STATISTICAL GRAPHS COMPLEXITY AND READING LEVELS: A STUDY WITH PROSPECTIVE TEACHERS

Pedro ARTEAGA[1], Carmen BATANERO[2], and Gustavo R. CAÑADAS

Statistique et Enseignement, 6(1), 3-23, 2015-04-26

GRAPHIQUES STATISTIQUES COMPLEXITÉ ET NIVEAUX DE LECTURE: UNE ÉTUDE AVEC DES ENSEIGNANTS

ABSTRACT

In this paper we relate the competence of 207 prospective teachers to build statistical graphs with their graph reading level. The graphs produced in their written reports in a statistical project, where the teachers had to compare three pairs of distributions are firstly classified in four different levels, according to their semiotic complexity. Secondly, the participants' interpretations of these graphs are classified according to Curcio's (1989) graph reading levels and thirdly the participant's conclusion about the research question posed in the project are examined. The semiotic complexity levels in the graphs produced by the prospective teachers was, in general, high enough to solve the task proposed; however, the correct conclusion on the question posed in the project was only reached by a minority of prospective teachers, because many of them were unable to read the graphs produced by themselves at the "reading behind the data" level. Higher levels of graph semiotic complexity favoured the teachers' interpretation of the graph, since higher semiotic level graphs also corresponded to higher combined percentage of "reading beyond data" and "reading between data" levels in the participants' written reports. Higher levels of graph semiotic complexity also favoured the reaching of a conclusion about the research question. No differences were observed according the variables compared.

Keywords: *statistical graphs, semiotic complexity, graph interpretation and comprehension*

RÉSUMÉ

Dans cet article, nous rapportons la compétence de 207 futurs enseignants à construire des graphiques statistiques avec leur niveau de lecture graphique. Les graphiques produits dans leurs rapports écrits dans un projet statistique, où les enseignants ont dû comparer trois paires de distributions sont classés en quatre niveaux différents, en fonction de leur complexité sémiotique. Deuxièmement, les interprétations de ces graphiques des participants sont classées selon Curcio (1989) après les niveaux de lecture graphique et, enfin, la conclusion du participant sur la question de recherche posée dans le projet est examinée. Les niveaux de complexité sémiotique dans les graphiques produits par les futurs enseignants est, en général, assez haut pour résoudre la tâche proposée; toutefois, la conclusion correcte sur la question posée dans le projet a été atteint seulement par une minorité d'enseignants, parce que beaucoup d'entre eux étaient incapables de lire les graphiques produits par eux-mêmes au niveau "lire derrière les données". Des niveaux supérieurs de complexité sémiotique graphique ont favorisé l'interprétation du graphique, puisque les graphiques de niveau sémiotique les plus élevés correspond également à un pourcentage supérieur des niveaux «lire au-delà des données» et «lire entre les données« dans les rapports écrits du participants. Des niveaux plus élevés de complexité sémiotique dans le graphique ont également favorisé l'atteinte d'une conclusion sur la question de recherche. Aucune différence n'a été pas observée selon les variables comparées.

Mots-clés : *graphiques statistiques, complexité ; niveaux de lecture, enseignants.*

[1] Universidad de Granada, Campus de Cartuja s/n, cp18071, parteaga@ugr.es
[2] Universidad de Granada, batanero@ugr.es

[4] Universidad de Granada, gcanadas@ugr.es

1 Introduction

Graphical language is essential in organising and analysing data, since it is a tool for *transnumeration,* a basic component in statistical reasoning (Wild & Pfannkuch, 1999) that consist in producing new information not available in the raw data with a change of representation (Arteaga et al., 2011). Being able to build and interpret statistical graphs is also an important part of statistical literacy (Watson, 2006), which is the union of two related competences: (a) Interpreting and critically evaluating statistically based information from a wide range of sources, and (b) formulating and communicating a reasoned opinion about such information (Gal, 2002).

Statistical graphs are included since the first grade of primary education in Spain (MEC, 2006) as well as in the Spanish textbooks for primary education, which introduce activities of reading, building and interpreting the following graphs: bar graphs (grade 1; 6-year olds), attached and stocked bar graphs (grades 2 and 3; 7-8 year-olds), simple and multiple line graphs (grades 3 and 4, 9-10 year-olds), pictographs and dot plots (grade 5, 11 year-olds), pie graphs and population pyramids as well as translations between different graphs (grade 6, 12 year-olds).

The success of such curriculum will depend on the extent to which school teachers are competent in graphical language; however, since statistics have only recently became an important part of the primary school mathematics curriculum, prospective primary school teachers often enter the Faculties of Education with a very limited statistical competence. The organization of didactical activities that serve to teach statistics to these teachers, while at the same time helping to bridge conceptualization and pedagogy (as suggested by Ball, 2000) should be, moreover, based on a previous assessment of teachers' knowledge.

In a previous paper (Batanero, Arteaga & Ruiz, 2010) we analysed the graphs produced by 93 prospective primary school teachers when comparing two distributions and defined a level of semiotic complexity in these graphs, in order to provide an indicator of the participants' competence in building statistical graphs. Using a more complete version of the same task (in which prospective teachers should compare three pairs of distributions) and a bigger sample of prospective teachers, we now relate the semiotic complexity of the graphs produced by 207 prospective teachers with the reading level they achieve when interpreting these graphs. We use Curcio's (1989) levels of graph understanding to classify the participants' interpretation of their graphs. A summarized version of this paper was published in Arteaga and Batanero (2011).

Below we describe the study background, method and results and finish with some implications for teacher education.

2 Background

2.1 Understanding statistical graphs

Results from previous research suggest that the competence related to building statistical graphs is not reached in compulsory education, since students make errors, such as using too large or small data range in the scale (Li & Shen, 1992), not including in the scale the graph origin (Hadjidemetriou & Williams, 2002), not including the axis labels or the graph title

(McClain, 1999), skipping data with zero values (Konold & Higgins, 2003) or wrong construction of specific graphs (Pereira Mendoza & Mellor, 1990; Lee & Meletiou, 2003; Bakker, Biehler & Konold, 2004).

Other authors analysed different possible levels in graph reading and suggested that some students also may have difficulty in interpreting statistical graphs. For this particular research we will use the following three levels defined by Curcio (1989) (see also Friel, Curcio & Bright, 2001):

(a) *Reading the data.* This is the first elementary level, where the student only focuses on extracting data from a graph (i.e., locating or translating information). This level describes the capacity of a student who is only able to read literally the direct factual information on the graph, or to answer simple explicit questions for which the obvious answer is right there on the graph. An example of this level is the student who is only able to read the graph title or single data values, but is unable to answer more difficult questions.

(b) *Reading between the data* is an intermediate level characterized by integrating, interpreting, interpolating and finding relationships in the data; this level requires that the student reading the graph centre his or her attention on two or more data points on the graph, and be able to apply operations to the data; often for comparison purposes and for finding relationships in the data presented in a graph. An example is the student who is able to determine visually the mode in a histogram or a bar graph.

(c) *Reading beyond the data* is an advanced level that requires extrapolating, inferring or predicting from the data and analysing the relationships implicit in a graph to answer questions related to the data tendencies. For example, a student who is given a scatter plot and is able to determine the type of relationship (linear or not, direct or inverse) between the two variables represented.

In our research we also take into account the study by Aoyama (Aoyama & Stephen, 2003; Aoyama, 2007) that classified the students who reach the "reading beyond data level" in the following categories, taking into account the students' critical competence to interpret the information in a graph and relate this information to a research problem (in all the levels the students correctly read the graph, are able to interpolate, detect the tendencies and predict, that is, reason at the "reading beyond data"):

1. *Rational/literal level.* At this level students make no critical reading of the information. They correctly read the graph and answer the question posed at the "reading beyond the data" level, but are unable to criticize the information in the graph or to provide alternative explanations. For example, when given a scatter plot that show a negative correlation between the birth rate and the life expectation in a group of countries they may not question the teacher' assertion that there is a causal relationship between both variables and think that an increase in the birth rate will automatically produce a decrease in the life expectancy.

2. *Critical level:* Students read the graph, understand the context and evaluate the information reliability; but they are unable to think in alternative hypotheses that explain the disparity between a graph and a conclusion. In the above example, the students will realize that the birth rate cannot directly influence the life expectancy, but will be unable to suggest other possible explanation for the negative correlation in the graph.

3. *Hypothetical level:* Students read the graph, interpret and evaluate the information, and are able to create their own hypotheses and models. In the example, the students may

suggest that the percent of women at work is higher in wealthy countries, where the life expectancy is higher; at the same time, they perceive that there is a direct relationship between the percent of women at work and the birth rate; consequently this could be a possible explanation for the negative correlation in the graph.

2.2 Graphical competence in prospective teachers

Only a few studies have focused on prospective primary school teachers' knowledge and competence related to statistical graphs. A few of them also concentrated on in-service school teachers (e.g. Tiefenbruck, 2007); in general the tasks given to the teachers were typical representation or reading tasks similar to the problems proposed to children in the school textbooks. Results from this research highlight a scarce graphical competence in prospective teachers (González, Espinel, & Ainley, 2011).

For example, in a study conducted with 29 prospective primary teachers in Spain, Bruno and Espinel (2009) found frequent errors when building histograms or frequency polygons. In another study with 190 prospective primary school teachers, Espinel, Bruno, and Plasencia (2008) observed lack of coherence between the graphs built by participants and their evaluation of tasks carried out by fictitious future students. Monteiro and Ainley (2007) in a sample of 218 Brazilian prospective teachers suggested that few of them had enough mathematical knowledge to read graphs taken from daily press. In Burgess's (2002) study with 30 prospective primary school teachers who were given each a set of 16 cards, with each card containing some data about one child and required them to examine and report any interesting features in the data. The author report that only six teachers were able to produce graphs from the data, and only about half of them were able to integrate the knowledge they could get from the graphs with the problem context or to generalise from data.

In our research we also gave the prospective teachers an open-ended task; however the task was more complete than that used by Burgess, as the teachers collected the data themselves and went through the complete investigative cycle described by Wild and Pfannkuch (1999) (research question, collect data, analyse data, interpret and answer the question). Our research is also based in a theoretical framework described in the next section.

2.3 Theoretical background

Different authors have emphasized the semiotic activity involved in the production and interpretation of graphs (e.g., Cleveland, 1987; Noss, Bakker, Hoyles, & Kent, 2007). When we build a graph, we encode the information using different elements, such as geometrical figures, or colours in the graph. When a different person read the graph, this information is visually decoded in a process of graphical perception (Cleveland, 1987). According to this author, a graph will be useful only if its visual decoding can be carried out accurately and efficiently.

Batanero, Arteaga and Ruiz (2010) defined a level of semiotic complexity in statistical graphs (described in detail in Section 4.1) using some ideas from the onto-semiotic approach to mathematics education (see, for example, Drijvers, Godino, Font, & Trouche, 2013; Font, Godino, & Gallardo, 2012; Godino, Batanero, & Font, 2007).

As it is accepted in mathematics education, meaning is a key notion in the ontosemiotic approach, and is basically conceived of in two ways. First, meaning is defined through the *semiotic function*. According to Hjemslev (1943) and Eco (1976) a semiotic function is a

correspondence between an antecedent (expression, signifier) and a consequent (content, signified or meaning) established by a subject according to a rule or habit. For example, when we use the word "median" we establish a correspondence between the expression "median" and the mathematical concept of median (meaning). Moreover, according to Font, Godino, and D'Amore (2007) all the different types of objects that intervene in mathematical practices: problems, procedures, concepts, language, properties and arguments can be used as either expression or content in a semiotic function. According to these authors, the semiotic functions, and hence the meaning, can be personal or institutional, unitary or systemic (Godino et al., 2007). In agreement with Peirce's semiotics (1978 /1965), the onto-semiotic approach assumes that both the expression (antecedent of a semiotic function) and content (consequent of a semiotic function) may be any type of mathematical object (for example, a definition, problem, procedure, argument, or a linguistic element).

Second, in the ontosemiotic approach, meaning can be understood in terms of use. From this perspective the meaning of the object is the set of practices in which the object plays a determining role. According to Godino et al., these two ways of understanding meaning complement each other, since mathematical practices involve the use of mathematical objects that are related by means of semiotic functions.

In our study we propose an open semi-structured project to prospective teachers, which is described in the next section. To complete the project, the participants have to solve some mathematical problems (comparing three different pairs of distributions), which required performing some mathematical practices. The focus in our research is the statistical graphs produced by the teachers and the mathematical practices linked to building and reading these graphs. When the teachers produced each graph they had to perform a series of actions (such as deciding the particular type of graph or fixing the scale), they implicitly used some concepts (such as variable, value, frequency, range) and properties (e.g., perpendicularity; proportionality) that vary in different graphs. Consequently the number and complexity of semiotic functions implicit in building each graph also vary. We therefore should not consider the different graphs as equivalent representations of the data distribution but as different configurations of interrelated mathematical objects that interact with that distribution.

Using the above ideas, in Batanero, Arteaga & Ruiz (2010) we defined a level of semiotic complexity in statistical graphs, in order to provide an indicator of graphical competence. We also observed that, although about two thirds of the participants in that research produced a graph with high semiotic complexity, only one third of them were able to get a conclusion in relation to the research question. A conjecture was that the prospective teachers' graph reading competence was low and that their reading level (in Curcio's, 1989 and Aoyama's, 2007 categorizations) was related to their competence in building the graphs (measured by their graph's semiotic complexity).

The aim of this research was to explore this conjecture analysing the responses to a variation of the statistical activity used by Batanero, Arteaga and Ruiz (2010) in a bigger sample of 207 prospective teachers. Below we first describe the project, data and instructions given to participants; secondly we classify the graphs produced according to their semiotic complexity; we then analyse the prospective teachers' reading of the graphs produced and classified them according the Curcio's (1989) levels; we then cross both classifications of graphs to study their possible association and study the relationship between graph complexity level and conclusion reached. Finally we present a discussion of our results and some implications for teacher education.

3 Method

A total of 207 prospective primary school teachers in their second year of study at a Spanish university participated in this study; in total, 6 different classroom groups (35-40 prospective teachers by group) participated in the research. All of them were following the same mathematics education course and had studied descriptive statistics at secondary school level, and in the previous academic year (for about 20 teaching hours). They also had experience in working with statistical projects, since they solved some projects the previous school year.

3.1 Tasks given to participants

The data were collected from the participants' solution to a statistical project taken from Godino, Batanero, Roa, & Wilhelmi (2008), in which participants were asked to perform a random experiment, collect data, compare three pairs of distributions, and come to a conclusion about the intuition of randomness in the group, basing their conclusion on the analysis of the data collected in the classroom. The sequence of activities in the project was as follows:

1. *Introducing the project and carrying out an experiment.* We suggested the prospective teachers to carry out an experiment and use the data collected in the experiment to decide whether the majority of the classroom had good intuitions on randomness or not. The experiment had two parts: a) In the first part (simulated sequence) each participant wrote down apparent random results of flipping a coin 20 times (without really throwing the coin, just inventing the results) in such a way that other people would think the coin was flipped at random. In the second part (real sequence) each participant flipped a fair coin 20 times and wrote down the results.
2. *Collecting data.* Each prospective teacher performed both experiments and recorded his or her results in a recording sheet provided by the lecturer. Afterwards, the lecturer started a discussion in the classroom; he asked for suggestions about what features of the simulated and real sequences could be compared to get a conclusion about the intuitions of randomness. A few prospective teachers suggested collecting data from the number of heads or for number of runs and comparing these numbers in both sequences for the entire classroom. After some discussion, the lecturer and the participants agreed in the interest of comparing the following statistical variables: a) number of heads, b) number of runs and, c) length of the longest run obtained by each prospective teacher in both sequences. Each prospective teacher provided his/her results for each of these six variables, and all these results were recorded on a recording sheet (See example of data collected by 10 participants in Table 1). At the end of the session each prospective teachers was given a printed copy of the data set for the whole group of students. These data contained six statistical variables: number of heads, number of runs and length of length of the longest run for each of real and simulated sequences from each participant in their classroom group. Since the number of students varied in each group the data to be analyzed contained 35-40 rows in the data sheet, depending on the group.

TABLE 1 – *Example of data collected (10 participants)*

Participant	Simulated sequence			Real sequence		
	N. heads	N. runs	Longest run	N. heads	N. runs	Longest run
1	10	14	4	11	9	4
2	12	9	4	11	16	2
3	11	12	4	11	16	2
4	10	9	4	8	9	4
5	11	11	3	7	11	4
6	9	13	3	8	10	5
7	10	12	3	9	9	4
8	11	14	3	11	4	7
9	9	13	3	10	12	3
10	10	8	5	9	9	5

3. *Instructions:* The prospective teachers were asked to produce a written report where they should summarise their statistical analyses of the data collected in the classroom and obtain a conclusion about the research question in the problem. That is, they should give their conclusion (and justify this conclusion with the statistical analysis) about whether they believed that the intuitions on randomness of the participants in the classroom were good, or to which extent they believed their intuitions were correct. There was no restriction in the report length; moreover, the prospective teachers were given freedom to use any statistical method or graph they wished and were allowed to use computers if they preferred. They were given a week to complete the reports.

TABLE 2 –*Summary statistics for number of heads, longest run and number of runs*

	Number of heads		Longest run		Number of runs	
	Real	Simulated	Real	Simulated	Real	Simulated
Mean	10.14	10	4.35	3.32	10.78	10.04
Mode	10	10	5	3	10	12
Median	10	10	4	3	10	12
Standard Deviation	2.05	1.06	1.6	0.8	2.06	2.9
Range	8	5	10	5	7	12

In Table 2 we include the averages and measures of spread in the six variables included in the project computed from the data collapsed from all the groups; results were very similar for

each group. We can see a strong coincidence in the averages for the number of heads in the real and simulated sequences (and then, a good perception of these averages in the group). As regards the other variables the group's intuitions are weaker as they tended to simulate more runs (and shorter) that happened in the real sequences. We expected that the participants would deduce the modes and ranges of the distribution from their graphs and used them to get a conclusion about the intuitions in the group.

A remark about the experiment proposed to the participants is that it is an example of a typical task frequently used in research directed to evaluate people's perception of randomness. We have used related tasks in our own research with prospective teachers; for example, in Batanero, Godino and Roa (2004) we describe a process of instruction in probability with teachers; as part of the instruction, the teachers are given an item in which two sequences of results of flipping a coin 40 times should be compared to decide which of the sequences is random. Other authors use sequences of different lengths ranging from 5 elements (e.g., Chernoff, 2009) to 150 (e.g., Green, 1983). In spite of these differences in length results, results concerning people's perception of randomness coincide that there is a good perception of the expected number of heads in the sequences and a poor perception of the independence of trials and the length of runs. In our research we do not focus on the teachers' perception of randomness (although the experiment may serve to assess this point) but in the teachers' graphical competence to build and interpret graphs as a part of their statistical analyses. Another difference is that in research dealing with people's perception of randomness it is the researcher who analyses the data produced in the experiments; while in this paper the analysis is carried out by the participants themselves.

4 Results and discussions

Once the prospective teachers' written reports were collected, we performed a qualitative analysis of these reports. Although many participants also computed the summary statistics (similar to those presented in Table 2) in this paper we only concentrate in the analysis of the graphs produced as a part of their analyses and in the conclusions they obtained directly from these graphs.

From a total of 207 participants in the study 181 (87.4% of participants), 146 (70.5%) and 128 (61.8%) produced some graphs to analyze the number of heads, number of runs and length of the longest run, even if the instructions given by the lecturer did not explicitly ask them to construct a graph. These high percentages suggest that the prospective teachers felt the need of building a graph to reach, through a transnumeration process (Wild & Pfannkuch, 1999) some information that was not available in the raw data; for example, finding the mode. These percentages also show the relative difficulty of the variables to be analyzed, as the number of heads was more familiar to the prospective teachers than the number of runs or the length of the longest run.

4.1 Level of semiotic complexity in the graphs produced

The graphs produced by the prospective teachers were firstly classified according to their semiotic complexity as defined by Batanero, Arteaga, and Ruiz (2010) (see examples in Figure 1). In this classification, the higher semiotic complexity level of the graph, the more

complex are the mathematical objects implicit and the higher is the number of interpretative processes involved in its construction. Below, we briefly describe each of these levels.

- *L1. Representing only individual results.* Even if prospective teacher were given a complete data set (with 35-40 values for each of the variables), a few of them were unable to produce a representation for the whole data set. Instead, they produce a graph where they only represented some isolated data values; usually their own data, without considering their classmates' results. In the example shown in Figure 1, the prospective teacher only represented the number of heads and tails in his individual experiment using a pie chart. Although the participant correctly draw this graph for his isolated data, and used proportional reasoning, he did not use the statistical variable "Number of heads" or the distribution of the number of heads when producing the graph.
- L2. *Representing all the individual values for one or several variables, without forming the distribution.* Some participants did not form the frequency distribution of the variables, when they were given the data set. They did not group the similar values in either the real or in the simulated sequences. Instead, they represented the value (or values) obtained by each student in the classroom in the order the data were collected. Consequently they neither computed the frequency of the different values nor explicitly used the idea of distribution. The order of data in the *X*-axis was artificial, since it only indicated the arbitrary order in which the students were located in the classroom. In the example shown in Figure 1, the participant built a line graph for the longest run in the simulated sequence; other participants in this level built horizontal or vertical bar graphs. Notice that these graphs do not serve to visualize the averages of the distribution, and then it is difficult to visualize the distribution centre, although the graphs show the data variability.
- *L3. Producing separate graphs for each distribution.* When comparing a pair of distributions, some participants produced a different graph for each of the two variables. Consequently, these prospective teachers were able to go from the data set to the statistical variable "number of heads in each sequence" (or else number of runs or length of the longest run) and were able to produce its distribution; therefore they used the ideas of frequency and distribution. The order in the *X*-axis in these graphs (see example in Figure 1) is the natural order in the real line. Teachers performing graphs in this level produced a separate graph for each variable to be compared; often they used different scales in both graphs and even different graphs for both distributions, and therefore the comparison was not easy for them.
- *L4. Producing a joint graph for both distributions.* This level corresponds to those participants who formed the distributions for the two variables to be compared and represented them in a joint graph, which facilitated the comparison. These graphs were the most complex, since the graph producer had to select a common scale and type of graph to represent both distributions. In the example in Figure 1 the participants used combined line graphs with a common scale to compare the distributions for the number of heads in the real and simulated sequences.

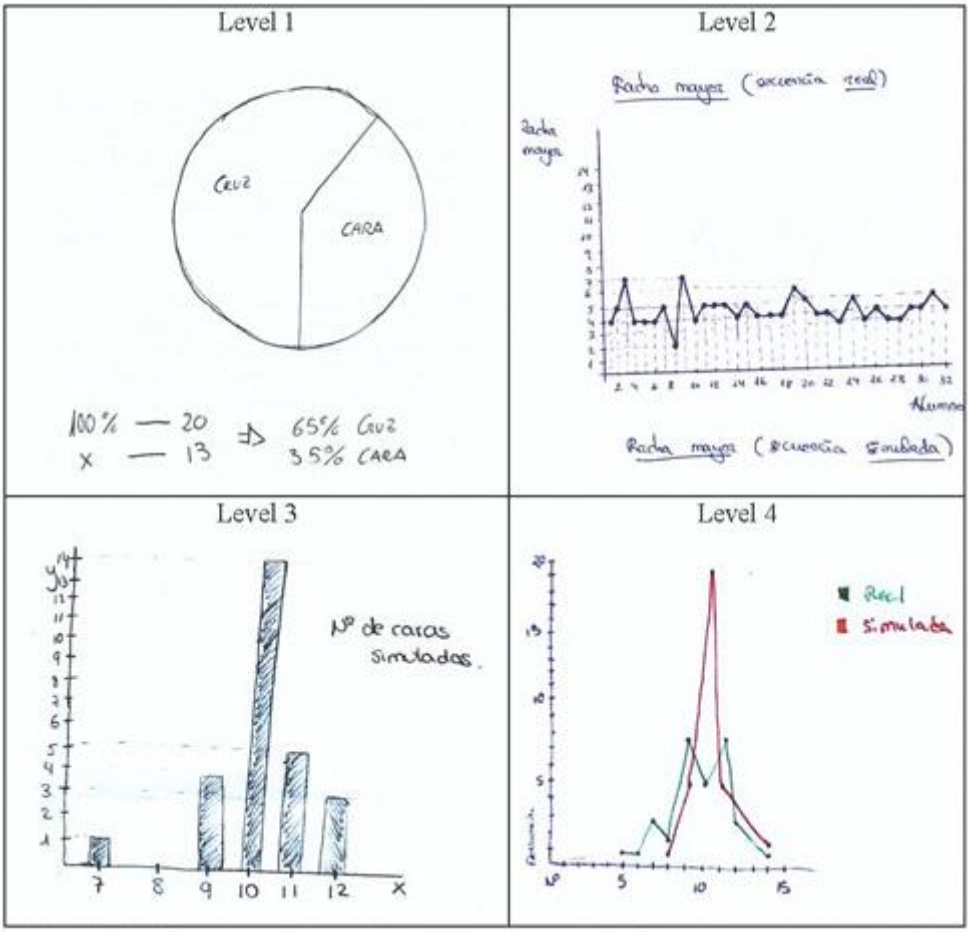


FIGURE 1 – Examples of graphs at different semiotic complexity level

In Table 3 we classify the graph produced by the participants in our study to compare each pair of distributions according to the graph semiotic levels. Few participants produced level L1 graphs, that is, where they only analyzed their own data, and less than 20% of participants represented the data list in the same order given in the data sheet, without making an attempt to summarize the data, and produce the variables distributions.

Taking into account all the graphs produced by the prospective teachers, we can observe that most graphs were built at level L3 (51.6%); when we add the participants who produced level L4 graphs (26.6%), we observe that 78.2% of the graphs produced by the prospective teachers had adequate semiotic complexity level to solve the problem posed, since they represent the statistical distributions of the variables to be compared. Consequently the concept of distribution seemed natural for the majority of participants, since most of them built a distribution to compare each pair of variables, although the instructions given to them in the task proposed did not require explicitly that they build the distribution.

TABLE 3 – *Frequency (percentage) of participants producing graphs in each semiotic level and pair of variables*

Graph semiotic complexity level	N. of heads	N. of runs	Longest run	All variables combined
L1. Representing only individual data	6 (3.3)	6 (4.1)	3 (2.3)	15 (3.3)
L2. Representing the data list	40 (22.1)	23 (15.7)	21 (16.4)	84 (18.4)
L3. Producing separate graphs for each distribution	91 (50.3)	77 (52.7)	67 (52.3)	235(51.6)
L4. Producing a joint graph to compare both distributions	44 (24.3)	40 (27.4)	37 (28.9)	121(26.6)
Number of participants producing graphs	181	146	128	455(100.0)

With the aim of checking whether there is influence of the variable analyzed (number of heads, number of runs, and length of the longest run) in the graph semiotic complexity level, we performed a Chi-square test of independence between variable analyzed and graph complexity level. We obtained a value Chi = 3.58, with 6 degrees of freedom, which was not statistically significant (p = 0.7339). Consequently, the semiotic complexity level of a graph in this research does not depend on the variable analyzed. Consequently, we concluded that the semiotic complexity level of a graph in this research *does not depend on the variable analyzed,* and therefore it is rather an indicator of prospective teachers' competence in graph building.

4.2 Participants reading levels of the graphs

In spite of Friel, Curcio and Bright's (2001) warning about the importance that students develop skills to interpret statistical graphs, reading and interpreting graphs is not easy, since *"the interpretation of statistical graphs is not a technical process, but an activity in which a wide range of knowledge, experiences and feelings are mobilized"* (Monteiro & Ainley, 2004, p. 8).

Besides building the graphs, the participants in the study had to read and interpret these graphs, in order to obtain the information they needed to reach a conclusions on the question proposed in the project. For each graph we indentified the participants' reading level according to Curcio's (1989) categorization from their reports, where we interpreted these levels as follows:

R0. Not reading the graph or incorrect reading: Some participants that produced a graph in their reports, made no attempt to read the graph; in addition, other participants made an incorrect reading of the graph they produced. For example, they made an incorrect identification of the mode or of the maximum in the graph. Par of these failures were produced by errors in the graphs that reproduced those described by Bruno and Espinel

(2009); for example, some participants made an incorrect choice of the type of graph that was then useless for finding the information required. Other students made conceptual errors, as in the following example; although the means in both distributions (as well as in the particular graph produced by ER) were close to 10, the prospective teacher argues that the means are close to zero. Consequently he does not reach Curcio' (1989) *reading between the data* level (extracting trends in the data) Moreover, ER incorrectly links the idea of representativeness to the standard deviation, though this is a property of the measures of centre.

> *"In analyzing the "number of heads" I noticed that the range of the simulated sequence (9) suggest the uniformity of the mean in the real sequence. Moreover both standard deviations suggest a high representativeness, since both means are close to zero"* (ER).

R1. *Reading the data*: Some participants made a correct literal reading of graphs labels and scales, but they only were able to read isolated information without making a global interpretation of the graph or being able to visualize summaries, such as measures of centre or spread. For example, they compared isolated values of the two variables, provided the frequency for a given value or made a general comment about the shape of the graph with no consideration of tendencies or variability in the data. In the following example, in which MLA correctly built a semiotic complexity L3 pie chart, which allows a more complete interpretation, this participant only made a literal reading of the frequencies associated to each number of heads.

> *"The (simulated) graph helps us to see the percentage of students who obtain x heads. These percentages are as follows: 45% students obtain 10 heads, 21.21% obtain 12 heads, 15.15% 11 heads, 9.09% 9 heads, and 3.03% 7 heads "* (MLA).

R2. *Reading between the data*: In this level, the participants showed made a more global reading of the graph; they made comparisons or found relationships between different data subsets within the graph. Prospective teachers at this level either compared averages (means, medians or modes) alone in both distributions (with no consideration of variation in the data) or else compared spread (without comparing averages). This implies a higher level of difficulty, than reading isolated frequencies. In the following example, AG, interprets a line graph; despite the imprecise language, he correctly perceives the graph tendency (in a line graph, when the graph increases, the frequency also increases) and identifies the mode of the distribution:

> *"In the line graph you can see the differences between the simulated and the real sequences: when the line rises, the frequency increases, the highest point in the graph corresponds to the variable value whit the highest frequency"*(AG).

R3. *Reading beyond the data:* This level involves making inferences and drawing conclusions from the graph. Prospective teachers at this level performed more complex reading of the graphs than those classified in the above levels, as they were able to analyze both the spread and centres in the distributions and to conclude about the distribution differences, when relating both measures, which were identified from their reading of statistical graphs. In the following example RC extracted information about the mode and range of the number of runs in both sequences from the graphs he built. Even if the language

is imprecise, in his analysis he compares the differences between the mode and range in both distributions.

> *"Based on the graph I produced, I can say that in the simulated sequence the modes are 12 and 13 while in the real sequence there is only one mode (11). The range in the first sequence is 6 and in the real sequence is 8"* (RC).

TABLE 4–*Frequency (percent) of participants producing graphs according to Curcio's reading levels*

Graph reading level	N. of heads	N. of runs	Longest run	All variables combined
R0. Do not read the graph	51 (28.2)	45(30.9)	42(32.8)	138 (30.3)
R0. Incorrect reading	21(11.6)	17(11.6)	18(14.1)	56(12.3)
R1. Reading data	41(22.6)	34(23.3)	32(25)	107(23.5)
R2. Reading between data	44(24.3)	32(21.9)	21(16.4)	97(21.3)
R3. Reading beyond data	24(13.3)	18(12.3)	15(11.7)	57(12.5)
Number of students producing graphs	181	146	128	455

In Table 4 we classify the participants who produced graphs, according to the reading level they used when interpreting the graphs built in their reports (in Curcio's categorization). We point to the high percentage of participants (30.3%) building graphs without interpreting them, in agreement with Burgess's (2002) results. Only 21.3% of the prospective teachers who built graphs reached the Curcio's (1989) intermediate level (reading between the data) and only 12.5% reached the upper level. The difficulty of reading the data slightly increased for variables related to runs that were less familiar to participants. However, the differences were not statistically significant in the Chi-square test of independence between reading level and variable analyzed (Chi = 3.56, df = 8, P = 0.8948). This result suggests that the reading ability of participants did not depend on the variable analyzed and was consequently an indicator of their graph reading competence.

These poor results may be explained by the wide range of skills required to read and interpret particular graphs according to Friel, Curcio and Bright (2001) that includes: recognizing the different elements that conform the graph (axis scales, titles, labels, mathematical content, etc...), perceiving the influence of each of these elements in the graph and making a synthesis of the information presented. Monteiro and Ainley (2007) also express concern that traditional teaching may offer prospective teachers enough opportunities to engage in activities that take into account the complexity of elements and process involved in the interpretation of graphs.

4.3 Relating the graph semiotic complexity and the teachers' reading level

Since we have interpreted both the level of complexity in the graph and the reading levels as indicators of the teacher's graphical competence, a question is whether these two levels are related. In order to study this question we crossed both variables

In order to study this question, in Figure 2 we classify the prospective teachers that produced some graphs, according to their graph semiotic complexity level and their reading level for each pair of variables and for all the graphs combined.

Comparing the different graphs, we can see that those prospective teachers who produced semiotic level L1 graphs (in any of the variables or in all the variables combined) either made an incorrect reading or only reached the literal "reading data" level. The percent of incorrect reading substantially decreased in Level L2, and is very similar in Level L3 and increase a little bit in level L4, because more complex graphs were harder to be interpreted correctly by the participants in the study. The highest percent of "reading beyond data" level (R4), when teachers are able to analyze both the tendency and spread and get a conclusion from the data, as well the combined percent of "reading between data" (R3) and "reading beyond data" (R4) levels were reached with semiotic level L4 graphs. Therefore level 4 graphs provided more opportunity for participants to perform a higher level reading. The same tendency appears in each variable and in all the variables combined.

The Chi-square test to check the independence of reading levels and semiotic complexity levels for all the graphs combined was statistically significant (Chi=40.4, dg=9, $p<0.0001$) and therefore we can accept that these two types of levels were related in our study. This point to the relevance to teach students to build level L4 graphs that will help them reach a higher level of reading of the same.

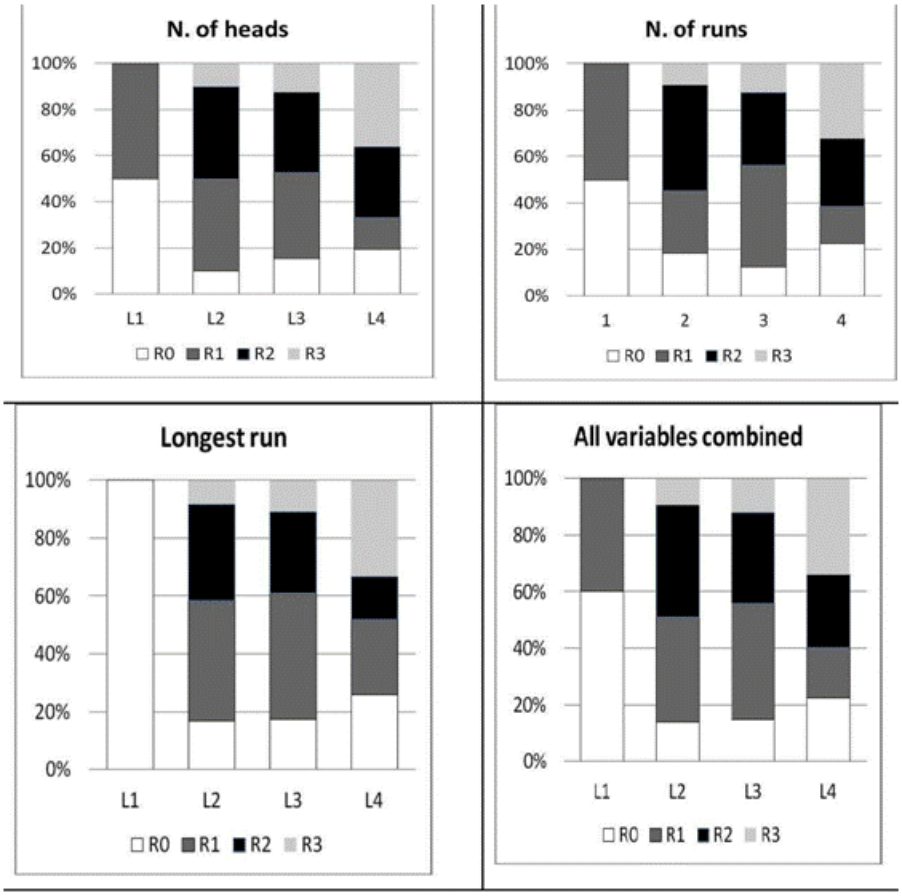


FIGURE 2 – Reading levels by semiotic complexity in the graphs

4.4 Graph production and conclusions reached about the research question

In the project proposed the participants not only should read the graph; they were asked to reach a conclusion regarding the collective intuition about randomness. Only a few prospective teachers in the study were able to get a correct conclusion, succeeding in performing an informal inference process (Rossman, 2008) and in relating the results from their statistical analyses to the problem posed in the project (the intuitions in the group). These prospective teachers concluded that the group had good intuitions as regards the average number of heads, while at the same time their perception of the average number of runs and of the average length of the longest run was weaker. In the same way the perception of variability in all the variables was in general poor. An example for a complete conclusion about the number of heads is given below (CG provided similar responses for the other variables):

> *"As regards the number of heads, the intuitions in the classroom were very close to what happen in reality; but not complete. The means in the real and simulate sequences are very close; the medians and modes are identical; however the standard deviations suggest that the spread in both distributions is different and then we failed in perceiving the variability of random sequences" (CG).*

This student is able to read the graph as Aoyama's (2007) hypothetical level; since, in addition to making a reading of the data at "beyond the data" level in Curcio's (1989) classification, he can deduce an explanation for the apparent differences in the graphs.

Other participants reached a partial conclusion, being only able to assess the part of intuitions related to either the central tendency or to the spread in the variables analyzed. In the following example, TG argues about the correctness of intuitions in relation to the average number of heads, but do not realize that the intuition about the variation is poor:

> *"Observing the table, I think that my friends have good intuition; since the most frequent values for the number of heads in the simulated sequence coincide with those in the real sequence; 10 and 11 are the most frequent values in both cases. The means are close to 10 in both sequences; therefore the intuitions are good" (TG).*

The remaining students either were unable to conclude or reached an incorrect conclusion. Part of them could not connect the results of their statistical analyses to the students' intuitions; that is, they did not see the implications of the results provided by the statistical analysis to the solution of the problem posed (assessing the students' intuitions). An example is given below:

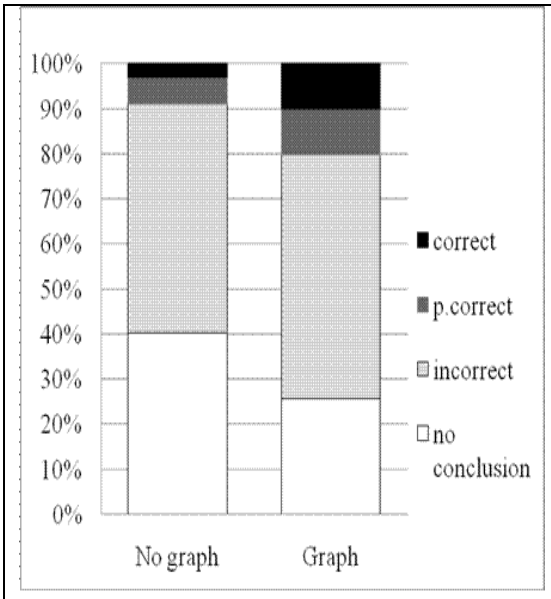
> *"When I compare the data, I realize that many students coincided in their results. In spite of this, I still think there is mere chance; since in the simulated sequences we invented the results" (EL).*

In Figure 3 we classify the participants according to the conclusions reached in the project and according to whether they produced or not a graph (Figure 3a) and according to the graph semiotic level complexity (Figure 3b in those who produced a graph). All the different

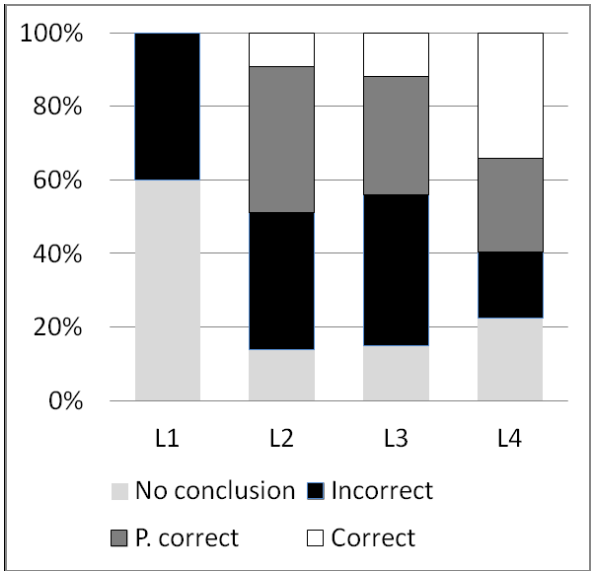
variables are mixed in Figure 3 to obtain a global synthesis of the effect of graph production and graph complexity level on the quality of the conclusions.

Only a few participants in the study reached the hypothetical level in reading the graphs, as they got the correct conclusion about the group's intuitions. These prospective teachers realized that the group had correct intuitions about the average number of heads but poor intuitions about the spread. Some more got a partly correct conclusion; for example, they assumed the intuitions were good because the averages of the number of heads had similar values in the simulated and real sequences.

The majority performed a mathematical comparison of the variables but got an incorrect conclusion or got no conclusion about the intuitions in the classroom (e.g. they correctly compared averages but did not understood what were the implications in relation to the students' intuitions). If we interpret these results according the Aoyama's (2007) levels, we conclude that the majority of prospective teachers only read the distribution data at a rational/literal level, without being able to read the results of their analises at a critical or a hypothetical level.



a. Conclusions according production of graph



b. Conclusions according graph semiotic level

FIGURE 3 –Conclusion according to (a) production of graph; (b) graph semiotic level

When comparing the results about the conclusions obtained in participants producing graph or not producing graphs (Figure 3.a) we see that only 3% of those prospective teachers building no graph got a complete conclusion and only 6% a partly correct conclusion. These percentages increased to 9.2% and 9,4% in those teachers that produced graphs as part of their analyses The Chi-square test to study the association between graph production and correctness of conclusion was statistically significant (Chi=18,72, df=3; p=0.007), so that we can accept that building a graph helped the teachers in their conclusions.

In Figure 3b we see that the percentage of correct conclusions increased to 34% in teachers producing level L4 graphs, because, on one hand, at that level the teachers read the graph at a higher reading level and; on the other hand, in L4 graphs the perception of both

tendencies and spread is easier. The percentage of partly correct conclusions was higher at level L2 graphs (39.5%), because the perception of variability in these graphs is possible, but not the perception of tendencies. Teachers who got a complete conclusion on both variation and tendency in this level usually had also computed some summaries statistics to help them in their conclusion.

The Chi-square test to check independence of type of conclusions and the semiotic complexity levels for all the graphs combined was statistically significant (Chi=40,45; df.= 9; p= 0,0000) and therefore we can accept that these two variables are related and that complete conclusion is easier with level L4 graphs.

In general (Figure 2), our participants interpreted correctly or partially correctly the graphs at least at the basic level in Curcio's (1989) classification; many of them reached the reading intermediate level (reading between the data), increasing this number with the graph complexity. However, an important part of participants in the graph levels 3 and 4, even when they reached the "reading between the data" level, could not achieve a complete conclusion on the research problem (Figure 3).

5 Implications for teacher education

In summary, our research suggest that building and interpreting graphs are complex activities and confirms part of the difficulties described by Li and Shen (1991) and Wu (2004) in students and by Bruno and Espinel (2005); Espinel (2007); Espinel, Bruno and Plasencia (2008) in prospective teachers. We agree with these authors in the relevance of improving the prospective teachers' levels of competence in both building and interpreting graphs, so that they can later transmit these competences to their own students.

Many participants in the study limited their analyses to producing graphs with no attempt to get a conclusion about the problem posed. Improving the teaching of statistics in schools should, consequently, start from the education of teachers that should take into account statistical graphs. These results confirm the opinion by Monteiro and Ainley (2007) that teacher education programmes should encourage prospective teachers to reflect on their own interpretations of graphs and that a teaching which is based on isolating statistical knowledge from out-of school knowledge and experiences is ineffective.

Prospective teachers in our research were asked to solve a problem and complete a modelling cycle. According to Chaput, Girard and Henry (2011), modelling consists of describing an extra- mathematical problem in usual language and building up an experimental protocol in order to carry out an experiment (in this research the problem consisted in checking the intuitions on randomness and a particular experiment was chosen to get data on the teachers' intuitions). This description leads to setting some hypotheses which are intended to simplify the situation (in the example, the length of the sequences to be produced was fixed and the equiprobability of heads and tails in the coin was assumed).

Next, the second step of the modelling process is translating the problem and the working hypotheses into a mathematical model in such a way that working with the model produces some possible solution to the initial problem. The teachers translated the question (what conceptions they had on randomness) and the working hypotheses to statistical terms (they compared three pairs of distributions: the number of heads, number of runs and length of the longest run in both sequences and for the whole classroom). Consequently participants in our

sample built and worked with different statistical models (each student chose and produced particular graphs, tables or statistical summaries to compare these pairs of distributions).

The third and final steps consist of interpreting the mathematical results and relate these results to reality, in such a way that they produce some answers to the original problem. Although the majority of participants in our research correctly completed steps 1 and 2 in the modelling cycle, few of them were capable to translate the statistical results they got to a response about what the intuitions of the classroom on randomness were like. That is, few of them could understand what the statistical results indicated about the intuitions in the group and therefore, these prospective teachers failed to complete the last part of the modelling process.

Dantal (1997) suggests that in our classroom, we concentrate in step 2 ("the real mathematics" in the modelling cycle), since this is the easiest part to teach to our students. However all the steps are equally relevant for modelling and in learning mathematics, if we want our students understand and appreciate the usefulness of mathematics. It is therefore very important that teachers' educators develop the prospective teachers' ability to model in statistics and the capacity to learn from data if we want succeed in implementing statistics education at school level.

Consequently, prospective teachers need more training in working with statistical projects, since working with projects is today recommended in the teaching of statistics from primary school level (NCTM, 2000; MEC, 2006).

Acknowledgment. Project EDU2013-41141-P (Ministerio de Economía y Competitividad) and group FQM126 (Junta de Andalucía).

Références

Aoyama, K. (2007). Investigating a hierarchy of students' interpretations of graphs. *International Electronic Journal of Mathematics Education 2(3).* Online: www.iejme/.

Aoyama, K., & Stephens, M. (2003). Graph interpretation aspects of statistical literacy: A Japanese perspective, *Mathematics Education Research Journal* 15(3), 3-22.

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. University of Rzeszów. Poland: ERME. Online: www.mathematik.uni-dortmund.de/~erme/doc/cerme7/CERME7.pdf

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

Ball, D. L. (2000). Bridging practices: Intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education, 51,* 241-247.

Bakker, A., Biehler, R., & Konold, C. (2004). Should young students learn about box plots? In G. Burrill & M. Camden (Eds.), *Proceedings of the International Association for Statistical Education (IASE) Roundtable on Curricular Development in Statistics Education* (pp. 163-173). Auckland: IASE. Online: www.stat.auckland.ac.nz/~iase/publications/.

Batanero, C., Godino, J. D. & Roa, R. (2004). Training teachers to teach probability. *Journal of Statistics Education* 12(1). Online: www.amstat.org/publications/jse/v12n1.

Batanero, C., Arteaga, P., & Ruiz, B. (2010). Statistical graphs produced by prospective teachers in comparing two distributions. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education.* Lyon: ERME. Online; www.inrp.fr/editions/editions-electroniques/cerme6/.

Bruno, A., & Espinel, M. C. (2005). Recta numérica, escalas y gráficas estadísticas: un estudio con estudiantes para profesores (Number line, scales and statistical grapas: a study with prospective teachers). *Formación del Profesorado e Investigación en Educación Matemáticas VII*, 57-85.

Bruno, A., & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training. *International Journal of Mathematical Education in Science and Technology, 40*(4), 473-493.

Burgess, T. (2002). Investigating the "data sense" of preservice teachers. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa: International Association for Statistics Education. Online; ww.stat.auckland.ac.nz/~iase/publicatons.

Chaput, B., Girard, J. C., & Henry, M. (2011). Modeling and simulations in statistics education. In C. Batanero, G. Burrill, & C. Reading (Eds.) (2011). *Teaching Statistics in School Mathematics- Challenges for Teaching and Teacher Education. A Joint ICMI/IASE Study* (pp. 85-95-83). New York: Springer.

Chernoff, E. (2009). *Subjective probabilities derived from the perceived randomness of sequences of outcomes.* Ph.D. Ottawa: Simon Fraser University.

Cleveland, W. S. (1987). Research in statistical graphics. *Journal of the American Statistical Association*, *82* (398), 419-423.

Curcio, F. R. (1989). *Developing graph comprehension.* Reston, VA: N.C.T.M.

Dantal, B. (1997). Les enjeux de la modélisation en probabilité [The betting of modelling in probability]. In M. Henry (Coord.), *Enseigner les probabilités au lycée* (pp. 57-59). Reims: Commission Inter-IREM.

Drijvers, P. Godino, J. D., Font, V. & Trouche, L. (2013). One episode, two lenses. A reflective analysis of student learning with computer algebra from instrumental and onto-semiotic perspectives. *Educational Studies in Mathematics*, 82:23–49. DOI: 10.1007/s10649-012-9416-8.

Eco, U. (1979). *Tratado de semiótica general* (General semiotics). Barcelona: Lumen.

Espinel, C. (2007). Construcción y razonamiento de gráficos estadísticos en la formación de profesores. *Investigación en Educación Matemática 11,* 99-119.

Espinel, C., Bruno, A., & Plasencia, I. (2008). Statistical graphs in the training of teachers. In C. Batanero, G. Burrill, C. Reading & A. Rossman (2008). *Proceedings of the Joint ICMI /IASE Study Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education.* Monterrey, México: ICMI &IASE. CD ROM.

Font, V., Godino, J. D., & Gallardo, J. (2012). The emergence of objects from mathematical practices. *Educational Studies in Mathematics*. In press, DOI: 10.1007/s10649-012-9411-0.

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). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review 70*(1), 1-25.

Godino, J. D., Batanero, C., & Font, V. (2007). The onto-semiotic approach to research in mathematics education. *Zentralblatt für Didaktik der Mathematik*, *39*(1-2), 127-135.

Godino, J. D., Batanero, C., Roa, R., & Wilhelmi, M. R. (2008). Assessing and developing pedagogical content and statistical knowledge of primary school teachers through project work. In C. Batanero, G. Burrill, C. Reading & A. Rossman (2008). *Proceedings of the Joint ICMI /IASE Study Teaching Statistics in School Mathematics*. *Challenges for Teaching and Teacher Education*.[CD] Monterrey, México: ICMI e IASE.

González, T., Espinel, C., & Ainley, J. (2011). Teachers' graphical competence. In C. Batanero, G. Burrill & C. Reading (Eds.). *Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. A Joint ICMI and IASE Study*. New York: Springer.

Green, D. R. (1983). A Survey of probabilistic concepts in 3000 pupils aged 11-16 years. En D. R. Grey et al. (Eds.), *Proceedings of the First International Conference on Teaching Statistics* (Vol.2, pp. 766-783). Universidad de Sheffield: Teaching Statistics Trust

Hadjidemetriou C. & Williams J.S. (2001). Children's graphical conceptions: assessment of learning for teaching. *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education, 3,* 89-96

Hjelmslev, L. (1943). *Prolegómenos a una teoría del lenguaje* [Introduction to a language theory]. Madrid: Gredos.

Konold, C., & Higgins, T. (2002). Working with data: Highlights of related research. In S. J. Russell, D. Schifter & V. Bastable (Eds.), *Developing mathematical ideas: Collecting, representing, and analyzing data* (pp. 165-201). Parsippany, NJ: Dale Seymour Publications.

Lee, C., & Meletiou, M. (2003). Some difficulties of learning histograms in introductory statistics. Paper presented at the Joint Statistical Meeting. *Section on Statistical Education.* Retrieved from http://www.statlit.org/PDF/2003LeeASA.pdf

Li, D. Y., & Shen, S. M. (1992). Students'weaknesses in statistical projects. *Teaching Statistics 14* (1), 2-8.

McClain, K. (1999). Reflecting on students' understanding of data. *Mathematics Teaching in the Middle School, 4(*6), 374-380.

MEC (2006). *Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación primaria.*

Monteiro, C., & Ainley, J. (2004). Interpretation of media graphs and critical sense: implications for teaching and teachers. Paper presented on the *10th International*

Congress on Mathematics Education. Copenhague. Online http://www.icme-organisers.dk/tsg11/Papers/

Monteiro, C., & Ainley, J. (2007). Investigating the interpretation of media graphs among student teachers. *International Electronic Journal of Mathematics Education, 2* (3), 188-207. Retrieved from http://www.iejme/.

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

Noss, R., Bakker, A., Hoyles, C., & Kent, P. (2007). Situating graphs as workplace knowledge. *Educational Studies in Mathematics* 65, 367–384

Pereira-Mendoza, L., & Mellor, J. (1990). Student's concepts of bar graph: Some preliminary findings. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* Voorburg: International Statistical Institute. Online; www.stat.auckland.ac.nz/~iase/publications.php

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, *7*(2), 5-19. Online: http://www.stat.auckland.ac.nz/serj.

Tiefenbruck, B. (2007). *Elementary teachers conceptions of graphical representations of categorical data.* Unpublished Ph.D. University of Minnesota.

Watson, J. M. (2006). *Statistical literacy at school: growth and goals*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review 67* (3), 223-265.

Wu, Y. (2004, July). Singapore secondary school students´understanding of statistical graphs. Paper presented at *10th International Congress on Mathematics Education*. Copenhagen, Denmark.