ANALYSIS OF TWO INTERNET INTERACTIVE APPLETS FOR TEACHING STATISTICS IN SCHOOLS

Juan D. Godino [(a)], Francisco Ruiz [(a)], Rafael Roa[(a)], Juan L. Pareja[(a)] and Ángel M. Recio[(b)]

[(a)] University of Granada (Spain)
[(b)] University of Córdoba (Spain)
[(1)]

Summary

The aim of this paper is to provide some information to guide the use of Internet interactive resources to teach Statistics in primary and secondary school. We identify key variables, possible cognitive effects and potential semiotic conflicts in two "stat-applets" from the NCTM Standards server. We conclude with some hypotheses about the positive effects of using these resources as computing and display tools, that might favour changes in the types of data analysis tasks, help discussions in the classroom, and amplify conceptual understanding. Even recognising these improvements, we should not forget the statistical objects' complexity, and the students' potential symbolic conflicts, which to overcome will require the teacher's awareness, the planning of appropriate interaction pattern and the improvement of the software.

1. INTRODUCTION

The number and quality of Internet interactive resources to facilitate the teaching and learning of statistics is progressively increasing. An important set of resources and applets for primary and secondary school levels has been made available by the National Council of Teachers of Mathematics (NCTM) at the web page http://illuminations.nctm.org/index2.html, where we can find useful tools for different mathematical content, and particularly for data analysis and probability.

Although the designers of these resources suggest different uses in the mathematics classroom, we still need theoretical and experimental studies of their didactical features, and of the learning environment where they can be implemented. Information about each tool's didactical variables, ways to overcome potential semiotic conflicts, and cognitive effects of different interactions between teacher, students and instructional context are also required.

In this paper we analyse two NCTM applets, by using theoretical tools from the ontological and semiotical model to mathematics cognition developed in our previous works (Godino, 2002a; Godino, 2002b; Godino and Batanero, 1998). This model might be included in the educational perspective on mathematics as semiosis (Anderson et al., 2003). Our theoretical tools provide a general methodology to analyse other similar applets, as well as teaching and learning processes.

For each applet, we will first include a brief description. Then we analyse the different elements in the institutional meanings of statistical objects involved in using the applet (situation-problems; language; actions; concept-definition; properties; and arguments) and will identify the student's potential semiotic conflicts. Finally we conclude with an evaluation and some implications for teaching.

[1] *Edumat-Maestros Team*, University of Granada (Spain). Research Project BS2002-02452. (Ministerio de Ciencia y Tecnología, Madrid, Programa de Promoción General del Conocimiento) and "Acción Hispana-Alemana", HA 2002-0069.

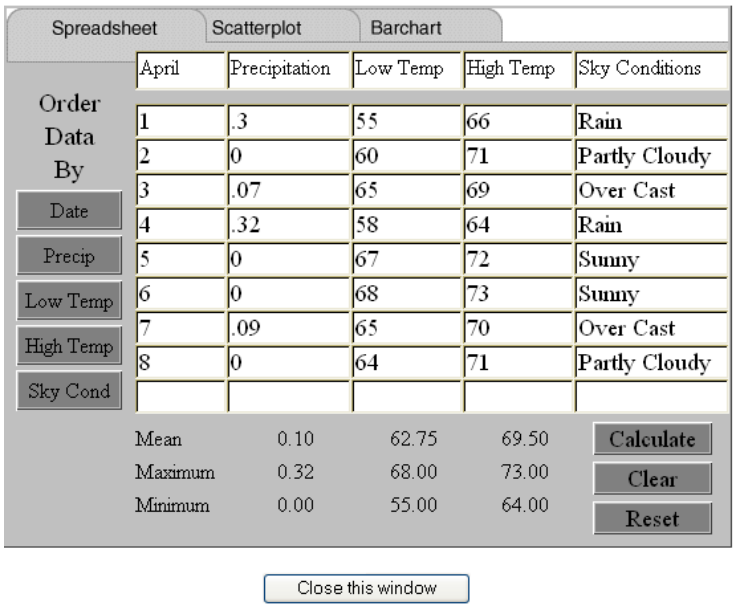
2. TWO STAT-APPLETS FOR ELEMENTARY STATISTICS INSTRUCTION

2.1. Spreadsheets and graphing software

This applets, recommended for primary school grades 3 to 5, its description and suggestions of use, are available at the web site: http://standards.nctm.org/document/eexamples/chap5/5.5/index.htm

2.1.1. Description

Spreadsheets and graphing software are tools for organizing, representing, and comparing data. This applet illustrates the use of a very simple spreadsheet. In the first part, Collecting and Examining Weather Data, students organize, then examine data that have been collected over a period of time and entered in a spreadsheet (Figure 1). In the second part, Representing and Interpreting Data, students use the graphing functions of a spreadsheet to explore these data (Figures 2 and 3). Working on similar activities, students learn to work with a simple spreadsheet and to use it in posing and solving problems, exploring data, and investigating patterns.



April	Precipitation	Low Temp	High Temp	Sky Conditions
1	.3	55	66	Rain
2	0	60	71	Partly Cloudy
3	.07	65	69	Over Cast
4	.32	58	64	Rain
5	0	67	72	Sunny
6	0	68	73	Sunny
7	.09	65	70	Over Cast
8	0	64	71	Partly Cloudy
Mean	0.10	62.75	69.50	
Maximum	0.32	68.00	73.00	
Minimum	0.00	55.00	64.00	

Figure 1: Spreadsheet

As a possible task to carry out with this applet**,** the students may be asked to collect meteorological data from their own city, to organize and record the data on the spreadsheet. The following questions are suggested:

- Observe how the "summary data" change when new data are introduced. What do the summary data tell you?
- After collecting data for a month, explore and develop answers for questions such as, What was our weather like in January? How would we describe the weather in January in our town to a visitor? What was the temperature generally like this month? How much rainfall did we get?

Similar questions might be posed for data corresponding to different months, or different cities. Students can also develop similar questions for other types of data collected in the classroom, newspapers or Internet databases, which can be introduced in the spreadsheet and later analysed.

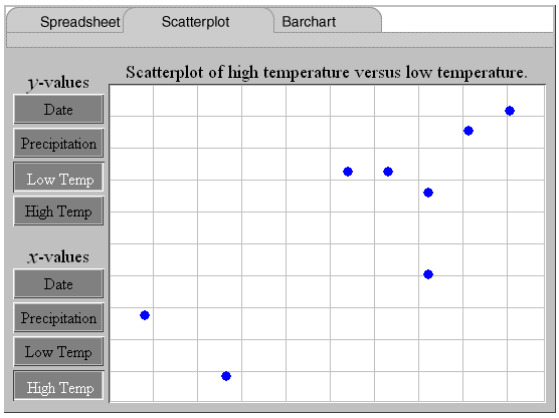


Figure 2: Scatterplot

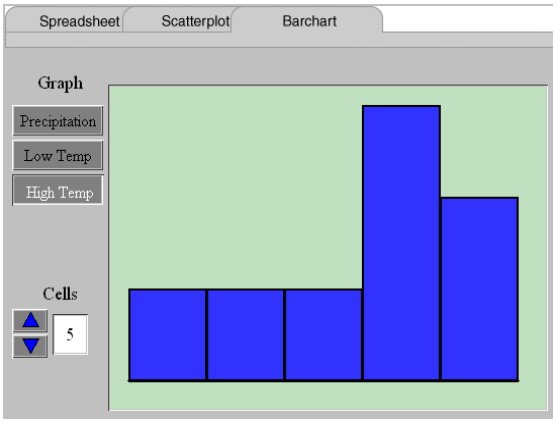


Figure 3: Barchart

2.1.2. Statistical objects and meanings

Our focus of interest is analysing the complexity of the applet tasks (based on computing and display tools), and identifying possible conflicts among the meanings of statistical concepts and techniques that can be achieved in using the applet (local meaning) and their statistical meaning..

Situation-problems

The applet presents a particular data analysis situation: information about 5 meteorological variables in a particular city (date, precipitation, low temperature, high temperature and sky condition) for an 8 day sample. Mean, maximum and minimum are automatically computed for each data series. Likewise, a barchart and a scatterplot for a pair of quantitative variables are presented. The problem posed and solved is the numeric and graphical (partial) *reduction* of four data series. This is an example of a very general field of problems, even when there is a severe restriction in the number of registers that can be processed (a maximum of 9 statistical individuals). This strongly limits the practical possibilities of this software, which, in fact is considered to be an early stage of the use of a professional spreadsheet.

The student can change the variable names (first line) and the values assigned, and it is therefore possible to solve problems of comparing two or more samples by comparing their arithmetic mean, maximum and minimum. The pair of variables represented in the scatterplot axes can be changed, so we can pose questions about the relationship between numeric variables.

The data analysis situations to be proposed to students have a multivariate nature, since five variables are simultaneously represented on the spreadsheet. This applet feature, together with the collecting and recording data activities, support a potential genuine and authentic statistical activity where it is possible to propose situations with some level of systemic complexity.

This applet is a tool that supports a very wide class of data analysis problems, therefore, enlarging the situational component of the meaning of descriptive statistical objects in an extraordinary way. Since the maximum and minimum are computed, its difference (range) can be used as a measure of the data spread. The fact that students collect their own data about contexts in which they are interested, gives the situation-problems a personal meaning for them.

Languages/ representations

The applet presents three screens called Spreadsheet, Scatterplot and Barchart. Each of these screens is a linguistic object whose components and rules should be analysed in order to foresee the interpretative processes involved and to identify possible semiotic conflicts.

Spreadsheet

The cells have a tabular disposition, where the first row is reserved for variable names (which can be changed) and the other 9 rows are (numeric or alphabetical) data cells. Since each column contains the variable values, there is a semiotic function between the first row cell (and their name) and the content of the other 9 cells. There is also a dependence between the variable name and the variable values (disposition in columns), while rows correspond to individuals.

The distinction between a statistical variable and its values is materialized here in the tabular disposition. We have here examples of variables and their respective values; the possibility of changing the first row values and names creates conditions to generalize the concepts of variable and value.

The terms "Mean", "Maximum", "Minimum" are related to the rows of values calculated for each data column. There is a double dependence "expression - content": to each data vector column which corresponds to the computed value; these numeric values also correspond to the terms “mean”, “maximum” and “minimum”.

We remark on the novelty of this way of expressing mathematical identities, such as, "the mean of the statistical variable Low Temp is 62.75" (where there is a syntactic dependence between the positions (terms) "Low Temp." and "Mean"). The teacher should explain these new conventions that some student might not discover spontaneously.

The terms "Mean", "Maximum", "Minimum", remain unchanged when the numeric values are changed, which indicates that they are designating not just the particular statistics values for each column, but the corresponding conceptual entities.

In the “order data” section, the user should know that by pressing each of the suggested keys the data columns are ordered, which permits the identification of the maximum and minimum. A correspondence is set between "Order data by" and each key, as well as with each data column, before and after pressing the key. A similar situation is produced with the orders "Calculate", "Clear", "Reset".

The atypical character of some elements in this spreadsheet should be emphasized, in particular the arrangement of the ordering and computing commands, which are given by specific conventions that students may find confusing when starting to use a professional spreadsheet (e.g. Excel). These local meanings are conflicting with other institutional meanings. Other specific linguistic elements in this spreadsheet appropriately represent the type of tabular arrangement used to introduce data in spreadsheets and statistical software (individuals in rows; variables in columns).

It is possible to modify the names of the variables in the spreadsheet (first line), and hence we can use it for other contexts. However, the ordering keys labels are fixed. This can produce another semiotic notational conflict.

Scatterplot

This part presents a 10x10 grid where points corresponding to two quantitative variables (x, y) are plotted. The coordinate axes and their scales are not marked, which means that the plot is an atypical representation. The pair of corresponding values appears when the cursor approaches the points.

There is an option menu to select the pair of variables to be represented as "y-value" and "x-value". The objects selected are the pair of variables, and its values. It is necessary to

connect the selection of a pair of variables to the distribution of points in the grid (referential dependence), and to interpret the meaning of this distribution in terms of association between the variables. A well-known syntactic rule is that the "x-values" are arranged on the horizontal axis and "y-values" on the vertical one, and therefore, the disposition of the "x-values" on the vertical axis might cause semiotic conflicts.

The applet allows a dynamic and interactive study of 9 pairs of statistical relationships. Because it is also possible to change the spreadsheet data, the number of scatterplots is potentially high and varied. The spreadsheet and the scatterplot cooperate to produce the new object, namely, stochastic dependence between variables, its direction and graduation.

Barchart

This part of the applet allows the display of barcharts for the variables Precipitation, Low Temperature and High Temperature (in fact, these barcharts are histograms because the applet produces interval groupings for numerical variables). We can change the number of cells (class intervals) between 1 and 10 and observe the effect on the bar heights. The representation is atypical, because axes and scales are not marked, although it is possible to read the point of the coordinates in the abscissas axis and the bar heights when the cursor approaches the points. This lack of axes and scales and the way of giving the number of intervals by pushing a button are particular syntactic rules used here.

The statistical objects involved here are the frequency distributions for quantitative statistical variables grouped in class intervals and their dependence of the grouping criteria (class intervals). The aim of changing the number of intervals is to study the characteristic properties of the histogram.

We can formulate the hypothesis that primary school children need complementary explanations to use and interpret the Barchart and Scatterplot tools, since they incorporate diverse conventions.

Actions

The students are supposed to collect data from their own city at different periods of time. This action is part of the general design of the didactic situation, and it is not required by the applet, since it offers data collection for 8 days in April. Once the data has been collected the students record them in the corresponding cells. This is a regulated action since it is necessary to introduce the values corresponding to each variable in the same column. The operations of computing the statistics, ordering the data, clearing the display, drawing a scatterplot, a barchart, and restoring the initial data is radically simplified compared to the operations carried out with other means, in particular with pencil, paper and calculator. The local meanings of these actions / operations are reduced to pressing a key and reading the result. The effort should be focused now on the interpretation of these results. Computing the mean, for example, it is now an elementary act, while it has a certain level of complexity when done by hand or with calculator: to add the data and to divide by the number of addends (should 0 data be counted?)

The "expression-content" dialectic it also involved with the operative objects. The action of pressing the keys is the antecedent while the result of the action is the consequent (content or meaning). When changing data in the cells and pressing "Calculate" the nine values of the matrix written in the lowest part of the spreadsheet change: these numbers are the result of calculating, and what is calculated are the values of the statistics for the four statistical variables. A possible operative semiotic conflict appears because when pressing "return", or going down to another lower cell with the cursor these statistics are also calculated. Are the results of these two different actions the same?

Concepts

We identify the following conceptual entities involved in the use of this applet: data, variable, value, mean, maximum, minimum, data ordering.

The possibility of ordering the data creates conditions for studying the maximum and minimum concepts, as the first and last element of the corresponding column.

The concepts of Spreadsheet, Scatterplot, and Barchart are here exemplified as compound objects whose main components can be shown. In the Scatterplot the concept of statistical association for quantitative data is involved, and in the Barchart (histogram in fact) the frequency distributions for data grouped in intervals.

Properties:

- The mean is a representative value of a data set and therefore can be used to compare two or more samples.
- The maximum and minimum can indicate the dispersion (range).
- The scatterplot indicates the type and intensity of the relationship between two variables.
- The change of the histogram shape is studied by changing the number of intervals. The amount of data is so small that grouping is unnecessary, which creates a rich discussion situation concerning the use of the histogram and the barchart.

Arguments:

The applet does not provide any justification or explanation of the actions and results that are carried out. However, we can create learning validation situations by using the results provided by the resource. For example, weather circumstances of two cities or periods of time can be compared, and it is possible to justify statements about them using the results of the applet. We can argue about the type of relationship between two variables by observing the scartterplot.

2.1.3. Positive and negative feature of this software

The analysis carried out suggests that the language and statistical objects implemented in this spreadsheet have certain level of complexity. All these elements cooperate to produce a new object: a spreadsheet with some specific local features. Our hypothesis is that working with this spreadsheet and interpreting its results, in particular the semiotic correspondences between different terms and expressions, require a learning process that will not be free of difficulties for primary school children and that teachers should be aware of.

Moreover this applet has some particular connotations which require certain effort to understand. The restriction to a maximum of 9 values is also excessive, because it prevents the analysis of a reasonable amount of data. Our analyses also raises the doubt if it would be preferable to directly use a professional spreadsheet, even at a later age.

On the other hand, with the information provided by our analyses, this interactive micro-program allows us to design didactic situations for primary school level with a high potential degree of validity for the statistical objects: statistical variable, mean and range, and to a smaller degree for statistical association and frequency histograms. The use of this software in ~~a~~ real teaching - learning process requires, however, the teacher to implement an explanatory and justifying discourse to facilitate the generalization of the techniques and results beyond the particular examples presented.

There is a high degree of students autonomy to solve the tasks, after the teacher has been able to explain how to use the device, whenever the students have some previous understanding of the concepts of mean, maximum and minimum.

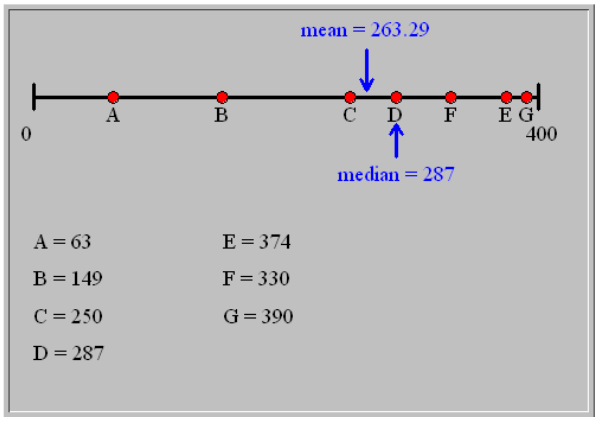
2.2. Comparing the mean and median properties

This applet recommended for school grades 6-8, its description and suggestions of use, are available at the web site:

http://standards.nctm.org/document/eexamples/chap6/6.6/index.htm

2.2.1. Description

Using this interactive micro-program we can explore and compare properties of the mean and median. When changing the data values, the relative positions of mean and median in the new data set are immediately shown in a lineal diagram (Figure 3).



Figures 3: Relative position of mean and median

The numeric data represented in the straight line represent the flying distances of a paper airplane when thrown in the air; these distances can vary in the interval [0, 400]. The task consists of exploring how the mean and median of this data set are affected when one or several values are changed, and analysing the representativeness of both statistics.

Some questions that can be experimentally explored are suggested on the web page:

- Can you find ways to move the data points that keep the median the same but change the mean?
- Can you find ways to move the data points that keep the mean the same but change the median?
- How do the mean and median change when you keep the points in the same order but just change their positions on the number line?
- What happens if you pull some of the data values way off to one extreme or to the other extreme?
- By moving data points, can you construct data sets in which the mean seems to be a typical value but the median is not? Vice versa? For what types of data sets, if any, is the mean not very representative? When is the median not very representative?

2.2.2. Statistical objects and meanings

Situation-problem

The situation is an experimental and dynamic study of the properties of the arithmetic mean and median. Although this is not suggested in the description, it is possible to imagine

several suppositions about the airplane design and to investigate which is more effective. This involves comparing two or more series of launchings. What centralization measure is preferable to use (mean or median)? Deciding which airplane is more effective motivates the data reduction and the election of a representative value.

We cannot change the number of data, fixed at 7 (odd number), but we can freely change the values in the interval [0, 400]. This allows us to propose symmetrical and non symmetrical distributions, introduce atypical values or not, with the purpose of studying the representativeness of mean and median and their relative positions in each case. In what types of data is the mean or the median preferable? It can be discovered that if the data distribution is symmetrical both statistics coincide. It is not possible to consider ordinal measures, where median is unavoidable or unordered data (where students might have difficulties with the idea of median).

The questions to be explored are in some sense "internal" to mathematics, since they refer to number handling. However, it is easy to attribute an external meaning to the numbers, such as measures of distances or of another magnitude. Representing the measures on the number line helps to visualize their relative positions.

The applet easily allows simulation of the data set, its representation on the number line and automatic calculation of mean and median. This can help to communicate the functional character of both statistics and to explore their properties.

Language

There is a display of the rational number line in the interval [0, 400] with 7 points marked and labelled with letters (variables). The positions of the points can be changed by displacing them on the number line, which is translated as changes in the variable values. The points are ordered on the number line, and this order corresponds to the alphabetical order of the variable's label.

The fact that "mean" (or "median") designates a particular value is frequently unnoticed. For example, 263.29 is the image that corresponds to a function applied to a data set. Mean or median are therefore functions of R^n to R, although in the usual statistical work, and in the teaching practice, this functional nature of summary statistics is unnoticed. Here the terms "mean" and "median" are names of statistics that can take diverse values as we move the points. As we can quickly and interactively change the data set, these statistics are produced and vary, thus showing the functional nature of the mean and median. Although the fact that these functions are applied to vectors (data set) is not perhaps so clear from the applet.

The position of A in the number line is supposed to be related to the origin, so that the numbers assigned to the letters are the measures of the segment OA taking one unit segment (which is not visible in the display). The writing A= 63 is characteristic of mathematical syntax; this is the usual way to assign a value to a variable. A particular syntactic rule in this graph is using arrows to mark the mean and median positions.

Actions

- Changing the data (assigning values to the variables) by displacing points on the number line. Each action (antecedent, expression) corresponds to a content (consequent, meaning), which is the numeric value corresponding to the new point position on the number line and to the new variable value. It also corresponds to a change in the value assigned to the terms "mean" and "median". Each elementary action corresponds, therefore to a three item image (data, mean, median). It is supposed that the user will produce a group of actions (series of data) to which the values of the mean and median with a certain relational structure will correspond.

- Representing real numbers in the number line, where the measure unit is not shown, because it is neither needed nor possible. This technique here is transparent.
- The calculations of mean and median are carried out automatically by the applet.

Concepts

In addition to the mean and the median the following concepts are also involved: data set; ordering; representativeness of a statistics for data set.

The reason for representing a data set by a unique value, such as the mean, emerges in the comparison of two or more samples. The relative advantages of the mean or the median, as best representantion have to be shown in comparison activities. Therefore, it is necessary to design a systematic plan of selecting different data sets to first make clear the use of data summaries, and later the relative advantages of one or the other statistics. However, the fact that in this applet the values introduced are not stored can hinder the appreciation of the representative qualities. This recording must be done by hand.

Properties

- The arithmetic mean changes when at least one data value changes.
- The mean can substantially change when one or several atypical values are deleted.
- The median does not change if changes to the data correspond to individuals located to the right or to the left of the individual located in the centre of the distribution, and the individuals who change stay on the same side of the centre.
- The median is a better representative of the data set if the distribution is asymmetric since it will be located in a position nearer to most of the data than the mean (it makes the sum of the distances in absolute value of each point to the median minimum). This preference supposes acceptance of that criteria as a good quality for representation.

Arguments

The argumentation is visual or empirical: the data and the position of the statistics are visually shown. As the number of data is 7 and they are ordered, the user can quickly calculate the median; this is not the case for the arithmetic mean, which is calculated by the program and it requires, therefore, a certain act of faith. A potential semiotic conflict might appear when the children later are given unordered data to compute the median.

The properties of these statistics are not deduced from the definitions, but are rather visually shown. This might create obstacles to understand the deductive nature of mathematical proof, contributing to confusion in the discovery context (where any argumentative resource is allowed) and the context of rational justification.

2.2.3. Positive and negative feature of this software

The applet have positive features to support the teacher's explanations and to stimulate the collective discussion of properties of the mean and median; however it is not a tool for personal exploration, because if the students do not know what to look for, it will be difficult to find. The learning of the idea of representativeness of statistics will require the teacher's contribution.

The discussion and orientation that accompany the applet do not mention the crucial point of the degree of asymmetry of the distribution and the presence of atypical values as key element in the decision to use the median. In this restricted context of data analysis the median will always be preferable to the mean, or indifferent when the distribution is symmetrical.

This applet provides computing and display tools to support the practical component of the objects "statistical representation", and the properties of the mean and median. But the generality of these properties and their rational justification should be carefully implemented by the teacher.

3. CONCLUSIONS

Our analysis suggests the level of complexity involved in the use of these interactive micro-programs, applets, given the variety of mathematical objects that intervene and the semiotic correspondences among these objects. These objects and knowledge have to be taught in the time assigned to the study process in the different scenarios where it takes place. The applet is certainly a resource that serves to support new and richer systems of mathematical practices and new meanings. However, their management and coordination with other didactic resources and scenarios pose new challenges to the teacher's work. We think that analyses similar to the one carried out in this investigation provide useful information to guide the work of the teachers' educator.

These analyses can raise awareness about the transparency illusion we frequently hold about mathematical (in this case, statistical) and didactical knowledge. The naïve use of the new technological resources available for the teaching of statistics, can reinforce this illusion, and obscure the even more delicate and complex teacher's work. The appropriate use of these new technologies will require changes that are not obvious in the teachers' and learners' roles, as well as new patterns of didactic interaction.

The applet facilitates the construction of the student's personal exploration scenarios. However, progress in this exploration as supported by interactive programs, necessarily requires the teacher's contribution. This is a new challenge for the teacher who need simultaneously to help students who are at different points of the learning process.

REFERENCES

Anderson, M., Sáenz-Ludlow, A., Zellweger, S. and Cifarelli, V. (2003). *Educational perspectives on mathematics as semiosis: From thinking to interpreting to knowing*. Otawa: LEGAS.

Godino, J. D. (2002a). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, 22, (2/3): 237-284.

Godino, J. D. (2002b). Studying the median: a framework to analyse instructional processes in statistics education. En B. Phillips (Ed.), *ICOTS-6 papers for school teachers*. Cape Town: International Association for Statistics Education (CD Rom).

Godino, J. D. and Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area of research in mathematics education. En, A. Sierpinska and J. Kilpatrick (Eds.), *Mathematics Education as a Research Domain: A Search for Identity* (pp. 177-195). Dordrecht: Kluwer A. P.