CONSTRUCTING AN UNDERSTANDING OF DATA GRAPHS[1]

Dani Ben-Zvi
The Weizmann Institute of Science, Rehovot, Israel

I describe episodes of two 13-year-old students working on Exploratory Data Analysis (statistics) developed within an innovative curriculum. I analyze the microevolution of their incipient understandings of some features of graphs as data representations. The description includes the role of the instructional materials, the students' discussions and collaborative attempts to solve the tasks, and the teacher's intervention. Although her intervention seemed to be a miscommunication, it appears to have helped the students to make sense of their tasks.

BACKGROUND

The teaching of *Exploratory Data Analysis* (statistics) is mostly based on: (a) organization, description, representation and analysis of data, with a considerable use of visual displays (Shaughnessy et al., 1996); (b) a constructivist view of learning (Garfield, 1995); and (c) incorporation of technological tools for making sense of data and facilitating the use of various data representations (Biehler, 1993).

With these perspectives in mind, we developed a middle school statistics curriculum[2] (Ben-Zvi & Friedlander, 1997a), implemented it in schools and in teacher courses, and undertook research on learning (Ben-Zvi & Arcavi, 1997; Ben-Zvi & Friedlander, 1997b). The curriculum is characterized by: (a) a use of extended real (or realistic) problem situations; (b) collaboration and communication in the classroom; and (c) a view of the teacher as “a guide on the side” (Hawkins et al., 1992). The students pose, collect, analyze, interpret data, and communicate (Graham, 1987) using a spreadsheet. The classroom activities are semi-structured investigations, in which students, working in pairs, are encouraged to hypothesize about possible outcomes, choose tools and methods of inquiry, design or change representations, interpret results, and draw conclusions.

THE STUDY

A pair of 13-year-old students (*A* and *D*) was videotaped at different stages of their learning statistics (20 hours of tapes). I focus here on 15 minutes of their work with brief teacher interventions. The students were considered by

1 Ben-Zvi, D. (1999). Constructing an understanding of data graphs. In O. Zaslavsky (Ed.), *Proceedings of the 23rd Annual Conference of the International Group for the Psychology of Mathematics Education, II,* 97–104. Haifa, Israel: Technion.

2 The project is part of *CompuMath*, an innovative and comprehensive curriculum (Hershkowitz & Schwarz, 1997).

their teacher to be both very able and very verbal. They were asked to talk aloud and explain their actions.

The purpose of the following analysis is to study how students construct their understanding of graphs as displays of real life data, and learn to design them to support certain claims. I used interpretive microanalysis (see, for example, Meira, 1991, pp. 62-3) to try to understand students' discussions, considerations, difficulties and solutions. In this analysis I consider socio-cognitive aspects, taking into account verbal, gestural, and symbolic actions, in the context in which they emerged -- comparing and contrasting the data with other pieces of data, written records, and conversations with the teacher.

The Problem Situation

The extended (four lesson) activity - *The same song, with a different tune* - occurs early in the curriculum. The context is the Olympic 100 meters race. The students were given, in a spreadsheet, the men's 100 meters record times, and the years in which they occurred (from 1896, the first modern Olympiad, to 1996). In the first part of the activity, the students were introduced to the context of the investigation and were asked to describe the data graphically and verbally. In the second part, the students were asked to manipulate data graphs, i.e., change scales, delete an outlier, and connect points by lines. In the third part, they were asked to design graph to support the following claim: *"Over the years, the times recorded in the Olympic 100 meters improved considerably"*.

In the following, I present and analyze the students' work through the activity.

DEVELOPING UNDERSTANDING OF DATA GRAPHS: THE 'STORY' OF *A* AND *D*

In this section, I present three parts of the activity chronologically: (a) getting acquainted with the context, (b) acquiring tools, and (c) designing graphs.

(A) Getting Acquainted with the Context

In the first part of the activity, *A* and *D* analyzed the table of results, compared the records of consecutive Olympiads, considered the issue of extreme data, sorted the data, and created a graph with a spreadsheet (Figure 1). In their written summaries, they wrote that (a) the best record is 9.48 sec. and the worst is 12 sec., (b) the greatest improvement is from 10.25 to 9.48 sec., and (c) the differences between records are not constant. The first two conclusions are wrong: the best record is 9.84 sec., and the greatest improvement is from 12 to 10.8 sec. When requested to describe the data *patterns*, they did not seem to understand the meaning of the question. With the teacher's help, they concluded correctly that "the record times seem to improve, yet there was occasionally a lower (slower) result, than the one achieved in previous Olympiads".

Although *A* and *D* seemed to notice the general trend of improvement in the records, their view was mostly local and focused on discrete data points, or, at most, on two consecutive records. I claim (based on data not detailed here) that their difficulty to discern general data *patterns* was caused by: (a) the students' lack of experience with the notion of *pattern*; (b) the discrete nature of the graph; (c) the non-deterministic and disorganized nature of statistical data, which is very different from the deterministic formulae, they had met in algebra.

(B) Acquiring Tools

In the second part of the activity, the students became acquainted with three strategies for manipulating graphs (changing scales, deleting an outlier, and connecting points), and considered the effect of these changes on the shape of the graphs. The objective was to prepare for the *design* task (Part C below).

<u>Changing Scales</u>

The following transcript describes the students' comments on the effect of changing the vertical scales of the original graph from 0-12 to 0-40 (Figures 1& 2):

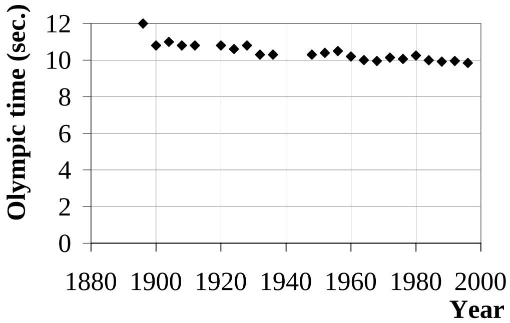

Figure 1: Given graph

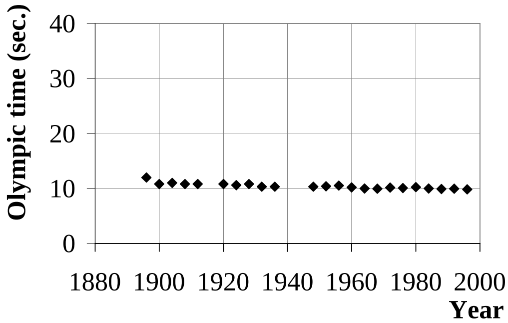

Figure 2: Manipulated graph

A. *Now, the change is that… that the whole graph stayed the same in shape, but it went down…*

D. *The same in shape, but much, much lower, because the column* [the y-axis] *went up higher. Did you understand that?* [*D* uses both hands to signal the down and up movements of the graph and the y-axis respectively.]

A. *Because now the 12, which is the worst record, is lower. It used to be once the highest. Therefore the graph started from very high. But now, it* [the graph] *is already very low.*

The students' perception of the change is restricted to the overall relative position of the graph; they considered the shape itself as remaining "the same". Their description includes: global features of the graph ("The whole graph … went down"), an interchange of background and foreground (the graph went down and/or the y-axis went up), and local features (12 as a "starting point" of the graph). These descriptions are linked and complement each other. *A* wrote the following synthesis in his notebook: “The graph remained the same in its shape, but moved downward, because before, 12 - the worst record - was the highest number on the y-axis, but now it is lower”.

<u>Deleting an Outlier</u>

In this task, the students were asked to delete an outlying point (the record of 12 sec. in the first Olympiad, 1896) from the graph (Figure 2), and describe the effect on its shape. First, *D* justified why 12 can be considered an outlier:

D. It [the record of 12 sec.] *is pretty exceptional, because we have here* [in the rest of the data] *a set of differences of a few hundredths, and here* [the difference is] *a whole full second.*

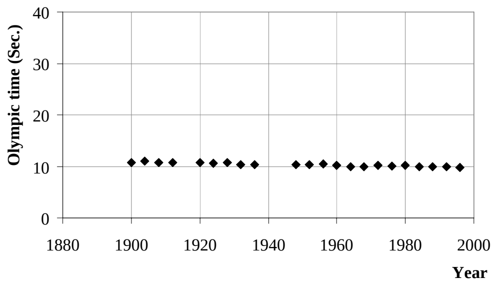

Figure 3: Outlier deletion

Then, they struggled to interpret the effect of the deletion on the graph (Figure 3):

D. The change is not really drastic … Now, however, the graph looks much more tidy and organized.

A. One point simply disappeared. The graph… its general shape didn't change.

They wrote in their notebooks different descriptions of the change: “The graph became straighter” (*D*); “One point in the graph disappeared” (*A*). Thus, the students struggled between different views of the effect: global and significant change (the graph is tidy and organized), no change at all (the general shape didn't change); or just a mere description (one point disappeared).

Although the dispute about the outlier was not resolved, it served another purpose: it drew *A*’s attention to a mistake in their conclusions in the first part of the activity, and corrected it: “the greatest improvement is from 12 to 10.8 seconds”.

<u>Connecting Points</u>

In the third task, they were asked to connect the points to obtain a continuous graph. The new graph (Figure 4) elicited many comments from the students, who tried to make sense of what they saw. They were particularly intrigued by the fact that the connected graph included both the original points, and the connecting line.

D. OK. You see that the points are connected by lines. Now, what’s the idea? The graph did not transform to one line. It transformed to a line, in which the points are still there. It

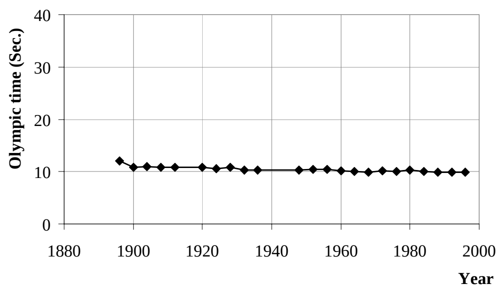

Figure 4: Connecting line

means that the line itself is not regarded as important.

A. *This line is OK. We previously thought that if we connect the points with a line, they might disappear. But now, there is a graph, and there are also the points, which are the important part.*

In their view, the connecting line (as provided by the spreadsheet) not only did not add any new meaning, but also contradicted the context, as *D* observed: “Olympiads occur only once in every four years” (namely, there is no data between the points). The students did not see the line as an aid to detect or highlight patterns in the data, and this is consistent with their previous difficulties in recognizing data patterns.

So far *A* and *D* were practicing manipulations (changing scales, deleting an outlier, and connecting points), and discussing their effect on the graph's shape. The intention was to provide students with the means to design a graph, in order to support a particular claim. In the following section, I discuss in what sense this preparation helped them achieve this purpose.

(C) Designing Graphs

I present here a fragment of the students' work on the third part of the activity. The students were asked to design a graph to support the statement: "*Over the years, the times recorded in the Olympic 100 meters improved considerably"*. I bring first a teacher intervention, which eventually helped the students understand the task. Then, I focus on five attempts (*Stages 1-5* below) to obtain a satisfactory form of the graph.

The Teacher Intervention

A and *D* did not understand the task and requested the teacher's (*T*) help:

T. [Referring to the 0-40 graph displayed on the computer screen -- see Figure 4.] *How did you flatten the graph?*

A. [Surprised] *How did we flatten it?*

T. *Yes, you certainly notice that you have flattened it, don’t you?*

D. *No. The graph was like that before. It was only higher up* [on the screen].

The teacher and the students are at "loggerheads". The teacher assumes that the students (a) had made sense of the task, but just did not know how to perform it, (b) had acquired the necessary tools, and understood their global effect on the graph's shape to be used to support the claim. Thus, her hint consisted of reminding them of what they had already done (scale change). However, the students did not regard what they had done, as changing the graph's shape. Although this intervention seemed to be a case of miscommunication, it apparently had a catalytic effect, as reflected in the dialogue, which took place immediately afterwards:

T. *How would you show that there were very very big improvements?*

A. [Referring to the 0-40 graph -- see Figure 4.] *We need to decrease it* [the maximum value of the y-axis]. *The opposite…*[of what we have previously done].

D. *No. To increase it* [to raise the highest graph point, i.e. 12 sec.].

A. *The graph will go further down.*

D. *No. It will go further up.*

A. *No. It will go further down.*

D. *What you mean by increasing it, I mean - decreasing.*

A. *Ahhh… Well, to decrease it… OK, That's what I meant. Good, I understand.*

Even though their use of language is not completely clear, their previous perception that the graph shape remains the same was not mentioned at this stage. Moreover, *D* expressed what appears to be a new understanding:

D. *As a matter of fact, we make the graph shape look different, although it is actually the same graph. It will look as if it supports a specific claim.*

At this point, *D* seems to discern that a change of scales may change the perceptual impressions one may get from the graph. Thus, they seemed to understand the purpose of the activity, and started to focus on its goal. In the following, the students' five attempts to design corresponding graphs are presented.

<u>*Stage 1 (The scales are changed to x: 1880-2000; y: 0-5)*</u>

D suggested (Figure 4) changing the scale on the y-axis to 0-11. It seems that he chose 11, since he had previously deleted the outlier, making 11 the maximum data point. They didn't implement this change, because he immediately proposed another scale change: 0-5. This suggestion seems to be based on his assumption that the smaller the range the larger the decline in the record time would look *(Idea I).* However, when they implemented this change, the graph disappeared.

A. *We don't see the graph at all, since there is no graph in 5.*

<u>*Stage 2 (x: 1880-2000, y: 0-12)*</u>

Having failed to present a new graph, they returned to the 0-12 range (see Figure 1):

A. *The graph looks more curved, because the difference between records is much bigger, since we increased the… now the "Olympic time in seconds"* [y-axis] *is from 0 to 12, and every record– as much as it descends – it is bigger than the record… the line is more…*[*D.* interrupts] *Wait a second, the line is bigger than it used to be from 0 to 40.*

The effect of changing scales on the graph's global features (straight, curved), which were not noticed initially, and started to be considered after the teacher's intervention, were now being fully considered. Still, the students struggle to verbalize and explain what they do, or want to do.

<u>*Stage 3 (x: 1896-1996, y: 0-12)*</u>

At this stage, it seemed that *A* and *D* had exhausted the changes on the y-axis. So they turned to the x-axis. *D* suggested changing the upper limit of x from 2000 to 1000 (*Idea I* above). They realized, however, that this would cause the graph to disappear again (the year's range is 1896-1996). Thus, *D* proposed using 1996 (instead of 2000) as the upper limit of x. Although the effect[3] was marginal, *D* commented:

D. *One can really see, as if there are bigger differences in the graph… Very interesting!*

Although they had presumably understood how changing scales effects the graph's shape, *D*'s wrong impression of this horizontal change, seems to originate from his ambiguous distinction between vertical or horizontal "differences" and/or distances. However, having focused their attention on the x-axis, they realized that it does not start at zero, which triggered the following idea *(Idea II)*.

<u>*Stage 4 (x: 1896-1996, y: 8-12)*</u>

A transferred attention from the x-axis to the y-axis, and suggested changing the lower limit for y from 0 to 8 (to get a scale of 8-12). Observing the resulting sharp visual effect, he reacted immediately:

A. *It looks much bigger*.

<u>*Stage 5 (x: 1896-1996, y: 9.48-12)*</u>

D suggested applying *Idea II* to the x-axis, but withdrew, when *A* indicated that it already started at a non-zero value. Instead, *A* suggested using the minimum record time (9.48 sec.) as the lower limit of y *(Idea III)*. The resulting graph (Figure 5) satisfied them, and they made the following final comments:

D. *This way we actually achieved a result* [graph] *that appears as if there are enormous differences.*

A. *To tell you the truth, this booklet is lovely.*

D. *Right, it is nice!*

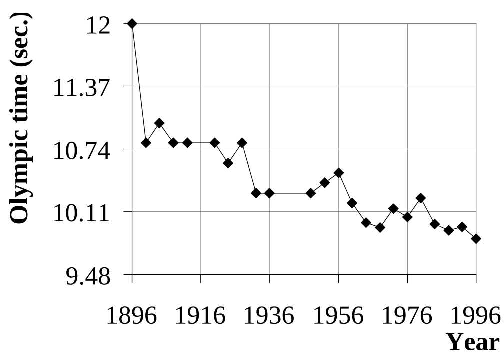

Figure 5: Final design

[3] The lower limit for y changed automatically to 1896, resulting in a final range of 1896-1996, instead of 1880-2000, which were the default values provided by the software.

DISCUSSION

This 'story' of *A* and *D* traces the microevolution of incipient understanding of some features of graphs as displays of real life data (see also Bright & Friel, 1997). It describes the students' perceptual development from a stage in which they did not understand the requirements of the task and the notion of data *pattern*, to the final successful completion of the *design* task. The following elements seem to have contributed to the construction of students' understanding of certain characteristics of data graphs.

Careful instructional engineering. The students worked with semi-structured guidance to solve open-ended questions. First, they acquired tools to modify graphs and then, they employed these tools in the design of graphs, to support a certain claim.

Close collaboration between the pair of students. The students:

a) verbalized almost every idea that crossed their minds. At times this spontaneous verbalization produced mere descriptions, but later served as stepping stones towards a new understanding, and at times, it served as self-explanation (Chi et al., 1989) to reinforce ideas;
b) complemented and extended each other's comments and ideas, which seems to have "replaced" some of the teacher's role in guiding their evolution;
c) decided to request the teacher's help when faced with a difficulty, which could not be resolved among themselves; and
d) transferred and elaborated, in iterative steps, ideas of changing scales, from one axis to the other.

The teacher's main intervention. At a first glance, the teacher's intervention to help the students make sense of the task, can be considered unfortunate. She did not grasp the nature of their question, misjudged their position, and tried to help by reminding them of their previous actions. The students, however, did remember the acquired tools, but perceived them differently.

Nevertheless, this miscommunication itself contributed to their progress. At first, *A* and *D* were surprised by her use of the notion of *flattening the graph* as a description of what they had done. Then, they started to direct their attention to the shape of the graph, rather than to its relative position on the screen. Although puzzled by the teacher's language, the students appropriated (Moschkovich, 1989) her point of view on what to look at. Their previous work and their "struggle" with language seems to have prepared them for the reinterpretation of what they had done, triggered by their teacher's comments.

In sum, the microevolution of the students' understanding of data graphs was influenced by the instructional engineering, the students' ways of making sense (descriptions, self-explanations, questions to a colleague and the teacher, transfer of ideas, etc.), and the teacher's intervention and the use students made of it.

Acknowledgment. I thank Abraham Arcavi and Alex Friedlander for their helpful comments and suggestions.

REFERENCES

Ben-Zvi, D., & Arcavi, A. (1997). Towards a characterization and understanding of students' learning in an interactive statistics environment. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics, Vol.* ***2***, 647-653. Voorburg, The Netherlands: International Statistical Institute.

Ben-Zvi, D., & Friedlander, A. (1997a). *Statistical Investigations with Spreadsheets* (in Hebrew). Rehovot, Israel: Weizmann Institute of Science.

Ben-Zvi, D., & Friedlander, A. (1997b). Statistical thinking in a technological environment. In J. B. Garfield & G. Burrill (Eds.), *Proceedings of the IASE Round Table Conference on Research on the Role of Technology in Teaching and Learning Statistics*, 45-55. Voorburg, The Netherlands: International Statistical Institute.

Biehler, R. (1993). Software tools and mathematics education: The case of statistics. In C. Keitel & K. Ruthven (Eds.), *Learning from Computers: Mathematics Education and Technology*, 68-100. Berlin: Springer.

Bright, G. W., & Friel, S. N. (1997). Students' (grades 6-8) understanding of graphs. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics, Vol.* ***2***, 655-661. Voorburg, The Netherlands: International Statistical Institute.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science,* ***13***, 145-182.

Garfield, J. (1995). How students learn statistics. *International Statistical Review,* ***63*** (1), 25-34.

Graham, A. (1987). *Statistical Investigations in the Secondary School*. Cambridge, UK: Cambridge University Press.

Hawkins, A., Jolliffe, F., & Glickman, L. (1992). *Teaching Statistical Concepts*. Harlow, UK: Longman.

Hershkowitz, R., & Schwarz, B. B. (1997). Unifying cognitive and sociocultural aspects in research on learning the function concept. In E. Pehkonen (Ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education, Vol.* ***1***, 148-164. Lahti, Finland: University of Helsinki.

Meira, L. R. (1991). *Explorations of Mathematical Sense-Making: An Activity-Oriented View of Children's Use and Design of Material Displays*. An unpublished Ph.D. dissertation. Berkeley, CA: University of California.

Moschkovich, J. D. (1989). Constructing a Problem Space Through Appropriation: A Case Study of Guided Computer Exploration of Linear Functions. Unpublished manuscript, available from the author.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International Handbook of Mathematics Education, part* ***1***, 205-237. Dordrecht, The Netherlands: Kluwer.