WHAT MEANING DO FUTURE ELEMENTARY SCHOOLTEACHERS ATTACH TO THE NUMBERS THAT APPEAR IN A EUROSTAT DATABASE?

Francisca M. Ubilla and Núria Gorgorió
Autonomous University of Barcelona
francisca.manriquez@e-campus.uab.cat

If we consider it important to prepare young people to become critical citizens, then we must begin by training teachers to cope with the complexity of databases in order to integrate them as resources for their future teaching. In the context of civic statistics, we carried out an activity structured around the basic cycle of learning from data with a group of first-year students on a primary teacher-training program. The activity was based on data extracted from Eurostat. In this paper, we analyze the way the students identified variables, values and their typology in order to show how they interpreted the "numbers" that appear in the tables, graphs and infographics. We conclude that understanding the numerical representations of civic statistics is a complex task that requires the learning of a decoding process.

INTRODUCTION

The current global context reinforces the need to understand scientific data and how it interacts with different social environments. The understanding of concepts and statistical representations that appear in scientific reports enables both governments and citizens to make decisions for everyday life. However, the current scenario has also shown that, depending on who analyses the data, the interpretation of a same social phenomenon can vary enormously. Therefore, bringing civic statistics to the classroom not only makes sense but turns out to be essential if we are to teach citizens to be critical thinkers when dealing with statistical data on social issues. Accordingly, mathematics teachers must be able to interpret statistical analyses and representations of social issues before presenting them in the classroom (Engel, 2019). Furthermore, "to teach civic statistics, teachers need not only statistical content knowledge but also pedagogical content knowledge, technological knowledge and a positive stance towards civic statistics (Frischemeier, Podworny & Biehler, 2018)".

In particular, Arteaga, Batanero, Cañadas and Contreras (2011) establish the need to learn about statistical tables as cultural objects since they are used to synthesize information both in the school environment and in the media. However, the tables in the databases are much more complex than those normally encountered by students, and their correct interpretation is not guaranteed. For this reason, civic statistics teaching must be incorporated into the school context. Furthermore, the training of future primary school teachers provides a key scenario for discussion on how to read these tables. This last point is particularly important, given that difficulties have already been identified among future primary school teachers as regards reading double-entry tables. They confuse absolute frequencies and percentages and do not distinguish between joint, conditional and marginal frequencies (Gea, Gossa, Batanero & Pallauta, 2020). Molina-Portillo et al. (2020) conclude that when prospective teachers are confronted with graphics extracted from mass media, distracting elements make it difficult for students to interpret and critically evaluate the information.

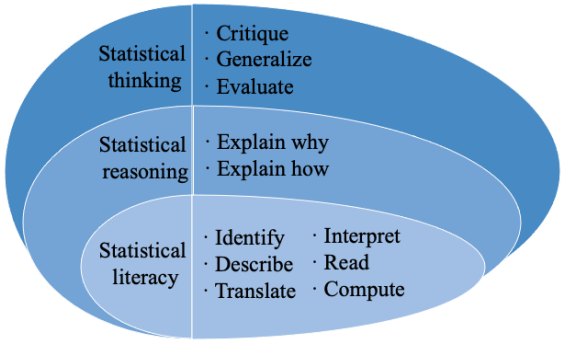
In the context of civic statistics, we asked ourselves what meaning student teachers might attach to the "numbers" that appeared in the tables, graphs and infographics of a study carried out with data from Eurostat. We answered this question by finding out what variables they identified in tables and graphs, what values they attributed to these variables, and how they classified them.

STATISTICAL SENSE

Garfield and Franklin (2011) state that the goal of statistics education is to produce statistically literate people who are also capable of reasoning and thinking statistically. Taking this same perspective, we consider that the statistical education of future teachers should be designed to develop their statistical sense, understood as an amalgam of statistical literacy (SL), reasoning (SR) and thinking (SR) (Ubilla, 2019). When setting the goals of statistical education, it is important to distinguish these three concepts, explaining their similarities and differences (Figure 1). Garfield, delMas and Zieffler (2010) establish a hierarchy, with literacy, reasoning and statistical thinking in ascending order. In conjunction, they propose a series of actions linked to the tasks that can be used to assess statistical literacy, reasoning

In: R Helenius, E Falck (Eds.), *Statistics Education in the Era of Data Science*
Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),
Aug-Sept 2021, Online conference. iase-web.org/Conference_Proceedings.php

and thinking. Drawing on these authors' ideas, Ubilla and Gorgorió (2021) propose Figure 1 to represent this hierarchy and the interaction of these three concepts understood as the components of statistical sense.



ST: knowing when and why to use certain statistical procedures and being capable of recognizing the limitations of statistics (Garfield and Ben-Zvi, 2008)
SR: knowing how to explain when or why statistical concepts and representations are connected (delMas, 2004)
SL: knowing what the basic statistical terms mean, using simple statistical symbols, and recognizing and being able to interpret different ways of representing data (Garfield and Ben-Zvi, 2008)

Figure 1. Actions associated with the assessment of statistical literacy, reasoning and thinking (Garfield, delMas & Zieffler, 2010). Figure drawn up by Ubilla and Gorgorió (2021)

CIVIC STATISTICS

Civic statistics is a sub-discipline derived from the intersection between statistics, education, politics and the social sciences (ProCivicStat Partners, 2018). Engel, Ridgway and Weber (2021) explain that civic statistics focuses on understanding the information about social issues supplied by the media, statistical institutes and other organizations, handling it through critical assessment and reflection on the data. According to these authors, the purpose of civic statistics is to enable people to see beyond the data so that they are capable of identifying the political and social implications of statistical information – for which purpose a contextual understanding of society is essential. The ProCivicStat project report (ProCivicStat Partners, 2018) lists six characteristics of the materials or data typical of civic statistics:

- The data on social variables does not exist in isolation. Therefore, its description and understanding implies a connection with other variables, resulting in *multivariate phenomena*.
- Statistics on social phenomena are presented as a combination of data components, such as percentages and coefficients, so there is a presence of *aggregate data*.
- Civic statistics is based on multiple data sources, so the results and messages may vary depending on who analyses the data, thereby making it *dynamic data*.
- The providers of official statistics issue press releases, summaries and visualizations that support a particular point of view. There is therefore a need for a critical interpretation of these *enriched texts*.
- Given that data and conclusions about phenomena are multivariate, dynamic and aggregate, their description over time or across multiple variables requires the use of *various visualizations*.
- Civic statistics involves matters of importance to society in general. For its proper interpretation, familiarity with the *social context* of the data is needed and attention must be paid to the global context.

In view of the complexity of the data and materials present in civic statistics, students need to take an inquiring approach to statistical information that describes social phenomena (Engel, Ridgway & Weber, 2021). This means that students must be able to question the nature of the data and the credibility of statistical messages, and identify the limitations of the statistical processes and representations that appear in different media. In practice, civic statistics should allow the generation of educational instances for the study of social phenomena in the students' surroundings and encourage the development of their statistical sense. On this basis, we are interested in knowing the meaning attributed by future primary school teachers to tables and graphs from Eurostat.

METHODOLOGY

We designed a classroom activity (see Figure 2) based on the cycle of learning from data (IDSSP Curriculum Team, 2019) and on the actions that allow the assessment of statistical literacy, reasoning

and thinking (Garfield, delMas & Zieffler, 2010). The implementation and development of this activity is described in detail in Ubilla and Gorgorió (2021).

ACTIVITY: FROM A DATABASE TO A BRIEF NEWS REPORT

1. Approach: Discuss the data in the table and the graphics you have been provided with. What do they refer to? What do they suggest to you?
2. Identification of variables: What variables can you identify in the table and graphs? What are the values of the variables? What kind of variables are they? Why?
3. Topic: Establish four topics that you can work on with the data provided. Choose one to work on during practice and justify your choice.
4. Questions: From the chosen topic, make a list of questions that can be answered with the data. Choose one (or more than one) that goes beyond a direct reading of the data. Justify your choice.
5. Plan: How will you approach the process to answer the question asked? Make a list of the steps you are going to follow, detailing the necessary actions in each step.
6. Data: Indicate the variables you will use to answer your question and how they will help answer it. If you need a new variable, introduce it, define it and indicate what values it has.
7. Analysis: Explain your calculations and why you did them. Draw the necessary graphics and explain why you chose this type of representation.
8. Results: Write out the partial results of the various steps you took until you found the answer to the question you asked.
9. Conclusions: Interpret the results obtained and write out the conclusions of your research, thus answering the question asked.
10. Reflection: Reflect on the development of the process. Identify strengths and weaknesses and suggestions for improvement.
11. News: In the seminar you have to present a news report linked to your practice work. The news report must include a headline, subtitle, basic ideas box, photographs and accompanying graphics.

Figure 2. Activity "*From a database to a brief news report*" (Ubilla & Gorgorió, 2021)

Together with the activity, we prepared four data packages that included graphs, tables and infographics extracted from Eurostat and from a study completed by the Spanish National Institute of Statistics titled "The life of women and men in Europe – a statistical portrait". The four data packages revolved around four social themes: education and work; work and family; habits and health; and life expectancy, health and retirement.

The activity was implemented during the first year of the degree in primary education at the Autonomous University of Barcelona, as part of the "Mathematics for Teachers" subject. A total of 134 students took part, organized into 38 working groups with three to four members. The activity was carried out in six face-to-face sessions lasting 90 minutes each. During the first session, the statistical concepts and representations studied during compulsory primary and secondary education were reviewed and the meaning of some social indicators present in the data packages was discussed. In the following four sessions, each group developed the activity shown in Figure 2, considering one of the four defined themes. In the final session, each group presented the news report it had produced to the rest of the class, and the teacher of the subject opened up a discussion around the concepts and procedures used by each group.

The data in this study consists of the reports written by the 38 groups to describe the development of the activities, distributed by theme as follows: education and work (nine groups), work and family (nine groups), habits and health (ten groups), and life expectancy, health and retirement (ten groups).

ANALYSIS

We made an inductive analysis of the written responses to task 2 of the activity (see Figure 2), focusing on the variables that the future teachers identified, the values they attributed to them and how they classified them. Based on these three axes, we looked for similar responses among the groups, thus generating a first system of categories based on their interpretation of the numbers appearing in the tables and graphs. In this paper we present this preliminary analysis. We hope to be able to present the complete category system shortly.

To make the analysis clear, we begin by presenting some examples of the tables and graphs that the students worked with. Then we give the meaning of the "numbers" that appear in each numerical representation and examples of the meanings attributed to them by some groups. The table in Figure 3 shows the population by gender of each EU country, the GDP per capita in PPS – whose meaning is described in the footnote – and the regions coded by numbers.

	Population		GDP per capita in PPS*	Region**
GEO/SEX	Men	Women		
European Union (EU)	248,232,994	260,307,109		
Belgium (BE)	5,524,068	5,713,206	118	1
…	…	…	…	…
Spain (ES)	22,826,546	23,623,019	91	3

*GDP is a measure for economic activity. It is defined as the value of all the goods and services produced less the value of any goods and services used in their production. The volume of GDP per capita in Purchasing Power Standards (PPS) is expressed in relation to the European Union (UE28), with the average set at 100. If the index of a country is higher than 100, this country's level of GDP per capita is higher than the EU average, and vice versa. PPS (EU28) = 100
**1 = Western Europe; 2 = Northern Europe; 3 = Southern Europe; 4 = Eastern Europe.

Figure 3. Part of the table distributed to all the work groups

We observe that 30 of the 38 groups considered the GDP per capita variable in PPS. Twenty-three of them thought that GDP per capita is described by the "numbers" that appear in the said column, i.e. they did not take into account the explanation of the meaning of the PPS currency. On the other hand, 16 groups identified the region variable. Twelve of them identified the "numbers" 1, 2, 3 and 4 as the values given to this variable and, despite indicating that the values of the variable are numerical, eight of these 12 groups indicated that it is a qualitative variable, which makes us think that they understood the meaning of this coding.

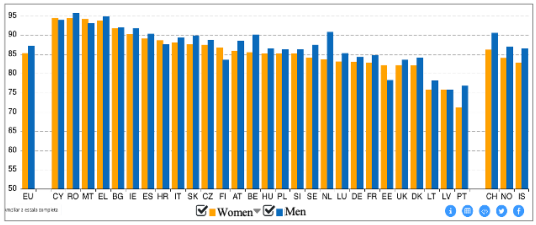
Figure 4 shows part of the table that provided information related to the self-perception of health variable, and the corresponding bar graph. In total, seven groups interpreted this information, with four reading from the graph and three from the table. We observe that the three groups that took a reading from the table and two of those that did so from the graph identified the self-perception of health variable as a quantitative variable because they affirmed that the percentages appearing in the table and graph were the values taken by this variable. The other two groups that identified this variable in the graph indicated that the "numbers" referred to a qualitative variable whose values are "good or very good".

Thematic table "Satisfaction with life"

	Percentage of people whose self-perception of their health is good or very good					
	16 to 44 years old		45 to 64 years old		65 years old and over	
GEO/SEX	Men	Women	Men	Women	Men	Women
European Union	87.1	85.3	63.8	60.7	41.1	34.5
Belgium	90.1	85.5	73.7	67.6	55.9	49.9
…	…	…	…	…	…	…
Spain	90.3	89.1	71.4	68.6	45.4	35.0

Percentage of people whose self-perception of their health is good or very good according to age (Source: Eurostat)

Thematic graph "Habits and health"



Percentage of people between 16 and 44 years old whose self-perception of their health is good or very good (Source: INE)

Figure 4. Part of the table and graph showing self-perception of health

Figure 5 presents the part of the table that shows the workforce in thousands and the percentage of unemployed in the workforce. Six of the nine groups that worked with this table considered the concepts of the workforce in thousands and the unemployed workforce. Two groups identified the percentages shown in the table as values of the unemployed workforce variable, while one group identified the "numbers" "15 to 24, and 25 to 74" as values of the said variable, when, in fact, these refer to the age groups. In other words, these three groups interpreted the unemployed workforce variable as a continuous quantitative variable. The other three groups identified the workforce variable in thousands, indicating that the values taken by this variable are numerical.

Figure 5 also presents the graph given to the groups who considered the theme of "work and family". Among the nine groups that worked with this graph six of them interpreted the percentages

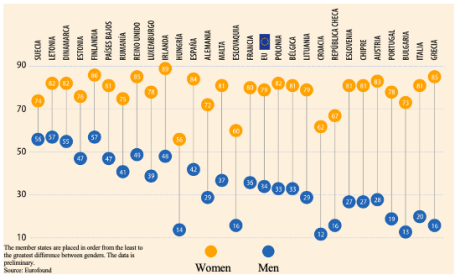
displayed in the graph as the values taken by a variable that they denominated "adults who do housework", and they identified this variable as continuous.

Thematic table "Education and Work"

	Workforce* in thousands				Unemployed workforce (%)			
	15 to 24 years old		25 to 64 years old		15 to 24 years old		25 to 74 years old	
GEO/SEX	Men	Women	Men	Women	Men	Women	Men	Women
European Union	12,404	10,437	115,692	99,350	21.1	19.5	8.1	8.5
Belgium	219	176	2,421	2,105	23.8	20.0	7.8	6.8
…	…	…	…	…	…	…	…	…
Spain	831	723	11,401	9,812	48.6	48.0	18.8	21.8

*The workforce is defined as the total of the employed plus the unemployed in search of work (registered as jobless)

Thematic graph "Work and Family"



Percentage of adults aged 18 or over who cook and/or do housework every day (Source: INE)

Figure 5. Part of the table showing "Education and Work" and the thematic graph of "Work and Family

INITIAL RESULTS

One of the initial results of our analysis reveals that the future primary school teachers in the sample were not always capable of interpreting the table titles and/or the labels on the graphs and infographics. Owing to this lack of understanding, they interpreted all the "numbers" that appear in these representations as the count of something that they consider to be a variable. In other words, in general the future teachers treated the headings as variables and the "numbers" in the columns under each heading as the values they take. Consequently, they interpreted the percentages as absolute frequencies, in this way classifying a considerable part of their variables as quantitative.

DISCUSSION AND CONCLUSIONS

Citizen empowerment begins with an understanding of the information and messages that we read and hear in different circumstances. Civic statistics provides a context for establishing connections between the teaching of statistics and relevant, urgent social issues. However, if we are to meet the challenge of introducing this sub-discipline into the classroom, teachers must, at the very least, be able to understand the information appearing in different types of representations (Engel, 2019). Therefore, it is vital that future primary school teachers are introduced to the study of statistics in social contexts during their training.

Specifically, the activity carried out by the participating students permitted a data analysis from the standpoint of statistical literacy (Garfield and Ben-Zvi, 2008) and this showed us how they read and interpreted the "numbers" in tables and graphs extracted from a database (Eurostat) and a statistical study. The tables and graphs used in the activity display characteristics typical of civic statistics materials (ProCivicStat Partners, 2018). Our analysis, although preliminary, confirms that understanding the numerical representations of civic statistics is a complex challenge and requires the learning of a decoding process, which is in line with the results of Molina-Portillo et al. (2020).

In particular, the way some of the participants misunderstood the headings in the tables brings to light the difficulties derived from one of the typical characteristics of civic statistics, the need to work with aggregate data. Owing to the inclusion of aggregate data in the tables, the participants understand the column headings in the tables as a variable whose values were the "numbers" in the respective columns. Although most of these "numbers" correspond to percentages, the students interpreted them as absolute frequencies, as similarly observed by Gea et al. (2020). This phenomenon prompts us to question the utility of the type of tables and graphs appearing in most current textbooks, which tend to offer a direct reading and mainly contain absolute frequencies.

Finally, we would insist on the importance of working with qualitative variables. Our analysis suggests that future teachers always tend to want to use "numbers", even quantifying qualitative variables. This may be due to the fact that the statistical tools and procedures they learnt at school were essentially focused on the calculation of statistics. In this respect, in the context of civic statistics, we see a need to introduce qualitative thinking into statistics teaching and give it greater relevance

(Ograjenšek & Gal, 2016), with the purpose of developing a type of statistical sense that includes both quantitative and qualitative components.

ACKNOWLEDGEMENTS

The research was supported by funding from the National Agency for Research and Development (ANID)/Grants Program /DOCTORAL GRANTS CHILE/2018 - 72190313, and by the project EDU2017-8247-R funded by Spanish Ministry of Science, Innovation and Universities.

REFERENCES

Arteaga, P., Batanero, C., Cañadas, G. y Contreras, M. (2011). Las tablas y gráficos estadísticos como objetos culturales. *Números. Revista de Didáctica de las Matemáticas*, *76*, 55-67.

delMas, R. (2004). A comparison of mathematical and statistical reasoning, in The challenge of developing statistical literacy, reasoning and thinking, D. Ben-Zvi, and J. Garfield, Eds., Springer, Dordrecht, 79-95.

Engel, J. (2019). Cultura estadística y sociedad: ¿Qué es la estadística cívica? En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en www.ugr.es/local/fqm126/civeest.html

Engel, J., Ridgway, J., & Weber Stein, F. (2021). Educación Estadística, Democracia y Empoderamiento de los Ciudadanos. *PARADIGMA*, *41*(e1), 01-31. https://doi.org/10.37618/PARADIGMA.1011-2251.2021.p01-31.id1016

Frischemeier, D., Podworny, S. and Biehler, R. (2018). Activities for promoting civic statistical knowledge of preservice teachers. Péter Kovács (ed.): Challenges and Innovations in Statistics Education. Multiplier Conference of PROCIVICSTAT: Szeged (Hungary), September 7-9, 2017. Conference Proceedings.

Garfield, J., & Ben-Zvi, D. (2008). Developing students' statistical reasoning: Connecting research and teaching practice. Springer Science & Business Media.

Garfield, J., del Mas, R. y Zieffler, A. (2010). Assessing important learning outcomes in introductory tertiary statistics courses. En P. Bidgood, N. Hunt, y F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 75-86) Chichester: John Wiley & Sons Ltd.

Garfield, J., & Franklin, C. (2011). Assessment of learning, for learning, and as learning in statistics education. In C. Batanero, G. Burrill, and C. Reading (eds.), Teaching Statistics in School 133 Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study, DOI 10.1007/978-94-007-1131-0_16, © Springer Science+Business Media B.V. 2011

Gea, M. M., Gossa, A., Batanero, C. y Díaz-Pallauta, J. (2020). Construcción y lectura de la tabla de doble entrada por profesores de Educación Primaria en formación. *Educação Matemática Pesquisa: Revista do Programa de Estudos Pós-Graduados em Educação Matemática*, *22*(1), 348-370. https://doi.org/10.23925/1983-3156.2020v22i1p348-370

IDSSP Curriculum Team (2019). Curriculum frameworks for introductory data science. Available at http://idssp.org/files/ IDSSP_Frameworks_1.0.pdf

Molina-Portillo, E. M., García, J. C., Salcedo, A., & García, J. M. C. (2020). Evaluación de la postura crítica de futuros profesores de Educación Primaria como componente de la cultura estadística. *Educación MatEMática*, *32*(3).

ProCivicStat Partners (2018). *Engaging Civic Statistics: A Call for Action and Recommendations.* A product of the ProCivicStat Project. http://IASE- web.org/ISLP/PCS

Ograjenšek, I. y Gal, I. (2016). Enhancing statistics education by including qualitative research. International Statistical Review, *84*(2), 165-178.

Ubilla, F. (2019). Componentes del sentido estadístico en un ciclo de investigación estadística desarrollado por futuras maestras de primaria, in Investigación en Educación Matemática XXIII, J. M. Marbán, M. Arce, A. Maroto, J. M. Muñoz-Escolano and Á. Alsina, Eds., Valladolid: SEIEM, 2019, 583-592.

Ubilla, FM. & Gorgorió, N. (2021). From a source of real data to a brief news report: Introducing first-year preservice teachers to the basic cycle of learning from data. *Teaching Statistics*, *43*, S110-S123. https://doi.org/10.1111/test.12246