ROLF BIEHLER, BIELEFELD

STUDENTS' DIFFICULTIES IN PRACTICING COMPUTER-SUPPORTED DATA ANALYSIS - SOME HYPOTHETICAL GENERALIZATIONS FROM RESULTS OF TWO EXPLORATORY STUDIES[1]

1. The context and methodology of the studies

In this paper, I will report and summarize some preliminary results of two ongoing studies. The aim of this paper is to identify problem areas and difficulties of students in elementary data analysis based on preliminary results from the two ongoing studies.

The general idea of the two projects is similar. Students took a course in data analysis where they learned to use a software tool, used the tool during the course, and worked on a data analysis project with this tool at the end of the course. The course covered elementary data analysis tools, such as variables and variable types, box plots, frequency tables and graphs, two-way frequency tables, summary measures (median, mean, quartiles, interquartile range, range), scatterplots, and line plots. The grouping of data and the comparison of distributions in the subgroups defined by a grouping variable was an important idea related to studying the dependence of two variables. The methods for analyzing dependencies differed according to the type of variables: for example, scatterplots were used in the case of two numerical variables, and two-way frequency tables and related visualizations were used in the case of two categorical variables.

I have been interested in students' knowledge and competence in using the software tool for working on a data analysis task. For this purpose, students were provided with data and given related tasks. The two studies differed in their basic design. In the "*Barriers project,*" students were directly interviewed with regard to the data with which they were familiar from the course and which they had used as basis for a class project. This design allowed the researchers to focus on preconceived problem areas. In the *"CoSta project,"* students were allotted approximately one hour for working in pairs on the data and the task before interviewers entered and discussed the results of their inquiry with them. This design provided more room for exploration of the data by the student pairs. However, the subsequent discussion was very dependent on the students' results. In both studies, the interviewers adopted a tutorial or teacher role to an extent that was not intended in the interviews' original design.

The *Barriers project* is a collaborative project between C. Konold (University of Massachusetts, Amherst) and H. Steinbring (University of Dortmund, Germany). The students involved were 12th graders at an American high school who had completed a statistics course that used the software *DataScope* (Konold & Miller, 1994) and was partly based on material with activities developed by Konold. The dataset contained more than 20

---

[1] This paper is identical with the pages 169 - 179 of the paper: Biehler, R. (1997). Students' difficulties in practicing computer supported data analysis - Some hypothetical generalizations from results of two exploratory studies. In Research on the Role of Technology in Teaching and Learning Statistics. J. Garfield and G. Burrill. (eds.) Voorburg, International Statistical Institute**:** 169-190. The pages from 179 - 190 of the original paper are summarized in this paper by providing the major theses only without detailed explanation. (Reprinted with permission.)

variables related to a questionnaire that was administered to approximately 120 students. The questionnaire asked the students how they spend their time outside school, about their family, their attitudes, and so forth. The anonymous data contained responses from the students in this class as well as from other students in their school. Students were interviewed at the end of the course about a project they had completed during the course, as well as about other aspects of data analysis. During the interview, the students continued to work on the data. The interviewer adopted a tutorial role by directing the students' focus and questioning their choice of method and result interpretation. The students worked in pairs and the process was videotaped and transcribed.

In the second project, "Cooperative statistical problem solving with computer support" (*CoSta*), I observed student teachers who had attended my statistics course where the emphasis was on descriptive and exploratory statistics. The software *BMDP New System for Windows* (BMDP, 1994) was used in the course. As part of the course assessment, all students were required to complete an oral and written presentation. After the course, four pairs of students volunteered for an extra session where they worked on a statistical problem. The dataset given to these students concerned the number of traffic accidents in Germany in 1987. Frequencies were provided for every day of the year, with differentiated information concerning the various street types and the type of accident (with or without injured victims). The daily number of injured or killed was also provided. The entire process-working on the task, presenting the results to the interviewers, the interview, and discussion-was videotaped.

We are currently analyzing the interviews, video tapes, and transcripts from different perspectives, including (1) the role of difficulties with elementary statistical concepts and displays, (2) the type of statistical problem solving, and (3) how the students' work is influenced by the computer as a thinking tool.

How the students' work is influenced by the computer as a thinking tool can be analyzed by identifying interface problems with the software, by observing how students cope with the weaknesses of the software, and by analyzing how the computer influences their thinking and behavior in detail. The results with regard to the software are interesting because they partly confirm but also partly contradict or add clarification to our current understanding of requirements for software tools designed to support the learning and teaching in an introductory statistics course (see Biehler, 1997). In this paper, I will not discuss results with regard to the third perspective, but will instead concentrate on the first two perspectives (i.e., the role of difficulties with elementary statistical concepts and displays and the type of statistical problem solving).

I will use some aspects of the videotaped episodes to demonstrate and argue for a basic problem; that is, the intrinsic difficulties of "elementary" data analysis problems that we give students or that they choose to work on. Analyzing what students do while at the same time reflecting on the possible solutions "experts" would consider may bring us a step closer to determining what we can reasonably expect from our students in elementary data analysis and where we can expect to encounter critical barriers to understanding. The videos from the *Barriers project* are currently being analyzed from other perspectives, such as from a psychological point of view (Konold, Pollatsek, Well, & Gagnon, 1997) and from the perspective of an epistemologically-oriented transcript analysis perspective (Steinbring, 1996). Preliminary joint discussions on the transcripts have influenced the following analysis.

In the following analysis, I will mainly concentrate on one task and one part of a recorded interview (episode) from the *Barriers project.* The generalizations I offer are also shaped by experiences and preliminary results from other episodes and the *CoSta* project. I will identify 25 problem areas related to elementary data analysis. The "expert view" on exploratory data analysis (EDA) and the task analysis are based on an analysis of important features of EDA for school teaching (Biehler, 1992; Biehler & Steinbring, 1991; Biehler & Weber, 1995).

2. Curfew, study time and grades in school: an annotated episode

The episode analyzed in this section is taken from two student pairs of the *Barriers project*. I shall concentrate on one episode to provide examples for my analysis. The analysis compares elements from the work of two student pairs and compares this to what we as "statistical experts" would have considered a "good" solution to the problem. I try to identify "obstacles" that students encounter. The extent to which these obstacles are generalizable and adequately explained is not known, although experiences and results of other studies have contributed to shaping the formulation presented here.

One of the problems the students of the *Barriers Project* selected to investigate was **"Does having a curfew make you have better grades?"** This formulation has a "causal flavor." The result of such an analysis may be relevant to parents' decision making or for students who want to argue about curfew with their parents. As part of their analysis, the variable *hours of homework* was grouped with the binary variable of having a *curfew* (no/yes). The students compared the distributions under the two conditions with several graphs and numerical summaries and found no "essential" difference. They combined their statistical analysis with common-sense hypotheses about why curfews are imposed and on the role curfews might play in academic achievements.

2.1 Defining the problem

The students' own formulation of this problem contains a "causal" wording (i.e., "make you"). It is not atypical for students to be interested in causal dependencies and in concrete decision making (e.g., can we argue against parents who want to impose a curfew?). Similarly, causal relations are present in the media where (statistical) research studies are quoted that seemingly support such claims.

It is important to study how students conceptualize and define the problem they want to analyze, before they use the computer to arrive at some (partial) answer. One student of the *Barriers project* expressed a revealing causal-deterministic chain of reasoning to support her interest in the curfew hypothesis:

> "I mean if you had a curfew, would you study more, would you have more time to sit down and like actually have an hour. Say okay, you have two hours and in those two hours, I just do my homework and nothing else and if you didn't have a curfew, you have more liberty, so would do more as you please and less homework, less studying. So that's kind of what I meant like. I, so what diff--I wanted to see what happened. So, if you studied more, did you have better grades, if you studied less, did you have--you know like, I was assuming that if ...you had a curfew, you were doing more studying, if you didn't have a curfew, you were doing less studying."

From the research question, the students derived a plan to compare the study time of those who have a curfew with those who do not have a curfew. They expected that a difference in

study time would support the hypothesis that curfew has an "effect" on study time and vice versa. A statistical expert would know that such a rush to conclusions is problematic in an analysis of observational data, because other possibly interfering variables may also be relevant. A difference would point to indications, which would increase the evidence, but definite conclusions cannot be drawn.
We can formulate the first problem area as:

(1) Students seem to expect that results of analyzing observational data can directly be interpreted in causal terms. However, results of a statistical analysis may be much weaker, especially if we analyze observational data. A reflection on the status of expected results should be part of defining a problem and of interpreting results.

The way of conducting data analysis in the classroom may be partly responsible for this obstacle. If students are given data analysis tasks with observational data the talk of "effects" of one variable on another one may be nothing more than a *façon de parler* introduced by the teacher for group comparisons. It is likely that students may interpret this as meaning "effect" in the causal sense if this is not discussed in the classroom.
The propositions stated by the female student (presented above) do not show any probabilistic or stochastic elements; that is, there are no formulations such as "will tend to," "are more likely," or "in general." She may have had something like that in mind, but used more common language for the sake of simplicity. Common language does not support statistical reasoning as well as it supports deterministic reasoning. However, other interviews show that students sometimes said "‘tend to’ do more homework.” A more elaborated way of describing a possible relation is as follows: Study time is dependent on many factors, one of them could be *curfew*. Imposing a curfew may have very different effects on the study time of different students, however. Even if students think that imposing a curfew may increase the *tendency* to study and that this tendency would reveal itself in a different *distribution* of study time in the curfew group, this would also be a superficial conceptualization.

(2) Students use common language and the idea of linear causal chains acting on individual cases to make sense of the situation. They do not use the idea of a multiplicity of influencing factors where an adequate design has to be chosen to find out the effects of imposing a curfew. Why should a comparison of groups with and without curfew throw light on this question at all? This critical question is not posed by the students.

It may be necessary to help students develop qualitative statistical-causal cognitive models (Biehler, 1995). Mere data analysis may only provide superficial insights. What may be required in "upgrading" students' cognitive models is a problem that has not yet been sufficiently analyzed.
In the next step, the students used the data to gather information in order to answer their question. The students examined the data base that contained the two relevant variables: the binary variable *curfew (yes/no)* and the variable *HW: hours of homework*, a numerical variable that contains an estimate of the number of hours devoted to homework weekly. The students used several data analytical methods for studying "dependencies" (e.g., scatterplots for two numerical variables or grouped box plots or frequency displays for studying the dependence of a numerical variable on a categorial variable).

In this step, the students replaced studying the original complex question with studying the differences in the distribution of *HW* grouped by the variable *curfew*. This replacement was probably not a conscious refinement and reduction but rather may have been suggested by the situational constraints of the experiment. The situation reduced the problem space in several ways: (1) students used the data given instead of thinking what data they would like to collect to answer their question, and they did not notice the limitations of the observational data for their causal question; (2) students searched the available variables in the data base for a match with their verbally-formulated question (actually, the question was chosen with regard to the variables available); the process of transforming words into statistical variables was cut short; and (3) nobody questioned whether a statistical analysis was reasonable at all. Other methods such as interviewing parents or students may be better methods. Teachers and students should be aware of the limitations of using statistical methods. If we apply qualitative interpretative methods in our educational research we should also be especially aware of these alternatives when we teach statistics to our students. Moreover, global differences between student groups with and without curfews may not matter to parents who have to decide whether to impose a curfew on their child under very specific circumstances.

The replacement of the subject matter question by a statistical question remained partly unnoticed and became a source of misunderstandings between the interviewer and the students. This indicates a general obstacle that is raised in the classroom, too: whereas the teacher may be thinking in terms of variables and statistical relations, the students may use the same words, such as "curfew," without thinking in terms of a "binary variable." Obviously, an operationalization of the verbal formulation of "having a curfew" could be different from a yes/no definition. Weekend or nonweekend curfews could be distinguished, or we could take into account the time when students have to be at home. In teaching mathematical modeling, we frequently emphasize the importance of distinguishing between the real world situation/problem and the mathematical model/problem. This clarification may also help in the statistical context. The scheme shown in Figure 1 illuminates necessary transformations between the stages, and the necessity to evaluate results in the light of the original problem. The system of variables collected in the data base is comparable to a reduced idealized model of a real situation.

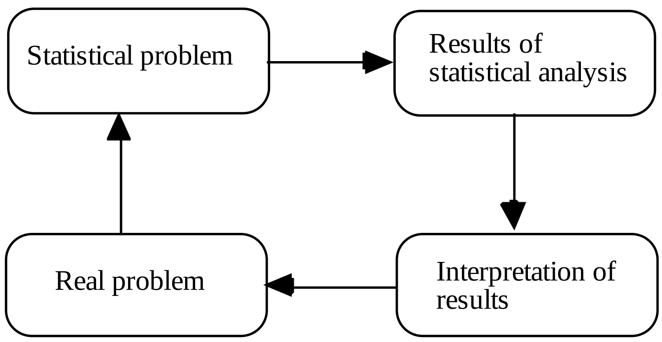

Figure 1: Cycle of Solving Real Problems With Statistics

(3) Genuine statistical problem solving takes into account and deals with the differences and transformations between a subject matter problem and a statistical problem and between the

results of a statistical analysis and their interpretation and validation in the subject matter context. When these differences are ignored, misunderstandings and inadequate solutions become likely.

I have already argued that the situational constraints of a task given to students may not be optimal for promoting a development of metacognitive awareness of this difference (i.e., the difference between a real problem and a statistical problem). These limitations are reduced when students are involved in the entire process of defining (constructing) variables and collecting data (Hancock, Kaput, & Goldsmith, 1992). How to cope with this problem when students are only asked for an analysis of available data is currently unknown.
The above problem is not limited to educational situations. For instance, Hornung (1977) admonished analysts to distinguish between experimental and statistical hypotheses and between the level of the statistical result (significance) and what this may say about the original real problem. It often remains unclear whether "rejecting a hypothesis" is a proposition on the level of the statistical problem or on the level of the real problem. More generally, we find a widespread simplistic view about the relation of formal mathematical (statistical) methods to subject matter problems (see Wille, 1995, for a critique). Some people think that formal mathematical methods can completely replace subject matter methods; however, frequently formal mathematical methods only deserve the status of a "decision support system." At one extreme, we find people in practice who use statistical methods for solving real problems as if they were solving artificial textbook problems in the classroom. However, the relation between subject matter knowledge and statistics is a difficult problem. Different traditions in statistics, such as the Neyman-Pearson school versus the tradition of EDA, differ with regard to this problem; for example, EDA allows context input in a more extensive flexible way (Biehler, 1982).

2.2 Producing statistical results

During the interview segment, all the displays and tables the software *DataScope* offers for comparing the *yes* and *no* curfew groups were produced; that is, frequency tables, bar graphs (histogram), box plots, and a table with numerical summaries (these were all grouped by the variable curfew). Our interview and video documents show that the process of selecting the first method or display and of choosing further methods and displays varies among students-some superficially trying out everything, others making reflective choices on the basis of knowledge and insight they had acquired. Most often though, students seemed to jump directly to particular methods offered by the software tool (means, box plots) without much reflection. The research problem here is the reconstruction of different patterns of software use in the context of a data analysis problem. Two basic problems can be summarized as follows:

(4) Superficially experimenting with given statistical methods is a first step. But how can we improve the degree of networking in the cognitive repertoire of statistical methods? In particular, students have to overcome the belief that using one method or graph "is enough."

(5) Software tools with ready-made methods influence the way a subject matter problem is conceived of and is transformed into a "statistical problem" and into a "problem for the software." This phenomenon can be exploited for developing students' thinking. However, later it is also

> necessary to reflect on these limitations and transcend the constraints of the tool. How can we achieve this step?

Let us think about what a good model of use would be. What would (or should) an "expert" do? The expert will conceptualize or classify our problem as "comparing distributions." For this purpose, several comparison tools are cognitively available: box plots, frequency bar graphs with various resolutions, numerical summaries, one-dimensional scatterplots (and probably other displays such as cumulative frequency plots or QQ-plots, as well as tools from inferential statistics). An expert will have knowledge and experience about the relation of these tools, especially about their relative virtues and limitations. Generally, an expert will know to experiment with several tools because each tool shows different aspects of the data or aspects of the data in different perspectives. Using only one tool will be not sufficient.
Experts operate within a *networked cognitive tool system* and recognize the *model character of a tool* or display. For instance, experts will know that several outliers with the same value will be shown in a box plot as only one point and that box plots cannot directly show big gaps in the main part of the data. An expert would also be aware of the differences of his/her cognitive statistical tool system and the tool system that a concrete software tool offers. For example, an expert may think that a jitter plot would be the best display for a certain distribution. If this were not available, an expert would use a combination of box plot, histogram, and dot plot or generate a jitter plot by using the random number generator together with the scatterplot command. An expert would also be aware that there may be differences in defining a certain concept or procedure in statistics in general and in a software tool in particular [e.g., the various definitions and algorithms for quartiles that are in use (Freund & Perles, 1987)]. Basically, we have to be aware of the subcycle shown in Figure 2.

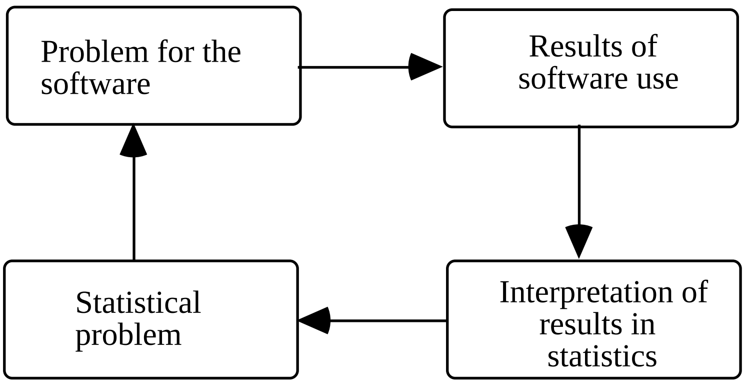

Figure 2: Subcycle of Computer-Supported Statistical Problem Solving

Experts would probably conceptualize the situation as "comparing distributions," reflecting their cognitive tool system, and then use the computer-based tool system in a reflective way (i.e., they would understand when the computer tools are not adequate and understand the possible distortions and changes when progressing from a real problem to a statistical problem to a computer software problem). In contrast, we can often reconstruct in our students a direct jump from a real problem to a problem for the software without an awareness of possible changes. Again, students are sometimes satisfied with producing computer results that are neither interpreted in statistical nor subject matter terms. Such a degenerate use of software for problem solving, where it only counts that the computer "does it," has also been reconstructed in other contexts (Krummheuer, 1988).

The degree of networking in some students' cognitive tool system seems to be rather low, otherwise the trial and error choice of methods that we observed quite frequently would be difficult to explain. Moreover, some students seem to look for one best display, when more than one display may be required.

Sometimes we can reconstruct episodes that show that students feel the need for a display not available in the software; that is, they try to transcend the system of available computer-implemented tools. Students express such needs fairly vaguely, probably because they have no command of a language necessary to express the design of new graphs. This could be due to the habit of teaching them the use of only those graphs that are already computer implemented, without sharing with the students why and how these specific graphs have come to be constructed.

2.3 Interpreting results

A characteristic feature of exploratory data analysis is the multiplicity of results.

> (6) Students have to overcome the obstacle that a data analysis problem has a unique result. However, it is difficult to cope with the multiplicity of results even at an elementary level.

Even if we compare two distributions, we can use various displays and numerical summaries, there may be contradictions, and students have to relate the various results and make some kind of *synthesis*. The term "data synthesis" was introduced by Jambu (1991) to emphasize that a new phase of work begins after the production of a multitude of results. However, even a single display such as the box plot contains an inherent multiplicity: It allows the comparison of distributions by median, quartiles, minimum, maximum, quartile range, and range. The selection and synthesis of these various aspects is not an easy task for students. An even simpler example of dealing with multiplicity is when distributions are compared by means and by medians-Should we choose one of them? Are both measures relevant? How can we understand differences if they occur? These questions are difficult for students (and teachers).

The difficulties that writing statistical reports pose to students are well-known; however, it is not only the limited verbal ability of high school students that is responsible for these problems. Not only superficial reading or writing will lead to distorted or wrong results. Our documents suggest that the description and interpretation of statistical graphs and other results is also a difficult problem for interviewers and teachers. We must be more careful in developing a language for this purpose and becoming aware of the difficulties inherent in relating different systems of representation. Often, diagrams involve expressing relations of relations between numbers. An adequate verbalization is difficult to achieve and the precise wording of it is often critical.

> (7) There are profound problems to overcome in interpreting and verbally describing statistical graphs and tables that are related to the limited expressability of complex quantitative relations by means of common language.

I now return to our interview to show some interpretation problems with elementary graphs. In the course of one interview in the *Barriers project*, the students produced a frequency bar graph (see Figure 3), but did not find it very revealing ("It is confusing").

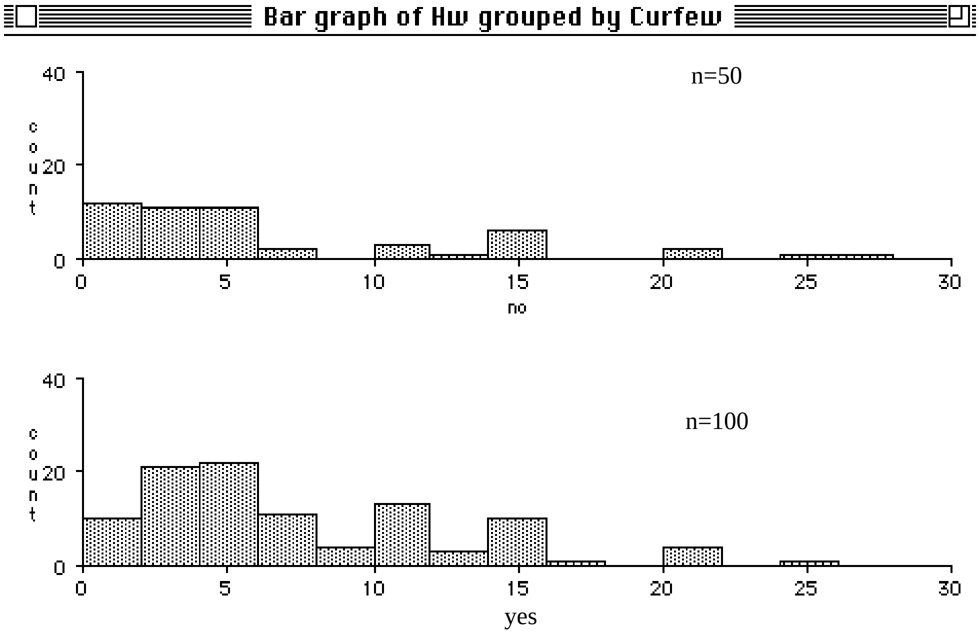

Figure 3: Histograms with Absolute Frequencies of *Homework* (in hours)

Some students even had difficulty in "reading off" basic information. The histogram for continuous variables in Figure 3 has an underlying display scheme that is different from the categorial frequency bar chart. In the histogram, the borders of the classes are marked, whereas in the categorial bar chart the category name (which could be a number) is shown in the middle below the bar. It seems that some of the students interpreted the above graph with this "categorial frequency bar chart" scheme in mind. For example, the “5” under a bar was interpreted in the sense that there are only data with the value “5” *in* the bar. Bars with nothing written below were difficult to interpret. There was a similar confusion of graphical construction schemes with regard to box plots. We may conclude that, independent of the newly taught schemes, students attempt to make sense of graphs by using graph construction schemes from other contexts. Thus, the notion that we must be more careful about our instruction of distinguishing among different types of axis in elementary graphs is reinforced. The software *TableTop* (Hancock, 1995) offers a carefully designed possibility here for changing among different types of axis that may be very helpful for beginners.
However, not only the high school students had problems here. Most of the student teachers in the *CoSta* project felt more "uncomfortable" with the continuous variable histogram than with the categorial frequency bar chart. The student teachers had various difficulties related to the relation between relative and absolute frequencies, and the various resolutions when changing the interval length of the grouping system. It could be a good didactical idea to distinguish "maximum resolution bar graphs" that show the entire raw dataset from "histograms" that are based on grouping the data and are thus only a summary of the data.
The fact that the computer hid the grouping of the data from the user could be hypothesized as a source of difficulty. The histogram is a very simple case from the expert's view.

However, the problem that users of a mathematical tool forget the "meaning" of a certain display or method is a general one.

(8) Students tend to forget the meaning of statistical graphs and procedures, and, often, the software tool does not support them in reconstructing this meaning.

Thus, perhaps the software we use needs to be improved: Some possibilities include adding hypertext explanations including prototypical uses and pitfalls for every graph or method, offering related linked methods (e.g., showing what is inside a histogram bar by clicking on it), highlighting the data of one bar in other displays or tables, or suggesting "related methods" to be combined with the histogram. We must, however, improve teaching and resist the temptation to take implemented statistical algorithms and displays "as given" in the machine, forgetting that students have to construct the meaning of the methods in their minds. Students produced a box plot display of *HW* grouped by *curfew* (see Figure 4).

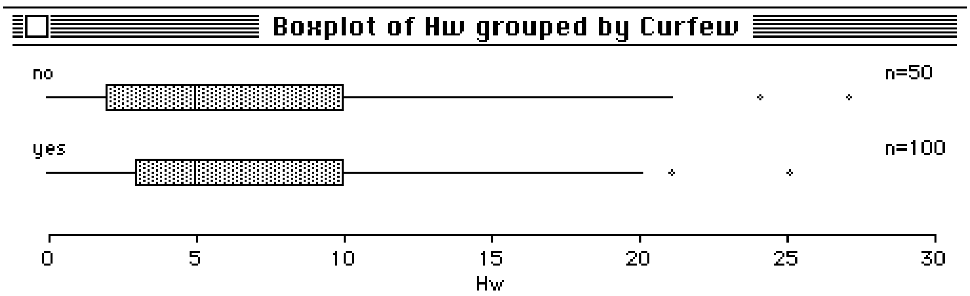

Figure 4. Boxplots of Homework (Weekly Hours) With and Without Curfew (Same Data as in Figure 3)

In this case, it was the interviewer's initiative combined with the easy availability of the box plot in the software that was largely responsible for choosing this display. Ideally, we want our students to know that one of the reasons whey the box plot was developed was that comparing frequency bar graphs can often be "confusing," especially if we have more than two bar graphs (see Biehler, 1982). Thus, it may be helpful to emphasize that the invention of the box plot was a solution to a problem.

(9) Students do not seem to appreciate that statistical methods and displays were constructed for serving a certain purpose.

Some of the students needed help in reconstructing which numerical summaries are displayed in the box plot and how they are defined. The different graphical conventions, namely that area and lines were both used to indicate the location of data and that area is not proportional to the amount of data, were a source of confusion.

(10) The graphical conventions underlying the definition of the box plot are very different from conventions in other statistical displays. This can become an obstacle for students. Moreover, a conceptual interpretation of the box plot requires at least an intuitive conception of varying "density" of data. This is a concept that often is not taught together with box plots.

After the interviewer clarified what the basic elements of the box plot represent, students faced further difficulties in interpreting the box plots shown in Figure 4. The dominant feature is that the box plots are the same with the following exceptions: the lower quartile is one hour less in the *no* group than in the *yes* group. There are two outliers in each group (maybe more, overplotting!). The end of the right whisker (signifying the maximum of the data without outliers) is one hour higher in the *no* group. The box as the visually dominant feature in the display conveys the impression that the spread (interquartile range) in the *no* group is higher than in the *yes* group. Which of the differences are relevant for the question of effects of having a curfew? This question was discussed by the students and the interviewer.

The students regarded the difference in the outliers to be irrelevant for the comparison ("Just because one studies 27 hours, the rest could study only 1 or 2 hours"). An expert would agree. But one student also rejected the difference in the lower quartile as relevant, because it ignores "the rest of the data." The equality of the median is accepted as an indication of no difference. Why? "Because, by average. You know on average, people studied 5 hours on both, with a curfew or without a curfew. So that would kind of be the median. That's right, yeah. Or, if you look at the mean..." (Note that the means are 6.44 hours for no curfew and 6.995 hours for curfew.) Reacting to the question of whether the mean uses all the data for comparison, one student said: "You're not using all the data but you're looking kind of averaging out, you know like looking at the average time that people spend studying, so you're using the whole data because you got to find one average."

We can see the interesting point that "comparison by average" seems to be a basic acceptable choice for the students; intuitive conceptions like averaging out seem to play a role in this. It would be interesting to explore this further. The students were asked to comment on mean or median but only referred to the mean; thus, we suspect that they may have less confidence in using medians for comparison. This observation was also made with the *CoSta* students. Moreover, the possibility that box plots offer--the simultaneous comparison according to different criteria-is not really used and accepted by the students as a part of their tool system.

(11) Establishing the box plot as a standard tool for comparing distributions is likely to conflict with "acceptable everyday heuristics" of comparing distributions or groups by arithmetic means (averages).

3. Further theses and problems related to students' statistical thinking

In this section, the inherent difficulties and obstacles in the above problem will be analyzed further. This complexity must be taken into account when designing problems and assessing students' performance and their cognitive problems. We summarize by providing a list of further problems.

(12) Choosing among various summaries in a concrete context requires knowledge of relations between distributional form and summaries, and of a functional interpretation of summaries (how they will be affected by various changes). Thinking about summaries only with regard to their value in empirical data distributions and not as properties of distributions as abstract entities may become an obstacle in data analytical practice.

(13) Even with the relatively elementary box plots, students will encounter a variety of unforeseen patterns in graphs in open data analysis tasks. Interpretation often tends to be difficult, may depend on the specific context, and may require substantial time before a satisfactory interpretation is achieved. Often, graphs will

be confusing even to experts. The search for interpretable patterns is natural but may not be successful, because they may not exist. The fact that many textbooks present easily interpretable box plots (or graphs in general) may serve to mislead students to expect that all plots are easy to interpret.

(14) Interpretations of summary statistics such as those represented in a box plot must take into account their different "reliability" and "robustness.” Sample size is important even when the data do not come from a random sample. Students generally lack the flexible knowledge and critical awareness of experts, which guides their behavior in such situations.

(15) Box plots can be used to see "properties of distributions" such as symmetry and skewness that cannot be well-defined in empirical distributions. Moreover, the concepts of symmetry and skewness are related to a classification of distribution types-the rationale of which is difficult to teach in elementary data analysis. For instance, experts will probably expect skew distributions for the variable homework, although this expectation would not be easily explainable.

(16) Conclusions depend on the statistical methods and displays that have been considered. Experts, aware of the limitations inherent in many summaries and the hermeneutic circle in data interpretation, consider alternative approaches. Students whose experience has consisted of well-defined textbook problems in a methods-oriented statistics course will not be prepared to appreciate this problem.

(17) Studying dependencies and possible "effects" in observational data is part of the agenda in elementary data analysis courses-but how do we cope with the problem of "lurking variables"?

(18) Statistics establish propositions about differences between "groups.” The relevance of group differences to evaluating individual cases is often not clear. If students are not able to distinguish between the group and individual level, they may run into problems when trying to interpret results. Statistical results and common sense judgments may become difficult to relate and integrate.

(19) Students have difficulties in relating abstract models of linear statistical-causal chains to studying frequency distributions under various conditions. Students conduct the data analysis study as they have learned in the classroom, but the classroom learning has not (yet) upgraded their cognitive statistical-causal modeling capability.

(20) The way teachers and students casually talk about box plots may come into conflict with frequency information that students read from histograms.

(21) The reasoning between "frequency "and "range for this frequency" in the case of the box plot is inverse to the corresponding reasoning with regard to histograms. This conceptual difficulty is exacerbated because it is difficult in common language to express the two different numerical aspects of a proposition such as "the frequency between 5 and 7 is 30%."

(22) Comparing multiple graphs such as box plots or histograms requires coordinated use of the defining concepts as well as the development of new concepts that are specifically adapted to the comparison of distributions.

(23) Statistics is concerned with empirical numbers. The question of how many digits should be taken seriously depends on the context. Metaknowledge is necessary for guiding data analysts. However, the orientation towards exact numbers in traditional mathematics instruction may become an obstacle for adequate behavior in statistical applications.

(24) If students have only learned a number of specific graphs, they may run into difficulties in various situations where more general knowledge of principles of good statistical graph construction is required.

(25) Interpretation of graphs and tables that are more than a mere reading off of coded information requires a rich conceptual repertoire.

4. Perspectives

We hope that the further analysis of our documents will contribute to a reshaping and sharpening of the 25 problem areas that I have defined above. A further clarification and identification of adequate didactical provisions for overcoming these difficulties or for redefining goals for teaching elementary data analysis is a task for future research and development projects.

Acknowledgment

Part of the work on which this paper is based was supported by the National Science Foundation (RED-9452917) who gave a grant for the common *Barriers project* with Cliff Konold and Heinz Steinbring. I am grateful to Heinz and Cliff with whom I discussed episodes of the transcripts of the *Barriers project* that helped to shape my hypothetical generalizations I have presented in this paper. I wish to thank an anonymous referee for his/her helpful suggestions. I am indebted to Cliff Konold for his detailed comments and constructive suggestions on an earlier version that helped much to improve and revise this paper. However, I am still responsible for all the weaknesses and partly unwarranted speculations and hypotheses.

References

BEHRENS, J. T. (1996). Using GEESC: A graphical environment for exploring statistical concepts. In C. Batanero (Ed.), *Research on the role of technology in teaching and learning statistics. 1996 IASE Round Table Conference Papers* (pp. 119-130). Granada: University of Granada.

BIEHLER, R. (1982). *Explorative Datenanalyse - Eine Untersuchung aus der Perspektive einer deskriptiv - empirischen Wissenschaftstheorie. IDM Materialien und Studien 24* [Exploratory data analysis - an analysis from the perspective of a descriptive-empirical epistemology of science]. Bielefeld: Universität Bielefeld, Institut für Didaktik der Mathematik.

BIEHLER, R. (1992). *Intendierte Anwendungen und didaktische Begründungen zu einem Softwarewerkzeug zur Explorativen Datenanalyse und stochastischen Simulation für Schule und Ausbildung* [Intended applications and pedagogical rationale for a software tool for exploratory data analysis and stochastic simulation for educational purposes.] München: Inst. f. Film und Bild in Wissenschaft und Unterricht (FWU).

BIEHLER, R. (1995). Probabilistic thinking, statistical reasoning, and the search for causes-Do we need a probabilistic revolution after we have taught data analysis? In J. Garfield (Ed.), *Research papers from ICOTS 4, Marrakech 1994.* Minneapolis: University of Minnesota.

BIEHLER, R. (1997). Software for learning and for doing statistics. *International Statistical Review*, *65(2)*. 167-189.

BIEHLER, R., & STEINBRING, H. (1991). Entdeckende Statistik, Stengel-und-Blätter, Boxplots: Konzepte, Begründungen und Erfahrungen eines Unterrichtsversuches [Statistics by discovery, stem-and-leaf, box plots: Basic conceptions, pedagogical rationale and experiences from a teaching experiment, an English versionis available from the author]. *Der Mathematikunterricht*, *37(6)*, 5-32.

BIEHLER, R., & WEBER, W. (Eds.). (1995). Explorative Datenanalyse. *Computer + Unterricht, 17*.

BMDP (1994). New System 1.0. BMDP statistical software, 1440 Sepulveda Blvd, Suite 316, Los Angeles, CA 90025, (MS-Windows)

FREUND, J. E., & PERLES, B. M. (1987). A new look at quartiles of ungrouped data. *The American Statistician, 41*, 200-203.

HANCOCK, C., KAPUT, J. J., & GOLDSMITH, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, *27*, 337-364.

HANCOCK, C. (1995). *TableTop Software* (Ver. 1.0) [Computer program]. Novato, CA: Brøderbund Software Direct.

HORNUNG, J. (1977). Kritik des Signifikanztests [Critique of significant testing]. *METAMED, 1*, 325-345.

JAMBU, M. (1991). *Exploratory and multivariate data analysis*. London: Academic Press.

KONOLD, C., & MILLER, C. D. (1994). *DataScope* (Ver. 1.4).) [Computer program]. Santa Barbara, CA: Intellimation.

KONOLD, C., POLLATSEK, A., WELL, A., & Gagnon, A. (1996). Students analyzing data: Research of critical barriers. In: J.Garfield & G. Burrill (Eds) Research on the role of technology in teaching and learning statistics. Voorburg, The Netherlands: International Statistical Institute

KRUMMHEUER, G. (1988). *Die menschliche Seite am Computer* [The human side with computers]. Stuttgart: Teubner.

STEINBRING, H. (1996). *The epistemological analysis of interactive mathematical processes of communication--theoretical background and example of analysis.* Unpublished manuscript, Universität Dortmund.

TUKEY, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

WILLE, R. (1995). Allgemeine Mathematik als Bildungskonzept für die Schule [General mathematics as a conception for teaching mathematics in schools]. In R. Biehler, H.-W. Heymann, & B. Winkelmann (Eds.), *Mathematik allgemeinbildend unterrichten - Impulse für Lehrerbildung und Schule* [Teaching mathematics with a view towards goals of school education in general - New impacts for teacher education] (pp. 41-56). Köln: Aulis.

Dr. Rolf Biehler
Universität Bielefeld
Universitätsstr. 25
33615 Bielefeld