

WORKING STYLES AND OBSTACLES: COMPUTER-SUPPORTED COLLABORATIVE LEARNING IN STATISTICS

Rolf Biehler

University of Kassel, Germany
biehler@mathematik.uni-kassel.de

Pairs of students use the computer software Fathom for working on problems from Exploratory Data Analysis. The exploratory study was interested in identifying how the software as a tool supports or hinders students' thinking. Working styles of students related to distributional thinking in the context of group comparison tasks were studied.

INTRODUCTION

We have been designing and evaluating several experimental courses in probability and statistics at high school level (Biehler, 2003). Our students use the software *Fathom*TM as a tool and teacher-prepared Fathom worksheets as part of their learning environments. The courses emphasize Exploratory Data Analysis, simulation and modeling and simulation – supported inference statistics. We are interested in observing the micro-processes in front of the computer. We pick out a couple of pairs in a course and record their collaborative work. A screen recording software with a microphone is used for capturing human-human and human-machine interactions.

Our research interest focuses on how the tool fosters or hinders statistical problem solving, what mental models students construct about the software tool and which styles of use are emerging. We are interested in obstacles and breakdown points where the students have to extend and adapt their mental models. The obstacles may be related to the statistical content, the software or the interface between both.

CONTEXT OF THE STUDY

The pairs of students whom we carefully observed had taken part in a 4 weeks long experimental course on Exploratory Data Analysis at the JGS School in Kassel in grade 13. We used one large data set in the teaching, the so-called MUFFINS data set that we collected in various schools. The data set included data on 671 students. One hundred thirty three students were from the JGS school and 538 were from a survey of students from schools in the federal state of Nordrhein-Westfalen (NRW). They served for comparative purposes. The JGS data were collected in 2003 in grades 11 to 13 whereas the data in NRW were collected in 2000 and 2001 in grade 11. The data contained several variables related to leisure time activities. Among others, students were asked several questions concerning their computer use. As part of the course, the students had to work on a little project concerning a self-chosen question related to this data.

After this course, we worked with 5 pairs of students about one hour. The average age of these students is 19 years. We gave them some open data analytical questions for solving in pairs, afterwards they had to present to the interviewers and were given some more structured items to work on. We were interested in their distributional thinking and in their strategies of solving group comparison tasks. A preliminary analysis of part of this data was done by Schäfer (2003).

OUR TASKS FOR THE STUDENTS

The students of the MUFFINS data set had been asked among others to estimate the weekly time (in hours) they spend on various activities with the computer such as total amount, doing homework, programming activities, games, and internet.

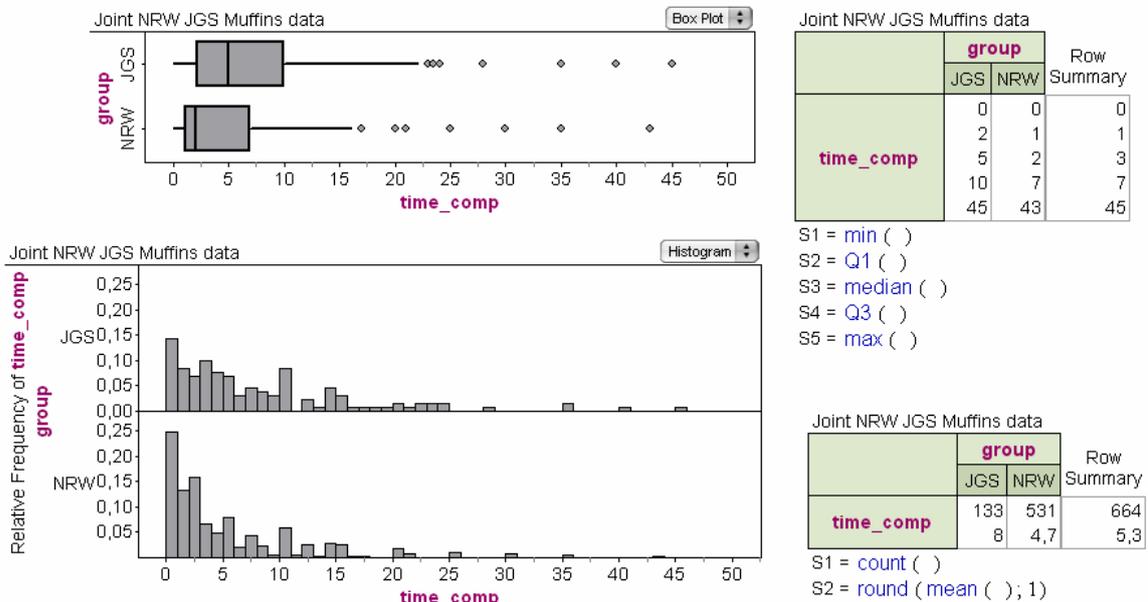
Item 1. Which differences do you expect between JGS and NRW students related to the different attributes? Which differences do you expect between the distributions and why?

A table with 9 attributes was presented to the students. They were asked to write down their expectations, and conjectures about factors that may have caused the expected differences.

Item 2a. Analyse the attribute *time_comp* (weekly hours of computer use) and compare the JGS and the NRW students.

Item 2b. Do the comparison for the JGS and the NRW students for each sex separately and compare the results. Use filters for selecting subgroups if appropriate.

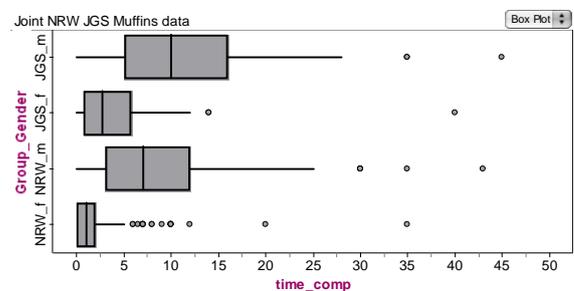
The students got 10 minutes for item 2a, for instance, and were asked to present the results to the interviewers afterwards. The following set of diagrams could have been produced by the students for working on item 2a. They show the differences of the distributions.



Item 2b required new statistical thinking with 3 attributes and a more complex use of the software from the students. The display we used as a starter for item 3 below shows the changes and differences students would see, when using box plots.

Item 3a. How would the box plot of the young females of JGS (*JGS_f*) look like if they all just spend 1,75 more at the computer than the young women of NRW [1,75 is the median difference]?

Item 3b. Why does the real box plot look differently than the hypothesized box plot.



DISTRIBUTIONAL THINKING

Results of Item 1. We explicitly asked about distributions to evoke distributional concepts. The results in this respect were disappointingly clear. Statistical concepts were used in none of the 45 responses (5 pairs, 9 variables), neither mean or average nor formulations such as “tend to” were used. 4 of the 5 pair expected that the JGS students will spend “more” time on using the computer, one pair expected no change. Although the formulations differed in details, none of the group changed their everyday language style. The pairs differed considerably with regard to the reasons they gave for the expected change. The major arguments were related to the time difference between 2000/01 and 2003. In 2003 as compared to 2001, we have better access to internet and wider availability of computers at home. Another factor could be that JGS students were older than the NRW students. We interpret this result as an indication of anchoring distributional thinking in everyday language practice.

STRATEGIES IN GROUP COMPARISON

Results of Item 2a. Group comparison was practiced in the classroom but not standardized with regard to the approach. We were interested what working styles had emerged and whether distributional thinking can be observed in a more elaborated way.

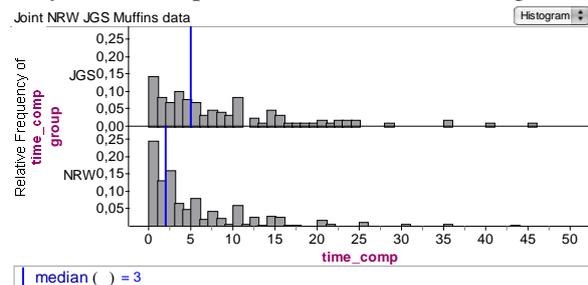
The software as a supportive tool. The software *Fathom* makes group comparison with one diagram very easy. All four of the above diagrams can be easily generated by putting the analysis variable on one of the axes and the grouping variable on the other axis that is perpendicular to it. All students used this facility. Defaults of the software influenced the analysis in various respects. When a grouping variable is added to a histogram with absolute frequencies, *Fathom* by default generates histograms with absolute frequencies. Too often students used these histograms instead of changing to relative frequencies. Different from younger students, our students became immediately aware of the inadequacy of using absolute frequencies for comparative purposes, when prompted. *Fathom* was supportive with regard to interactively changing the display type. Students often used this facility and practiced “graph hopping.”

Organizing before analyzing. However, nobody did arrange multiple diagrams for comparison similar to our above graph although the software supports the spatial arrangement of various diagrams in one interactive worksheet. The changing (and replacing) of displays obviously was an easier default for our students than producing multiple diagrams. Readable tables with rounded numbers and where information is split into several diagrams were used in our above diagram. We split the table into two, rounded the value of the mean to one decimal digit, and arranged the group categories horizontally to make the *Fathom* tables more readable than in the version where the group is assigned vertically to the summary table. We did not observe such refinements in students’ work: tables and graphs often tended to become overloaded.

Data analytical working styles. Although several examples of data analysis were presented in the classroom the strategy or working style was not something that was explicitly to be learned. This partly explains the large variety of different styles we observed.

Pair 5 adopts a strategy of *successive interactive refinement in the analysis of distributional differences* with a high level knowledge of relations between summary statistics and diagrams. First they calculate the mean and observe the mean difference between both groups. Then they add the median and conjecture that there may be outliers explaining the lower medians. They argue: “... a median of nearly 5 and a mean of 8 means that many people say that they spend not much time at the computer but that some spend very much time.” The students seem to correctly construct and interpret the difference between mean and median as a measure of skewness, which was not part of the teaching unit. After this, they check the distribution shape with histograms. Detailed comparisons are not successful because they compare absolute frequencies. They become aware of this limitation and intend to compare “proportions” and “quartiles.” Although they do not remember how to display relative frequencies in the histogram, they manage to use the summary table for this purpose. They “see” a higher spread in the JGS histogram and check this with box plots in the next step. They carefully analyse the box plots and recall the meaning of the outliers, whiskers and borders of the box.

Although these constructions are partly not correct the episode shows the students’ careful meaning construction activities. The carefulness of meaning construction is also indicated by the following episode. They plotted the median into the histograms as in the diagram to the right. *Fathom* shows the value of the median of the whole data set



whereas the medians are plotted separately for each group. The students try to understand this and conjecture that the total median is the average of the two group medians. Although this is not precisely true, this shows a successful constructive attempt in the correct direction. Later on, they also carefully compare computed proportions to proportions displayed in box plots and histograms.

Pair 1 had an elaborated “*picking out the best*”- strategy. From each display they picked up what they thought was best displayed in the particular graph: they used summary tables for exact numbers especially mean and median, box plots for spread and outliers, histograms for popular values. They discovered the high frequency of students in the class from 0 hours to less than 1 hour in the histogram. In contrast to pair 5, a very detailed comparison of different displays could not be observed. Pair 1 invented themselves an adequate strategy of group comparison,

namely using cut-off points and comparing in both groups the relative frequency left or right from these cut-off points. The display to the right is quoted from their work. The NRW group uses the computer less, because 54% of them use them less than 3 hours whereas only 29% of the JGS group uses computers less than 3 hours, and so on. This on the one hand shows how the openness of the software supported students' constructions. On the other hand we may interpret this type of comparison as a certain resistance to work with the usual statistical concepts of quartiles, measures of spread and location (Bakker *et al.*, 2005). Despite the their elaborate strategy the students did not use box plot and histogram as global distributional displays.

Joint NRW JGS Muffins data

	group		Row Summary
	JGS	NRW	
	133	531	664
time_comp	0,29323308	0,53672316	0,48795181
	0,48120301	0,27495292	0,31626506
	0,13533835	0,047080979	0,064759036

$$S1 = \frac{\text{count}(\text{time_comp} < 3)}{\text{count}(\text{time_comp})}$$

$$S2 = \frac{\text{count}(\text{time_comp} > 5)}{\text{count}(\text{time_comp})}$$

$$S3 = \frac{\text{count}(\text{time_comp} > 15)}{\text{count}(\text{time_comp})}$$

Three pairs (2, 3, 4) more or less consider their tasks finished after they had calculated and compared means and medians. We can reconstruct their strategy as a “*hypothesis testing strategy*.” If the means or medians already show the direction and amount of the difference, what else should they study to confirm their conjecture? Describing the group difference as a difference of means or medians (increase of 3 hours) was natural and not a problem as it was observed with earlier children (Konold and Pollatsek, 2002). Further diagrams were generated mainly in reaction to an interviewer prompt. Either they were regarded as merely ornamental or they were just analyzed with regard to the expected differences. Typical is the following episode from pair 2 who conceives their task more as a hypothesis testing task: “it is clear already (!) from the numbers that the computer use has strongly increased in the recent two years ... the means show it even clearer. The difference is 3.3 [and the median difference is 3 only].” Further displays are questioned with regard to whether they confirm the hypothesis of the difference. Pair 3 refers to the median difference indicating that the JGS students use the computer much more, and then: “This is confirmed by the histogram that shows that a large bulk has nearly no computer use.” Again, the students do not see this information as a different aspect of the distributional difference to be explored but just as further evidence for their qualitative hypothesis.

CONTEXTUAL INTERPRETATIONS

Centre. For our students, location and centre is the most natural group property that can be related to qualitative changes that they expect due to various factors. Hypotheses such as “JGS students spend more time at the computer” can be seen as supported by the shift of location and centre. The amount of change is measured by the difference of means or medians.

Spread. Spread differences are interpreted as a by-product of differences in location only by some of the pairs. If the quartiles change differently than the measures of centre – as in this example – a different spread is just resulting. For instance, pair 3 comments the box plot after the interviewer has prompted them. “Here we are seeing a larger spread [in the JGS group], because (!) there is a shift to the right, ... the large bulk is shifted to the right... The spread is larger in the JGS group. This is due to the fact that the JGS data are more distributed, not so concentrated between 0 and 2...” In a descriptive manner, increased spread can be just interpreted as “more diversity” in a group. At an explanatory level one may speculate about reasons for increased diversity. One could argue that the increased availability of computers and internet has increased diversity as some do make hardly no use of it whereas others take up the opportunities. A descriptive and explanatory interpretation of spread was not evoked at all in this item. We had to prompt students in item 3 (below) in order to get better results.

Outliers. Outliers can be interpreted as due to exceptional circumstances, in this case by extreme habits as extended internet surfing, chatting or playing computer games. This hypothesis could have been further explored by checking how these extreme students' had answered to other questions such as time for computer games. This is only done by one pair who attributed the outliers to “computer freaks” namely to King of Case players. This remains merely speculative. The students didn't take up this opportunity for a deeper contextual analysis.

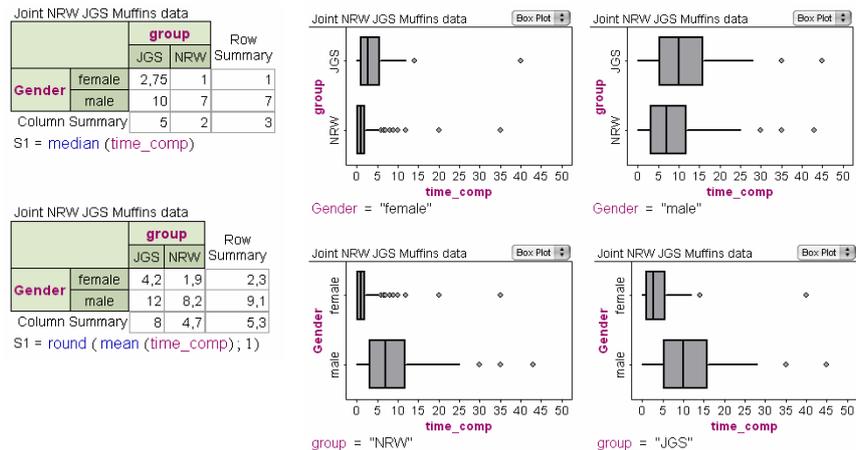
Shape as such was not a concept taught in the unit. This explains that some of the pairs struggle with phenomena that we would describe as skewness or varying density without being able to conceptualize them adequately. The data are skewed to the right. If the data were distributed according to the lognormal distribution (they are very close) one could hypothesize that the data come from a stochastic process with many small multiplicatively interacting causes just as symmetrical distributions that are close to the normal can be hypothetically explained as being generated by a process of many small additively interacting little causes. It is clear that we did not observe such interpretations as this was not taught in the classroom.

Changes in shape. It at all, this is brought to the attention to our students by comparing box plots. In other studies we had found the notion of shift of the box plot helpful, as well as the distinction of uniform and non-uniform shifts. But these concepts were not part of the teaching unit at school, partly because the teacher was hesitating to introduce too much new terminology.

A contextual interpretation of shape changes is possible but was not observed with our students. The non-uniform shift from the NRW to the JGS data might be interpreted as due to something like the following mechanism. Assume that the underlying boundary conditions in the year 2000/01 have always produced data of computer use similar to the NRW data. If these conditions change there maybe a subgroup of the students who change their computer habits only a little bit whereas others change it dramatically. Therefore we expect more diversity, and we may expect more diversity especially in the relative heavy users of computers as compared to the low users. This may explain the shift of the median, means, and upper quartile by three hours per week, whereas the shift of lower quartile is only 1 hour and the minimum has not changed at all.

COPING WITH STATISTICAL AND COMPUTATIONAL COMPLEXITY

Item 2b introduced complexities. In statistical respect we have a three variable problem and with regard to the software there are several solutions and the situation is new for the students and they have to adapt their software knowledge to it. Working on item 2b could have led to the diagrams.



The software as a supportive tool. We get the first pair of graphs by starting with a box plot split according to group. Then filters for male or females are added. The box plots can be produced similarly. We can easily generate a two-way table of Gender and group with frequency counts. Then we can replace the formula $\text{count}()$ by any other formula. This formula will be evaluated in the four subgroups separately. The latter option requires a clear and more advanced knowledge of the behaviour of two-way summary tables that was not discussed in the classroom. This task made a lot of implicit problems observable, especially limited mental models of the software and limited experience with complex comparisons.

Filters. Filters have to be added locally to the graph. Students tried filters on the data table (no effect), on the collection (is o.k. but it is impossible to produce two graphs with different filters), on the summary table (option not available in the context menu). *Drag and drop facilities.* Students tried to just drag and drop the third variable into a graph or a summary table. This is an *overgeneralization* of the experience they had made with *Fathom* so far, however, this does not give satisfactory results here. A third variable in a histogram is interpreted as a legend attribute, in the box plot it is not possible, a third variable in a summary table has different effects, depending on where it is dropped. Such overgeneralizations are quite common in other uses too.

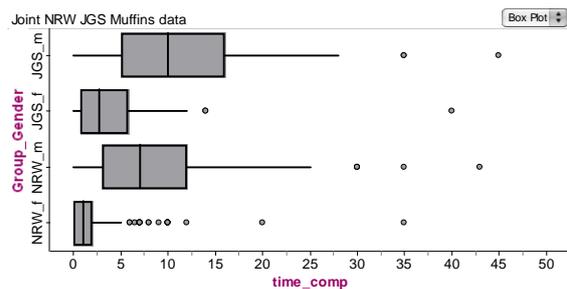
Organizing before analyzing; data analytical working style. Whereas the working style of interactive switching between default displays was at least of some success in earlier items, in this

item this style led to a complexity nearly all pairs could not cope with. 4 out of 5 pairs focus on two summary tables with means, medians and other statistics, one for females and one for males. Graphs are used in a second step, but usually not systematically analyzed and compared with results of the summary tables. Most pairs finally used only one table.

Interpretations. Contextual interpretations are nearly absent. For challenging students' interpretative capabilities it would have been better to choose an item where the pre-organized graphs and table would have been given to them so that they could have concentrated on interpretation. Interpretation activities focussed on selectively picking out information from the tables. A reasonable strategy would be concentrating on one summary statistics first, for instance on the means, and afterwards do some refinements with the median, measures of spread and further statistics. We did not observe this in pure form; all pairs focused on their table and then made some unsystematic refinements. Comparisons can be made additively or multiplicatively. The result of the comparison of comparisons will depend on this choice. If we notice that the ration of the mean males to the mean of the females is 3:1 with JGS but 4:1 with NRW, then the gender discrepancy is higher in NRW. But if we take an additive stance we notice that the mean difference of males and females is 8 in JGS and only 6 in NRW. Students argued either multiplicatively or additively and some pairs did not make a comparison of comparisons but only single comparisons. There is another option, namely comparing males in NRW to males in JGS and doing the same with the females. This again is another perspective, and our students generally chose only one the perspectives. The choice between multiplicative and additive comparisons and the issue of comparison of comparison, and different possible perspectives of comparison turned out to be a complicated issue, to which more learning should be devoted in advance.

CONCLUSIONS: PROMPTS FOR DISTRIBUTIONAL THINKING

In item 3a the students got a second graph with a missing box plot where they had to sketch the hypothesized one. This prompt enabled 4 of the 5 pairs to generate reasonable explanations why we may see a non-uniform shift. Pair 5 has the clearest statement: "it is as follows, maybe some are really working 1,75 more, but others work 2 hours more, others only half an hour more, and therefore the spread is higher with the JGS women, although one says that all spend 1,75 hours more on the computer on average." We have to include more of these tasks in our teaching for developing distributional thinking. Moreover we may wish to separate interpretation tasks from software related tasks in order to better develop the respective competencies.



REFERENCES

Bakker, A., Biehler, R., and Konold, C. (2005). Should young students learn about box plots? In G. Burrill and M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education 2004 Roundtable*, (pp. 163-173). Voorburg, the Netherlands: International Statistical Institute, www.stat.auckland.ac.nz/~iase/publications/rt04/4.2_Bakker_etal.pdf.

Biehler, R. (2003). Interrelated learning and working environments for supporting the use of computer tools in introductory courses. *Proceedings IASE conference on Teaching Statistics and the Internet*, Berlin, www.stat.auckland.ac.nz/~iase/publications/6/Biehler.pdf.

Konold, C. and Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33, 259-289.

Schäfer, T. (2003). *Schülerkompetenzen beim Arbeiten mit statistischer Werkzeugsoftware (Fathom) nach einem Unterrichtsversuch zur Explorativen Datenanalyse*. Masters thesis, University of Kassel.

Fathom Dynamic Statistics™ Software. Emeryville, CA: Key Curriculum Press.