Zetetiké

DOI: 10.20396/zet.v28i0.8656257

A semiotic analysis of the object statistical table in Chilean textbooks

Un análisis semiótico del objeto tabla estadística en libros de texto chilenos

Jocelyn D. Pallauta[1]

María M. Gea[2]

Carmen Batanero[3]

Abstract

In this paper we perform a semiotic analysis of the primary mathematical objects implicit in the work with different types of statistical tables, using elements of the ontosemiotic approach. This analysis is used to extend the levels of semiotic complexity defined by Arteaga for statistical graphs to statistical tables. The distribution of these types of tables in a sample of twelve Chilean school books, directed from 5th to 8th grade (10 to 13 years) is also presented through a content analysis. The tabular representation with greater presence in these textbooks is the distribution table, which corresponds to level 3 of semiotic complexity, subdivided into three levels, and within them the most frequent is the first, since the last sublevel only appears in 8th grade. The contingency table, associated with the maximum level of complexity, appears scarcely and paradoxically its use decreases with the school year.

Keywords: Statistical tables; semiotic analysis; level of semiotic complexity; textbooks.

Resumen

En este artículo se realiza un análisis semiótico de los objetos matemáticos primarios en el trabajo con diferentes tipos de tablas estadísticas, utilizando elementos del enfoque ontosemiótico. Dicho análisis, se utiliza para extender los niveles de complejidad semiótica definidos por Arteaga para los gráficos estadísticos a las tablas estadísticas. También se aborda, a través de un análisis de contenido, la distribución de estos tipos de tabla en una muestra de doce libros escolares chilenos, dirigidos de 5º a 8º Básico (10 a 13 años). La representación tabular con mayor presencia en estos textos es la tabla de distribución, que se enmarcaría en un nivel 3 de complejidad semiótica, subdividido en tres niveles, y dentro de ellos el más frecuente es el primero, apareciendo el último sólo en 8º curso. La tabla de contingencia, asociada al máximo nivel de complejidad, aparece con poca frecuencia y paradójicamente disminuye su uso con el curso escolar.

Submetido em: 18/08/2019 – **Aceito em**: 06/01/2020 – **Publicado em**: 16/01/2020

[1] Master in Mathematics Education, University of Granada, Spain. Email: jocelyndiaz@correo.ugr.es.

[2] PhD in Education, University of Granada. Teaching Assistant at the University of Granada, Spain. Email: mmgea@ugr.es.

[3] PhD in Mathematics, University of Granada. Professor of Mathematics Education, University of Granada, Spain. Email: batanero@ugr.es.

DOI: 10.20396/zet.v28i0.8656257

Palabras clave: Tablas estadísticas; análisis semiótico; niveles de complejidad semiótica; libros de texto.

Introduction

The school curriculum in different countries includes the study of statistics from an early age (e.g. MECD, 2014; MINEDUC, 2015; 2018), in recognizing its relevance in the training of citizens, which has been highlighted by several authors as a tool for understanding our society (Engel, 2019; Gal, 2002). These curricular guidelines highlight the need that students develop skills in data representation and interpretation, in order to reinforce their understanding of everyday world.

Statistical tables are an important content in the study of statistics as a tool to represent and communicate information (Arteaga, Batanero, Cañadas, & Contreras, 2011). Feinberg and Wainer (2012) indicate that the construction and interpretation of these tables are not simple, since interpreting the variables and their categories (rows and columns), in addition to understanding the type of information that is represented is required. It is also possible to raise a variety of tasks related to these tables, such as representing information, reading and interpreting the table, or translating between different representations (Koschat, 2005). Moreover, Estrella (2014) suggests that along teaching, the technique (routine guidelines) is given more priority than the understanding of fundamental aspects needed to construct or read the table.

We can find a wide variety of types of statistical tables, which were classified into three main groups by Lahanier-Reuter (2003). The purpose of this work is to expand this study, by analysing the mathematical objects implicit in the work with these types of tables. Additionally, we intend to adapt the semiotic complexity construct, described by Arteaga (2011) for graphs to statistical tables, and show the different difficulty involved in working with each of them. Finally, we analyse the distribution of these types of tables in a sample of Chilean Basic Education textbooks, in order to provide the teacher with detailed information on the level of complexity required in the work with the statistical tables in the different grades where its teaching is included.

Foundations

Below, we include the bases of the work, which are of two types: the theoretical framework and the background in which we base our analysis, and the subsequent exposition of the results obtained in the investigation.

Theoretical framework

We rely on the ontosemiotic approach to mathematical knowledge and instruction (Godino, 2002; 2017), in which the problem-situation and the practices that are carried out in its resolution serve to define the mathematical object and its meaning. Two interdependent

DOI: 10.20396/zet.v28i0.8656257

dimensions are considered for the meaning of a mathematical object (Godino, Batanero & Font, 2007; 2019): the personal (subjective, mental) dimension, and the institutional (objective, contextual) dimension. According to this framework, the meaning of the statistical table is the set of mathematical practices performed to solve problems related to this object.

In these practices, mathematical objects of different nature are involved: symbols, graphs, texts, among others. If the objects come from an institution they are "institutional objects" and if they arise individually, "personal objects". In Godino et al. (2007; 2019) the following types of primary object are proposed: language (terms, expressions, notations, graphs, etc.); problem situations (problems, extra-mathematical or mathematical applications, exercises, etc.); concepts, given by their definition or description; propositions, properties or attributes; procedures (operations, algorithms, techniques); and arguments or reasoning used to validate and explain the propositions and procedures (deductive, inductive, etc.).

We also consider the semiotic complexity levels, which is a model proposed by Arteaga and collaborators (Arteaga, 2011; Arteaga & Batanero, 2011; Batanero, Arteaga & Ruiz, 2010). These authors define four levels of complexity in the graphs, depending on the type of objects represented in them, because the understanding of graphs requires different interpretative processes, both of each element represented, and of the graph as a whole. The proposed levels are as follows:

- *Level 1 (N1).* Represent only some isolated data of one variable. At this level the idea of variable or distribution is not used.
- *Level 2 (N2).* Represent a set of data associated with a variable, without forming the frequency distribution. The idea of a variable and its values are used, but not that of frequency or distribution.
- *Level 3 (N3).* Represent a frequency distribution of a variable. In this case, the concept of frequency and distribution already appears.
- *Level 4 (N4).* Represent a frequency distribution of two or more variables. All the above objects are used and in addition, generally, the same scale is used to represent the variables.

The purpose of our work is to apply and extend these levels to the statistical tables, whose construction, reading and interpretation, also implies the mathematical objects described above.

Background

In addition to the theoretical framework, we will consider research that theoretically analyse the statistical table. Among this research, we find the study by Estrella (2014), who identified the following components in the statistical table:

- *Title:* that should describe the variables represented, and when possible, the context in which the study was conducted or the information collected.

DOI: 10.20396/zet.v28i0.8656257

- The *labels displayed in the lateral margin or first column*: show the different modalities that the variable takes, according to its classification scale.
- The *labels that are presented in the top heading or first row*: where the content of each columns is explained. For example, the type of frequencies, or the modalities of a second variable (two-way tables). Additionally, the totals are located in the last row and / or column.
- The *body of the data*: is the set of cells that are formed inside an intersection of rows and columns. These cells generally contains numerical information of different type, such as absolute or relative frequency, percentage and even, in some cases, it can contain more than one element in the same cell.

Types of tables

Lahanier-Reuter (2003) identified three types of tables, present in the teaching of statistics: data table, distribution of a variable, and contingency table. Each one has specific functions that give it its meaning, and are analysed below.

- *Data table*. It is the first organization of a data set. It is organised as a matrix and contains, for each individual in the sample, the values of one or several variables.
- *Distribution of a variable.* This table is constructed from the unclassified data set or from a data table. It describes the distribution of a variable, since it associates each modality of the variable with the number of individuals in the sample (frequency) that present this modality.
- *Two-way or contingency table*. Represents the data by crossing two statistical variables. In the upper part of the table (first row), the modalities of one of the variables are indicated, while the modalities of the second variable are included in the first column. The body of the table is formed by the joint frequencies corresponding to the modality of the row for the first variable and the column for the second variable.

In addition, we will consider the research of Díaz-Levicoy and collaborators (Díaz-Levicoy, Morales & López-Martín, 2015), who analysed the statistical tables in four 1st and 2nd year Basic Education (6 and 7 years) Chilean textbooks. They considered, among other variables, the type of table (data table, counting table, frequency and two-way table) in two different editorials. In one of them, the counting tables (a type of table in which the frequency counts are recorded, through marks or symbols within a cell) reached around 83% of all the tables in the book, while in the other, only 42 %. In the latter, a more balanced distribution was found with frequency tables (50%) and two-way tables (17%), which were not considered by the first editorial. This study was subsequently complemented (Díaz-Levicoy, Ruz & Molina-Portillo, 2017), by analysing the tasks related to statistical tables in three of third grade Basic Education (8 years) textbooks, also in Chile. The analysis showed that the type of table that appeared most strongly was counting tables (38.9%), followed by frequency table (32.2%), and then data table (25,6%). To a lesser extent they found some tables that combined two of the types mentioned above, such as counting and data table, or counting and frequencies table. It

DOI: 10.20396/zet.v28i0.8656257

is striking that, in this educational level, the two-way table had no presence in the explored textbooks, unlike what happened in the study aimed at 1st and 2nd grade (Díaz-Levicoy et al., 2015).

Our research completes the above papers, by using the semiotic complexity levels and the analysis of the mathematical objects implicit in the different types of tables, to show that, although visually these tables may seem very similar, different cognitive demands are required to work with each of them, which are not always explained or addressed by teachers in the school (Koschat, 2005). In addition, we analyse the distribution of the types of tables in the textbooks and compare the results with the studies by Díaz-Levicoy and collaborators, which were directed to lower educational levels in Chile.

Methodology

This is a descriptive research, where content analysis is used as a technique in a sample of school textbooks, following systematic steps (Mayring, 2000), which allows to obtain analysis categories cyclically and inductively.

The sample consists of twelve school textbooks, published in 2017 (see Appendix), which are aimed at students in grades 5-8 (10 to 13 years) of Basic Education in Chile, and attend the current curriculum framework (MINEDUC, 2015; 2018). For each educational level three textbooks were analysed: the student's book, the exercise book, and the teaching guide for the teacher. We highlight that these textbooks in the sample are a primary educational resource in the teaching and learning of mathematics in Chile, and were chosen because they are the most used, since they are freely distributed to students in the public system and schools supported by the state.

In these textbooks a semiotic analysis has been carried out, based on the study of the curricular guidelines by the NCTM (2000), the GAISE project (Franklin et al., 2005) and the Chilean curricular guidelines (MINEDUC, 2015; 2018), as well as with consult to specialized Statistical Education papers. This process makes it possible to assess to the planned instruction in a textbook, and its result can be considered as the intended institutional meaning (Godino, 2002; 2017) for the educational levels examined.

To carry out the analysis, the primary elements considered in the ontosemiotic approach are studied for each type of table, since the work with statistical table requires activating a set of elements that constitute an epistemic configuration, in each of them (Godino, 2002; 2017), which is composed of the following objects: problem situations, language, arguments and rules (concepts, propositions and procedures).

Results

We present below the results of the analysis: First, we describe the semiotic complexity

DOI: 10.20396/zet.v28i0.8656257

levels corresponding to the different types of statistical table described by Lahanier-Reuter (2003), and adapt to them the levels of semiotic complexity defined for statistical graphs by Arteaga (2011). Next, for each type of primary object considered in the EOS theoretical framework (Godino, 2002; 2017), we describe those associated with each type of table. We finish with a study of the distribution of the types of tables in the sample of textbooks.

Semiotic complexity levels associated to different types of tables

The semiotic complexity levels defined for the statistical graphs by Arteaga et al. (Arteaga, 2011; Arteaga & Batanero, 2011; Batanero et al., 2010), can be adapted to work with statistical tables. We need to adjust the categorization, since the lowest level is not considered and, in addition, it is necessary to subdivide the two upper levels into other more specific levels. These levels are closely related to the different types of tables described by Lahanier-Reuter (2003):

- *Data table.* They are characterized by their large size, since they have as many rows as individuals in the sample. Generally, at the top of the table, a heading is included with the description of the variables, collected in each individual. The first column on the left contains the list of subjects, while in each of the following columns the list of the data associated with each variable are displayed. Sometimes the variables appear obeying a certain hierarchy or order; For example, in Figure 1, the individuals are dates that appear in chronological order, on which data has been collected in two years. In our proposal, this type of table corresponds to the semiotic level N2, described by Arteaga (2011), since in this representation the idea of variable and value of the variable appears, but no frequencies associated with each modality are shown and, therefore, the concepts of distribution of statistical variable does not emerge.

b leen e interpretan la información correspondiente a las temperaturas máximas de los diez primeros días del mes de octubre en los años 2009 y 2010

	2009	2010
01/10/2010	22°	24°
02/10/2010	25°	25°
03/10/2010	26°	20°
04/10/2010	21°	18°
05/10/2010	18°	19°
06/10/2010	20°	21°
07/10/2010	25°	18°
08/10/2010	26°	20°

Figure 1. Data table proposed for 5th grade
Source: MINEDUC (2013, p.147)

- *Distribution of a variable.* Usually, this type of table has several columns. Generally, the type of frequencies represented in the table (absolute, relative, cumulative or percentages) are indicated in the upper part (first row). In the first column the different

DOI: 10.20396/zet.v28i0.8656257

modalities or values that the variable takes are described. And in each of the following columns, the cells in the table contain the frequency that corresponds to that value or modality (of the type indicated in the first row for that column). Sometimes, the last row records the total values. If the number of values of the variable or categories is small, it is possible to exchange rows and columns in the table.

An example taken from a textbook is shown in Figure 2. In this table, in addition to absolute frequencies, cumulative frequencies are included, and a row is added with the computation of quartiles. We suggest that these tables correspond to the semiotic level N3 (Arteaga, 2011), since it represents the frequency distribution of a variable. In addition, this level could be divided into three sub-levels, depending on whether cumulative frequencies and class intervals are considered. This classification takes into account the mathematical objects involved (Cobo & Batanero, 2000), by giving each table different meaning and complexity for the student.

- *Complexity level N3.1:* Ordinary frequency distribution tables: absolute, relative or percentage.
- *Complexity level N3.2:* Frequency distribution tables, which also include cumulative frequencies (absolute, relative or percentage). Its level of complexity is higher, because it involves the use of inequalities.
- *Complexity level N3.3:* When considering the grouping of the values of the variable in intervals, for any type of frequency, both ordinary and cumulative. The ideas of interval, its extremes and class mark are added.

Calcula los valores de los cuartiles a partir de la información de la tabla.

Cantidad de hermanos de un grupo de estudiantes		
Cantidad de hermanos	Frecuencia absoluta	Frecuencia acumulada
0	3	3
1	9	12
2	6	18
3	18	36
4	3	39
5	2	41
Cuartil 1 (25%)	Cuartil 2 (50%)	Cuartil 3 (75%)
$0{,}25 \cdot 41 = 10{,}25$ $Q_1 = 1$	$0{,}5 \cdot 41 = 20{,}5$ $Q_2 = 3$	$0{,}75 \cdot 41 = 30{,}75$ $Q_3 = 3$

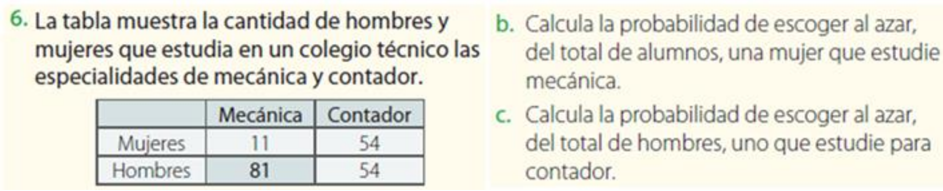
Figure 2. Distribution table for 8th grade
Source: Maldonado, & Castro (2017, p. 317)

- *Contingency table.* As in the previous case, the frequencies of the cells can be ordinary (absolute, relative or percentage), but in this case, we find different types of relative frequencies (with respect to the row or the column). The last row and the last column are intended to register subtotals, of rows and columns respectively. We include in Figure 3 an example for 7th grade (12 years). These tables correspond to the highest level of semiotic complexity: N4 according to Arteaga (2011), since they allow

DOI: 10.20396/zet.v28i0.8656257

comparing the distribution of two unidimensional statistical variables and relating them to each other. These tables, in turn, could contain class intervals, although it is not usual to consider cumulative frequencies. Therefore, this level could be classified into two sub-levels, depending on whether class intervals are considered or not:

- *Complexity level N4.1:* Contingency tables of ordinary frequencies: absolute, relative or percentage.
- *Complexity level N4.2:* When considering the grouping of the values of the variable in interval, for any type of frequency.



6. **La tabla muestra la cantidad de hombres y mujeres que estudia en un colegio técnico las especialidades de mecánica y contador.**

	Mecánica	Contador
Mujeres	11	54
Hombres	81	54

b. Calcula la probabilidad de escoger al azar, del total de alumnos, una mujer que estudie mecánica.

c. Calcula la probabilidad de escoger al azar, del total de hombres, uno que estudie para contador.

Figure 3. Contingency table for 7th grade
Source: Merino, Muñoz, Pérez, & Rupin (2016, p. 348)

Once our classification of the tables is proposed, according to semiotic complexity levels, we proceed to the semiotic analysis of these tables, in considering successively the different types of mathematical objects linked to the work with the statistical tables.

Problem-situations linked to work with statistical tables

In the EOS, the problem situation has a relevant role, since mathematical objects intervene and emerge from personal or institutional practices when solving problems (Godino et al., 2007; 2019). The following problem-situations (SP) have been found in the textbooks:

- SP1. The first of these problems is the *organization of the data,* which entails the registration of the same. It occurs when data is collected for some purpose and it is necessary to organize the data in order to interpret it. As Estrella (2014) suggests, this use is very ancient and has been found in different civilizations. An example is shown in Figure 1.
- SP2. Another different problem is the *construction of the distribution of a statistical variable* through the table, which entails the calculation of the frequency for each modality or value and its organization in the distribution table (See Figure 2). It involves the action of organizing a data set, which can be delivered or collected by the student, and is associated with a one-dimensional statistical variable, in a table in which one or more types of frequencies (ordinary or accumulated) are presented. It appears with the need to summarize a list of data, considering that the same values of the variable are repeated in the list. Subproblems could be differentiated, depending on the type of variable, if one works with more than one variable or according to the frequencies to

DOI: 10.20396/zet.v28i0.8656257

be calculated. Depending on the length of the data list, in some cases, it is necessary to distribute the modalities of the variable intervals.

- SP3. *Translation between representations.* This type of problem-situation is divided into four categories, to differentiate each of the translation processes linked to statistical tables: from table to graph or vice versa, from table to table, from table to text or vice versa, and from table to statistical summary. An example of a graph-to-table translation is presented in Figure 4. Such translation may involve a transnumeration process (Wild, & Pfannkuch, 1999), when new information is deduced from the data. This change in the representation of information and its subsequent analysis involves a series of tasks such as: distinguishing and classifying data, computing central tendency or spread measures in one variable, or computing frequencies. In the transnumeration process with statistical tables, Chick (2004) identifies different stages prior to the process, such as representing the data in tables or graphs, or computing data frequencies, among others.

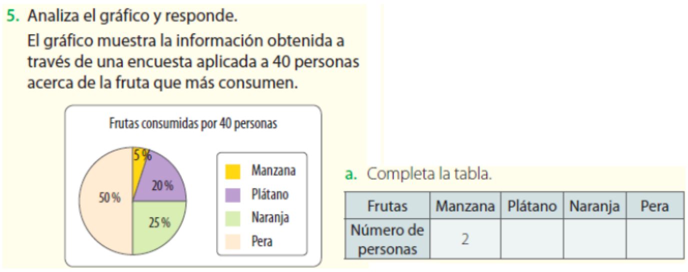


Figure 4. Translation of graph to a table for 7th grade
Source: Merino, Muñoz, Pérez, & Rupin (2016, p. 348)

- SP4. *Performing a cross classification of two variables.* The two-way table allows to organize the frequency distribution of a two-dimensional statistical variable, through as many rows and columns as modalities present the variables that make it up. At each crossing of the respective variables, the joint frequency (absolute, relative or percentage) of the data is represented. Likewise, we highlight the importance of the two-way table in the study of association between two-dimensional statistical variables (See Figure 3). Gea, Batanero, Cañadas and Contreras (2013) classify this problem situations in two-dimensional studies in three large types:
    - SP4.1. Organising information from a bivariate data set is an important aspect in the subsequent study of topics such as correlation and regression (Gea et al., 2013). It would correspond to the example shown in Figure 3, since in order to respond to the proposed tasks, it is necessary to calculate the totals in the table presented in the problem-situation.
    - SP4.2. Analysis of the variables that make up the two-dimensional statistical variable. When analysing the two-dimensional statistical variable, it is necessary to delimit each of the variables that constitute it, such as the situation-problem shown in Figure 5, where the student after describing the variables

DOI: 10.20396/zet.v28i0.8656257

must construct an appropriate graph to represent the information presented in the table. In this category the descriptive analysis of each variable is considered, as well as the steps required to complete a double-entry table known the marginal averages, and the analysis of the functional or statistical dependence of the variables, their intensity and meaning (Gea, 2014).

Define la variable en estudio en cada tabla y representa en un gráfico adecuado de acuerdo al tipo de variable.

a. Tabla 1: Número de funciones realizadas por un grupo de teatro en Chile, Brasil y Argentida entre el 2010 y 2012.

	2010	2011	2012
Chile	5	7	3
Brasil	1	4	2
Argentina	10	8	7

Figure 5. Analysis of the variables in a contingency table for 7th grade
Source: Merino, Muñoz, Pérez, & Rupin (2016, p. 361)

- SP4.3. Study of the association between the variables that make up the table, which pursues the analysis of the dependence of the variables and, in the case that there is intense dependence, establish an adjustment model with predictive purpose (Gea, 2014). We did not find examples in the analysed textbooks.

Table 1. Problem fields linked to the different types of statistical tables

PROBLEM AREAS	Data table	Distribution table			Contingency table
		Ord.	Cum.	Group.	
SP1. Summary and organization	x	x	x	x	x
SP2. Construction of the distribution of a variable		x	x	x	
SP3. Translation of:					
1. Table to graph, or vice versa	x	x	x	x	x
2. Table by table	x	x	x	x	x
3. Table to text, or vice versa		x	x	x	x
4. Table to statistical summary	x	x	x	x	x
SP4. Cross classification of two variables:					
1. Organization of information					x
2. Analysis of the variables that make up the two-dimensional statistical variable					x
3. Study of the association between the variables that make up the table					x

Source: Own elaboration

In Table 1 we present a summary of the problem situations (SP) that implicitly or explicitly characterize the three main types of statistical tables. The Distribution table has been subdivided into three categories (ordinary: Ord; cumulative: Cum; grouped data: Group) according to the mathematical objects involved, and the sub-levels of semiotic complexity, but we have not subdivided the contingency tables having not found in the textbooks tables of this type where class intervals are used.

DOI: 10.20396/zet.v28i0.8656257

We can see that the first three types of problem are associated with all types of tables, so these problem fields can be worked with any of them, while the SP4 problem field, with its different subproblems, would be limited to the contingency table.

Language associated with work with statistical tables

The language allows to represent a mathematical object. In relation to the statistical tables, independent of its type, different types of mathematical language are used, both when working with them and to communicate the definition, procedure or the solution of a problem-situation. In summary, we identify the following language typology, which is applicable to all types of tables:

- *Verbal*, present in the title, modalities of the variable (when qualitative), in the labels indicating the type of frequencies (e.g. absolute, relative, percentages) or last row labels (totals). It also includes words that refer to the magnitudes represented (for example, prices, heights, distances, etc.).
- *Symbolic*, depending on the type of table, mathematical symbols may appear on the labels of the margins, upper, lower rows, first column and last column; symbols appear in these spaces, for example: x_i (value of an X variable); n_i (absolute frequency); f_i (relative frequency), and operations such as $\sum$ (summation, with or without subscripts).
- *Numerical*, in the body of data different types of numbers appear as integers and rational (decimals, fractional, percentages).
- *Diagrammatic*. The statistical table itself constitutes a spatial positioning of one or two dimensions (right, left, up, down, intersection, diagonal), through its rows and columns, allowing transformations, combinations and constructions, obeying certain syntactic or semantic rules (Godino, Giacomone, Wilhelmi, Blanco & Contreras, 2015).

Arguments

As for the arguments, also common to the different tables, they are used to justify the resolution of various problem situations, and also to communicate the results obtained.

The arguments not only are used for the formal deductive demonstrations, characteristics of mathematics, but in any statement used to justify (Godino & Recio, 2001). For example, they can be used to validate whether a series of statements about the information presented in a table are true or false; interpret a certain context, why it is interesting to consider a measure as the most representative of a distribution presented in a table, based on the study of central tendency measures.

The arguments that are used around the table are determined by the concepts, procedures and properties used to respond to the problem-situation that arises. Thus, we find a variety of ways of arguing as they are (Recio, 1999):

- *Examples or counterexamples.* It concerns the use of concrete examples to validate a

DOI: 10.20396/zet.v28i0.8656257

certain property or procedure.

- *Use of graphical representations* from a table, as it is sometimes useful to graphically represent the data in the table to argue a statement.
- *Deductive verbal arguments,* referring to deductive proofs, using axioms or theorems that make up this argument.
- *Algebraic deductive argument,* are those that demonstrate deductive arguments based on algebraic language, associated with the table. These are algebraic manipulations used to argue a property, the use of some formula, etc.

Concepts and properties

A mathematical concept is recognized by its definition and its properties, which may vary according to the institution, which implies that they have a relative character. The definitions of concepts are evoked by the student when solving a problem situation, so it is important to analyse the treatment of concepts in teaching, since the progressive construction of their meaning will influence the concepts that are learnt and used (Godino, 2017). In the statistical tables, the concepts and their depth obey the curricular framework (MINEDUC, 2015; 2018) for the educational levels to which our study is directed. Some of the concepts linked to the tables that appear in the Chilean curricular bases are: population, individual, census and samples; variables and their values; class intervals, extremes and class mark; frequency distribution of one-dimensional statistical variables (absolute, relative, percentage, cumulative); odds; frequency distribution of two-dimensional statistical variables (double, marginal, conditional); In addition, other concepts such as centralization, position and dispersion measures appear in the textbooks.

In Table 2, a summary of the concepts that implicitly or explicitly characterize the three main types of statistical tables is presented. We can observe the greater complexity of the distribution and contingency table, regarding the smaller number of concepts that it is necessary to use in the work with the data table. From this semiotic diversity, the levels of complexity that we propose in the study of statistical tables are also justified.

In relation to properties, these refer to specific characteristics of the concepts or relationships between the concepts. Each property is associated with a mathematical object or the relationship between objects, so it regulates its use and contributes to the growth of the meaning of the object in question (Godino, 2002; 2017). For the statistical table (See Table 2), the main properties refer to the relationships between frequency types and depend on whether it is a frequency distribution, or a contingency table and the type of frequencies (cumulative or not, in grouped data or not). It should be noted that the contingency table covers practically all of the properties, as well as other specific ones. This demonstrates its complexity in terms of the high level of understanding required for the treatment of this type of tabular representation.

DOI: 10.20396/zet.v28i0.8656257

Table 2. Concepts and properties linked to the different types of statistical tables

CONCEPTS	Data table	Distribution table Ord.	Ac.	Cum.	Contingency table
Population, individual	x	x	x	x	x
Census, sample	x	x	x	x	x
Variable, values	x	x	x	x	x
Range, maximum, minimum	x	x	x	x	x
Class intervals, extremes, class mark				x	x
Ordinary frequencies		x	x	x	x
Cumulative frequency			x	x	
Joint frequencies					x
Marginal and conditional frequencies					x
Simple probabilities		x	x	x	x
Compound and conditional probabilities					x
Central tendency measurements, dispersion	x	x	x	x	x
Covariance, correlation, regression					x
PROPERTIES					
Variable types and scale of measurement	x	x	x	x	x
Sum of absolute frequencies and sample size		x		x	x
Sum of relative frequencies		x		x	x
Relative and absolute frequency proportionality		x		x	x
Increasing value of accumulated frequencies, the last of which is the sample size			x		
Relationship between cumulative and non-cumulative frequencies in aggregate data			x	x	
Dependent and independent variable, association between variables					x
Relationship between conditional, joint and marginal relative frequencies					x
Expected frequencies in case of independence					x

Source: Authors

Procedures used in working with statistical tables

The procedures refer to the algorithms or strategies that are applied in a table and allow to operate with the data. For the tables we identify three processes:

- *Table reading*, if the table is built and some data should be obtained from it, an interpretive process of each element that forms it (title, labels, values) and the table as a whole is required. The reading can be more or less complex, since you can consider the reading levels for graphs proposed by Friel, Curcio and Bright (2001), which involves four procedures, which can be transferred to the statistical tables: *Read the data:* the reading of the data conforms to what is explicitly shown in the table, such as reading a frequency for a given value of the variable, or vice versa; *Read within the data:* it consists in the use of the explicit information of the table, by means of simple calculations with the data to obtain information that is not directly available in it; *Read beyond the data*: that implies inferring information that is not exposed, nor is it possible

DOI: 10.20396/zet.v28i0.8656257

to extract it in an arithmetic way; and *Read behind the data:* it implies a critical assessment of the content of the table, as the source from which the information has been extracted from the table, in addition to relating that information to the context in which it is proposed.

- *Construction of tables.* The student must form the table based on the raw data. Duval (2003) indicates that this is a procedure that involves the deployment of a variety of cognitive functions, necessary to carry out this task; The categorization of data and the calculation of frequencies, when necessary, stand out.
- *Other procedures*, for example, data count operations, calculations of different types of frequency, determine statistics.

In addition, we could consider that translation between representations involves both reading and construction.

Table 3. Procedures linked to the different types of statistical tables

PROCEDURES	Data table	Distribution table			Contingency table
		Ord.	Ac.	Group.	
Table reading	x	x	x	x	x
Construction of tables	x	x	x	x	x
Other procedures	x	x	x	x	x

Source: Authors

In Table 3 we can see that the procedures described above are common for all types of statistical tables. However, it should be taken into account that each of these actions requires specific skills and abilities, depending on the type of tabular representation in question, because the tables, although visually similar, each has a different structure that characterizes it with Different meaning and function.

Distribution of the types of tables in the textbooks analysed

To complete the work, a statistical analysis of the distribution of the different types of tables analysed in the sample textbooks has been carried out. The classification has been divided into the three types of statistical tables described above: data table, distribution table of a variable and contingency table. In turn, remember that the Distribution Table has been subdivided into three categories: ordinary frequencies (absolute, relative and percentage), cumulative frequencies and distribution of the variable through grouped data.

As a result of the analysis, we observe (See Table 4) that globally more than half of the tables present in school textbooks correspond to distribution tables, first, those that correspond to ordinary frequencies (51.8%), followed by data tables (19.2%), then contingency tables (13.1%). Those of distribution of grouped data frequencies (3.9%) are the least frequent in the textbooks, as well as cumulative frequencies (12%). The tendency of our results is similar to

DOI: 10.20396/zet.v28i0.8656257

that evidenced in the studies by Díaz-Levicoy et al. (2015; 2017), although in the aforementioned investigations only ordinary frequency tables appear, due to the lower age of the students. In their study, most of the tables are counting table, not considered in the grades we have analysed. The frequency table appears around 32% in the third grade and there is no presence of the contingency table, although it is rarely found in the 1st and 2nd Basic Education grades. In our study, we were also able to detect a low presence of the contingency table, although it appears at all levels explored and, paradoxically despite its complexity, it is used more at the lower levels.

Table 4. Frequency (and percentage) of table type, by education level

TYPE OF TABLE		5º Level	6º Level	7º Level	8º Level	Total
Data table		27(18,1)	9(8)	65(14,3)	90(32,5)	191(19,2)
Distribution table	Ordinary	81(54,4)	83(73,4)	272(59,9)	78(28,1)	514(51,8)
	Accumulated			77(17)	42(15,2)	119(12)
	Grouped				39(14,1)	39(3,9)
Contingency table		41(27,5)	21(18,6)	40(8,8)	28(10,1)	130(13,1)
Total		149	113	454	277	993

Source: authors

We also observe variation with the school level. In the 5th and 6th grades, the study with the tables mainly focuses on those of distribution with ordinary frequencies. As of 7th grade, the statistical tables that incorporate the accumulated frequencies appear, and in a very low percentage in the 8th grade the frequency distribution with data grouped in intervals is studied. We also observe that the 5t h level is the one that shows a greater presence of contingency tables (27.5%).

Conclusions

In the presentation, we can appreciate the cognitive complexity linked to the statistical table, which implies the understanding of a variety of mathematical objects necessary to successfully carry out tasks that involve these tables. Each of these mathematical objects has different levels of formalization as the educational level progresses, since it is not the same to deal with a statistical table in 5th grade than in 8th grade, where the tasks require more knowledge and mathematical skills.

The semiotic analysis of the statistical table, through the primary elements of the ontosemiotic approach (Godino, 2017), becomes a useful methodological tool to deepen and identify the various mathematical objects that are required to know to work with the different types of tables proposed by Lahanier-Reuter (2003), given that each of them has a specific structure and function, thus achieving to characterize the intended institutional meaning of the statistical table.

The study of the distribution of the tables in the analysed textbooks, allows to detect that the contingency tables, of the highest level of semiotic complexity (N4), are the least

DOI: 10.20396/zet.v28i0.8656257

frequent in the textbooks, a fact that attracts attention given that this type of tables is the one that presents a greater wealth of mathematical objects and possible relationships to establish between them, being mostly used in 5th grade and with a decreasing frequency as the educational level increases. On the other hand, most of the activities analysed are based on N3 complexity level tables.

With this study, we confirm the applicability of complexity levels, initially proposed by Arteaga (2011) for statistical graphs to tables, where an adjustment of the levels was made, each of them were associated with a type of statistical table. This in order to deepen and better describe the semiotic complexity, especially levels N3 and N4, which has offered a complement to the model described by Arteaga, in proposing three sub-levels of complexity for N3, and two sub-levels for the highest degree of complexity N4. They obey to the different types of mathematical objects that can be identified in the table, such as the consideration of ordinary or accumulated frequencies and class intervals, which allows a greater degree of difficulty, within the same level of semiotic complexity. Thus, an enrichment of the levels of complexity associated with statistical representations is offered.

Acknowledgement:

Project EDU2016-74848-P (AEI, FEDER), Research group FQM-126 (Junta de Andalucía) and grant CONICYT: 72190280

References

Arteaga, P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores* (Tesis PhD). University of Granada, Granada, Spain.

Arteaga, P. & Batanero, C. (2011). Relating graph semiotic complexity to graph comprehension in statistical graphs produced by prospective teachers. En M. Pytlak, T. Rowland, & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp. 368-377). Rzeszów: ERME.

Arteaga, P., Batanero, C., Cañadas, G. & Contreras, J. M. (2011). Las tablas y gráficos estadísticos como objetos culturales. *Números 76*, 55-67.

Batanero, C., Arteaga, P. & Ruiz, B. (2010). Análisis de la complejidad semiótica de los gráficos producidos por futuros profesores de educación primaria en una tarea de comparación de dos variables estadísticas. *Enseñanza de las Ciencias, 28*(1), 141-154.

Chick, H. (2004). Tools for transnumeration: Early stages in the art of data representation. In I. Putt, R. Faragher, & M. McLean (Eds), *Mathematics education for the third millennium: Towards 2010, Proceedings of the Twenty-seventh Annual Conference of the Mathematics Education Research Group of Australasia* (pp.167-174). Sydney: MERGA.

Cobo, B. & Batanero, C. (2000). La mediana en la educación secundaria obligatoria: ¿un concepto sencillo? *UNO, 23*, 85-96.

DOI: 10.20396/zet.v28i0.8656257

Díaz-Levicoy, D., Morales, R. & López-Martín, M. M. (2015). Tablas estadísticas en libros de texto chilenos de 1º y 2º año de Educación Primaria. *Revista Paranaense de Educação Matemática, 4*(7), 10-39.

Díaz-Levicoy, D., Ruz, F. & Molina-Portillo, E. (2017). Tablas estadísticas en libros de texto chilenos de tercer año de Educación Primaria. *Espaço Plural*, *18*(36), 196-218.

Duval, R. (2003). Comment analyser le fonctionnement représentationnel des tableaux et leur diversité ? *Spirale-Revue de recherches en éducation*, *32*(32), 7-31.

Engel, J. (2019). Statistical literacy and society. En J. M. Contreras, M. M. Gea, M. M. López-Martín, & E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Granada: Grupo FQM-126. Disponible en: www.ugr.es/local/fqm126/civeest.html.

Estrella, S. (2014). El formato tabular: una revisión de literatura. *Revista Actualidades Investigativas en Educación*, *14*(2), 1-23.

Feinberg, R. & Wainer, H. (2012) Extracting sunbeams from cucumbers. *Journal of Computational and Graphical Statistics*, *20*(4), 793-810.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M. & Scheaffer, R. (2005). Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K- 12 curriculum framework. Alexandria, VA: American Statistical Association. Disponible en: www.amstat.org/Education/gaise/.

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.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International statistical review*, *70*(1), 1-25.

Gea, M. M. (2014). *La correlación y regresión en bachillerato: Análisis de libros de texto y del conocimiento de los futuros profesores* (Tesis doctoral). Universidad de Granada, Granada, España.

Gea, M. M., Batanero, C., Cañadas, G. R., & Contreras, J. M. (2013). Un estudio empírico de las situaciones-problema de correlación y regresión en libros de texto de bachillerato. En A. Berciano, G. Gutiérrez, A. Estepa, & N. Climent (Eds.), *Investigación en Educación Matemática XVII* (pp. 293-300). Bilbao: Sociedad Española de Investigación en Educación Matemática (SEIEM).

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactiques des Mathematiques, 22*(2-3), 237-284.

Godino, J. D. (2017). Construyendo un sistema modular e inclusivo de herramientas teóricas para la educación matemática. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M. M. Gea, B. Giacomone, & M. M. López-Martín (Eds.), *Actas del Segundo Congreso International Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemática*. Disponible en: http://enfoqueontosemiotico.ugr.es/civeos/godino.pdf.

Godino, J. D., Batanero, C. & Font, V. (2007). The onto-semiotic approach to research in

DOI: 10.20396/zet.v28i0.8656257

mathematics education. *ZDM. The International Journal on Mathematics Education, 39*(1-2), 127-135.

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), 37-42.

Godino, J. D., Giacomone, B., Wilhelmi, M. R., Blanco, T. F. & Contreras, A. (2015). Configuraciones de prácticas, objetos y procesos imbricadas en la visualización espacial y el razonamiento diagramático. *Departamento de Didáctica de la Matemática*. Universidad de Granada, España.

Godino, J. D. & Recio, Á. M. (2001). Significados institucionales de la demostración. Implicaciones para la educación matemática. *Enseñanza de las ciencias: Revista de investigación y experiencias didácticas*, *19*(3), 405-414.

Koschat, M. (2005). A case for simple tables. *The American Statistician, 59*(1), 31-40.

Lahanier-Reuter, D. (2003). Différents types de tableaux dans l'enseignement des statistiques. *Spirale-Revue de recherches en éducation*, *32*(32), 143-154.

Maldonado, L., & Castro, C. (2017). *Texto del estudiante Matemática 6º básico*. Santiago: Grupo Santillana de ediciones.

Mayring, P. (2000). Qualitative inhalts analyse. *Forum: Qualitative Sozialforschung/Forum: Qualitative Social Research*, *1*(2). Art. 20.

Merino, R., Muñoz, V., Pérez, B., & Rupin, P. (2016). *Texto del estudiante Matemática 7º básico*. Santiago: Ediciones SM.

MECD (2014). *Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación Primaria*. Madrid, España: Autor.

MINEDUC (2013). *Matemática. Programa de Estudio Quinto Año Básico*. Santiago, Chile: Unidad de currículum y evaluación.

MINEDUC (2015). *Bases curriculares 7º Básico a 2° Medio*. Santiago, Chile: Unidad de currículum y evaluación.

MINEDUC (2018). *Bases Curriculares Primero a Sexto Bási*co. Santiago, Chile: Unidad de currículum y evaluación.

NCTM (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Recio, A. M. (1999). *Una aproximación epistemológica a la enseñanza y aprendizaje de la demostración matemática* (Tesis doctoral). Universidad de Granada, Granada, España.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223-248.

APÉNDICE

Books used in the analysis: 5th grade

Kheong, F. H., Soon, G. K., & Ramakrishnan, C. (2017). *Texto del estudiante Matemática 5º*

DOI: 10.20396/zet.v28i0.8656257

básico. Santiago: Marshall Cavendish Education.

Kheong, F. H., Soon, G. K., & Ramakrishnan, C. (2017). *Cuaderno de ejercicios Matemática 5º básico.* Santiago: Marshall Cavendish Education.

Kheong, F. H., Soon, G. K., & Ramakrishnan, C. (2017). *Guía didáctica del docente – Tomo 2 Matemática 5º básico*. Santiago: Marshall Cavendish Education.

Books used in the analysis: 6th gradel

Maldonado, L., & Castro, C. (2017). Texto del estudiante Matemática 6º básico. Santiago: Grupo Santillana de ediciones.

Castro, C. (2017). Cuaderno de ejercicios Matemática 6º básico. Santiago: Grupo Santillana de ediciones.

Juillet, I., & Martínez, M. (2017). Guía didáctica del docente – Tomo 2 Matemática 6º básico. Santiago: Grupo Santillana de ediciones.

Books used in the analysis: 7th grade

Merino, R., Muñoz, V., Pérez, B., & Rupin, P. (2016). Texto del estudiante Matemática 7º básico. Santiago: Ediciones SM.

Santis, M. (2016). Cuaderno de ejercicios Matemática 7º básico. Santiago: Ediciones SM.

Raydoret del Valle, J. (2016). Guía didáctica del docente Matemática 7º básico. Santiago: Ediciones SM.

Books used in the analysis: 8th gradel

Catalán, D., Pérez, B., Prieto, C., & Rupin, P. (2016). Texto del estudiante Matemática 8º básico. Santiago: Ediciones SM.

Muñoz, V., & Chacón, A. (2016). Cuaderno de ejercicios Matemática 8º básico. Santiago: Ediciones SM.

Muñoz, V., & Manosalva, C. (2016). Guía didáctica del docente Matemática 8º básico. Santiago: Ediciones SM.