International Electronic Journal of Mathematics Education
2023, 18(2), em0733
e-ISSN: 1306-3030
https://www.iejme.com

MODESTUM

Review Article

OPEN ACCESS

# Statistical knowledge of primary schoolchildren: An overview of study approaches

Daniel Londoño [1], Ángel Alsina [1*]

[1] University of Girona, Girona, SPAIN
*Corresponding Author: angel.alsina@udg.edu

**Citation:** Londoño, D., & Alsina, Á. (2023). Statistical knowledge of primary schoolchildren: An overview of study approaches. *International Electronic Journal of Mathematics Education, 18*(2), em0733. https://doi.org/10.29333/iejme/12984

**ARTICLE INFO**

Received: 09 Dec. 2022

Accepted: 19 Feb. 2023

**ABSTRACT**

A review of studies analyzing the statistical knowledge of primary schoolchildren (6-12 years old) is carried out. Based on a review in JCR/SSCI, Scopus, Eric, Google Scholar, Science Direct, World Scientific, Springer, and Wiley Online library, 18 articles (2003-2021) have been identified and analyzed based on two objectives: (i) to identify the different study approaches and (ii) to analyze the elements of statistical knowledge. The results show that almost half of the investigations were carried out based on one of the following approaches: the Toulmin approach (TM), the statistical mathematical working space (SMWS), the structure of observed learning outcomes (SOLO) taxonomy and Curcio's graph reading levels (CGRL). It is concluded that CGRL is the most common approach and statistical graphs are the most analyzed statistical objects.

**Keywords:** statistical knowledge, statistical learning, statistical literacy, research approaches, primary school

## INTRODUCTION

In recent times, statistical knowledge has become a fundamental aspect within the integral education of citizens and their participation in contemporary society (Alsina et al., 2023; Muñiz-Rodríguez et al., 2020). Statistics are taught today at all levels of education, as part of the culture of a society, which is increasingly interested in obtaining information and in participating broadly in the administration and investigation of different fields of knowledge (Batanero et al., 2011). An interest in the analysis of statistical learning (and probability) at young ages has led certain researchers in recent years to initiate studies, which reveal significant factors for consideration among primary schoolchildren (the stage of education upon which this study focuses). In the opinion of Ruiz López (2015), for example, this component of school mathematics makes it possible to develop basic skills for collecting, organizing and tabulating data, comparing phenomena, analyzing variables, building and interpreting tables and graphs, taking decisions and contributing towards the development of cognitive skills, with a tendency towards reinforcing logical reasoning and problem-solving. In addressing such knowledge, Batanero et al. (2013) use the fundamental types of statistical reasoning and involve students in a complete research cycle and in statistical modelling.

Different studies have analyzed primary students' statistical knowledge from different focus. Some of the most common ones are the Toulmin approach (TM), based on this author's contributions on argumentation in mathematics (Toulmin, 1958); the statistical mathematical working space (SMWS), developed by Vidal-Szabó et al. (2020); the structure of observed learning outcomes (SOLO) taxonomy of Biggs and Collis (1982); and Curcio's graph reading levels (CGRL) (Curcio, 1987; Friel et al., 2001).

Considering these approaches, we aim to explore the following questions: What studies have been published on statistical knowledge among primary schoolchildren? How are the analytical approaches implemented in these studies? What tools are being used? What topics are students learning about? How are students acquiring knowledge within the statistical cycle? What level of learning do they achieve? What results are being obtained?

To answer these questions, it is necessary to go to the work carried out by the researchers, which, due to their high degree of importance and rigor, are published in different scientific documents. As an effect of the above, this study involves a review of articles on statistics education published by indexed journals *in journal citation reports* (JCR), which provides clear information on journals available within the research community and can also be of assistance in terms of organization and assessment; *Scopus*, a database of abstracts, bibliographic references on science, technology, medicine, the social sciences and humanities, with quality web content, including analytical tools for scientific research; the impact indices Scimago journal rank and source-normalized impact; the *Scientific Electronic Library Online* (SciELO) database, an electronic library, which encompasses a select collection of scientific journals from all areas of knowledge for Latin America, the Caribbean, Spain, and Portugal; *Latindex*, a database encompassing several disciplines and including the collaboration of institutions with publications of Ibero-American origin; *Dialnet*, a bibliographic website focusing on Hispanic scientific literature in the fields of the human, legal and social sciences; *JSTOR*, which, since 1994 and on the initiative of William G. Bowen, assists university libraries in transforming printed academic journals into electronic formats; *Google Scholar*, a specialized Google search engine for academic documents such as scientific papers, theses, dissertations, book citations, articles and news from scientific magazines; and, finally, the *Directory of Open Access Journals* (DOAJ) in which open access journals can be found covering the areas of science, technology, medicine, the social sciences, arts and humanities.

In this review, two fundamental objectives are proposed:

(i) to analyze the different approaches to study the statistical knowledge of primary education students (6-12 years old) and

(ii) to determine and analyze the elements of statistical learning of primary schoolchildren identified by researchers based on the implementation of the tools and on their results.

## RESEARCH APPROACHES FOR THE ANALYSIS OF THE STATISTICAL KNOWLEDGE

As mentioned in the introduction, this paper addresses the analytical approaches of students' knowledge described in the different studies found on the issue. It focuses particularly on TM, which integrates representations with argumentation, highlighting the verbal and gesture elements originating from the oral, numerical and geometrical skills of the students (Estrella et al., 2017; Toulmin, 1958); SMWS, which provides an analytical approach from a theoretical and methodological point of view for studying the work carried out by students when confronting a mathematical activity, in this case in terms of statistical knowledge and related cognitive skills possessed by a student when carrying out a statistical task (Vidal-Szabó et al., 2020); SOLO taxonomy of Biggs and Collis (1982), which aims to characterize the answers, which arise in the process of interaction; and CGRL (Curcio, 1987; Friel et al., 2001), which consider elements such as the words, which make up the title, the labels of the axes and scales on statistical graphs, the implicit mathematical content represented therein and the symbology used in their creation. Such components provide the keys to understanding the context of the data, the variables, and their relationships (Díaz-Levicoy et al., 2019a).

As far as TM is concerned, Toulmin (1958) considered that when stating an idea in accordance with certain data, a responsibility is created for the student with that declaration and this event should be justified. In other words, students draw conclusions based on their representations of data and there should be a guarantee to support these conclusions. This approach takes into consideration three elements, which play a role in argumentation: the *conclusions*, which are the statements used by the subjects of the study with the aim of achieving the acceptation of their ideas; the *data*, which are the elements, which justify and ground the conclusions; and the *warrants*, which are the processes, which validate the conclusions based on the data (Estrella et al., 2017). In order to clarify the difference between data and warrant, it can be said that the former obeys the explicit manifestation of the ideas, whereas the latter occurs in a more hidden way. The study carried out by the author via the analysis of argumentation is noteworthy, based on the contributions made by students in verbal, written and graphic form to, then, be able to determine, in accordance with the approach, the main characteristics of their learning.

In second place is the SMWS, which is an adjustment of the MWS (Mathematical Working Space), a theoretical and methodological approach, which supports the analytical aspects, which should be taken into consideration when studying students' performance when carrying out a mathematical activity within certain environmental conditions, which seek to facilitate the work being carried out (Flores-González & Montoya-Delgadillo, 2016; Kuzniak & Richard, 2014). This approach encompasses the comprehension of the personal or institutional meaning within the study of a mathematical object in a pragmatic way via the completion of activities by students or in a hypothetical way when applying the planning of topics and tasks carried out by teachers. In this way, both epistemological and cognitive aspects are considered in the development of the mathematical object (Vidal-Szabó et al., 2020). A fundamental aspect of the SMWS is the context, which provides meaning to the interpretation of the variability of the data. Thus, from the SMWS focused on the exploratory analysis of data, arises a new determination of the qualities of *semiotic* origin, which confronts the semantic and syntactic aspects of the data in context, from their meaning and representation, making it possible to analyze the characteristic features of the reading and writing of the meaning of the different statistical elements presented; *instrumental*, which expresses the carrying out of activities, which require the handling of manipulative materials, which can be considered as tools for statistical production in data representation tasks by way of compendiums or synthesis; and *discursive*, which bases its analysis on the search for explanations grounded in the properties and definitions of statistics. The capacity for statistical reasoning is demonstrated via the evidence in the data when taking decisions in the specific context and the theoretical components of statistics are used in coherence with the contextual knowledge provided by the data (Vidal-Szabó et al., 2020).

Thirdly, Díaz-Levicoy et al. (2017a) present the theoretical elements of SOLO taxonomy of Biggs and Collis (1982), which defines a way to specify the students' category of cognitive development in accordance with their interconnection with the situations proposed in the classroom. Here, five levels of complexity of knowledge can be differentiated: *pre-structural* in which the student takes part in the task but does not pass beyond simple aspects; *unistructural* in which the student focuses on a single component of the information; *multistructural* in which the student manages to absorb a greater amount of information albeit in an isolated manner; *relational* in which he/she manages to integrate several characteristics correctly; and *extended summary* in which a higher level of cognitive development is generated incorporating new characteristics. Thus, it is explained that the levels of learning range from handling data fluently and accurately to referring to more advanced relationships, of deep knowledge and with other contents and contexts. These are grounded in the type of response given by the student in the teaching process (Biggs & Collis, 1982).

**Table 1.** Keywords

| Block | Keywords |
|---|---|
| Block 1: Statistics (and data) | Statistics, statistical literacy, data analysis, & data literacy |
| Block 2: Teaching and learning | Statistical knowledge, statistical learning, statistical education, & statistical-research cycle |
| Block 3: Levels | Primary education & elementary education |

Finally, there is the CGRL approach (Curcio, 1987; Friel et al., 2001), which presents four levels of graph reading describing students' advancement from the basic to the complex. *Level 1* refers to "reading the data", which is thus identified as the student only requires a literal reading of the data presented in the representations and does not need to carry out any kind of interpretation of the data, for example, reading the frequency, the title, the variable represented, etc. (Díaz-Levicoy et al., 2020). *Level 2* corresponds to "reading between the data", which requires making comparisons and/or simple mathematical operations in order to find intermediate values and, thus, reveal clues about the trend of the data to generate certain interpretations of the data contained in tables and/or graphs (Díaz-Levicoy et al., 2019b). *Level 3* refers to "reading beyond the data", with the student having to interpret, carry out calculations and predict values, which cannot be found directly in the representation, by way of estimation, extrapolation and interpolation of the data (Díaz-Levicoy et al., 2017a). *Level 4* deals with "reading behind the data", which is the last step in the evolution of the reading of graphs and/or tables in which the student investigates the procedure used to gather the data and create the representations, interpret and critically analyze the quality of the data and relate mathematical knowledge and the context in order to take up a clear position towards the affirmations generated in the problem studied (Díaz-Levicoy et al., 2020).

## METHOD

In order to identify studies on the statistical knowledge of primary schoolchildren, a search is made in JCR, Scopus, SciELO, Latindex, Dialnet, JSTOR, Google Scholar, and DOAJ, using keywords that are directly related to the objectives of the study and are helpful as a starting point for preliminary exploration. Initially, the publications, which contain the keywords were considered in accordance with the organization presented in **Table 1**.

In accordance with the initial Boolean combinations, 125 articles were found with a direct link to the keywords. Subsequently, the titles and abstracts were reviewed with the aim of selecting those publications, which involve the analysis of statistical and probability knowledge among primary schoolchildren aged between six and 12 years old. The decision was taken to exclude from the analysis any articles on the topics of teacher training, early years learning, and secondary and university education, because the study focuses specifically on the statistics knowledge of primary education students. A review of the bibliography of the articles identified was also carried out via keywords, with the aim of selecting those with appropriate study aspects for their possible review.

## RESULTS

Initially, an overall analysis of the studies considered within the review is presented. Later, an analysis is made of the tools and results obtained from the different studies.

### Overall Results

As a result of the selection process, a total of eighteen studies were taken into consideration for analysis, highlighting basic information such as: title, author, analytical approach, tools employed, and statistical object studied. **Table 2** shows a synthesis in accordance with the analytical approach employed by the researchers. Based on the synthesis of **Table 2**, a quantitative description of the articles is made in which the following components are taken into consideration: statistical object, analytical approach of learning, and tools employed for data collection.

**Table 2.** Summary of the articles analyzed

| Author(s) | Title of article/thesis | Analytical approach employed | Tools employed | Statistical object studied |
|---|---|---|---|---|
| Aoyama and Stephens (2003) | Graph interpretation aspects of statistical literacy: A japanese perspective | CGRL-SOLO | Questionnaire & interviews | Reading & interpretation of graphs |
| Arteaga et al. (2021) | Primary school students' reading levels of line graphs | CGRL | Questionnaire | Reading & interpretation of graphs |
| Díaz-Levicoy et al. (2017a) | Lectura de pictogramas por estudiantes Chilenos de educación primaria [Reading pictograms by Chilean primary school students] | CGRL | Questionnaire | Reading & interpretation of pictograms |
| Díaz-Levicoy et al. (2017b) | Niveles de estudiantes Chilenos de primaria en lectura de pictogramas [Levels of Chilean primary school students in reading pictograms] | CGRL | Questionnaire | Reading & interpretation of pictograms |
| Díaz-Levicoy et al. (2019a) | Construcción de gráficos de barras por estudiantes Chilenos de educación primaria [Construction of bar graphs by Chilean primary school students] | | Questionnaire | Construction of bar diagrams |

**Table 2 (Continued).** Summary of the articles analyzed

| Author(s) | Title of article/thesis | Analytical approach employed | Tools employed | Statistical object studied |
|---|---|---|---|---|
| Díaz-Levicoy et al. (2019b) | Extrapolación de valores en un gráfico. Un estudio con escolares Chilenos [Extrapolating values in a graph. A study with Chilean school children] | CGRL | Questionnaire | Interpretation of bar diagrams |
| Díaz-Levicoy et al. (2020) | Conocimiento sobre tablas estadísticas por estudiantes Chilenos de tercer año de educación primaria [Knowledge of statistical tables by Chilean students third year of primary education] | CGRL | Questionnaire | Reading, creation, & interpretation of statistical tables |
| Díaz-Levicoy et al. (2021) | Lectura de un diagrama de puntos por estudiantes Chilenos de educación primaria [Reading a dot plot by Chilean primary education students] | CGRL | Questionnaire | Reading & interpretation of graphs |
| Dolores and Cuevas (2007) | Lectura e interpretación de gráficas socialmente compartidas [Reading and interpretation of socially shared graphs] | | Audio-visual recordings & interviews | Reading of graphs |
| English (2010) | Young children's early modelling with data | | Interview & questionnaire | Organization & representation |
| Estrella et al. (2017) | Argumentaciones de estudiantes de primaria sobre representaciones externas de datos: Componentes lógicas, numéricas y geométricas [Primary school students' arguments regarding external data representations: Logical, numerical, and geometric components] | TA | Interview & questionnaire | Organization, representation, & interpretation of data |
| Evangelista (2013) | Atividades de interpretação de gráficos de barras e linhas: O que sabem os alunos do 5° ano? [Bar and line graphs interpretation activities: What do fifth grade students know?] | | Questionnaire | Interpretation of bar & line graphs |
| Fielding-Wells (2010) | Linking problems, conclusions and evidence: Primary students' early experiences of planning statistical investigations | | Questionnaire | Collection, organization, representation and interpretation of data |
| Frischemeier (2019) | Primary school students' reasoning when comparing groups using modal clumps, medians, and hatplots | | Questionnaire | Reading & interpretation of pictograms |
| Kawakami and Aoyama (2018) | Ingenuity and challenges to incorporate statistical-inquiry process into statistics lessons in primary schools: The case of Japan | | Questionnaire | Collection, organization, representation, & interpretation of data |
| Ruiz (2015) | Un estudio de caso sobre errores y dificultades observadas en la elaboración de algunas gráficas estadísticas [A case study about errors and difficulties encountered in the development of some graphical statistics] | CGRL | Questionnaire | Interpretation of qualitative & quantitative data in statistical graphs |
| Santibañez and Vásquez (2021) | Construcción de tablas y gráficos estadísticos: Análisis de una experiencia en contexto de pandemia [Construction of statistical tables and graphs by students of third year of basic education: Analysis of an experience in the context of pandemic] | SMWS | Audio-visual recordings & interviews | Collection, organization, representation, & interpretation of data |
| Vidal-Szabó et al. (2020) | Análisis cualitativo de un aprendizaje estadístico temprano con la mirada de los espacios de trabajo matemático orientado por el ciclo investigativo [Qualitative analysis of early statistical learning with the perspective of mathematical working spaces oriented by the research cycle] | SMWS | Audio-visual recordings and interviews | Organization, representation, & interpretation of data |

### *Statistical object*

**Figure 1** shows that, in the studies identified, the analysis of statistical knowledge on issues relating to the production and reading of graphs is predominant, as eleven out of eighteen articles (61.11%) deal with this matter. As far as the organization of data is concerned, six studies were identified within the total number of articles in this review (33.33%). Last of all, as regards the reading, production and interpretation of statistical tables, only one study was identified (5.56%).

### *Analytical approach*

**Figure 2** highlights that a notable number of investigations included in this analysis (38.89%) do not use a specific approach. Nevertheless, within the studies that show a defined approach, there are seven studies that are examined based on the *CGRL* (38.89%), with the *SMWS* two studies are analyzed (11.11%), whereas each of the other approaches are employed with a low percentage (5.56%).

**Figure 3** shows, in a detailed and clearly-articulated manner, the relationship between the analytical approach of the research and its respective objects of statistical knowledge.

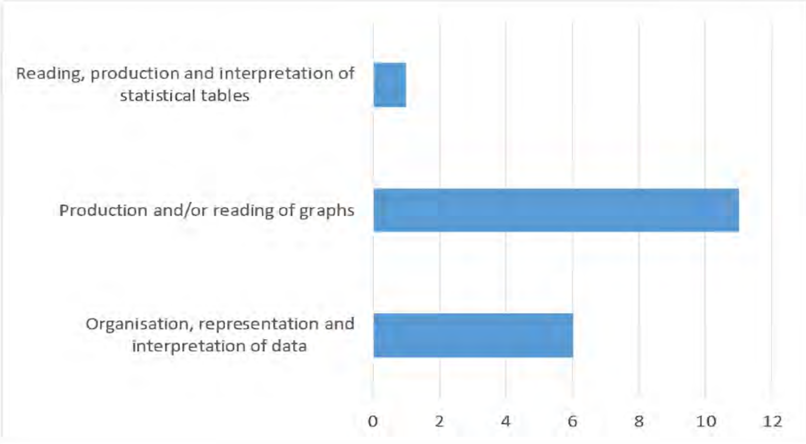


**Figure 1.** Statistical object (Source: Authors' own elaboration)

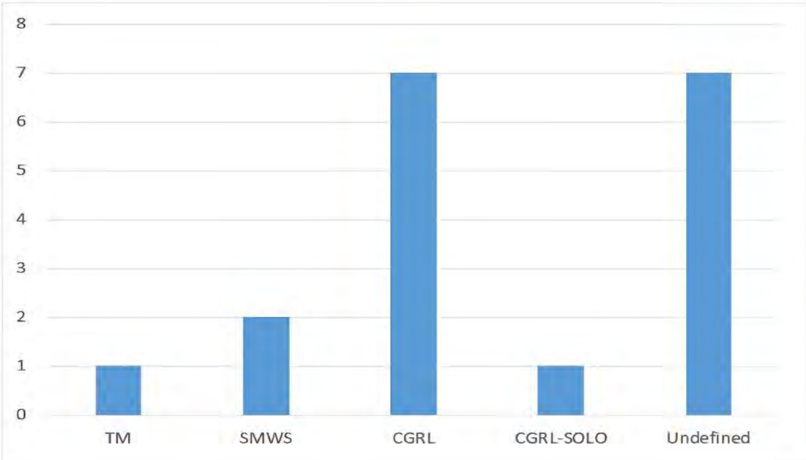


**Figure 2.** Analytical approaches of the research (Source: Authors' own elaboration)

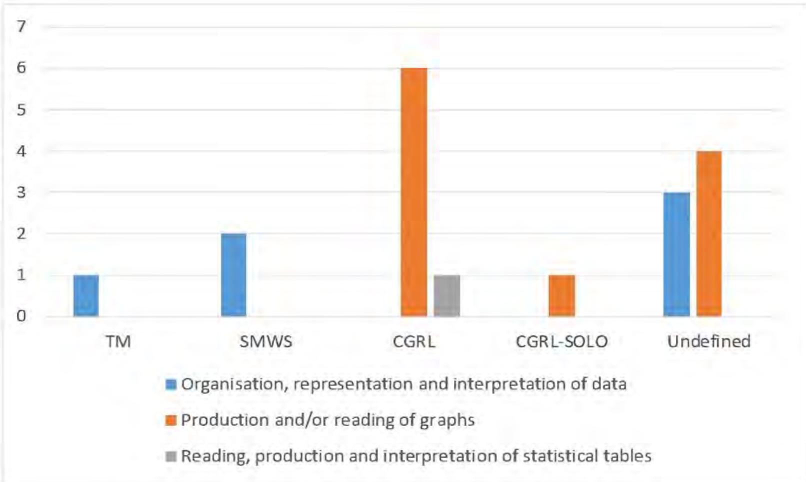


**Figure 3.** Articles represented in accordance with their analytical approach and their object of statistical knowledge (Source: Authors' own elaboration)

Based on the above information, it can be determined that researchers have been attracted by the study of statistical learning with a preference for an approach towards the production and reading of graphs for which the majority of the studies employ the *CGRL* and other non-defined approaches for their analysis (33.33%).

### *Data collection tools employed in the research*

**Figure 4** shows the classification of the research according to the tool employed for the generation of data. Questionnaires are the most frequently-used tool in research processes on statistical learning among primary education students (n=12), followed by mixed tools (33%).

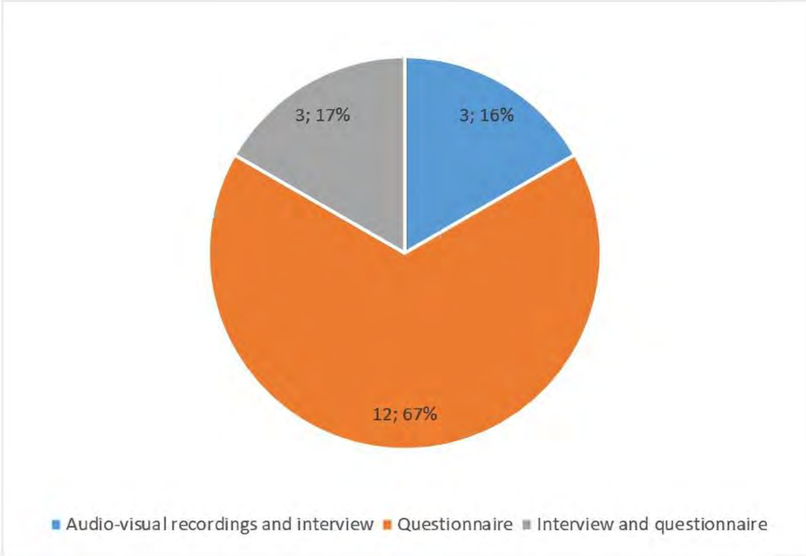


**Figure 4.** Data collection tools (Source: Authors' own elaboration)

Examples of this component of the research include questionnaires with tasks in which students choose the option, which they consider correct whilst carrying out interviews to explore the reasons why such responses are expressed. Furthermore, situations are proposed in which data is organized, making it possible to analyze the representations created by students in order to then carry out individual semi-structured interviews, including complimentary questions and video recordings of the activities carried out with the intention of recording their responses and analyzing the verbalization of their ideas. Adaptations of the items included in primary education textbooks are also presented related to the reading of pictograms and the determination of missing values in frequency tables. Audio and video recordings are also used as means for collecting the data, which students generate when expressing their interpretations of the graphs proposed in their research.

### The Perspectives of the Studies Carried Out on the Statistical Knowledge of Primary Schoolchildren

This review has revealed that there are few studies, which analyze the statistical knowledge of primary schoolchildren, although, in the last decade or so, certain researchers have developed an interest in exploring, describing and characterizing the different levels of comprehension of students during the statistical cycle. As previously mentioned, in this section the main data found based on the different approaches employed are described, with the aim of providing researchers with a clearer view of the processes, which are being developed.

Employing the CGRL approach, Aoyama and Stephens (2003) carried out a study with seventeen fifth-year students, the aim of which was to analyze the students' basic capacity to read graphic information by way of a questionnaire with a multiple-choice format in which they were asked to express their reasons for selecting their answers. In relation to this, and with the interviews carried out with the students at the time of studying their responses, the authors employed the CGRL to analyze the reading of (bar and line) graphs and determined that 83% of the students achieved level 3 ("reading beyond the data"), as it was evident that they had carried out calculations and comparisons in order to select the correct answer. It may be that the students' performance was influenced by the technical nature of the two topics represented (environmental contexts). In the analysis of the students' explanations, the SOLO taxonomy was applied to categorize the responses. It was found that the majority of them (almost 83%) are located in the *pre-structural* and *unistructural* levels, as they provided a correct explanation in accordance with the reading made of the graphs. However, the students were involved in the *relational* level to a much lesser extent (close to 17%) by way of more elaborate presentations in accordance with the information provided in the tasks. None of the students achieved the level of *extended summary*, which demonstrates that primary schoolchildren present difficulties in understanding and connecting the information they are presented with, although the study shows a clear improvement in their capacity to evaluate statistical information in year 5. This may be caused by cognitive development, some kind of training in critical thinking or experience with statistical information from other academic or real-world contexts. These are important issues for future research, the challenge for which is to confirm the existence of complementary levels related with the responses of *extended summary* in primary education with the aim of describing planning in statistical education in more detail.

The analysis made by Ruiz (2015) is aimed at the identification of certain errors and difficulties in the production of statistical graphs presented by 31 fifth-year students via the design and implementation of three activities centered on the determination of data from a graph provided (interpretation of qualitative and quantitative data), the construction of a graph representing the data provided and the identification of patterns of change in a set of data. In this process, it was found that the students experienced difficulties in elaborating statistical graphs in terms of handling an appropriate scale for the representation of the data, thereby hindering interpretation. In addition, it was determined that the students included in the research were, on the whole, located in the level of "reading the data" proposed by Curcio (1987), as they perceived the data in a general way, as shown by their textual reading of the graph without attributing a logical meaning to the information presented. It was also observed that their graphs lacked appropriate labelling, which would contribute to their comprehension.

It should be highlighted that some students succeeded in passing to the level of "reading between the data" (Curcio, 1987), as they demonstrated that they were able to compare quantities and carry out mathematical calculations in order to interpret and articulate the data contained in the graph. With regard to what is mentioned above, the author demonstrates that it is of prime importance to resolve these shortcomings via the correct production of graphs and the continuous evaluation of the statistical learning process. Still under the same approach, Díaz-Levicoy et al. (2017b) focus on determining students' reading levels and the strategies they employ in order to interpret pictograms (in which each icon represents a certain number of units) and represent them in tables. Tasks were set for 69 students of 11 and 12 years old, taken from Chilean primary education textbooks. The authors found that the vast majority of the students achieved level 2, as more than 90% of them were able to read such pictograms correctly, represent them in tables and determine their values appropriately. They also found that students were able to achieve level 3 ("reading beyond the data") according to Curcio's (1987) categorization, albeit at a low percentage (almost 10%), as, when asked to refute or defend a claim, a lack of critical reading and argumentation skills became evident. This confirms that the comprehension of graphs does not only depend on the capacity for visualization, but also on the reasoning skills, which students may possess.

In another study with 380 students aged 11-12, Díaz-Levicoy et al. (2017) demonstrated evidence of level 4 ("reading beyond the data") (Curcio, 1987), as they considered that some students, in addition to carrying out the required calculations to determine frequencies and correctly interpret the pictogram, succeeded in making a critical reading as they were able to provide a coherent argument in order to support the statements or objections generated in the different tasks. In order to complement the previous studies, Díaz-Levicoy et al. (2019c) analyzed a task set for 105 students of 11 and 12 years old, which consisted of elaborating a bar graph using a squared template based on a list of data, whilst also calculating frequencies. In this analysis, the graphs were classified as basically correct, partially correct and incorrect. Errors were identified relating to the proportion of the scale of the graph, the labelling of the axes, the representation of the numerical values and the construction and spacing of the bars. 6.67% of the students who participated in the study created graphs characterized as "basically correct", while the majority of the graphs (50.5%) were classified as "partially correct". Therefore, the conclusion can be drawn that these students knew how to produce a bar graph, although they made the afore-mentioned mistakes. This indicates a significant level of difficulty of the task for these students. Consequently, it is of prime importance to reinforce concepts such as arithmetic and geometrical proportionality, the distribution and calculation of the frequency of data, the order of numbers and their representation in the number line. Along the same line of research, Díaz-Levicoy et al. (2019a) set 380 students aged between 11 and 13 years old an additional task to those of their previous studies, which consisted of extrapolating values from a bar graph from which they determined that very few students (8.4%) reached Curcio's (1987) third reading level in this activity. They concluded that this level is too high for students of these ages. However, it may occur, for example, that a student responds to the question related to this level but does not make a correct reading of the graph, thus corresponding to level 1. On the other hand, Díaz-Levicoy et al. (2020) carried out a study with 79 students (7 to 10 years old) with the aim of analyzing their knowledge when working on activities concerning statistical tables by way of tasks extracted from textbooks, which address, for example, the identification of the title, the reading of data, the calculation of values and the completion of the table. In accordance with Curcio's (1987) reading levels, the results show that 68.6% of the students were better able to complete the tasks relating to reading and completing a table and showed better results in tasks requiring a literal reading (level 1). 55.5% of the students responded correctly in the activities in which they were required to carry out calculations and comparisons (level 2). These results prove that students handle basic components when operating with statistical tables.

Afterwards, Díaz-Levicoy et al. (2021) carry out a study involving 380 students (11 and 12 years old), in order to evaluate critical reading competence in a point graph, according to the level of "reading behind the data" (Curcio, 1987; Friel et al., 2001; Shaughnessy et al., 1996). In the task, students are asked to read the frequency that appears in two graphs (read the data), then they are asked to determine the frequency according to the comparison of specific data (read within the data) and, from these readings, they must indicate if they agree with a statement and then justify their answer. The results reveal that 45% of the students fail to answer the task or the reading is incorrect (level 0) and 6.6% of the students provide a correct literal reading of the graphs; Likewise, almost half of the students (47.4%) manage to "read within the data" and manage to perform calculations and comparisons with them, although only 1.1% of the participants in the study are at the level 4 (read behind the data), who manage to answer the initial questions correctly through valid arguments that arise from their correct analysis of the diagrams shown. In this regard, the authors deduce that the critical reading of data where students must interpret a graph to obtain information, discuss assertions and make decisions are complex tasks; therefore, they suggest reinforcing this type of activity in the classroom and conducting more research in this field. Also, Arteaga et al. (2021) carry out a study with 380 students (11 and 12 years old), which aims to describe errors and reading levels (Curcio, 1987; Friel et al., 2001; Shaughnessy et al., 1996) that students achieve in a primary education statistics course, through the approach of two tasks: reading a line graph, where about 95% of students manage to be at the level "read the data" when they are asked about the title of the graph, frequency of some data and inverse reading of data, but they are located at 38% when they must identify the variables that intervene; the second consists of choosing the appropriate line graph for a proposed situation, which does not happen with 29.2% of the students, who cannot read values on the graph. It should be noted that 53.4% of the students, who are located at the "read within the data" level, carry out a literal reading of the information provided by the graph, identify trends and look for relationships between the data), but they cannot make critical reading. Only 11.6% are at the "read behind the data" level, although with various inaccuracies in their arguments. Given these results, the authors conclude that there is a difficulty in reaching this last level, since a limited number of questions are presented in textbooks that develop competence in critical reading of graphics, an aspect that is quite relevant at the time to acquire an appropriate statistical culture.

As far as the TA is concerned, Estrella et al. (2017) explore and characterize the structure of arguments regarding the external representation built by three children of eight and nine years old, via organizational tasks with manipulative material (colored tokens) and the implementation of a semi-structured interview including additional tasks associated with the logical, numerical and geometrical components of the representation created by the students. In this analysis, the elements of TM (Toulmin, 1958) are presented and justified by way of the arguments provided by the students when asked how they had organized the data and about certain components of the representation of the data, which they had built. This research made it possible for the students to use grounds for classification, which are defined with the *variable* (logical pattern) of partition (cardinality, which is determined with the *frequency* [mathematical pattern]) and of relative position between lines, which are detailed with the *linear-base and graphic linearity* (geometrical pattern), in order to justify their productions. In accordance with these results, it was concluded that justifications appear between the ages of 8 and 9, given that the fluent verbalization of explanations appears in these cases.

Vidal-Szabó et al. (2020) employ the SMWS approach to characterize a statistics class given to 17 students of nine and 10 years old in the fourth year of primary education. In this study, an exploratory data analysis was carried out within the data problem-solving cycle known as PPDAC (problem, plan, data, analysis, and conclusion) (Wild & Pfannkuch, 1999) and the students' performance was studied. For this purpose, a qualitative analysis was carried out of statistical learning from the mathematical work of a task based on the investigative cycle. An overall perspective was presented, which describes and characterizes the teaching and learning process according to the stages of the investigative cycle based on the ideal and personal components, which explore: the semiotic genesis represented by the decoding of the data and their recoding (words or drawings); the instrumental genesis shown in the construction of new representations in the form of tables, diagrams, etc., which enable the student to make his/her own calculations; and the discursive genesis shown in the design of appropriate classification and counting strategies to identify the different statistical elements such as categories and frequencies. Based on these aspects, the student demonstrates the necessary knowledge to draw the relevant conclusions of the problem and, thus, obtain additional related results, for example, with the measurements of central tendency in which arithmetic judgement and graphic visualization are employed. Last of all, the authors demonstrated that students were able to link procedures, which are used to think about, discover, argue and communicate statistical ideas in their learning in accordance with the knowledge acquired in the statistical cycle, the authenticity of the problems posed, prior experience and the context proposed for its development. Following the same approach, Santibáñez and Vásquez (2021) investigate the way in which 13 third-grade students of primary education (eight and nine years old) develop their statistical learning process, focused on the execution of activities related to their diet at home, given the situation of confinement caused by the measures adopted to contain COVID-19. The design of activities is contemplated within the PPDAC statistical research cycle (Wild & Pfannkuch, 1999) and the performance is qualitatively analyzed within the SMWS to identify the genesis that are activated in the resolution of the proposed task and its vertical articulation: the semiotic genesis occurs when students identify the food they eat as data and codify them according to the types and quantity of food using symbols and written records, respectively, both graphically and in tabular form; next, the instrumental genesis is evidenced in the use of artifacts for the construction of graphs, tables, symbols (pencil, notebook, etc.); Finally, the discursive genesis is activated when the students classify the food as data, find their respective frequency, build tables and graphs, which allows them to reason, argue, justify and finally answer the questions that the teacher asks them, according to the context of the situation. According to the above, it can be affirmed that students put into practice various visualization, instrumentation and testing processes, by applying their knowledge to achieve classification, counting, elaboration of ordered lists, counting tables and frequencies, and construction of graphs, to subsequently make their appropriate interpretations. To sum up, the authors observe how students manifest the interrelation between semiotic, instrumental and discursive genesis, through the use of arguments and the communication of statistical ideas, which arise as a result of reasoning originating from the analysis of their own representations.

Lastly, this review has detected other studies in which the researchers did not employ a specific theoretical point of reference. Dolores and Cuevas (2007) carried out a descriptive-qualitative study with the aim of exploring the readings and interpretations made by 11-13-year-old primary schoolchildren relating to graphs published in out-of-school contexts, which were disseminated via the mass media. The authors found that the students made comparisons of a qualitative nature by way of explicit manifestations resulting from the question "how does it change?" and expressions such as "ascends", descends", "goes up", goes down", and "stays the same" via visual estimations. The students focused their interest on the reading of the maximum and minimum values and on the shape of the graphs more than on their contents. Therefore, they do not show evidence of carrying out mathematical calculations in order to discover how much the frequency of the variable increased or decreased. Fielding-Wells (2010), on the other hand, focused her study on 24 students of nine and 10 years old with the aim of determining relevant information regarding the observation and analysis of the initial phases of statistical learning within the PPDAC investigative cycle (Wild & Pfannkuch, 1999). Tasks were set, which involved the students in the administration of their own research. The authors concluded that the development of statistical reasoning is not an easy task for students as, in spite of having diverse paradigms aimed at the use of research, they continue to present difficulties when visualizing the statistical research technique. This is shown, for example, in the lack of clarity when generating a question, an impediment when predicting, which tests could be used for undertaking the process, drawbacks for defining the type of data to be summarized according to their source, difficulties in making inferences in accordance with the context and problems when determining the answer to the question posed. Therefore, the author specifies the need to explicitly teach and examine in depth the planning process of statistical research, as she has detected that it is insufficient for their understanding to permit students to develop their research projects autonomously.

English (2010) studied first-year primary schoolchildren (six years old) with the aim of analyzing their statistical reasoning in data modelling activities via tasks in which the most significant characteristics of a group of elements can be identified in order to, then, progress towards the organization, structuring, visualization and representation of data. The author explains that the students identify a multitude of attributes and classify the elements according to those attributes in order to guide their attention to these characteristics, which, often, are not so immediately evident. However, they discard qualities, which they do not consider to be representative within their analysis, thus leading to the production of suitable models of organization, structuring and representation of data in a relatively natural manner. When graphing, they value the importance of the fact that their images are built in rows or columns with the aim of facilitating counting and, in several cases, that they contain the names of the attributes with their respective frequency. Furthermore, Evangelista (2013) carried out a quantitative study with 60 fifth-year students (10 and 11 years old) in which the participants were required to respond to a test, which explored their reading and interpretation skills of information represented on different types of graphs. By analyzing the students' performance, it was found that 60% of them responded correctly to questions relating to the comprehension of the data represented in bar graphs, whereas only 43% of them did so with questions referring to line graphs. According to the author, this is due to their lack of familiarity with the latter. There were also correct answers at significant percentages (>60% of the students) when finding a frequency or category, possibly due to the explicitness of the values. Also, in questions relating to finding the mode, only 51% of the students responded correctly as there was an extremely common error consisting of locating the highest frequency without distinguishing the corresponding category (in the case of double-entry graphs). Furthermore, 59% of the students presented difficulties in questions in which it was necessary to join or compare values of the frequencies due to the fact that they did not understand the need to carry out calculations with the corresponding numbers. Frischemeier (2019) focused his mixed research on 11 students of nine and 10 years old from the fourth-year of primary education. They were required to analyze a point chart presenting the comparison between two individual groups and to find the qualities of the modal groups. More than 90% of the students demonstrated the capacity to define modal groups with the aim of determining trends in the data presented, albeit with erroneous explanations, which only focus on individual characteristics. The author recommends carrying out subsequent studies in contexts of larger and different-sized group comparisons. Lastly, Kawakami and Aoyama (2018) present two classes, which incorporate the collection of data from everyday life and from an experiment, via the use of the PPDAC investigative statistical cycle (Wild & Pfannkuch, 1999) with primary students aged between nine and 11 years old. The authors determined that the students became aware of the variation and of the source of the collected data due to their realistic origin, considering the importance of the appropriate orientation of the process of statistical research in accordance with the students' characteristics, their capacities and limitations. Therefore, the use of the experimental activity becomes one way to incorporate statistical literacy via simple research approaches in primary education, in such a way that it is possible to strengthen critical thinking at these ages.

Thus, there is a constant need to carry out more research regarding the knowledge and interpretation of data with primary education students, contemplating the situations of their daily lives, which may occur both inside and outside school and how they resolve statistical reasoning activities in which they ask their own questions, discuss how to address them, generate their own data and determine how to organize, structure, represent and communicate what they have found.

## FINAL CONSIDERATIONS

In this article, a review has been carried out of fifteen articles published between 2003 and 2021, which has made it possible to characterize the different approaches employed to study the development of statistical learning and the analysis of the most relevant elements of the statistical knowledge of primary education students.

This study has noted the growing interest of researchers in the study of statistical knowledge in primary education. They seek to analyze how students organize the data, which they obtain from a problem presented from real contexts with their active participation or from tasks set in textbooks. In this regard, it is of interest to highlight the use of the PPDAC cycle (Wild & Pfannkuch, 1999), which guides students towards carrying out their own research based on the handling of real data, promoting an environment of statistical literacy from the exploration of information provided by the context with data linked to the students' own reality (Vidal-Szabó et al., 2020). They also analyze the students' construction of arguments as a fundamental part of the comprehension of different statistical concepts. In this sense, the authors consider it to be of vital importance to reflect on such knowledge when designing and constructing curricular materials and guidelines for statistical literacy (Estrella et al., 2017).

However, the approach most frequently used to analyze student learning is CGRL (Curcio, 1987; Friel et al., 2001). This is probably due to the fact that researchers address the topics related with the production and interpretation of statistical graphs and this approach fits this approach well. Furthermore, as far as the analysis of students' knowledge is concerned, researchers' attraction for the analysis of the production and interpretation of statistical graphs stands out, possibly due to the great influence, which these representations generate in present-day society in different fields of knowledge. In this regard, Aoyama and Stephens (2003) consider that part of students' performance can be attributed to the contact, which they have with experiences in which statistical information is presented in other areas of education or in everyday life, which contribute towards clarifying the contexts of the activities carried out by students.

In conclusion, this review provides an initial overview of the research processes on the statistical knowledge of primary schoolchildren, as one of the main limitations has been not considering studies from other sources such as books from prestigious publishers, etc. In the future, a more systematic review will be needed to provide a more complete panorama and thus be able to guide researchers in developing new study perspectives on the levels of understanding of primary schoolchildren' statistical knowledge. Additionally, according to Rodrigues and da Ponte (2022) it will also be necessary analyze how the teacher's didactic knowledge to teach statistics influences the students' statistical knowledge.

**Author contributions:** All authors have sufficiently contributed to the study and agreed with the results and conclusions.

**Funding:** No funding source is reported for this study.

**Ethical statement:** Authors stated that the study did not require ethics committee approval, as it discussed approaches to analyse the statistical knowledge of primary school students. As no studies were conducted with children, no ethics committee approval was required.

**Declaration of interest:** No conflict of interest is declared by authors.

**Data sharing statement:** Data supporting the findings and conclusions are available upon request from the corresponding author.

## REFERENCES

Alsina, Á., Muñiz-Rodríguez, L., Rodríguez-Muñiz, L., García-Alonso, I., Vásquez, C., & López-Serentill, P. (2023). Alfabetizando estadísticamente a ninos de 7-8 años a partir de contextos relevantes [Statistical literacy based on relevant contexts with 7-8-years old children]. *Revista Complutense de Educación [Complutense Journal of Education], 34*(1), 95-108. https://doi.org/10.5209/rced.77186

Aoyama, K., & Stephens, M. (2003). Graph interpretation aspects of statistical literacy: A Japanese perspective. *Mathematics Education Research Journal, 15*(3), 207-225. https://doi.org/10.1007/BF03217380

Arteaga, P., Díaz-Levicoy, D., & Batanero, C. (2021). Primary school students' reading levels of line graphs. *Statistics Education Research Journal, 20*(2), 1-11. https://doi.org/10.52041/serj.v20i2.339

Batanero, C., Arteaga, P., & Gea, M. (2011). El currículo de estadística: Reflexiones desde una perspectiva internacional [Statistical curriculum: Reflections from an international perspective]. *UNO. Revista de Didáctica de las Matemáticas* [*UNO. Mathematics Didactics Magazine*]*, 59,* 9-17.

Batanero, C., Díaz, C., Contreras, J. M., & Roa, R. (2013). El sentido estadístico y su desarrollo [The statistical sense and its development]. *Números. Revista de Didáctica de las Matemáticas* [*Numbers. Mathematics Didactics Magazine*]*, 83,* 7-18.

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382-393. https://doi.org/10.2307/749086

Díaz-Levicoy, D., Arteaga, P., & Batanero, C. (2017a). Lectura de pictogramas por estudiantes chilenos de educación primaria [Reading of pictograms by Chilean primary school students]. In J. M. Muñoz-Escolano, A. Arnal-Bailera, P. Beltrán-Pellicer, M. L. Callejo, & J. Carrillo (Eds.), *Investigación en educación matemática XXI* [*Research in mathematics education XXI*] (pp. 217-226). SEIEM.

Díaz-Levicoy, D., Arteaga, P., & Batanero, C. (2017b). Niveles de estudiantes Chilenos de primaria en lectura de pictogramas [Levels of Chilean primary school students in reading pictograms]. In *Proceedings of the CERME 10*.

Díaz-Levicoy, D., Batanero, C. & Arteaga, P. (2019a). Construcción de gráficos de barras por estudiantes Chilenos de educación primaria [Construction of bar graphs by Chilean primary school students]. In J. M. Contreras, M. M. Gea, M. M. López-Martín, & E. Molina Portillo (Eds.), *Proceedings of the 3rd Virtual International Congress on Statistical Education*.

Díaz-Levicoy, D., Batanero, C., & Arteaga, P. (2019b). Extrapolación de valores en un gráfico. Un estudio con escolares Chilenos [Extrapolating values in a graph. A study with Chilean school children]. In J. M. Marbán, M. Arce, A. Maroto, J. M. Muñoz-Escolano, & Á. Alsina (Eds.), *Investigación en educación matemática XXIII* [Research in mathematics education *XXI*] (pp. 263-272). SEIEM.

Díaz-Levicoy, D., Batanero, C., Arteaga, P., & Gea, M. M. (2019c). Chilean children's reading levels of statistical graphs. *International Electronic Journal of Mathematics Education, 14*(3), 689-700. https://doi.org/10.29333/iejme/5786

Díaz-Levicoy, D., Morales, R., Arteaga, P., & López-Martín, M. (2020). Conocimiento sobre tablas estadísticas por estudiantes Chilenos de tercer año de educación primaria [Knowledge of statistical tables by Chilean students third year of primary education]. *Educación Matemática* [*Mathematics Education*]*, 32*(2), 247-277. https://doi.org/10.24844/EM3202.10

Díaz-Levicoy, D., Batanero, C., Arteaga, P., & Begué, N. (2021). Lectura de un diagrama de puntos por estudiantes Chilenos de educación primaria [Reading a dot plot by Chilean primary education students]. *Revista Conrado* [*Conrad Magazine*]*, 17(78)*, 61-67.

Dolores, C., & Cuevas, I. (2007). Lectura e interpretación de gráficas socialmente compartidas [Reading and interpretation of socially shared graphs]. *Revista Latinoamericana de Investigación en Matemática Educativa* [*Latin American Journal of Research in Educational Mathematics*]*, 10*(1), 69-96.

English, L. D. (2010). Young children's early modelling with data. *Mathematics Education Research Journal, 22*(2), 24-47. https://doi.org/10.1007/BF03217564

Estrella, S., Olfos, R., Morales, S., & Vidal-Szabó, P. (2017). Argumentaciones de estudiantes de primaria sobre representaciones externas de datos: Componentes lógicas, numéricas y geométricas [Primary school students' arguments regarding external data representations: Logical, numerical, and geometric components]. *Revista Latinoamericana de Investigación en Matemática Educativa* [*Latin American Journal of Research in Educational Mathematics*]*, 20*(3), 345-370. https://doi.org/10.12802/relime.17.2034

Evangelista, M. B. (2013). Atividades de interpretação de gráficos de barras e linhas: O que sabem os alunos do 5º ano? [Bar and line graphs interpretation activities: What do fifth grade students know?]. *Probabilidad Condicionada: Revista de Didáctica de la Estadística* [*Conditional Probability: Statistics Didactics Magazine*]*, 1*, 121-128.

Fielding-Wells, J. (2010). Linking problems, conclusions and evidence: Primary students' early experiences of planning statistical investigations. In C. Reading (Ed.), *Data and context in statistics education: Toward s an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics* (pp. 1-6). International Statistical Institute.

Flores-González, M., & Montoya-Delgadillo, E. (2016). Artefacto y espacio de trabajo matemático en la multiplicación de números complejos [Artifacts and mathematical working space in multiplication complex numbers]. *Educación Matemática* [*Mathematics Education*]*, 28*(2), 85-117. https://doi.org/10.24844/EM2802.04

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

Frischemeier, D. (2019). Primary school students' reasoning when comparing groups using modal clumps, medians, and hatplots. *Mathematics Education Research Journal, 31*(4), 485-505. https://doi.org/10.1007/s13394-019-00261-6

Kawakami, T., & Aoyama, K. (2018). Ingenuity and challenges to incorporate statistical-inquiry process into statistics lessons in primary schools: The case of Japan. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the 10th International Conference on Teaching Statistics* (pp. 1-6). International Statistical Institute.

Kuzniak, A., & Richard, P. (2014). Espacios de trabajo matemático: Puntos de vista y perspectivas [Spaces for mathematical work: Viewpoints and perspectives]. *Revista Latinoamericana de Investigación en Matemática Educativa* [*Latin American Journal of Research in Educational Mathematics*]*, 17*(4), 5-15. https://doi.org/10.12802/relime.13.1741a

Muñiz-Rodríguez, L., Rodriguez-Muñiz, L.J., y Alsina, Á. (2020). Deficits in the Statistical and Probabilistic Literacy of Citizens: Effects in a World in Crisis. *Mathematics, 8,* 1872. https://doi.org/10.3390/math8111872

Rodrigues, B. M. B., & da Ponte, J. P. (2022). Teacher education and didactics knowledge to teach statistics: A case study. *European Journal of Science and Mathematics Education, 10*(2), 225-242. https://doi.org/10.30935/scimath/11717

Ruiz López, N. (2015). La enseñanza de la estadística en educación primaria en América Latina [The teaching of statistics in Latin America primary education]. *REICE: Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación* [*REICE: Ibero-American Magazine on Quality, Efficacy and Change in Education*]*, 13*(1), 103-121.

Ruiz, A. (2015). Un estudio de caso sobre errores y dificultades observadas en la elaboración de algunas gráficas estadísticas [A case study about errors and difficulties encountered in the development of some graphical statistics]. *Góndola, Enseñanza y Aprendizaje de las Ciencias* [*Gondola, Science Teaching and Learning*]*, 10*(1), 26-39. https://doi.org/10.14483/jour.gdla.2015.1.a02

Santibáñez, L., & Vásquez, C. (2021). Construcción de tablas y gráficos estadísticos: Análisis de una experiencia en contexto de pandemia [Construction of statistical tables and graphs by students of third year of basic education: Analysis of an experience in the context of pandemic]. *Unión: Revista Iberoamericana de Educación Matemática* [*Union: Ibero-American Journal of Mathematics Education*]*, 17*(63).

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 205-237). Springer. https://doi.org/10.1007/978-94-009-1465-0_8

Toulmin, S. (1958). *The uses of argument.* Cambridge University Press.

Vidal-Szabó, P., Kuzniak, A., Estrella, S., & Montoya, E. (2020). Análisis cualitativo de un aprendizaje estadístico temprano con la mirada de los espacios de trabajo matemático orientado por el ciclo investigativo [Qualitative analysis of early statistical learning with the perspective of mathematical working spaces oriented by the research cycle]. *Educación Matemática* [*Mathematics education*]*, 32*(2), 217-246. https://doi.org/10.24844/EM3202.09

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *Internacional Statistical Review, 67*(3), 223-265. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x