BUILDING AND EXPERIMENTING: A MODEL FOR MEANINGFUL INSTRUCTION IN DATA ANALYSIS

Juan D. Godino and Carmen Batanero, University of Granada, Spain

In this report, we describe a model for designing and assessing the processes of teaching and learning mathematics, as well as its theoretical bases. This model is applied to analyse a University data analysis course, which was supported by the intensive use of computers.

INTRODUCTION

Teaching data analysis at University level poses didactic and cognitive problems, which have scarcely been analysed up to date (Shaughnessy et al., 1997). Statistical training at the University does not only requires the proper use of concepts, statistical methods and computer resources, but a practical sense to apply these tools in solving real data analysis problems. Therefore, an instructional model for achieving meaningful statistical learning, must consider the complex relationships generated in the classroom between the students' attitudes and previous ideas, the conceptual structures we try to teach them, the tools available, and the problem data analysis situations. In this presentation, we analyse a didactic model for statistical instruction which takes into account these aspects, and is based on our theory about mathematical objects, their personal and institutional meanings and understanding (Godino and Batanero, 1994; Godino, 1996). This model have been experimented in two courses at the University of Granada.

INSTRUCTIONAL MODEL AND EPISTEMOLOGICAL FOUNDATIONS

According to our view, mathematical knowledge is built by the subject's interaction with specific problem fields. Different types of practices are carried out by subjects to solve mathematical problems, communicate the solutions to others, validate and generalise the solutions to other contexts and problems. These practices, from which mathematical abstractions emerge, may be classified in three types, which constitute the structural elements of meaning of the mathematical objects: a) *Intentional elements*: Concept definitions, propositions, procedural descriptions; b) *Extensional elements*: Situation-problems, exercises, tasks; c) *Notational (or instrumental) elements*: Systems of symbols, computing and representational resources (tables, graphs, texts, etc.)

As these practices depend on the institutional and personal contexts in which a problem field is solved, we differentiate between the institutional and personal meanings of mathematical objects, which are interrelated. The problem fields in which mathematical objects are applied also produce their different senses or partial meanings. For example, the meaning of averages is different in exploratory data analysis, and in inference. Affective factors and time assigned to the study process are two additional factors conditioning the meanings.

As a consequence, a meaningful study of mathematical objects requires a representative sample of those practices which constitute the systemic meaning of the object within a given institution. Students should have the opportunity to explore relevant problems, formulate hypotheses and conjectures, compare different representational systems, communicate and validate the problems solutions, and to confront them with mathematical meanings. However, due to the limitations of time and previous knowledge, the meanings of mathematical objects built by participants in an instructional process is always partial and relative to the institutional, material and temporal contexts in which the process takes place.

ANALYSING A PROCESS OF STUDYING STATISTICS WITHIN A COMPUTER ENVIRONMENT

Institutional and temporal contexts

The course (80 hours long) was optional and was taught in the academic year 1996-97 to 34 students and in 1997-98 to 58 students from different specialities (Pedagogy, Psychology, trainee Teachers, Management) from the School of Education, University of Granada. We only analyse the part devoted to Exploratory Data Analysis (about 30 hours).

Institutional meaning implemented and its evaluation

Problem situations and data files (extensional elements): The teaching was based on five data files, which contextualised statistical concepts and techniques: ATTITUDES (students' attitudes towards statistics); MEASURES (studying the relationship between different anthropometrical measures, gender, and the practice of sport), MFF20 (effect of impulsiveness / reflexiveness on children's reading capacity); NEURONS (relating the size of two rodent species' neuron components and their location in the brain), and TESTP (relating children's age, probabilistic reasoning, and mathematical achievement).

Instrumental elements: Statgraphics is an integrated package, with a menu and icon system. We can compare it to a new statistical language, where a number of gestures, and complex symbolisation refer to calculation procedures. Solving the problems only requires understanding the data, providing suitable entry parameters and interpreting the outputs. Therefore, it stresses the separation of technical, technological and theoretical work in statistical problem solving. There is also the possibility of having various tabular and graphic displays simultaneously on screen, which the student can manipulate using a number of options. This allows him to relate the different displays, concepts and techniques. It is also important the distinction between data files and Statfolios where results and textual reports can be saved. The following options were studied: FILE, EDIT, DESCRIBE, COMPARE, RELATE.

Intentional elements: Concepts and statistical techniques: Though the students have previously studied some statistics, their knowledge was very poor. Therefore we made no formal presentation of the concepts. Students with sufficient knowledge of English, could also use the "Statistical Advisor" to reinforce their learning. The following statistical concepts and procedures were studied: Population, sample, statistical variables; Data collection, and recording. Frequency tables, bar charts and pie charts. Central position, dispersion and shape; Grouped frequency tables; stem and leaf; box-and-whisker plots; Crosstabulations; association coefficients. Scatter plots; correlation and regression.

The study of concepts and properties was supported by the use of computers. For example, the stochastic convergence of the frequency polygon to the density curve was experimentally presented, using a file with 1,000 simulated data about students' heights. However, there were difficulties in observing convergence, because of fast and high fluctuations in the frequency polygon, when the number of intervals was slightly increased.

The teaching process: What students learn depends on the sequence of tasks, their temporal organisation, the role attributed to the student's personal work, drill and practice work, and communication, validation, institutionalisation and assessment activities implemented (Brousseau, 1986). We used the following scheme: a) Posing a problem situation to introduce a Statgraphics option; b) Collective solving, guided by the lectures; c) Explaining and discussing the results, concepts and techniques involve, and d) Working in pairs to solve complementary problems, with the teacher's assistance when required. Because lessons took place in the computer lab and there was an extensive list of topics, communication, validation and institutionalisation situations were scarce.

Active participation, solutions to the tasks proposed, two intermediate written tests, a final examination using the computer, and optional individual projects were taken into account in the assessment. Statfolios printouts with the solutions to the tasks proposed, and final tests, permitted an individualised follow-up of the students' work and their final achievement.

Affective factors : Problem took into account the students' interest in education, and daily life situations. For example, the data in the file MEASURES were taken by the students themselves, who could also select the topic for their projects. The assessment favoured doing the exercises, but the lack of final examination on theoretical concepts strongly conditioned the process of study. Other motivational factors were the utility of statistics for the students in other areas, and the fact that many of the students had never studied a subject based on the use of computers.

Evaluating course design and development

The problems gave a meaningful context to the main statistical concepts and techniques, and learning of the computer software. Drill and practice activities were based on other applications and personal projects, such as “Overall situation of children”, “Qualities of ideal teachers according to students”, “ Profile of adult educator”, “Index promotional polices for women”, “Students' beliefs about assessment”, “Migratory movements in Andalusia”, and “Gender and age structure of population at work”. The experience showed the complexity of real data analysis problems formulated in the students' projects, which required the lecturer's assistance. The question of the extent to which these problems should be simplified to fulfil their training purposes remained open.

Our experience suggested the need to reduce the course contents, if we want the students' knowledge to surpass a mere technical level, and so reach technological and theoretical levels. Greater time is also needed to validate and institutionalise the different data analysis concepts and techniques. Projects should be shared by 2-3 students, and be systematically followed-up with different means, including students' presentations of the projects at certain stages of development. As a rule, students showed interest in the course, though computer slowness and mouse malfunctions failure sometimes caused frustration and discouragement.

Personal meanings and their dynamic

We collected a written initial assessment test about elementary stochastic ideas, and 2 written intermediate tests. Data was also recorded from a final test, and solutions to 8 intermediate exercise relations using computers for all the students. In addition, 7 students performed a personal data analysis project. From this information, in this section we analyse the knowledge acquired by a student (J), as well as his final data analysis capacity and the factors which conditioned his learning. J was in his first year of Psychopedagogy (4th year of University studies), and had previously studied most of the course contents, though he had not handled computers or statistical software before. He was extremely interested in the subject.

Problems posing ability: There were scant opportunities for students to state their own problems on data files during the course. We can assess this dimension of J's knowledge from his personal data analysis project on: *"Students' study habits and judgments about their teachers' evaluations"*. J showed great difficulty in stating his goals and hypotheses and in expressing research questions. His description of the research problem was confusing, the identification of variables lacking, and the analysis and results discussion poor. He mainly posed and solved routine questions, like the following one: *"How many students have a mean mark over 6? Which value takes 25% percentile of the score?"*. However, we should remember that his project was carried out without the help of the lecturer, except for some points in preparing the questionnaire, and that the data collection and report writing was carried out at the end of the course, during the period of final exams.

Degree of mastering the computer software: J has achieved a high degree in mastering the specific options studied, and, furthermore was also interested in using word processors in non-academic hours.

Understanding statistical representations: We observed misunderstanding of several graphical and tabular representations. For example, this is his explanation of the box-and-whisker plot: "*The box-and-whisker plot is a plot designed to inform us about the following statistics:, the average, which is given by the central line crossing the graph; the median, which is represented by the central perpendicular line crossing the average; the maximum or minimal range defined by the segments (whiskers) drawn from the rectangle"*.

There are some aspects of the box-and-whisker plot which have been well understood: *"In this graph we can see the symmetry of the sample if we look at the inside of the rectangular box. If the space between the median and one quartile is greater than the distance from the median to the other quartile, we may say that there is asymmetry"*. However, we point out the error of attributing such symmetry to the sample, and not to the frequency distribution of the statistical variable.

Intentional aspects (concepts and their properties): We found other conceptual misunderstandings in J's project. For example, when interpreting a contingency table he does not report the values of the corresponding variable*: "The marginal distributions of the gender variable are: 33.3 and 67.7;The conditional distributions of 'sport' according to 'gender=boy' are: 20.0, 45.0 and 35.0"*. As regards association, we detected a local conception of this, since he bases his association judgement on the frequency of only one cell in the contingency tables.

Process and affective elements: We have not collected enough systematic data to report on J's verbal and graphical expression, explanation and validation ability, for which we would have to plan clinical interviews with him. Nonetheless, the personal project carried out by J suggests important limitations in these components, for which it was not possible to organise specific didactical situations. We may conclude that the knowledge attained by J is mainly technical.

FINAL REMARKS

The traditional teaching of statistics at University level, with large groups of students and restricted computer resources, force the teacher to concentrate on the intentional component of statistics, reducing the extensional and instrumental aspects to routine exercises solved with paper, pencil and calculators. On the contrary, teaching small groups of students with adequate computer resources allows us to focus study on the extensional and instrumental aspects, putting the intentional aspects in the background. It is difficult to organise collective activities for communication, validation and institutionalisation, due to the dispersion of tasks and phases the students are performing when solving the exercises. Another problem is the very unequal level of students' previous knowledge, in particular when the subject is offered for a range of degree courses. Our experience has shown the difficulty of reaching a balance between the different elements of the meaning of statistical concepts in these circumstances.

The theoretical model we are developing for mathematics education intends to facilitate the study of the relationships of personal and institutional meanings about mathematical contents. Its aim is explaining learning difficulties by the structural elements, institutional factors, process, temporal and affective components involved, on which we can act. Secondly, we may search for learning difficulties in the students' intrinsic cognitive shortcomings. Given the complexity of mathematical concepts, their teaching and learning depend greatly on the selection made for their different components. Recognising that the knowledge attained by each student about mathematical content is always partial, our aim is to try to complete this knowledge as far as possible in each circumstance, facilitate future growth, and increase the students' personal interest.

The relationships between the institutional and personal meanings of mathematical objects have barely been studied. Our analysis shows the didactic systems' degrees of freedom to select the elements of meaning, and the variety of final states that personal meanings may achieve. More research into when, how, and how long we can introduce validation and institutionalisation moments to transform technical knowledge into technological and theoretical statistical knowledge is also needed.

ACKNOWLEDGMENT

This research has been funded by the DGES (MEC, Madrid), Project PS96-14111.

REFERENCES

Batanero, C., and Godino, J. D. (1998). Understanding graphical and numerical representations of statistical association in a computer environment. ICOTS 5, Singapore.

Brousseau, G. (1986). Foundaments et mÈthodes de la didactiques des mathÈmatiques. *Recherches en Didactiques des MathÈmatiques, 7(2)*, 33-115.

Godino, J. D. (1996). Mathematical concepts, their meanings, and understanding. In L. Puig, and A. GutiÈrrez (Ed.), *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education* (v.2, pp. 417-424). University of Valencia.

Godino, J. D., and Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactiques des MathÈmatiques, 14(3)*, 325-355.

Shaughnessy, J. M., Garfield, J. B., and Greer, B. (1997). Data handling. In A. Bishop et al. (Eds.), *International handbook of mathematics education*.(pp. 205-238). Dordrecht: Kluwer.