STUDYING THE MEDIAN: A FRAMEWORK TO ANALYSE INSTRUCTIONAL PROCESSES IN STATISTICS EDUCATION

Juan D. Godino
University of Granada
Spain

In this paper we describe the main ideas in a theoretical model that was developed for mathematics education research and is also applicable to statistics education. This model takes into account the three basic dimensions of teaching and learning processes: epistemic dimension (concerning the nature of statistical knowledge), cognitive dimension (concerning subjective knowledge) and instructional dimension (related to interaction patterns between the teacher and the students in the classroom). These theoretical notions are justified and applied to analyse a teaching process for the median in the introductory training of teachers.

INTRODUCTION

Research in mathematics education in general and in statistics education is mainly focused on one of the following two aspects involved in teaching and learning processes:

- The cognitive pole, that is, students' learning and mental processes involved (psychological focus);
- The instructional pole, that is, design and implementation of teaching experiments, curricular and new resources development (pedagogical focus).

In addition, recent research tendencies consider it to be crucial to extend the research to the mathematical /statistical content. Researchers in this group suggest that the analysis of the nature of knowledge should be the starting point of didactic problems. From an educational perspective this analysis should adopt appropriate epistemological models, or even develop new conceptualisations of mathematics (statistics).

Moreover, we think that an effective research paradigm should take into account the three facets mentioned (epistemological, cognitive and instructional) and should build adequate theoretical models to describe the interactions among the same.

In this paper we briefly describe the main notions of the systemic and integrative approach to research in mathematics education, in which we have been working for several years (Godino & Batanero, 1998a; 1998b). This theoretical framework is of general application in mathematics education and also in the study of the specific problems posed in statistics education. In fact the model is being developed, applied and tested in several works of investigation focused on the teaching and learning of statistical notions (Batanero & Godino, 2001). In this paper we will focus our attention on a teaching experiment about the median in the context of primary school teacher training. A more complete description of this research can be found in Godino (2001). Our theoretical model is based on the following epistemological and cognitive assumptions about mathematics:

i) Mathematics is a human activity involving the solution of problematic situations (external and internal), from which mathematical objects progressively emerge and evolve. According to constructivist theories, people's acts must be considered the genetic source of mathematical conceptualisation.
ii) Mathematical problems and their solutions are shared in specific institutions or collectives involved in studying such problems. Thus, mathematical objects are socially shared cultural entities.
iii) Mathematics is a symbolic language in which problem-situation and their solutions are expressed. The systems of mathematical symbols have a communicative function and an instrumental role.
iv) Mathematics is a logically organized conceptual system. Once a mathematical object has been accepted as a part of this system, it can also be considered as a textual reality and a component of the global structure. It can be handled as a whole to create new mathematical objects, widening the range or mathematical tools and, at the same time, introducing new restriction in mathematical work and language.

ANALYSING THE PROCESS OF STUDY: THE CASE OF MEDIAN

A study (teaching and learning) process implemented in a classroom for a mathematical or statistical content is for us the main unit for didactic and mathematical analyses. The observation, description and analysis of these processes will show the factors that affect such processes, and will serve to evaluate experiments on new teaching methods.

In the following, we assume that we are trying to help to introduce the median to a group of students for the first time. The aim is to teach them how to solve those problems where the use of the median is appropriate, to help students distinguish this use as compared with the use of mean and mode, and not to confuse the median with other statistics measures.

One of the first steps for teachers or researchers is planning the teaching. Within this planning, the elaboration or selection of "the knowledge" to be taught will be a reference for the action in the classroom and the students' personal study. Below we describe one such teaching experience, which was carried out with student teachers without previous statistical knowledge, and where we used as didactical resource the section about the median in a secondary school textbook. The aim of this description is to exemplify the theoretical tools we propose in a simple case.

EPISTEMIC FACET

The analysis of textbooks is an important research task, since textbooks are basic references for the knowledge to be taught, and they are main teaching and learning resources. It is true that teachers make a creative use of textbooks, and when planning their lessons they modify the tasks that will be carried out in the classroom, change the explanations, sequence and type of activities. The interactions in the classroom, which includes the students' active participation, also contribute to fix the exact meaning implemented.

On the other hand, there are important differences among textbooks concerning the examples, problems, definitions and arguments presented. The study and evaluation of such differences in meaning requires a global point of reference that should be determined by the educational level and context. The writing of textbooks or the teacher's selection of them for classroom use is implicitly or explicitly based on the institutional meaning of reference for the particular mathematics content to be taught

In our theoretical framework, we approach the epistemological problem of determining the features of institutional mathematical knowledge, by introducing the notion of *institutional meaning*. This is conceived as the "system of operative and discursive practices", that is, as "mathematical praxeologies" (or statistical praxeologies, if the problems posed refer to statistics). We also propose a classification of institutional meanings and distinguish among factual, potential and global meanings to study the relationships between the meanings implemented in the classroom in a specific teaching and learning process, and the variety of meanings in the different textbooks.

Sufficiently rich epistemological models that take into account the genesis and components of knowledge are needed to characterise these praxeological systems, and the relationships between them. In our model, we suggest that mathematical (and statistical) objects emerge from a problem solving activity, mediated by the linguistic tools available, and that they include conceptual, propositional and argumentative objects. The potential meaning, that is, the meaning proposed by the textbook used in the experience in our example about the median, can described in the following way:

Language (terms, notations, graphical representations):

- The term 'Median', and the notation 'Me '.
- Listing of data (horizontal and vertical, which serve to visualise the median as the central position of the ordering).
- Frequency tables.

Representations not used:

- Percentile 50%; 2º quartile; 5º decile.

- Abscissa of the absolute (relative) frequency point in the accumulative graph whose ordinate is n/2 (0.5).
- Accumulative diagrams; box and whisker display, steam-leaf etc.

Situation-problems (phenomenological components, promoting and contextualizing statistical activity):

Examples of the following types of problems:

- Finding a representative value for a data set from a quantitative statistical variable with atypical values (odd or even number of data).
- Finding a representative value for a data set from an ordinal statistical variable.
- Looking for a representative value for a data set from a quantitative statistical variable in listing form from a frequency table (situation non specific for the median).

Situations not studied:

- Non-symmetrical frequency distributions without atypical values.
- Continuous variables with data grouped in class intervals.
- Variables represented graphically.
-

Actions (operations, computing techniques):

Examples of the following techniques:

- Computing the median for an odd number of values.
- Computing the median for an even number of values.
- Computing the median in cumulative frequency tables of discrete variables.

Techniques not studied:

- Computing the median for data grouped in class intervals (interpolation).
- Graphical determination using the cumulative frequency polygon.
- Use of statistical software and calculators.

Concepts (definitions): *Four characterizations of the median (including the idea of central value):*

- Value that leaves the same number of data above and below it.
- The central data of an odd number of ordered data.
- The mean of the two central data of an even number of ordered data.
- The first value of the variable that corresponds to the cumulative absolute frequency, immediately superior to half the number of data.

Propositions (properties):

- Statistical property: the median is a more representative value that the mean in the case of atypical values.
- The median is a measure of central tendency.

Properties not studied:

Given the institutional context in which the text will be used (secondary school level) the properties of the median are not studied (with the exception of being a representative value in some circumstances).

Arguments (justifications, validations):

The fact that the median is sometimes more representative that the mean is justified by the mean of an example (when the series of data has an atypical value).

Validations avoided:

- Equivalence of the four definitions introduced.

- The notions of "central value" and "better representation" are not justified. It would be necessary to prove that the central value of the data is the value which is nearest the majority of data, that is, to prove the property that the sum of the absolute deviations of data to a value *a* is minimum when *a*= Me.

We have identified these components of the praxeological meaning of the median in the text as a result of applying a technique that we call "semiotic analysis." This technique allows us to characterise the systemic (or praxeological) meanings of a mathematical object and the elementary meanings involved in an act of mathematical communication. Likewise, it provides a tool to identify potential semiotic conflicts in the interpretation of a textbook in a teaching and learning process, that is, the conflicts that take place in the effective realization of a didactic interaction.

Semiotic analysis can be applied to any text where the mathematical activity developed by participants is registered: e.g., to the planning of instruction, to the transcriptions of the class development, to interviews and written answers to assessment tests. The analysis is based on the decomposition of the text in analysis units, on the identification of the mathematical entities involved, and of the semiotic functions (correspondences between an expression and a content) that are established by the subjects among these different entities.

In our example about the median, the author of this textbook prefers to anticipate discursive entities (concept definitions and properties) to actuative entities (problems, computations). He firstly defines the object and describes some of its properties, and later introduces the ways of acting to solve the tasks. It would be advisable to change this order, since discursive entities have their justification as ways to solve the problems. Moreover, for this teaching level, instruction could be carried out without putting so much emphasis on rigorous definitions, and with investing more time in justifying (to make reasonable) the representative character of the median.

It would be desirable to spend some teaching time in introducing graphical displays, such as the cumulative frequency polygon and the stem and leaf, which serve to visualize and compute the median. Being aware of epistemic gaps in a teaching and learning process is a key factor to make decisions about such a process, such as, delaying the study of the topic until the students have the necessary cognitive resources.

COGNITIVE FACET

In the previous section we have described the analysis of the epistemic facet (institutional knowledge) in mathematical and statistical instruction. However, didactics should also evaluate the type and quality of students' learning. As a result of the study process directed by the teacher and of personal study, each student should be able to: a) carry out the tasks posed, b) explain the way he solves the tasks, c) relate some objects to others (for example, discriminating the use of mean, median and mode). This competence and understanding (knowing how to solve a task and knowing why it is solved in this way) is more or less complete in each student. We need to know the students' idiosyncratic systems of operative and discursive practices to identify regularities in the students' answers.

To assess the students' learning we introduce the notion of "personal meaning" for a mathematical object, or meaning link to a type of problems. These meanings are also interpreted as praxeologies, in which we distinguish linguistic, situational, actuative, conceptual, propositional and argumentative components. The notion of semiotic conflict is used to describe disparities or disagreements between the meanings two people or institutions attribute to the same expression in a communicative interaction; semiotic conflicts are considered as potential explanations of learning difficulties and limitations.

In the assessment carried out in our teaching experience about the median we requested the students to explain what the median is and in what situations it is used. In Figure 1 we reproduce the answer given by a student:

The median is the value that leaves the same number of data above it and below it. It *represents the number closer to more number of data, that is to say, it is a most representative measure than the arithmetic mean.*

There are two types of median:

- *Median of an odd number of data, the median is the central data. Example,*

  80.000
  $\boxed{60.000}$
  40.000

- *Median of an even number of data, the median is the mean of the two central data. Example,*

  80.000
  60.000
  50.000

  90.000

  sum of the two central values divided by 2 = x

Figure 1. Student's Answers to Question's about the Median.

We can see in this response the semiotic conflict that the ordering of the data collection poses to the student. In the second case, the example is incorrect since the student does not order the data previously. The textbook used includes the following definition: "The median of an ordered group of data from a variable is the value that leaves the same number of data above and below it." It is not mentioned that the median is also applicable to non-ordered data, although in this case the data should be previously ordered before finding the median. The explanation of the students' errors is then found in the teaching and learning process and more specifically in the institutional meaning implemented.

INSTRUCTIONAL FACET

The theoretical tools described until now are insufficient to carry out a comprehensive didactical analysis for a teaching and learning process. Between the implemented knowledge (conceived as institutional praxeologies) and the knowledge built by the subject (personal praxeologies) an instructional process that conditions and determines the learning took place. The teacher carried out a series of actions. Students also carried out several actions, according to their own initiative or planned by the teacher. All these actions took place in a period of time, involved interactions among students themselves, the teacher, and the diverse components of the mathematical praxeology built in the classroom. Didactic interactions are not exempt of conflicts, which should be solved by negotiation of the meanings involved. The instructional facet is approached in our theoretical model by introducing the notions of *teacher functions*, *student's functions* and *interaction patterns*. A first list of such functions whose articulation defines a *didactic trajectory*, is the following:

Teacher's functions:

i) Planning: designing the teaching process, selecting contents and meanings to teach;
ii) Guidance: monitoring the teaching process, deciding changes of tasks, orientating and stimulating student's functions;
iii) Teaching: presenting the information;
iv) Assessment: evaluating the state of the learning achieved in critical moments;
v) Investigation: reflecting on and analysing the teaching process to introduce changes in future sessions thereof.

Student's functions:

i) Exploration: Inquiry, search of conjectures and ways to answer the questions posed (action situations).
ii) Formulation/communication: of solutions (formulation-communication situations),

iii) Validation: Argumentation and verification of conjectures (validation situations).
iv) Reception: reception of information about ways of making, describing, naming, validating the tasks posed (institutionalisation situations);
v) Drill and practice: making routine tasks to master the specific techniques (exercising situations).
vi) Application: applying the knowledge learned to solve real problems (non-didactic situations).

In our example, since the instructional process consists of the use of a textbook, the teacher's function of presentation of information prevailed in an almost exclusive way, not only in the tasks presentation, but also in the solution techniques, definitions and justifications. The writing of a textbook supposes an intense work of planning the instructional process by the author, which is carried out previously; this particular textbook does not include an assessment section. However, there is a section where knowledge is summarised and systematized. The student's functions are limited to, receive and retain the information presented and carry out the routine tasks posed. There are no moments of exploration of possible solutions and formulation and validation thereof. The information provided by the semiotic analyses of the epistemic and cognitive facets should guide the teacher's future actions.

FINAL REMARKS

The theoretical notions described in this paper define a semiotic approach to research in mathematics education and constitute analytic tools to describe the complexity of mathematics instructional processes, as previous step to understand and to intervene in the design and development of such processes.

A question is whether this model is directly applicable to statistics education research. We need to reflect about whether the specific nature of statistical knowledge and the teaching and learning statistics processes can be described with these theoretical notions. Due to the generality of the notion of mathematical praxeology we can give a positive answer to this question, although we recognise that statistical data analysis activities incorporate specific phenomenological elements. In any case, statistical activity, and the statistical objects involved, arise from the subjects' commitment with a certain class of problem-situation. This activity leads to ways of solving such situations (techniques), and to descriptions and justifications that organise and systematise the solutions. It would then be necessary to conceive statistical knowledge in terms of "systems of operative and discursive practices", that is, as statistical praxeologies, whose components will be different to those in geometry, algebra, etc. We conclude, therefore, that a semiotic approach to research is also appropriate for statistics education research.

ACKNOWLEDGEMENT

This research has been supported by the DGES grant BS02000-1507 (M.E.C., Madrid).

REFERENCES

Batanero, C., & Godino, J.D. (2001). *Developing new theoretical tools in statistics education*. 53 ISI session, Seul.

Godino, J.D. (2001). *A semiotic approach to research in mathematics education*. URL: http://www.ugr.es/local/jgodino/

Godino, J.D., & Batanero, C. (1998a). Clarifying the meaning of mathematical objects as a priority area of research in mathematics education. In A. Sierpinska and J. Kilpatrick (Eds.), *Mathematics education as a research domain* (pp. 177-195). Dordrecht: Kluwer, A. P.

Godino, J.D., & Batanero, C. (1998b). Building and experimenting a model for a meaningful instruction in data analysis. In L. Pereira-Mendoza et al. (Eds.), *Proceeding of the 5th International Conference on Teaching Statistics* (Vol. 2: 905-912). Singapore: ISIS.