SEVENTH GRADE STUDENTS' SENSE MAKING OF DATA AND DATA REPRESENTATIONS

Dani Ben-Zvi
Weizmann Institute of Science and University of Haifa
Israel

The paper describes 7th graders' cooperative work on a data assessment task in a computer-assisted environment. The task was administered at the end of a carefully designed Exploratory Data Analysis (EDA) course. The purpose of the study is to assess students' ability to make sense of data and their representations: a) use of data analysis skills, and understanding of basic statistical procedures and concepts; b) if and how they adopted the dispositions and points of view of certain aspects of the EDA culture. The "local-global lens" is used to assess students' formulation of research questions and hypotheses, and use and interpretation of data representations.

INTRODUCTION

This study presents a data investigation assessment task administered to seventh graders after their engagement with the Statistics Curriculum (*SC*). The *SC* is a carefully designed 10-week Exploratory Data Analysis course (EDA, in the sense of Shaughnessy, Garfield & Greer, 1996) using spreadsheets, that introduces students to statistical thinking and the "art and culture" of EDA (Ben-Zvi & Friedlander, 1997). In the assessment task, students were asked to examine real data on the yearly number of immigrants to Israel. The analysis is proposed to describe students' statistical understandings and dispositions while formulating research questions and hypotheses, using data representations, and interpreting the findings of their work.

THEORETICAL PERSPECTIVES

Enculturation processes in statistics education. Briefly stated, *enculturation* refers to the process of entering a community or a practice and picking up their points of view. The beginning student learns to participate in a certain cognitive and cultural practice, where the teacher has the important role of a mentor and mediator, or the "*enculturator*" (Schoenfeld, 1992; Resnick, 1988). This is especially the case with regard to statistical thinking, with its own values and belief systems, and habits of questioning, representing, concluding and communicating.

Statistics experts' points of view. Wild and Pfannkuch (1999) provide a comprehensive four-dimensional description of the processes involved in the practice of data based enquiry from problem formulation to conclusions. The focus in this study is on one perspective of the expert's view: a flexible and dynamic shift between local observations and global observations of data.

Local and global views of data. Local understanding of data involves focusing on an individual value (or a few of them) within a group of data (a particular entry in a table, a single point in a graph). Global understanding refers to the ability to search for, recognize, describe and explain general patterns in a set of data (e.g., change over time, trends). Ben-Zvi and Arcavi (2001) showed that learning to look globally at data can be a complex process for students.

METHODOLOGY

Students. 80 mixed ability seventh graders (13-year-old girls and boys) in a progressive experimental school in Tel-Aviv, taught by skillful and experienced teachers, who were aware of the spirit and goals of the *SC*. The students were used to work collaboratively and held positive dispositions towards statistics, mostly through exposure to statistics jargon in the media.

The experiment and the data collection. Two weeks after the end of the *SC* course, students were asked to analyze the immigration data (provided in a spreadsheet table). They worked in pairs for 90 minutes in a computer lab using the "*Immigration to Israel*" worksheet (Figure 1) and spreadsheets. Minimal instructions were provided, to allow students to display their knowledge and abilities. Students furnished their worksheets and computer files for analysis. The teachers reported that most students engaged seriously with the task, and were independent in their work.

Methods. An extended *performance task* was used to better illuminate students' capacity to apply statistics in a complex and new situation. The analysis goals were to infer the degree of students' statistical reasoning and dispositions, in particular, their use of local/global views of data and their representations. To maximize the value of assessment, the task was administered in similar settings to those of the *SC*, namely, open-ended, computer-assisted data investigation in pairs.

IMMIGRATION TO ISRAEL

In this activity you are asked to examine real data about the numbers of immigrants to Israel since the establishment of the state (in 1948). With the help of the spreadsheet you will look for trends and interesting phenomena in the data. The data was obtained from the Statistical Abstract of Israel (1995).

Research questions and hypotheses

1. Suggest a research question and an hypothesis concerning the immigration to Israel.
2. The number of immigrants in 1949 was 240,000, and 76,000 in 1995. Based on these data, suggest an additional research question and a related hypothesis.

Trends

Use the computer to graphically display the immigration data and highlight trends.

3. What do you learn from the graph about the immigration to Israel?
4. Based on the graph, explain the meaning of the notion "immigration waves".

Figure 1. Excerpts from the "The Immigration to Israel" Assessment Task (freely translated from Hebrew).

Students' responses were systematically categorized and analyzed taking into account the type of explanations associated with them. The categorization scales were not pre-determined, but rather the result of a careful evaluation of responses, in order to authentically reflect students' types of reasoning and dispositions. In case of doubt in grading a response, it was assigned to a "lower" level category. To validate the analysis, parts of the data were analyzed by expert researchers, who discussed, presented, advanced and/or rejected proposed hypotheses, interpretations, and inferences to reach a consensus. Several findings are presented in the following sections. Quotes are freely translated from Hebrew; therefore they may not sound as authentic as in the original.

FORMULATING RESEARCH QUESTIONS AND HYPOTHESES

After a brief introduction to the context, the students were asked to formulate research questions and hypotheses (Figure 1). While the first question is open-ended, the second is more directive – aimed to "gently" guide students to focus on the task's target, by providing two "special" cases. Proposing research questions and hypotheses was a familiar task to the students from their *SC* experiences. Three categories of research questions and related hypotheses were identified.

A contextual question – aims at extracting various kinds of contextual information concerning the given situation: a) descriptive (what is?) e.g. *What is the origin of most of the immigrants to Israel?* b) predictive (what will be?) e.g. W*hat will be the number of immigrants during the next year?* c) prescriptive (what can be done about it?) e.g. *Can the number of immigrants be controlled in order to avoid a population explosion*? d) causal (why?) e.g. *Why the immigration rate seems to decrease with time?* Such questions represent students' inclination to better understand the context prior to formulating well-defined statistical questions. Building meanings is strongly weaved into the understanding of the investigation context.

A local question – focuses on an individual value (or a few of them) within a group of data. For example, *Which year had the largest immigration to Israel?*

A global question – focuses on general patterns in a set of data, such as change over time or trends, e.g. *What is the trend in the immigrants' numbers over the years? Is the number of immigrants associated with the number of Israeli citizens?*

These categories are not fully distinct. Contextual responses may indicate global reasoning, such as causal questions concerning trends. Moreover, some of the global responses involve a combination of local and global utterances. Naturally, the *SC* course aimed at "pushing" students towards global reasoning (intertwined with relevant local observations) – away from irrelevant and local reasoning. Table 1 summarizes students' responses to this part. A

majority (57.5%) formulated *global* research questions in response to Question 1, focusing mainly on trends and association in the data. One fifth of the responses were *local*, focusing on the number of immigrants in a specific year, or on a maximal yearly number. The *contextual* questions (22.5%) dealt with immigration issues, but were related to variables not provided in the data or to causes for a hypothesized phenomenon, which cannot be derived directly from the data.

Table 1 *Types of Research Questions (and Hypotheses) Formulated by Students*

Type of Research Question	Question 1 (Non-guided)		Question 2 (Guided)	
	N	%	N	%
Contextual	9	22.5	11	27.5
Local	8	20.0	2	5.0
Global	23	57.5	27	67.5
Total	40	100.0%	40	100.0%

When prompted by two data entries in Question 2, more students (67.5%) were able to formulate *global* research questions, and the *local* responses almost vanished (5%). Most of the *contextual* responses to Question 2 were causal ("why" questions) that dealt with relevant variables (number of immigrants and year), and came from students who were global reasoners in Question 1. For example, a pair of students, who hypothesized in Question 1 that *the number of immigrants decrease with time*, responded to Question 2 by simply asking, *Why does the number of immigrants decrease since the establishment of the state?* Thus, these students perceived the two pieces of data in the "focusing" question as a confirmation of their preliminary hypothesis.

Almost half of the students (49%) in Question 2 were more *global* than in Question 1. The types of changes were: a) focusing on the two relevant variables (50%), b) transferring from a contextual or local question to a global question (37.5%), and c) re-wording the first global question (12.5%). This data suggest that the absence of data in Question 1 made it harder to formulate statistical questions and hypotheses; and that the "focusing" data in Question 2 supported students' reasoning towards global view of the data.

INTERPRETING DATA REPRESENTATIONS

When asked to display the data, most students were fluent and independent in using the computer to produce graphs (time plots or bar charts, Figure 2). Many pairs experimented with different graphs, before they made their final choice. Furthermore, one-third made substantial changes in their graph (e.g., scaling, adding a regression line) to investigate trends. These actions can be considered as an indication of students' *graphicacy*, that is, skillful manipulations of graphs, as well as awareness of their role in conveying ideas by the visual impression they may induce.

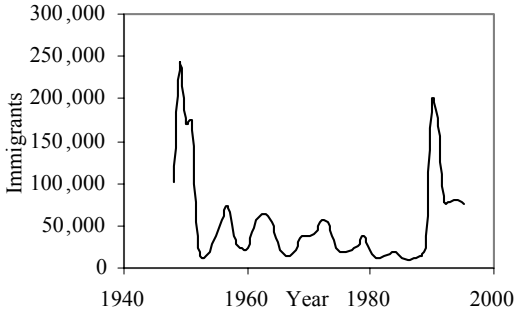

Figure 2. A Smoothed Time Plot Displaying the Immigration Data.

Students provided a variety of *local* or *global* graph interpretations (Question 3, Figure 1). The higher rate of global reasoners in this task, compared with the first task, could be related to the presence of a graphical representation that supports the construction of global meanings. The global interpretations were categorized to the following non-hierarchical categories.

Three-period interpretation - noticing, grouping and describing three immigration periods: "high", "low" and then "high" again. For example, *There were few periods in which the*

immigration thrived: the foundation of the state and 1990 onwards. In the rest - there was a stable, average immigration.

Four-period interpretation - noticing, grouping and describing four immigration periods in the data: "high", "low" , "high", and then "moderate". For example, *In the late forties, there was a mass immigration. The fifties saw a significant decrease; in the beginning of the nineties there was again a large immigration, which later became moderate*.

Focus on variability - interpretation that emphasizes the existence of variability in the data, without indicating any specific pattern in them, e.g. *The number of immigrants is not constant nor have any trends, but is rather variable*.

Cyclic interpretation - focuses on the cycles in the data. This category is discussed in the "*immigration waves*" task below.

General pattern interpretation - noticing and describing general patterns (trends or association), for example, *There is no association between the year and the number of immigrants. However, except for several outliers - it is a decreasing line*.

Most of the students (87.5%) were able to interpret the graph in a *global* way, attending to one of the following general features of the data: sub-groups (according to three or four subintervals), the existence of variability, cycles, trends or association (Table 2).

Table 2 *Types of Graph's Interpretations*

Type of Interpretation		N	%	Total
Global	Three periods	16	40.0	
	Four periods	4	10.0	
	Variability	3	7.5	
	Cyclic	4	10.0	
	General patterns (trends, association)	8	20.0	87.5
Local		5	12.5	12.5
Total		40	100.0%	100.0%

Some students combined local and global views of data in a "narrative" style, as illustrated in A and D's interpretation of their graph: *The number of immigrants per year is not constant. During the first years of the state, the number of immigrants was relatively high. Then, a long period followed, in which the number of immigrants increased and decreased between 50,000 and 70,000 - but did not arrive at the numbers of the first years. Later on, in recent years, the number of immigrants soared, then became moderate - higher than the previous years, but not as high as the first years of the state.* In this task, only five responses (12.5%) were strictly local/pointwise, e.g. *The number of immigrants in 1986 was the smallest*; *The largest number of immigrants arrived between 1948 and 1951*.

DEALING WITH CYCLES (" IMMIGRATION WAVES")

Given the distinct non-regular *cycles* in the data between 1952–1989 (Figure 2), students were asked to explain what "immigration waves" might mean, and suggest a graphical representation to highlight the phenomenon (Question 4, Figure 1). Their responses are presented in Table 3.

Table 3 *Types of Students Explanations of "Immigration Waves"*

Type	Interpretation of "Immigration Waves"	N	%	Total
Global	1. A succession of influxes of people immigrating, or successive ups and downs	11	27.5	
	2. The number of immigrants swells and dies away (a ridge) or vice versa	7	17.5	45.0
Local	3. A sudden rapid increase in the number of immigrants	12	30.0	
	4. Marked change in the number of immigrants	4	10.0	40.0
Other	5. Unclear or incorrect answer	4	10.0	
	6. Did not answer the question	2	5.0	15.0
	Total	40	100.0%	100.0%

The first response (27.5%) reflects an understanding of the circulatory nature of the given data. For example, *Immigration waves indicate that there are downs and ups in the immigration, a periodical phenomenon, for example, an increase in 1954–56, a decrease in 1956–58*. The second response (17.5%) focuses only on one complete cycle of an immigration influx. These types of explanations (45%) indicate students' *global* understanding, i.e., the ability to recognize and describe cycles on the basis of the variability in the data. Other responses handled the variation differently: In the third (30%) and fourth (10%) types, students referred to a sudden rapid increase, *like an ocean wave*, or a marked change in the number of immigrants, without explicitly stating the overall cyclic pattern. These responses were regarded as a partial understanding of the cycles with a local focus on the data.

DISCUSSION

Global views of data and data representations. Current studies in mathematics education show that students with a sound *local* understanding of certain mathematical concepts, struggle to develop *global* views (Monk, 1988). In statistics education, Konold, Pollatsek & Well (1997) observed that high school students - after a year long statistics course - still had a tendency to focus on properties of individual cases, rather than propensities of data sets. Our observations at the beginning stage of the *SC* (Ben-Zvi & Arcavi, 2001) show that students persistently emphasized local views of data in tables and graphs. They were attentive to the prominence of "local deviations", which kept them from dealing more freely with global views of data. Only later, the focus on certain pointwise observations, the gradual adoption of the notion of *trend*, and the exercise of scaling, helped them to direct their attention to the shape of the graph as a whole, taking into account the variability in the data. They gradually began to develop a global view of data and their representations, and flexibly integrate local and global views in data analysis.

The results of this study show that most of the students at the end of the *SC* held several aspects of experts' points of view on local-global approaches to data and data representations. In particular, in the context of a real and meaningful data investigation, students were able to formulate global research questions and hypotheses (67.5%), interpret graphs globally in a variety of ways (87.5%), and sensibly handle cycles in data, while independently learning about "immigration waves" (45%). Students were fluent in choosing a variety of graphs to display global features of data, and in manipulating representations (e.g., changing graph type, scaling), in order to better present their understanding of the data and argue their position. Global expressions were more frequent in the presence of graphical representations of data, and when prompted by "guiding" (yet open) questions. Even the minority that still remained attached to local views of data did better in the presence of supporting prompts and graphs.

I suggest that this relative success in adopting a *global view*, that is, the tendency to notice and describe generalities in data, was supported by: a) the *SC* emphasis on *enculturation* processes, i.e., entering and picking up the points of view of a community or culture, through interactions with a teacher, who plays an important role as an "*enculturator*"; b) the extensive and meaningful learning experiences in handling data within a purposeful context related to complex ideas in a social setting during the classroom activities and the preparation of a "research project"; c) the structure of the assessment task, which was similar in contents and style to the textbook's worksheets (but with fewer instructions); and d) the support provided by the computerized tool, which removed most of the technical load, to allow students focus on becoming *interpreters* of data and findings (Ben-Zvi, 2000).

Intertwining the local and global views of data. The above description may create a false impression that local and global views of data exist in sharp dichotomy, and that students' understanding mainly develops from the "less sophisticated" local view to the "more sophisticated" global view. However, this is *not* what this study tells us. The data shows that students functioned in various levels of understanding within their local or global views, and also variably combining both views. This was evident in the multiple ways they used their views to make sense of data and their representations, in formulating research questions and hypotheses, handling and interpreting graphs, and dealing with new concepts, such as immigration waves. I suggest that sophistication in students' understanding of data developed both within each point of view (local and global) and within the dynamic and flexible integration of those views. Thus, two

trajectories of development, which may occur simultaneously, can be traced: "*vertical*" – growth in sophistication within a view (local or global); and "*horizontal*" – growth in the webbing of the local and global views.

Assessment in a complex learning environment. Multiple challenges exist in the assessment of outcomes of students' work in such a complex learning environment: the existence of multiple goals for students, the mishmash between the contextual (real-world) and the statistical, the role of the computer-assisted environment, and the group vs. the individual work (Gal & Garfield, 1997). The assessment employed in this study consists in the use of an extended performance task in similar settings to those exercised during the learning, i.e., open, semi-structured questions, work in pairs, and use of computers. Although shown beneficial in many respects, this method still needs further investigation, in particular to find efficient ways to evaluate the knowledge and dispositions of the individual within a group (Hershkowitz, 1999).

Epilogue. Several lessons can be learned from this study. a) Evaluation of statistical reasoning must include cautious measures to uncover student's understandings that are not yet fully "expressible". The related pair of prompts (research question and hypothesis) and the "focusing" questioning technique may assist in this direction. b) Prior to dwelling on formulating questions and hypotheses and analyzing real data, students must be given opportunities to learn more about the context of the situation (e.g. by discussing, *What else would you like to know before you propose a research question?*). c) Tables' interpretation tasks must precede graphs' interpretation tasks to support the construction of meanings for data. The challenges in attending to the many issues raised here remain overwhelming. Supporting students' becoming familiar not only with statistical techniques, concepts and tools but also with the many nuances, considerations and points of view involved in generating, describing, analyzing and interpreting data and in reporting findings, is one of the greatest challenges in statistics education.

REFERENCES

Ben-Zvi, D. (2000). Towards understanding of the role of technological tools in statistical learning. *Mathematical Thinking and Learning*, *2* (1&2), 127-155.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, *45*, 35–65.

Ben-Zvi, D., & Friedlander, A. (1997). Statistical thinking in a technological environment. In J.B. Garfield and G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 45-55). Voorburg, The Netherlands: International Statistical Institute.

Gal, I., & Garfield, J. (Eds.) (1997). *The assessment challenge in statistics education*. Voorburg, The Netherlands: ISI.

Hershkowitz, R. (1999). Where in shared knowledge is the individual knowledge hidden? In O. Zaslavsky (Ed.) *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education*, *I*, 9-24. Haifa, Israel.

Konold, C., Pollatsek, A., & Well, A. (1997). Students analyzing data: Research of critical barriers. In J.B. Garfield and G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 151-168). Voorburg, The Netherlands: ISI.

Monk, G.S. (1988). Students' understanding of functions in calculus courses. *Humanistic Mathematics Network Journal*, *9*, 21-27.

Resnick, L. (1988). Treating mathematics as an ill-structured discipline. In R. Charles and E. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 32-60). Reston, VA: National Council of Teachers of Mathematics.

Schoenfeld, A.H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*, 334-370. New York, NY: Macmillan.

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*, *1*, 205-237. Dordrecht, The Netherlands: Kluwer.

Wild, C.J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67* (3), 223-265.