SECONDARY SCHOOL STUDENTS' INTERPRETATION OF THE FREQUENCY TABLE

Jocelyn D. Pallauta, María M. Gea, Pedro Arteaga, and Silvia M. Valenzuela-Ruiz
University of Granada, Spain
jocelyndiaz@correo.ugr.es

This research analyses Spanish secondary school students' reading capacity of elements in a frequency table. The study is based on the two reading levels of reading the data and reading between the data proposed by Curcio (1989) and applied to data displayed in a statistical table. The results of the evaluation study carried out with 149 students in secondary school first grade and 128 in third grade (12–14-year-olds) show that the majority of the sample performs a correct reading, especially in the first reading level. We also use the notion of semiotic conflict proposed by Godino et al. (2007, 2019) to describe mistakes in the students' responses, which were mainly comparing frequencies incorrectly and misunderstanding the questions posed.

INTRODUCTION

Citizens need to interpret statistical tables to function in the information society where statistical knowledge is essential for making better decisions. Statistical tables are one of the most common representation of data (Feinberg & Wainer, 2011). They are often included in the study of mathematics and other subjects to represent, summarize, and communicate information and for data analysis (Estrella & Estrella, 2020). In everyday contexts, tables appear, for example, on television, in the news, in advertising, and on the Internet. They are even more evident in the current crisis caused by COVID-19, in which the lack of statistical literacy of citizens caused many to misunderstand the situation (Rodríguez-Muñiz et al., 2020). Thus, the ability to read and to interpret statistical tables should be a component of statistical literacy for all citizens in the information society (Gal, 2002).

The Spanish curricular guidelines suggest working with statistical tables during primary education (6–11-year-olds). In the first and second grade of secondary education (Ministerio de Educación, Cultura, y Deporte [MECD], 2015), students organize data obtained from a population of qualitative or quantitative variables into tables, calculate absolute and relative frequencies, represent them graphically, and carry out the reverse process of translating graphs into tables. Cumulative frequencies are introduced in third grade.

The aim of this research was to analyse first and third grade students' reading capacity of one-variable distribution tables and to identify the semiotic conflicts that may arise in that interpretation for these two groups of Spanish secondary education students.

FUNDAMENTALS

Reading Levels of Tables and Graphs

Different questions at different levels of difficulty can be asked from a statistical table or graph. As a result, different authors have defined reading levels for graphs. In our work, we used the first two levels of reading statistical graphs established by Curcio (1989) and Friel et al. (2001), a hierarchy that is also valid for statistical tables.

- Reading the data (L1) corresponds to a literal reading of information that is explicitly presented in the table or graph; therefore, no calculations or other operations need to be performed on the represented data. An example is finding the frequency of a given value for a variable.
- Reading between the data (L2) involves not only literal reading but also comparisons or calculations of data represented in the table. This level is required, for example, to compare the frequencies of two variable values.
- Reading beyond the data (L3) involves a greater ability to infer unrepresented information that cannot be extracted arithmetically. This reading level was not considered in this study.
- Reading behind the data (L4) involves being able to judge the information represented in the table and the source from which was extracted. This reading level was also not considered in this study.

Semiotic Conflict

We based our work on the onto-semiotic approach to mathematical knowledge and instruction (Godino et al., 2007, 2019), which assumes that mathematical objects emerge from mathematical

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

practices. The meaning of a mathematical object can be considered from an institutional or personal perspective. We used the idea of semiotic conflict, which corresponds to a discrepancy attributed to the meaning of a mathematical object from an institutional and personal perspective. A conflict is conceptual if it refers to misunderstanding concepts or properties, procedural when steps of the procedure are confused, and notational when is related to mathematical language (symbols, specific terms, or another mathematical language).

Research Background

Research on the ability of secondary school students to read statistical tables has mainly focused on two-way tables. The hierarchical reading levels proposed by Curcio (1989) and Friel et al. (2001) were valid for the statistical tables studied by Gabucio et al. (2010), who gave a questionnaire containing questions about two-way tables to 112 Spanish students in the fifth and sixth grades of primary school (10–11-year-olds) and 88 students in the first and second grades of secondary school (12–13-year-olds). The questionnaire was designed to assess the following four levels of students' reading: (a) comprehension of the tabular structure, or understanding of the different elements of the table; (b) direct reading of explicit data from the table according to Friel et al.'s (2001) level L1; (c) data inference, or searching for data not directly represented in the table, which combines levels L2 and L3 (Friel et al., 2001); and (d) global interpretation, equivalent to reading behind the data at level L4 (Friel et al., 2001). Thus, questions requiring inference were more difficult than those requiring direct reading of data. The percentage of correct answers on direct reading items ranged from 80.5% to 89.5%, and correct answers on inference ranged from 29% to 47%. There was a very small improvement in the number of correct answers as grade levels increased.

Álvarez et al. (2020) conducted a study with 65 Colombian students aged 11 to 16 to analyse their difficulties in interpreting tables with different frequencies, some of them with grouped data. The main difficulties in the tasks were not knowing the meaning of frequency, confusing different types of frequencies, and confusing the value of the variable with the frequency.

METHOD

The sample consisted of 277 Spanish students: 149 in the first grade and 128 in the third grade of Compulsory Secondary Education from two schools in Andalusia (ages 12–14). The sample was a convenience sample because access to educational establishments was particularly complex due to the COVID-19 health crisis, but the teacher confirmed that students had worked with statistical tables in previous years. Students were given a questionnaire that included nine tasks, one of which assessed students' reading capacity of a one-variable distribution table (Figure 1). The questionnaire construction was based on a previous analysis of curricular guidelines and Spanish secondary textbooks (Pallauta et al., 2021); items were selected by expert judgment to ensure content validity in analysing Spanish secondary students' comprehension of statistical tables.

A group of children were asked about their favorite yoghurt flavor; they could only choose one. Their preferences are recorded in table below. Answer the following questions:

(a) How many children prefer coconut flavor yoghurt?
(b) Which flavor is preferred by exactly 6 children?
(c) Which yoghurt flavor do most of the children prefer?
(d) How many children in total were asked about their favorite yoghurt flavor?
(e) How many more children prefer vanilla yoghurt than pineapple yoghurt?

Flavor	Number of children
Strawberry	8
Vanilla	6
Apple	4
Coconut	4
Pineapple	2

Figure 1. Task proposed to students

The task was adapted from a textbook (Kheong et al., 2017) and selected by a panel of experts from three possible items. From the table of the frequency distribution, students had to: (a) find the frequency associated with a category, (b) identify the category that corresponds with a frequency, (c) find the mode, (d) determine the sample total, and (e) compare frequencies for two categories.

Data analysis was based on content analysis (Neuendorf, 2016). The primary unit of analysis was each response to each question, in which the authors systematically reviewed the categories established in the previous analysis. In discordant cases, a consensus was reached to ensure reliable coding. Evaluation of responses was discussed in two rounds, and responses were classified into categories of correct, partially correct, or incorrect. In a third round, we used prior research results to identify several semiotic conflicts that arose from interpreting the data presented in the table.

RESULTS

This section first describes the evaluation criteria for each question; then the results of correct, partially correct, and incorrect responses by grade; and finally, the semiotic conflicts identified along with their distribution.

Question (a): Finding the Frequency Associated with a Modality

In this question, L1 of reading the data is required (Curcio, 1989; Friel et al., 2001) because students only need to perform a direct reading of data in the table. Answering the question only requires identifying a category and reading its frequency.

- A correct response resulted when the student identified that four children prefer coconut flavored yoghurt.
- In a partially correct response, students reported relative frequency instead of absolute frequency. The students reached the L1 level in reading the table, but confused absolute and relative frequency, such as the response by E23: 4/24 children.
- In an incorrect response, the student answered with meaningless values, such as E180, who responded with “apple” (a category instead of a frequency).

Question (b): Identifying the Modality Corresponding to a Frequency Value

The second question also requires reading the data (L1). However, this task requires identifying the category of the variable associated with a given absolute frequency value. To answer the question, a student must identify the given value in the absolute frequency column and then respond with the corresponding variable category.

- In correct answers, students indicated that vanilla yoghurt is preferred by exactly six children.
- In partially correct responses, students gave the frequency (six children) instead of the category, confusing the frequency with the variable value, such as when E103 answered “6” children.
- For incorrect answers, students indicated a category that does not correspond with the frequency provided in the question, such as E153, who responded, “Strawberry,” and thus does not reach a table reading of level L1.

Question (c): Finding the Mode

The student must compare all of the absolute frequencies and select the category with the highest frequency. This corresponds to the level of reading between the data (L2).

- A correct response indicated that strawberry was the category with the highest absolute frequency.
- Partially correct answers reported the highest frequency and not the category associated with that frequency (mode), such as E153 who confused the variable modality with its frequency by responding with “8.”
- Incorrect responses appeared when other values of the frequency or variable were indicated, or the student indicated that the mode did not exist. An example is E151 who did not appear to read the labels in the table correctly with the response: “There is no yoghurt flavor.”

Question (d): Determining the Sample Total

This question corresponds to the level of reading between the data (L2) because it requires calculations to determine the total number of children in the sample. No partially correct responses were observed for this question.

- In correct responses, the students indicated that 24 children in total answered the question with a preferred yoghurt flavor.

- Incorrect answers involved wrong total values, showing procedural difficulties. For example, E21's response of "26 children" could have resulted from incorrectly adding the frequencies.

Question (e): Comparing Two Modalities' Frequencies

To answer the last question, the values of the absolute frequencies of two categories of the variable must be identified and their difference calculated. Therefore, a level of reading between the data (L2) is required. The responses were classified as follows:

- For a correct response, the student answered that four more students prefer vanilla yoghurt than pineapple yoghurt.
- Partially correct answers appeared when the categories along with their absolute frequency were identified, but no subtraction was performed. For example, E51 only showed a level L1 of reading the data, because the student identified the frequency for each category: "6 children prefer vanilla and 2 children pineapple." It is possible that the student did not translate the expression, how many more, into an operation.
- Incorrect response occurred when other values were given, such as the answer of student E129 who adds the values of the frequencies instead of subtracting them: "8 because 2+6 = 8."

Task Results

The results for the correctness of students' responses for different elements in the statistical table are presented in Table 1. A high percentage of students answered each question correctly. The best results were obtained for question (b) (identifying the category corresponding with a frequency value), with a similar percentage of correct answers in third grade (98.4%) and in first grade (96.6%), as well as for question (c) (finding the mode), with a similar percentage of achievement (correct answers) in the two grades (98.7% in first grade and 96.6% in third grade). In question (a) (finding the frequency associated with a modality), a high percentage of correct answers in first grade is observed (96%), which coincides with the study developed by Gabucio et al. (2010) in which the participants of this educational level showed only 18.6% of incorrect responses in a similar question.

Table 1. Frequency (and percentage) of responses to each question according to grade

	Question (a)		Question (b)		Question (c)		Question (d)		Question (e)	
	First	Third	First	Third	First	Third	First	Third	First	Third
Correct	143 (96)	119 (93)	144 (96.6)	126 (98.4)	147 (98.7)	124 (96.9)	130 (87.2)	118 (92.2)	115 (77.2)	106 (82.8)
Partially correct	2 (1.3)	6 (4.7)	1 (0.7)			1 (0.8)			5 (3.4)	3 (2.3)
Incorrect	3 (2.0)	1 (0.8)	3 (2.0)	1 (0.8)		2 (1.6)	13 (8.7)	8 (6.3)	26 (17.4)	17 (13.3)
Non-response	1 (0.7)	2 (1.6)	1 (0.7)	1 (0.8)	2 (1.3)	1 (0.8)	6 (4.0)	2 (1.6)	3 (2.0)	2 (1.6)

In question (d), determining the sample total, there was a lower percentage of correct answers (first grade: 87.2%; third grade: 92.2%), and the lowest percentage of correct answers was in question (e), comparing frequencies for two categories, which was especially lower for first grade (77.2%) than third grade (82.8%). These results are similar to those from a study carried out by Álvarez et al. (2020) with secondary school students, where some difficulties were found with the interpretation questions that required calculations, such as questions (d) and (e).

Semiotic Conflicts in the Interpretation of the Table

A second analysis of partially correct and incorrect answers for interpretating the table identified types of semiotic conflicts classified by notational, conceptual, and procedural conflicts.

Conceptual Semiotic Conflicts

- *C1. Confusing absolute and relative frequencies.* This conflict appears when the relative frequency is indicated rather than the absolute frequency, as was observed in the response of student E23 in question (a) (literal reading of data).
- *C2. Confusing the value of the variable and its frequency.* This conflict was also identified by Mayén et al. (2009) in their study on the median, and it happens when the value of the frequency is indicated, instead of the category; for example, E153 in question (c) (finding the mode).

Procedural Semiotic Conflicts

- *P1. Conflict in computing total.* This conflict is observed when the student incorrectly calculates the sample total in question (d), such as the answer of E21.
- *P2. Conflict in comparing frequencies.* This conflict is detected in question (e) (comparing frequencies for two categories); for example, the answer of the student E129 showed this conflict.

Notational Semiotic Conflicts

- *N1. Failure to reach the minimum reading level of the data.* This notational conflict is manifested in the incorrect reading for particular data in the table. For example, student E153 in question (b) and student E151 in question (c) showed this conflict.
- *N2. Conflict in the interpretation of the question.* This conflict occurs when the answer is related to the previous question. For example, student E87, in question (d) (determining the sample total) indicated the frequency of the mode instead of the total of the sample by responding with "8 children."

Table 2. Frequency (and percentage) of semiotic conflicts according to grade and question

Type	Question (a)		Question (b)		Question (c)		Question (d)		Question (e)	
	First	Third	First	Third	First	Third	First	Third	First	Third
C1	2 (1.3)	6 (4.7)				1 (0.8)				1 (0.8)
C2			1 (0.7)							
P1							7 (4.7)	7 (5.5)		
P2									15 (10.1)	15 (11.7)
N1	3 (2.0)	1 (0.8)	3 (2.0)	1 (0.8)		1 (0.8)	3 (2.0)		2(1.3)	
N2						1 (0.8)	3 (2.0)	1 (0.8)	14 (9.4)	4 (3.1)

The distribution of different semiotic conflicts identified in the students' responses is summarised in Table 2. Question (e) shows a higher percentage of conflicts, where the procedural conflict P2, comparing incorrectly frequencies, appears most frequently at both educational levels (first grade: 10.1%; third grade: 11.7%). Question (c) (finding the mode) has few conflicts. Regarding notational conflicts, N1 appears in all questions, and N2 appears from questions (c) through (e) in the third grade, reaching its highest percentage in the first grade (9.4%) in question (e).

CONCLUSIONS

Our research results showed that interpreting information displayed in a table was easy for the majority of the students, particularly when performing a literal reading of the data as was reported by Gabucio et al. (2010). However, reading that requires simple calculations or comparison of data represented in the table presented some difficulties. Thus, the reading between the data of level L2 in the model of Friel et al. (2001) raised a wide variety of semiotic conflicts, the most recurring being procedural conflicts when comparing frequencies and computing the sample total. Other conflicts involved an incorrect notational interpretation of questions or category labels in the statistical table.

These results suggest reinforcing students' basic reading capacity to ensure later achievement of higher levels (L3 and L4) in Friel et al.'s (2001) model because of the different types of conflict identified on the meaning of data represented in the table, which influences the ability to perform mathematical processes. This research is original because it provides new information on the interpretation of statistical tables by Spanish secondary school students, which had not previously been addressed.

ACKNOWLEDGEMENTS

Project PID2019-105601GB-I00 / AEI / 10.13039/501100011033 (MICIN), Group FQM-126 (Junta de Andalucía) and scholarship ANID Number: 72190280.

REFERENCES

Álvarez Alfonso, I., Guerrero Gutiérrez, Y. A., & López, Y. D. T. (2020). Taxonomía de errores y dificultades en la construcción e interpretación de tablas de frecuencia [Taxonomy of errors and difficulties in the construction and interpretation of frequency tables]. *Zetetike, 28*, Article e020012. https://doi.org/10.20396/zet.v28i0.8656553

Curcio, F. R. (1989). *Developing graph comprehension*: *Elementary and middle school activities***.** National Council of Teachers of Mathematics.

Estrella, S., & Estrella, P. (2020). Representaciones de datos en estadística: De listas a tablas [Data representations in statistics: From lists to tables]. *Revista Chilena de Educación Matemática, 12*(1), 21–34. https://doi.org/10.46219/rechiem.v12i1.20

Feinberg, R., & Wainer, H. (2011). Extracting sunbeams from cucumbers. *Journal of Computational and Graphical Statistics, 20*(4), 793–810. https://doi.org/10.1198/jcgs.2011.204a

Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education 32*(2), 124–158. https://doi.org/10.2307/749671

Gabucio, F., Martí, E., Enfedaque, J., Gilabert, S., & Konstantinidou, A. (2010). Niveles de comprensión de las tablas en alumnos de primaria y secundaria [Levels of comprehension of tables in primary and secondary students]. *Cultura y Educación, 22*(2), 183–197. https://doi.org/10.1174/113564010791304528

Gal, I. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review*, *70*(1), 1–25. https://doi.org/10.2307/1403713

Godino, J. D., Batanero, C., & Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM Mathematics Education, 39*(1), 127–135. https://doi.org/10.1007/s11858-006-0004-1

Godino, J. D., Batanero, C., & Font, V. (2019). The onto-semiotic approach: Implications for the prescriptive character of didactics. *For the Learning of Mathematics, 39*(1), 38–43.

Kheong, F. H., Soon, G. K., & Ramakrishnan, C. (2017). *Texto del estudiante Matemática 5º básico* [Mathematics 5th basic student text]. Marshall Cavendish Education. http://cipe.cl/V2/wp-content/uploads/2020/03/MATSA20E5B.pdf

Mayén, S., Díaz, C., & Batanero, C. (2009). Conflictos semióticos de estudiantes con el concepto de mediana [Semiotic conflicts of students with the concept of median]. *Statistics Education Research Journal, 8*(2), 74–93. https://doi.org/10.52041/serj.v8i2.396

Ministerio de Educación, Cultura, & Deporte (MECD). (2015). Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. *Boletín Oficial del Estado*, *3*, 169–546.

Neuendorf, K. (2016), *The content analysis guidebook*. Sage. https://doi.org/10.4135/9781071802878

Pallauta, J. D., Gea, M. M., Batanero, C., & Arteaga, P. (2021). Significado de la tabla estadística en libros de texto españoles de educación secundaria [Meaning of the statistical table in Spanish secondary education textbooks]. *Bolema*, *35*(71), 1803–1824. https://doi.org/10.1590/1980-4415v35n71a26

Rodríguez-Muñiz, L. J., Muñiz-Rodríguez, L., Vásquez Ortiz, C. A., & Alsina, Á. (2020). ¿Cómo promover la alfabetización estadística y de datos en contexto? ¿Estrategias y recursos a partir de la COVID-19 para Educación Secundaria [How to promote statistical and data literacy in context? Strategies and resources from COVID-19 for secondary education]. *Números*, *104*, 217–238.