watson, J., & Wright, S. (2019). The practice of statistics for STEM: Primary students and pre-service primary teachers exploring variation in seed dispersal. *Teaching Science, 65*(1), 38–47.

Snee, R. (1993). What's missing in statistical education? *The American Statistician, 47*(2), 149–154.

Torok, R., & Watson, J. (2000). Development of the concept of statistical variation: An exploratory study. *Mathematics Education Research Journal*, *12*(2), 147–169.

Watson, J. (2016). Linking science and statistics: Curriculum expectations in three countries. *International Journal of Science and Mathematics Education, 15*(6), 1057–1073.

SOLEDAD ESTRELLA
Blanco Viel 596, Cerro Barón
Valparaíso, Chile

APPENDIX

LEARNING ACTIVITIES	TEACHER INTERVENTION	EVALUATION OF THE LESSON PROGRESS	
1. Introduction to the topic. A) Do you remember what a tsunami is?	Pose foreseen questions to the class.	Are students motivated by the topic?	
Possible answers: a) a big wave, the tidal wave in Japan, more than one wave, an earthquake in the ocean, …			
b) What definition is appropriate to what happened?	Show a video about the tsunami. Generate a whole-class discussion, define tsunami and tidal wave.	Students understand that a tsunami is more than a wave, and use the term "tidal wave".	8 minutes
c) What towns were affected by the February 2010 tsunami in Chile?	**Possible answers:** c) "I was in….", "My grandfather felt it in…", in the south, Constitución, Dichato, …		
d) Did you read or see anything in the media about the tsunami? In a newspaper, how is data shown?	**Possible answers:** d) I heard it on the radio, I saw videos on TV, on the internet. In photos, graphs, maps, pictograms…		
2. Putting prior knowledge into play. e) What have you learned to represent data and communicate information?	**Possible answers:** e) bar graphs, line graphs, point graphs, pictograms, and tables…		
	Write on a section of the board the types of graphs mentioned.	Did students mention the expected graphical representations?	
3. Posing the problem. The teacher projects an image of the tsunami that affected Chile on February 27, 2010. An image of the tsunami is given to each student.	"I found this image about the 2010 natural disaster in a newspaper. It has a lot of information, so I want you all to help me to organize the data in a simpler way, and we'll send the material you create to the towns that were affected so it can be distributed and save lives."	Are students interested in the problem? Do students understand the task?	5 minutes
Challenge: Extract and arrange data to communicate it to the towns that were and could be affected by tsunamis.			
4. Solving the problem. Choose data and arrange it to communicate conclusions to the affected towns.	f) What do you observe? What data can you extract?	Were students able to identify the relevant data (height, place, time)?	
Possible answers: f) a big wave, compare the height of a person with the height of the wave, different heights of waves, the time between waves is not constant, there are several waves in the same town.			
[5 minutes of individual work, then work in pairs or trios for 10 minutes] Students observe and reflect individually and then share in groups. Students extract some data and create representations to communicate information using what they have learned.	The teacher observes students' outcomes at their desks, identifying those exhibiting different strategies they display and a variety of graphs and tables.	Did students work together? Do students' outcomes meet the given challenge?	20 minutes
Anticipating errors/difficulties in the representations: Identifying the variables at play (height, place, time…). The range of the variable. Graphs and/or tables without titles or without text in table headings. Absence of tick marks and labels on the axes. Choosing a line graph when the variable is discrete. Sums or averages without arguing the reason for their calculation. Interpreting the intersection of the waves in the image as if they were simultaneous waves in the same place.			
5. Sharing ideas. Students stick their work on the board. Students justify their outcomes and communicate the message to be sent to the towns.	Choose students to share their strategies with the class.	Were students able to communicate through their graphs and/or tables the message to be sent to the towns? Can students "read between the data"?	15 minutes
Anticipating student responses: They create tables and/or graphs of one or various towns according to the number of waves, height of the waves, or time they occurred. Line, point, or bar graphs. Simple tables with arranged data. Map of towns and records of the number of waves and/or their height. Calculation of some measures of central tendency, such as average, mode, or median height of the waves. **Anticipating student explanations:** "In one town the waves were of the same height", "In the same town there was more than one wave", "The waves were between 5 and 30 meters tall", "There was no regularity in the time between waves", "In one town there was only one wave", "There were four waves in the same town", "11 waves impacted during three hours and 26 minutes", "Between the first and second waves only five minutes passed", "The first wave arrived 16 minutes after the earthquake"….			
6. Synthesizing ideas. A) About the value of statistical representations. B) About natural disasters in Chile. What did you learn today?	The teacher poses discussion questions regarding: What is the usefulness of the tables and graphs created? Is the number of waves per town predictable? Is the time between one wave and the next predictable? Can lives be saved with this information?	Are useful to the towns the representations created by the students for making current and future decisions? Are students able to "read beyond the data"? Do students write in their notebooks what they learned today?	10 minutes

Figure 1. Lesson Plan