

REASONING ABOUT VARIABILITY IN COMPARING DISTRIBUTIONS

DANI BEN-ZVI

*University of Haifa, Faculty of Education
dbenzvi@univ.haifa.ac.il*

SUMMARY

Variability stands in the heart of statistics theory and practice. Concepts and judgments involved in comparing groups have been found to be a productive vehicle for motivating learners to reason statistically and are critical for building the intuitive foundation for inferential reasoning. The focus in this paper is on the emergence of beginners' reasoning about variation in a comparing distributions situation during their extended encounters with an Exploratory Data Analysis (EDA) curriculum in a technological environment. The current case study is offered as a contribution to understanding the process of constructing meanings and appreciation for variability within a distribution and between distributions and the mechanisms involved therein. It concentrates on the detailed qualitative analysis of the ways by which two seventh grade students started to develop views (and tools to support them) of variability in comparing groups using various statistical representations. Learning statistics is conceived as cognitive development and socialization processes into the culture and values of "doing statistics" (enculturation). In the light of the analysis, a description of what it may mean to begin reasoning about variability in comparing distributions of equal size is proposed, and implications are drawn.

Keywords: Variability; Comparing distributions; Statistical reasoning; Exploratory data analysis; Enculturation; Appropriation

1. SCIENTIFIC BACKGROUND

1.1. ENCULTURATION

Research on mathematical cognition in recent decades seems to converge on some important findings about learning, understanding, and becoming competent in mathematics. Stated in general terms, research indicates that becoming competent in a complex subject matter domain, such as mathematics or statistics, "may be as much a matter of acquiring the habits and dispositions of interpretation and sense making as of acquiring any particular set of skills, strategies, or knowledge" (Resnick, 1988, p. 58). This involves both cognitive growth and socialization processes into the culture and values of "doing mathematics or statistics". Many researchers have been working on the design of learning environments and teaching in order to "bring the practice of knowing mathematics in school closer to what it means to know mathematics within the discipline" (Lampert, 1990, p. 29). This study is intended as a contribution to the understanding of these processes in the area of Exploratory Data Analysis (EDA), focusing on reasoning about variability in comparing distributions.

One of the ideas used in this study is that of a process of *enculturation*, which is included in several recent learning theories in mathematics education (cf., Resnick, 1988; Schoenfeld, 1992). Briefly stated, this process refers to entering a community (or a practice) and picking up the community's 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*. This is especially the case with regard to statistical thinking, with its own values and

belief systems and its habits of questioning, representing, concluding, and communicating. Thus, for *statistical enculturation* to occur, specific thinking tools are to be developed alongside collaborative and communicative processes taking place in the classroom.

1.2. RESEARCH ON VARIATION

Bringing the practice of knowing statistics at school closer to what it means to know statistics within the discipline requires a description of the latter. Based on in-depth interviews with practicing statisticians and statistics students, Wild and Pfannkuch (1999) provide a comprehensive description of the processes involved in statistical thinking, from problem formulation to conclusions. They suggest that statisticians operate, sometimes simultaneously, along four dimensions: investigative cycles, types of thinking, interrogative cycles, and dispositions. They position *variation* at the heart of their model of statistical thinking as one of the five types of fundamental statistical thinking.

Pfannkuch and Wild (2004) further explain the centrality of reasoning about variation in data inquiry problems:

Adequate data collection and the making of sound judgments from data require an understanding of how variation arises and is transmitted through data, and the uncertainty caused by unexplained variation. It is a type of thinking that starts from noticing variation in a real situation, and then influences the strategies we adopt in the design and data management stages when, for example, we attempt to eliminate or reduce known sources of variability. It further continues in the analysis and conclusion stages through determining how we act in the presence of variation, which may be to either ignore, plan for, or control variation. Applied statistics is about making predictions, seeking explanations, finding causes, and learning in the context sphere. Therefore we will be looking for and characterizing patterns in the variation, and trying to understand these in terms of the context in an attempt to solve the problem. Consideration of the effects of variation influences all thinking through every stage of the [statistical] investigative cycle. (Pfannkuch & Wild, 2004, pp. 18–19)

According to Wild and Pfannkuch (1999), there are four aspects of variation to consider: noticing and acknowledging, measuring and modeling (for the purposes of prediction, explanation or control), explaining and dealing with, and developing investigative strategies in relation to variation. Reading and Shaughnessy (2004) suggest two additional aspects of variation that need to be considered; describing and representing. They claim that these six aspects of variation form an important foundation for statistical thinking.

Studies of reasoning about variation include investigations into the role of variation in graphical representation (Meletiou & Lee, 2002), comparison of data sets (Watson & Moritz, 1999; Watson, 2001; Makar & Confrey, 2004), probability sample spaces (Shaughnessy & Ciancetta, 2002), chance, data and graphs in sampling situations (Watson & Kelly, 2002), and variability in repeated samples (Reading & Shaughnessy, 2004). Hierarchies to describe various aspects of variation and its understanding have been developed by Watson, Kelly, Callingham, and Shaughnessy (2003) and by Reading and Shaughnessy (2004) in the context of repeated samples.

Noticing and understanding variability encompass a broad range of ideas. The basic form of variability in data is the variation of values *within* one distribution. Comparing distributions creates the impetus to consider other types of variability that exist *between* groups. Makar & Confrey (2004) discuss three different ways that teachers consider issues of variability when reasoning about comparing two distributions. They analyzed (1) how teachers interpreted variation *within* a group - the variability of data; (2) how teachers interpreted variation *between* groups - the variability of measures; and (3) how teachers *distinguished* between these two types of variation.

1.3. RESEARCH ON COMPARING DISTRIBUTIONS

Comparing groups provides the motivation and context for students to consider data as a distribution and take into account and integrate measures of variation and center (Konold & Higgins, 2003). At an advanced level, comparing distributions can stimulate learners to consider not only

measures of dispersion *within* each group, but comparisons of measures *between* groups, and hence to consider variation within the measures themselves (Makar & Confrey, 2004). Watson and Moritz (1999) suggest that comparing two groups provides the groundwork to the more sophisticated comparing of populations or two treatments in statistical inference. Without first building an intuitive foundation, inferential reasoning can become recipe-like, encouraging black-and-white deterministic rather than probabilistic reasoning.

There is some evidence however that the group comparison problem is one that students do not initially know how to approach and the challenge may remain even after extended periods of instruction. Students' difficulties may stem from the multifaceted knowledge necessary for comparing groups, such as understanding distributions (Bakker & Gravemeijer, 2004), representativeness (Mokros & Russel, 1995), and variability in data (e.g., Meletioui, 2002). Students also have difficulties in adopting statistical dispositions, such as tolerance towards variation in data, and integration of local and global views of data and data representations (Ben-Zvi & Arcavi, 2001; Ben-Zvi, 2002; Ben-Zvi, 2004).

Watson and Moritz (1999) observed two response levels in group comparison tasks completed by students during school years. In the first cycle, responses compared data sets of equal sizes, with or without success depending on the specific context. They did not recognize and/or did not resolve the issue of unequal sample size. In the second cycle, the issue of unequal sample size was resolved with some proportional strategy employed for handling different sizes.

There are a number of studies in which students who appeared to use averages to describe a single group or knew how to compute means did not use them to compare two groups (Bright & Friel, 1998; Gal, Rothschild & Wagner, 1990; Hancock, Kaput & Goldsmith, 1992; Konold, Pollatsek, Well, & Gagnon, 1997; Watson & Moritz, 1999). Konold et al. (1997) argue that students' reluctance to use averages to compare two groups suggests that they have not developed a sense of average as a measure of a group characteristic, which can be used to represent the group. Cobb (1999) proposes that the idea of middle clumps ("hills") can be appropriated by students for the purpose of comparing groups.

1.4. THE RESEARCH QUESTION

Based on these perspectives and studies, the following research question is used to structure the current study and the analysis of data collected: *How do junior high school students begin to reason about variability as part of an open-ended group-comparison task given in a rich and supportive classroom context?* Such a context involves a computerized environment, peer collaboration and classroom discussions, guidance of a teacher and curriculum-based tasks. The current study is different from some of the studies described above: It follows closely the dynamic behavior and discourse of two novice seventh grade students engaged with an EDA task. The students are observed within their classroom during an extended period of engagement with curriculum-based data investigation. A qualitative detailed analysis of the protocols is used, taking into account all kinds of actions, discussions and gestures within the situations in which they occurred. The goal is to trace the emergence of beginners' reasoning about variation in a comparing distributions situation, including the development of cognitive structures and the sociocultural processes of understanding and learning.

2. METHOD

Descriptions of the research setting, the statistics curriculum and the specific activity are followed by a profile of the students, technology used, and methods of data collection and analysis.

2.1. THE SETTING

This study took place in a progressive experimental school in Tel-Aviv, Israel. Skillful and experienced teachers, who were aware of the spirit and goals of the *Statistics Curriculum* (SC), taught

three classes. The SC was developed in Israel to introduce junior high school students (grade 7, age 13) to statistical reasoning and the “art and culture” of EDA (described in more detail in Ben-Zvi & Friedlander, 1997b; Ben-Zvi & Arcavi, 1998). The curriculum is characterized by the teaching and learning of mathematics using open-ended problem situations to be investigated by peer collaboration and classroom discussions using computerized environments (Hershkowitz et al., 2002). The design of the curriculum was based on the creation of small scenarios through which students can experience on their own, with limited teachers’ guidance, some of the processes involved in the experts’ practice of data-based enquiry. The SC was implemented in schools and teacher courses and subsequently revised in several curriculum development cycles.

The SC emphasizes student’s active participation in organization, description, interpretation, representation, and analysis of data situations on topics close to the students’ world, with a considerable use of visual displays as analytical tools (in the spirit of Garfield, 1995, and Shaughnessy, Garfield, & Greer, 1996). It incorporates technological tools for simple use of various data representations and transformations of them (as described in Biehler, 1993, 1997; Ben-Zvi, 2000). The scope of the curriculum is 30 periods spread over two to three months, and includes a student book (Ben-Zvi & Friedlander, 1997a) and a teacher guide (Ben-Zvi & Ozruso, 2001).

2.2. THE SURNAMES ACTIVITY

The *Surnames* activity, which is the focus of the current study, is the second full data investigation of the SC. It comes after an investigation involving the analysis of a time-series dataset with tabular data about Olympic 100 meters running times and a time plot of these data. The students are asked to compare the length of a set of surnames collected in their own class (35 Hebrew names) with a set of surnames from an American class that were given to them (35 English names). Equal sized data sets are used to simplify some aspects of the complex situation of comparing groups found in other studies (e.g., Gal, Rothschild & Wagner, 1990; Konold, Pollatsek, Well, & Gagnon, 1997), primarily students’ difficulties with proportional reasoning. It was expected that the Surnames activity will support the development of beginners’ reasoning about variability from the intuitive and simple to the more sophisticated and expert-like reasoning. The Surname data were given in a table (a part of it is presented in Figure 1).

Israeli Class (Hebrew names)				American Class (English names)			
Student’s Number	First Name	Surname	Surname’s Length	Student’s Number	First Name	Surname	Surname’s Length
1	מריה	אלקס	4	1	Kenneth	Auchincloss	11
2	מיכאל	אקרמן	5	2	Melinda	Beck	4
3	צפורה	בוסקילה	7	3	Edward	Behr	4
4	דולי	בטש	3	4	Patricia	Bradbury	8
5	רינה	בן שטרית	8	5	William	Burger	6
6	תמר	בריל	4	6	Mathilde	Camacho	7
7	איזבל	ברלין	5	7	Lincoln	Caplan	6

Figure 1. The upper part of the spreadsheet table displaying the raw data
(There were 35 students in each class.)

In order to understand the analytic and interpretive challenge faced by the students, the two distributions are presented graphically in Figure 2. As a background, it is important to note that the variability between the two groups of names is in part due to differences in the structure of English and Hebrew. In modern Hebrew, as in Arabic and some other Semitic languages, words are often written without some vowels, making Hebrew words shorter than English words. Vowels are usually optional and if needed are written as diacritical marks under, within or above the letters, using dots and dashes which signify different types of vowels. These diacritical marks are not displayed in the second and third columns of Figure 1. There are additional cultural and historical factors that contribute to the variability in name length within and between the two language groups.

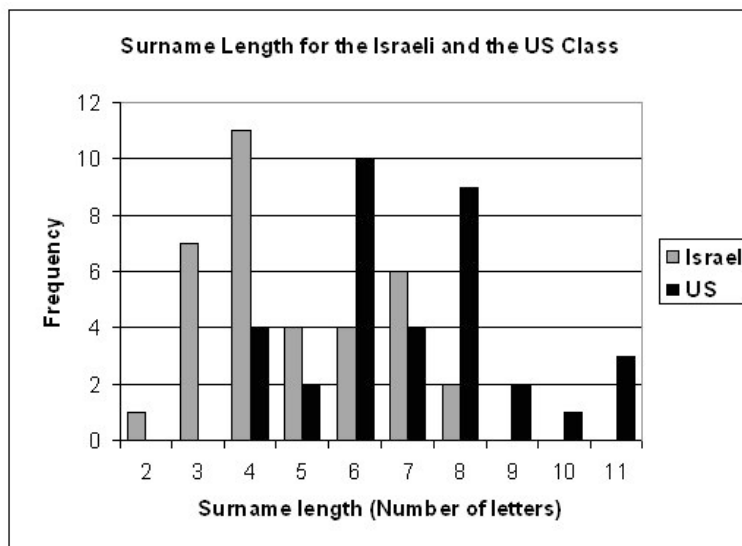


Figure 2. Double bar chart of the two surname groups

The whole Surnames activity took place during approximately three 90-minute lessons. Most of the time was spent on students' work in pairs in the computer lab, led by the student textbook. The teacher's interactions with the students were short and mostly occurred in reaction to their request for help. A session started with 5-10 minutes whole class introductory discussion and usually ended with a summary led by the teacher. In a preparatory lesson, students were asked: "What is the favorite shoe color and shoe size in your class? Compare the results to other seventh-grade classes". Students collected, organized, displayed and interpreted the data, compared the groups, and composed a summary report for a shoe company. Several statistical concepts and tools were informally introduced, or revisited, such as, statistical question and hypothesis, sample, categorical and quantitative variables, absolute and relative frequency, bar charts and frequency table.

In the following lessons, which are the focus of this report, three methods were offered by the curriculum-based materials to compare distributions: (a) absolute and relative frequency distributions presented in tables; (b) basic measures of variation and center, such as range, mode, mean, and median; and (c) graphical representations, such as a double bar chart. These statistical methods and tools were introduced to help students in describing and interpreting the surnames data and the variability in it, searching for trends and drawing conclusions on comparing the two groups. The purpose of the activity was to set the stage for students to consider data as a distribution and provide many opportunities to notice, acknowledge, intuitively deal with, and describe the variability within and between distributions.

2.3. PARTICIPANTS

This study focuses on two students, *A* and *D*, who were above-average ability students (grade 7, age 13), very verbal, experienced in working collaboratively in computer-assisted environments, and willing to share their thoughts, attitudes, doubts, and difficulties. They agreed to participate in this study, which took place mostly within their regular classroom periods and included being videotaped and interviewed (after class) as well as furnishing their notebooks for analysis. While not necessarily representing their classmates, verbal and able students provide a better opportunity for collecting valuable and detailed data on their actions, thoughts and considerations.

When they started to learn this curriculum, *A* and *D* had limited in-school statistical experience. However, they had some informal ideas and positive dispositions toward statistics, mostly through exposure to statistics jargon in the media. In primary school, they had learned only about the mean and the uses of some basic diagrams, such as bar and pie charts. Prior to, and in parallel with, the learning of the SC they studied beginning algebra based on the use of spreadsheets to generalize numerical linear patterns (Resnick & Tabach, 1999).

The students appeared to engage seriously with the curriculum, trying to understand and reach agreement on each task. They were quite independent in their work, and called the teacher only when technical or conceptual issues impeded their progress. The fact that they were videotaped did not intimidate them. On the contrary, they were pleased to speak out loud, address the camera explaining their actions, intentions, and misunderstandings and share what they believed were their successes.

2.4. TECHNOLOGY

During the experimental implementation of the curriculum a spreadsheet package (Excel) was used. Although Excel is not the ideal tool for data analysis (Ben-Zvi, 2000), there are several reasons for choosing this software. Spreadsheets provide direct access that allows students to view and explore data in different forms, investigate different models that may fit the data by, for example, manipulating a line to fit a scatter plot. Spreadsheets are flexible and dynamic, allowing students to experiment with and alter displays of data. For instance, they may change, delete or add data entries in a table and consider the graphical effect of the change or manipulate data points directly on the graph and observe the effects on a line of fit. Spreadsheets are adaptable by providing control over the content and style of the output. Finally, spreadsheets are common, familiar, and recognized as a fundamental part of computer literacy (Hunt, 1995). They are used in many areas of everyday life, as well as in other domains of the mathematics curricula, and are available in many school computer labs. Hence, learning statistics with a spreadsheet helps to reinforce the idea that this is something connected to the real world.

2.5. DATA COLLECTION AND ANALYSIS

A diverse body of data was collected to study the effects of the new curriculum. The behavior and reasoning of the two students on which the present study focused was analyzed using lengthy video recordings of whole class sessions, classroom observations, interviews, and students' notebooks and research projects. In addition, observational data and summative assessment data were also collected for the whole class to support other research objectives, but are beyond the scope of this paper.

The analysis of the videotapes was based on interpretive microanalysis (see, for example, Meira, 1991): a qualitative detailed analysis of the protocols, taking into account verbal, gestural and symbolic actions within the situations in which they occurred. The goal of such an analysis is to infer and trace the development of cognitive structures and the sociocultural processes of understanding and learning.

Two stages were used to validate the analysis, one within the SC researchers' team and one with four researchers in science education, who had no involvement with the data or the SC (triangulation in the sense of Schoenfeld, 1994). In both stages the researchers discussed, presented, and advanced and/or rejected hypotheses, interpretations, and inferences about the students' cognitive structures. Advancing or rejecting an interpretation required: (a) providing as many pieces of evidence as possible (including past and/or future episodes, and all sources of data as described earlier) and (b) attempting to produce equally strong alternative interpretations based on the available evidence. In most cases the two analyses were in full agreement, and points of doubt or rejection were refuted or resolved by iterative analysis of the data. In the presentation of transcripts, comments in block parentheses are clarifications suggested by the author, and were verified by a triangulation process.

3. RESULTS: STUDENTS' DEVELOPMENT OF REASONING ABOUT VARIABILITY

This paper describes how A's and D's novice views slowly changed and evolved towards an expert perspective while comparing two data sets of the same size. The focus is on how they began to notice and acknowledge variability in the data and make use of special local information in different ways as stepping-stones towards the development of global points of view of describing and explaining the variability between the groups. The study identifies seven developmental stages of their reasoning about variability (Figure 3).

Stage 1	On what to focus: Beginning from irrelevant and local information.
Stage 2	How to describe variability informally in raw data.
Stage 3	How to formulate a statistical hypothesis that accounts for variability.
Stage 4	How to account for variability when comparing groups using frequency tables.
Stage 5	How to use center and spread measures to compare groups.
Stage 6	How to model variability informally through handling outlying values.
Stage 7	How to notice and distinguish the variability within and between the distributions in a graph.

Figure 3. The seven suggested stages through which the two students progress

Stage 1. On What to Focus: Beginning from Irrelevant and Local Information

When the teacher introduced the whole class to the *Surnames* problem situation, she asked the students to hypothesize about interesting phenomena regarding names in general, without first providing them with any data. After a brief discussion about students' intuitive hypotheses, the teacher focused the discussion on name length in various cultures and countries, and presented the main task: Compare the surname length of the Israeli and the American groups. The teacher considered some sample quick responses (e.g., “*American surnames are longer than Israeli surnames*”, “*They are about the same*”) as an indication that the students had enough familiarity with the context of the task in order to engage meaningfully with the data. When the introduction was over, *A* and *D* moved to the school computer lab to work on the Surnames activity. Their work was guided by a list of questions that appeared in the Student Workbook which was part of the SC.

After *A* and *D* added the names of their classmates to the Excel table (a part of it presented in Figure 1), they started working on the first question in their Workbook, “*Look at the table and suggest a research question about length of surnames.*” The raw data, i.e., names, were displayed in a table on the computer screen. After a short discussion they agreed on posing the question, “*Which of the two countries has longer names?*” This initial focus on finding the “winning” group resembles the type of questions suggested in the introductory whole class discussion and was typical of students' questions in the experimental classes. This formulation, deterministic in nature and ignoring the complexity involved in comparing groups, is not surprising at this beginning stage of working on a complex data-analysis task.

In the second question, students were asked to formulate a hypothesis regarding interesting phenomena in the data. The question, which was proposed to ‘push’ students to look at the data and consider patterns and variability, provoked the following exchange between *A* and *D*. (The row numbers in the transcripts are provided to assist later in referring to specific sentences.)

- | | | |
|---|----------|--|
| 1 | <i>A</i> | We have to phrase now a hypothesis regarding interesting phenomena in the data. |
| 2 | <i>D</i> | Interesting phenomena, interesting phenomena. O.K., we should find interesting phenomena. We'll find interesting phenomena. [Reads the question again] “Formulate a hypothesis about interesting phenomena in the length of surnames”. I didn't understand what it exactly means. |
| 3 | <i>A</i> | O.K., lets skip this [question], since we don't have anything interesting at hand. We may shortly find something. |
| 4 | <i>D</i> | I don't think we should skip this, we'll simply ask what the precise intention is. I didn't really understand: Shall we hypothesize about ‘Mc's’? [There are three surnames in the American class, beginning with the letters Mc, such as McDaniel.] No! I don't understand. [Laughing] This isn't funny. I'll ask Michal [their teacher] to come and help us. |

Their remarks indicate that questions like “phrase a hypothesis regarding interesting phenomena in the data” may encounter an initial inability to focus attention on relevant (even informal) aspects of the data. *A* and *D* seemed to be unable to make full sense of the intention of the question and its formulation. Their focus on irrelevant features of the data, or their inability to focus on anything at all [row 3], is similar to their reaction at the beginning of the first problem situation in the SC–*Olympic*

Records (analyzed in detail in Ben-Zvi & Arcavi, 2001). In both activities, they were aware that their observations, such as names beginning with Mc, might not be relevant. They somehow recognized what not to focus on, but were uncertain about what may qualify as ‘interesting phenomena’ in this context, or how to reply to such questions, and finally requested the teacher’s assistance to help them overcome this difficulty.

In the above brief discourse the students did not notice global features of the data and the variability within it. Their initial local focus on what they saw as outstanding regularity in the data (the three “Mc” surnames) seems to restrict them from observing the distributions as a whole. Interestingly, this phenomenon was already observed when these students worked on their first activity of the SC (see section 2.2). There, *they* were similarly attentive to the prominence of “local deviations” in data and this appeared to keep them from creating more global interpretations of data. Only after the following teacher intervention were they able to start focusing on relevant information, taking into account the variability in the data.

Stage 2. How to informally describe the variability in raw data

When *A* and *D* requested the teacher’s help in answering the hypothesis task, the following dialog took place.

- 5 *A* [Asking the teacher] What does it mean?
6 *D* What does it mean to “phrase a hypothesis about interesting phenomena”?
7 *A* That there are many names beginning with ‘Mc’?
8 *T* About the length of surnames. OK?
9 *A* What is ‘interesting phenomena’?
10 *T* Are there no interesting phenomena in the data?
11 *A* [Cynically] It’s very interesting that there is a Michael...
12 *T* You are asked about length!
13 *D* About length ... An interesting phenomenon is that there is a [counting letters in the Hebrew name Levkowitz] 1, 2, 3, 4, 5, ... [7] letter name here and a 4 there [Cose in the American class].
14 *T* OK. You suggest that there are very short names and very long ones.
15 *A* Do we have to compare?
16 *D* So what’s the hypothesis?
17 *T* I don’t know [what the hypothesis is]. First, it’s a phenomenon. What do you think? Are there many long or many short [surnames]?
18 *A* There will be a lot more of the long in USA.
19 *D* More long than short.
20 *T* OK. You have a hypothesis: In the USA...
21 *A* But what is long, and what is short?
22 *T* That’s a different question.
23 *A* What should we write?
24 *D* Perhaps longer than this? Or...
25 *A* What name is considered long?
26 *T* OK. Longer than this – that’s a comparison. When you compare these groups, you say – I expect that there will be so and so here... That’s comparing two groups. That’s all right.

The students were uncertain about the intention of the question (“phrase a hypothesis”) as well as the meaning of the phrase “interesting phenomena”. The fact that a particular research question (comparing the two groups in terms of surname length) had been introduced at the beginning of the activity did not help them to focus and they seemed to be overwhelmed by the complexity of the data. Their initial observations are irrelevant and local (Mc’s, Michael). It seems that there are three factors interacting to produce the students’ inability to proceed: (a) the lack of understanding of the intent of the question, (b) the lack of understanding of the phrase “interesting phenomenon”, and (c) the complexity of the data. These factors played a role in causing confusion in other parts of the transcripts of these students (cf. Ben-Zvi & Arcavi, 2001).

The teacher's initial help consisted of calling their attention twice to the investigated variable, namely, the length of a surname. Only her second trial [row 12 in the transcript above] pushed *D* to compare the surname length of two students (one from each class) [13]. Thus, he began focusing on the correct variable and noticing one aspect of the variability in the data, but in a very local way. The teacher accepted his answer as being in the right direction, and suggested a generalization of his local observation [14]. This intervention represents a generalization 'jump' by the teacher not reflected in the students' previous comments. She then nudged them to quantify the variability in the data in a simple way [17].

In response to the teacher's direct question, the students suggested that the long surnames are more frequent in the USA group [14–15]. It is hard to determine at this point if *A* considered only the variability within the American group, or the variability between the groups. Whichever interpretation is taken here, this initial consideration of variability later became the foundation on which *A* and *D* developed an informal model of the variability within, as well as between, the two groups. The students' first attempts to describe the variability in the data by comparing long and short names raised a new concern about the borderline between long and short names [21], which was not resolved at this stage, and may be the beginning of an attempt to handle variability by grouping the data. The interaction with the teacher closed with her recommendation to focus on comparing groups.

Stage 3. How to formulate a statistical hypothesis that accounts for variability

The above interaction with the teacher helped the students to re-focus and propose a hypothesis. The following dialogue between *A* and *D* took place immediately after the teacher left them.

- | | | |
|----|----------|--|
| 27 | <i>D</i> | Our hypothesis about interesting phenomena in the length of surnames is: In the USA, surnames will be... |
| 28 | <i>A</i> | Will be longer... |
| 29 | <i>D</i> | Longer than in Israel... |
| 30 | <i>A</i> | Usually than in Israel... |
| 31 | <i>D</i> | Usually, not always, usually. |
| 32 | <i>D</i> | Let's see, we have Levkowitch here [in the Israeli class] and Cose there [in the American class] – that's different. |
| 33 | <i>A</i> | Enough, enough, come on. |
| 34 | <i>D</i> | OK, never mind. |
| 35 | <i>A</i> | So, in the USA... the surnames... |
| 36 | <i>D</i> | Will be usually longer. |
| 37 | <i>A</i> | Very nice! |

After the previous discussion with the teacher, the students were able to formulate a sensible hypothesis regarding the comparison between the two groups that took into account the variability in the data. They began with a deterministic proposal for a rule, 'surnames in the USA are longer than in Israel'. However, they noticed immediately that this assertion does not take full account of the situation presented by the data, and decided that variability should be included in their description by adding the constraint "*usually, not always*" to the rule. Understanding that some surnames can "behave differently", i.e., deviate from a general rule they formulated, can be considered an important step in the development of their acceptance of the existence of and tolerance to variability. In other words, they began to adopt the statistical perspective of trends that are generally true, but still have exceptions.

This new understanding is evident in *D*'s provision of an "opposite example", an Israeli name that is longer than a USA name, to show that the 'rule' holds even if there are opposite cases. *D* suggested this same example in the previous discussion with the teacher. While at that time it limited his ability to formulate a general hypothesis and view the data globally, here it is an expression of comfort with global views of the data that include variability. Hence, this opposite example, which derailed *D* from being on the right track on the first occasion, helped him adopt a statistical view of variability at this subsequent time.

Why might the students have initially focused on deterministic relationships between the variables and paid special attention to the unusual case? A possible explanation for their perspective can be found in their short-term learning history. *A* and *D* used spreadsheets in their algebra studies (immediately before they started to learn the EDA unit), to explore patterns, generalize, model mathematical problems, create and use formulae, and draw tables and graphs. Most of the tables investigated were linear correspondences between two sets of values. The students were accustomed to generating tables with the spreadsheet by ‘extending’ the pattern of constant differences between adjacent cells through the act of ‘dragging’ a pair of cells to duplicate this difference to the rest of the cells in the column resulting in long tables with clearly defined patterns. Using the same exploratory learning environment may have evoked for them the same deterministic nature of the relationship between variables found in algebra, which they incorrectly applied in statistics in order to make sense of data. Thus, their first focused observations referred to what was salient to them and a familiar part of their practices, the ‘differences’ between adjacent data entries not being constant. The only regularity they found in the data was a set of three Mc names. Maybe they implicitly began to sense that the nature of these data in this new area of EDA, as opposed to algebra, is disorganized, and it is not possible to capture it in a single deterministic formula, e.g., the previous “*Usually, not always*” comment.

At the end of this episode the two students seemed very satisfied with their answer. However, it was hard to appreciate at this stage how fragile their current understanding was. Additional difficulties with their abilities to acknowledge, explain, describe and deal with the variability in data in the context of this “noisy” and complex data situation unfolded in later stages of their work.

Stage 4. How to account for variability when comparing groups using frequency tables

After the students formulated a research question and hypothesis they were introduced by the student textbook to different concepts related to frequency in the context of the surnames investigation: frequency, relative frequency, and creating univariate frequency tables using spreadsheets. At this stage, *A* and *D* worked smoothly with the software and tasks, explaining every step and overcoming technical and conceptual hurdles. The following dialogue took place when they completed the production of two univariate frequency tables and were asked to use them to compare the two groups. See Figures 4 and 5, which are recreations of actual displays students generated on their own.

Israeli class		
Surname's length	Frequency	Relative frequency (%)
2	1	3
3	7	20
4	11	31
5	4	11
6	4	11
7	6	17
8	2	6
Total:	35	100%

Figure 4. Frequency table of surname's lengths in the Israeli class

American class		
Surname's length	Frequency	Relative frequency (%)
4	4	11
5	2	6
6	10	29
7	4	11
8	9	26
9	2	6
10	1	3
11	3	9
Total:	35	100%

Figure 5. Frequency table of surname's lengths in the American class

- 38 *D* [Reads the task] Use the frequency tables that you generated to compare the surnames' length in the two countries... Emm... They [the American surnames] are really a little bit longer. In the USA there are no 2 or 3-letter names...
- 39 *A* Yes. And in Israel...
- 40 *D* ... since they [the 2 or 3-letter names] are a bit short.
- 41 *A* The table [Figure 4] starts from...
- 42 *D* From 2 [letters] to 8 [letters].

- 43 *A* The [Israeli] surname length is from 2 to 8... And in the USA they're from 4 to 11... In other words, in the USA 2 or 3-letter names are not considered at all.
- 44 *D* They're considered, but there are simply none.
- 45 *A* There are none, or there is exactly one in the whole USA, something like that... And in Israel, names with 9, 10, and 11 letters are not considered, because there are none.
- 46 *D* Because they [American names] have vowels. For example, Raz, Itzik Raz [a student in their class]: Here [in Hebrew] it's R and Z, and there [in English] it's R, A, and Z – three letters, did you understand?
- 47 *A* In Israel, names with 9, 10, and 11 letters are not at all considered, because there are none. There may be one or two all over the country, yes, yes.
- 48 *D* Like Levkowitz.
- 49 *A* So, for example, we see that names with 8 letters are 6% in Israel.
- 50 *D* There – they are 26%.
- 51 *A* In the USA they are 26%.
- 52 *D* 20% more.
- 53 *A* 20% more, and it's a lot more, and...
- 54 *D* A lot more, interesting, lovely... Actually, emm... just a second... That's exactly all I'm saying...
- 55 *A* I assert that in the USA there are more... the names...
- 56 *D* There are longer names, right.
- 57 *A* Longer according to the comparison between these tables [Figures 4 & 5]. It may not be certain, but at least according to these tables... So, in the USA table, there are no 2 and 3 letter-names while there are 9, 10, and 11, but none in Israel. This means that the names are longer. [Writing this conclusion in his notebook.] Now, we also see here that in Israel, there are many more 4-letter names, which is considered pretty short.
- 58 *D* Having a 4-letter name is the coolest matter in Israel.
- 59 *A* So maybe because of that, there are more of those [surnames] in Israel, and in the USA – the names are longer. Therefore there aren't many names with 4 letters there. I brought up the 4 letters just as an example.

The students were faced with an unfamiliar and complex situation, presented in two separate frequency tables that included many values (Figures 4 & 5). Their purpose was to find ways to justify their hypothesis that surnames in the USA are usually longer than in Israel using the two frequency tables they had just created. On their own, they constructed a comprehensive argument, consisting of the comparison of two kinds of “special” values within the distributions: disjoint edge values – present in one distribution and absent from the other (and vice versa), and common edge values – the first and the last common values of the two distributions.

They began their argument by looking at the distributions' edges, moving from the lowest to the highest edge, and the range of values in between. *D* used the left “tail” (the shortest surnames in Israel that are missing in the USA group) as a justification for the claim that American surnames are “*a little bit longer*” [38]. They continued by noticing the different ranges of the groups; however, they did not make explicit use of them as measures of dispersion [41–43]. Then *A* argued symmetrically about the right “tail” of the USA distribution that is missing in Israel. While this opposite symmetry between the distribution edges seems to strengthen their confidence in the claim that the USA surnames are longer, it does not help them see the horizontal shift between the two generally-similar distributions.

Once the disjoint values were considered, the students moved on to compare the frequencies of the neighboring values, namely the last and the first common values of the distributions (8 and 4-letter names respectively). *A* suggested that the large differences in the relative frequencies of these values provided additional support to their hypothesis. They also informally acknowledged that 4-letter surname is the ‘mode’ in Israel [58]. These comments may represent first steps towards understanding density in a distribution.

A and *D* integrated contextual knowledge to support their understanding of, and in order to account for, the variability in the data. First, *D* suggested a causal explanation to account for the group differences, namely the use of vowels in English versus diacritical symbols in Hebrew. He also provided an example of one Israeli surname Raz, which has three letters in English but only two in Hebrew [46]. *A* further speculated that their sample implied the rarity of very short and very long

surnames in the USA and the Israeli populations respectively [47]. *D* supported him bringing up his frequently mentioned example of Levkowitz, a relatively long Israeli surname in their class. In these actions, *A* and *D* were trying to synthesize statistical and contextual knowledge to draw out what can be learned from the data about the context of the problem. The context of the problem supports their statistical reasoning by providing reasonable explanations to the emergent patterns in the variation. At the end of this dialogue they wrote the following synthesis in their notebooks.

- A* “~~In the USA, the names are longer than in Israel.~~ [This sentence was written and later erased by A.] In the American table, there are no names with 2 and 3 letters, and there are names with 9, 10, 11 (none in Israel). In Israel, short names are more frequent; In the USA, the long names are more frequent.”
- D* “In the USA, the names are longer than in Israel (according to the tables). In the American table, there are no names with 2, 3 letters, and there are of 9 to 11.”

Arriving at a general conclusion was not a straightforward process for both students; however, they seem to be in different positions. *D*, without much doubt, accepted that the conclusion “*In the USA, the names are longer than in Israel*” captured the essence of the situation, and was less disturbed by the presence of outlying values, or irregular patterns in the data. In contrast, *A* struggled more with the variability presented in the data, and was more attentive to the prominence of “local deviations”, which kept him from dealing more freely with global views of data. This could have been the reason for his erasing the general conclusion in his written summary. On the other hand, the rest of his conclusion is a beginning step to modeling variability and conceptualizing the use of ‘density’ in comparing distributions.

Stage 5. How to use center and spread measures to compare groups

In the second part of the *Surnames* activity the students were introduced to basic statistical measures of center (mode, mean, and median), spread (range) and outliers. They used the computer to find the statistical measures of the two groups and organized them in a table. See Figure 6 which is a recreation of the actual display the students generated. The next question was to use these measures to compare the groups. The students were uncertain how to answer the question and asked for help. After the teacher approved one answer as being in the right direction, *A* and *D* started to interpret the table.

Statistical Measures	Israeli Class	American Class
Number of Students	35	35
Mode	4	6
The maximal value	8	11
The minimal value	2	4
Range	6	7
Mean	4.83	7.06
Median	4	6
Outlying values	2, 8	5, 9, 10

Figure 6. Statistical measures of the two classes (The correct median of the USA group is 7. For the outliers, the students chose values with minimal frequency.)

Using the statistical measures table that they generated (Figure 6), the students started comparing the groups by noticing that both the maximal and the minimal values of the Israeli group are smaller than those of the American group. However, they erroneously concluded that the range is also smaller since the two extreme values are smaller in the Israeli names. While the range does happen to be smaller, it is not for the reason stated. This shows a misinterpretation on the part of the students. Once they noticed that the mean and the median also behaved in a similar way, they inferred that all the statistical measures of the Israeli distribution are smaller than those of the USA distribution. In spite of their fluent work at this stage, their actions seem to be merely procedural, missing both the meaning of measures as representative numbers (Mokros & Russel, 1995), and the distinction between center and spread measures.

Stage 6. How to model variability informally through handling outlying values

Dealing with information in the last row of the measures table (Figure 6) initiated the following dialogue about outliers.

- 60 *D* But in the outlying values...
- 61 *A* In fact here it's [different than the rest of the measures]... You expect that in Israel the outlying values will be higher [larger] than in the USA, since there are less high [long surnames in Israel]. But in fact you see here that in Israel the outlying values are not so high [large].
- 62 *D* I am confused now, I don't understand. Not correct, because if your data...
- 63 *A* If everything in Israel is smaller, then you would expect that the outlying values, yes, will be high [large] numbers, since there are few of them; and in the USA, the outlying numbers – will be lower [smaller], since there are few of the low [short surnames].
- 64 *D* Yes, but this is not correct.
- 65 *A* But in fact in the USA also - the high [large surnames] are the outliers.
- 66 *D* 9 and 10.
- 67 *A* Right, 10 and 9 are outliers, but 11 is really high [long].
- 68 *D* Correct.
- 69 *A* Well, let's not write about that.

So far, the comparing of the two groups using statistical measures had been a straightforward and monotonous task. However, the outliers in the last row of the measures table presented a new challenge to the students: how to compare sets of numbers (2 and 8 in Israel vs. 5, 9, and 10 in USA) that had no trivial pattern and meaning. Furthermore, *A*'s pre-conceptualization of outliers as unusual and least frequent values in a distribution made him predict that the outliers in Israel would be only the long surnames since the Israeli surnames tended to be short (and vice versa in the USA distribution).

A seems to deal with distributions' variability with a plain dichotomous model. In his mental model, he divides the distributions to two groups: The short surnames that include the majority of the Israeli values, and the long surnames - the minority (and vice versa in the USA). This model appears to have helped him deal, describe and quantify the variability by reducing the 'noise' within the distributions. He consequently predicted that the variability between the groups would be also straightforward [61]. Once the students realized that the outliers were telling them a conflicting, more complex 'story' of the variability in the data, they did not find an alternative explanation and gave up on the resolution of the conflict.

It appears that having to deal with the outlier as a concept (i.e., a principled class of observations, not just some specific data points) contributed to the complexity of the students' conceptual task and understanding at this stage. A few minutes before the above dialogue took place, they came across outliers and chose to define them as "*the highest and the lowest values*". The meaning of the Hebrew word for outlier is "exceptional or unusual" and may have influenced their definition choice. Thus, from their perspective, the modal value was also an outlier. The teacher's explanation that outliers are individual data points that fall outside the overall pattern of the distribution made them abandon the mode as an outlying value, but left them with the view of outliers as merely the least frequent values.

Through their dealing with the outliers, the students presented a simplistic view of the distributions in order to handle the variability in the data. In their model, resembling a skewed distribution, the majority of the distribution concentrates in one interval, while the less frequent values, the outliers, are positioned in a disjoint interval. This model helped them to present clearly the difference between the distributions, which followed opposing patterns. In their view, the selection of outliers is based on low frequencies, meaning they are exceptional, since they are rare. In that respect, the students' consistent use of "high" and "low" to describe the "long" and "short" surnames in all the dialogues can be attributed to their focus on the variability in frequencies and not only to a careless use of language.

Stage 7. How to notice and distinguish the variability within and between the distributions in a graph

In the third and final part of the activity, the students were guided to generate graphical displays of the data and were asked to use them to compare the distributions. The following dialogue took place after they created a double bar chart of the two groups (similar to the graph displayed in Figure 2).

- 70 D [Reading the task] Use the graph you generated (Figure 2) to describe the emerging trend in the surnames' length of the two countries.
- 71 A Let's see: The USA... usually... no... hold on...
- 72 D It seems that it's a lower trend in the USA.
- 73 A Not low, it seems about the same in the graph.
- 74 D Aha... No, higher trend.
- 75 A Hold on, the USA...
- 76 D Since you do not compare this to that, but rather this to that.
- 77 A [Cynically] Really!
- 78 D All right. [Unclear] ... seven.
- 79 A So it's higher here, it's higher here, here, here, and it's higher here; but in Israel it's higher here, here, here, and here.
- 80 D And here.
- 81 A And here.
- 82 D They balance each other.
- 83 A Look, the advantages [height differences] are bigger in Israel. No, not always. Let's ask someone [a teacher] what it means.
- 84 D I know what it means.
- 85 A What?
- 86 D It means that the emerging trend is...
- 87 A But it is not equal. Look, we said that the USA is longer... The USA leads in 8, 9, 10, and 11, while Israel leads only in 2, 3, 4, 5, and...
- 88 D We said that the USA names are longer, what's the big deal?
- 89 A That's right. So, the USA leads in the longer names. That's also not a big deal since 2 was not considered at all in the USA, while 11 was not considered at all in Israel.
- 90 D What's the big deal? They were not considered because there are none.
- 91 A OK, but...
- 92 D They did not ignore data... It appears that in Israel the lengths of the lower names are...
- 93 A No...
- 94 D The length of the names
- 95 A In Israel... In Israel...
- 96 D The lengths of the lower names are...
- 97 A No. In Israel, the lengths of names with fewer letters have a higher frequency, but in the USA, the lengths with... [having difficulties to complete the sentence]
- 98 D I know how to formulate this. Write down.
- 99 A No. I first want to hear what you have to say.
- 100 D OK. In Israel, the frequency of the names with low number of letters...
- 101 A Relatively low.
- 102 D ... is higher than in the USA.
- 103 A Just a second, low – let's say smaller than 5.
- 104 D Let's assume so. ...is higher than...
- 105 A No. But there is also one exception here.
- 106 D The frequency is higher than in the USA.
- 107 A But there is also one exception here.
- 108 D [Angrily shouting] OK, it's in general! It's a general trend! It's not the trend for the exceptional one.
- 109 A [Surprised by D's reaction] Buu ...
- 110 D OK. On the other hand, in the USA, the trend... the frequency of the long surnames is relatively higher than in Israel.

Although the students are familiar with generating and interpreting bar-graphs, handling this particular double bar-graph (Figure 2) is a complex task for them. Their challenge is figuring out the graph and understanding the variation embedded in the data. At first, the students provided conflicting interpretations of the graph; their rather unclear statements [72-73] are initial attempts to find one global description that accounts for variability by summarizing the difference between the bars in the two groups. This attempt can be considered a progress in comparison to their previous interpretations of graphs in the SC, which were mostly local, focusing on one or more individual values within the distributions (Ben-Zvi & Arcavi, 2001). *D* suggested that their disagreement arose from their different ways of reading the graph: ‘horizontal’ reading – comparing values, vs. ‘vertical’ reading – comparing heights of bars (density, frequency).

The students then began focusing on comparing the heights of adjacent bars from the two groups. Based on a method *A* suggested for summarizing the differences between the groups [76-79], they counted how many times the bar of one group was higher than the bar of another group for each surname value on the X-axis. For example, if for a surname length of 6 letters the bar for the Israeli group was of height 4 and for the US group of height 10, then the US group was “winning” there. However, this led them to an impasse: the number of “winning” Israeli and American bars was equal [82]. A second trial to compare the height differences between adjacent bars also proved fruitless.

Only when they began focusing on the *location* of the “winning” bars of each group, did they realize that the American bars are higher than the Israeli bars for the long names, while the Israeli bars are higher for the short names. Thus, they reduced the problem of comparing each pair of bars to comparing two subgroups, the relatively short and long surnames. Their previous success, in the frequency table task, in handling the variability between the groups by dividing the distributions to two groups seems to have helped the students out of impasse also here. This informal comparing method resembles Cobb’s (1999) finding that the idea of middle clumps (“hills”) can be appropriated by students for the purpose of comparing groups.

However, *A* was not completely satisfied with the above realization and was particularly concerned [103] about the distinction between short and long names. This issue, which worried him also at the beginning of the activity [21], was triggered here by the lack of clear-cut borderline between the groups: 5 and 7-letter names are more frequent in Israel and the 6-letter names are more frequent in the USA (see Figure 2). While *A* could not ignore the presence of this deviation in favor of a global summary of the variability between the groups, *D* was not disturbed by the ‘noise’ in the data. He claimed that their comparison is general and therefore they must ignore the one exception [108].

They requested the teacher’s approval before they wrote a summary in their notebooks: “*The emerging trend is that the frequency of relatively short names (up to 5 letters) is higher in Israel than in the USA, but the frequency of relatively long names is higher in the USA than in Israel.*” Thus their final description of the variability between the groups was based on comparing the frequencies of two subgroups ignoring the deviation from the trend in the center.

4. DISCUSSION

This study was undertaken to contribute to our understanding of the process through which students develop ways to reason about variability within and between distributions. The study examined the first steps of two students who worked on a group-comparison task in a rich technology-based environment. In this environment, as happens in regular classes, students’ work and intuitions are supported by formal curricular materials and ongoing instructional activities. The results illustrated several aspects, discussed below, of students’ emerging understanding of variability in comparing groups and the role of supporting factors in that process, in particular the teacher’s role. Conclusions and implications are discussed further below.

4.1. STAGES IN DEVELOPMENT OF REASONING ABOUT VARIABILITY

A and *D* started by trying to make sense of general questions normally asked in EDA tasks. Their learning trajectory included coming up with irrelevant answers and feeling an implicit sense of

discomfort with them, asking for help, getting feedback, trying other answers, working on a task even with partial understanding of the overall goal, and confronting the same issues with different sets of data and in different investigation contexts. This problem-solving process is consistent with several other research findings (see, for example, Moschkovich, Schoenfeld, & Arcavi, 1993; Magidson, 1992): novices may be either at a loss (when asked these kinds of questions) or their perceptions of what is relevant are very different from the experts' view.

When looking at raw data (stages 1-2), the students initially did not notice global features of the data and the variability within them. Their initial focus on what they saw as outstanding regularity in the data, the three "Mc" surnames, was based on attention to local features and seems to have restricted them from observing global features of the distributions. As noted in an earlier activity of the SC, *A* and *D* were attentive to the prominence of "local deviations" in data and this kept them from dealing more freely with global views of data. It is interesting that they did not benefit from this earlier experience. Only after the teacher's intervention they started focusing on relevant information and took into account the variability in the data. Their reasoning about variability evolved then from observing differences between two values, to distinguishing between long and short names, to noticing and informally describing the variability between the groups. They finally arrived (stage 3) at a formulation of a rule or hypothesis that took into account the variability in the data ("*usually, not always*").

In the frequency table task (stage 4), *A* and *D* focused on individual edge values, not noticing the global features of the distribution and ignoring the center interval of the distributions (5 to 7 letters). Possible sources of their difficulties could have been their being novices in the new area of EDA, and the type of representation used, two single frequency tables, which seems complex to analyze and less supportive in terms of displaying general trends. Their initial focus on distribution edges is consistent with other studies, for example, Biehler (2001). Novice students tend to focus on the "least" and the "most" while describing the variability between two distributions using box plots.

The students' insignificant and monotonous use of statistical measures (stage 5) to compare the groups ("*Everything is smaller*") resembles students' reluctance to use averages meaningfully to compare two groups in other studies. There are a number of studies in which students who appeared to use averages to describe a single group or knew how to compute means did not use them to compare two groups (e.g., Bright & Friel, 1998; Watson & Moritz, 1999). Konold et al. (1997) argue that students' reluctance to use averages to compare two groups suggests that they have not developed a sense of average as a measure of a group characteristic, which can be used to represent the group (see also Mokros & Russell, 1995). In addition, students in this study may be seeing averages as only representing middles and having nothing to do with variation.

Throughout their dealing with and comparing the outliers between the groups (stage 6), the students presented a simplistic view of the distributions in order to handle the variability in the data. In their model, resembling a skewed distribution, the majority of the distribution concentrates in one interval, while the less frequent values, the outliers, are positioned in another interval. This model helped them to compare the distributions as following opposing patterns. In their view, the selection of outliers was based on low frequencies, meaning they are exceptional, since they are rare. In that respect, the students' consistent use of "high" and "low" to describe the "long" and "short" surnames in all the dialogues can be attributed to their focus on the variability in frequencies and not only to a careless language flow.

They finally struggled (stage 7) with reading and interpreting the graph they generated (double bar chart, Figure 2). They first practiced their reading of the graph, trying 'vertical' (density) and 'horizontal' (variation in values) interpretations of the variability presented in it. Then they used different local methods to describe the variability in the data. Information they gained in handling the frequency table task helped them in developing a dichotomous model to compare the groups.

The students' development of reasoning about variability in comparing the groups was accompanied by somewhat parallel development of global perception of a distribution as an entity that has typical characteristics such as shape, center, and spread. This perception seems to be a precondition to being able to describe the two distributions as generally similar in shape and variability, but horizontally shifted (USA distribution shifted to the right of the Israeli distribution).

Similar difficulties were demonstrated by eight-grade students working on “prediction” questions about comparing groups (Bakker & Gravemeijer, 2004). These students did not shift a whole shape of a distribution, but reasoned about just the individual bars or the majority (see also Biehler, 2001).

4.2. SUPPORTING FACTORS

The study describes the difficulties and successes of what *A* and *D* did and how they reasoned about variation in the presence of supporting factors that are part of the learning environment in many classes: carefully-planned curricular materials, computer tools, peer collaboration and teacher interventions. It is difficult to tease out, however, what was “naturalistic” about students’ actions, and what was an outgrowth of these external factors of the learning environment. What students can and cannot do or think regarding variation is not merely a series of simple natural steps, but rather reactions to and struggles with the challenges and tools (including computer tools, two frequency displays, bar graphs, etc.) that were presented to them at each successive stage of an EDA journey. In particular, students’ statistical reasoning and actions were developed throughout by introduction to new cognitive tools and statistical concepts in a supportive learning environment.

Several factors appear to have helped the students develop their statistical reasoning about variability:

- a) Students repeatedly experimented with using different informal tools and methods, mostly local in nature (e.g., comparing heights of adjacent bars in a graph) or invented simple models (e.g., dividing the distributions to two subgroups) that partially capture the variability in the data within and between the groups.
- b) Students were helped by previous experiences with these data and other sets of data. For example, the dichotomous interpretation of the graph (stage 6) outgrows of previous handling of the statistical measures table.
- c) The context of the *Surnames* problem (e.g., the difference between Hebrew and English names) supported *A*’s and *D*’s reasoning in the statistical sphere and provided reasonable explanations to the patterns they observed in the variation. Integration of statistical knowledge and contextual knowledge is considered a fundamental element of statistical thinking (Pfannkuch & Wild, 2004).
- d) The incorporation of technological tools enabled students to simply and directly explore data in different forms and experiment with and alter views or displays of data.
- e) The interactions with the teacher helped students to adopt a statistical perspective but did not instruct them in exactly what to do. A detailed description of the teacher’s role is provided in the following section.

4.3. APPROPRIATION: A LEARNING PROCESS THAT PROMOTES UNDERSTANDING

The data show that most of the learning took place through dialogues between the students themselves but also after brief conversations with the teacher. Of special interest were the teacher’s interventions at the students’ request (additional examples of such interventions are described in Ben-Zvi & Arcavi, 2001; Ben-Zvi, 2004). These interventions, though short and not necessarily directive, had catalytic effects. They can be characterized in general as “negotiations of meanings” (in the sense of Yackel & Cobb, 1996). More specifically, they are interesting instances of appropriation as a nonsymmetrical, two-way process (in the sense of Moschkovich, 1989). This process takes place, in the zone of proximal development (Vygotsky, 1978, p. 86), when individuals (expert and novices, or teacher and students) engage in a joint activity, each with their own understanding of the task. Students take actions that are shaped by their understanding; the teacher “appropriates” those actions, into her own framework, and provides feedback in the form of her understandings, views of relevance, and pedagogical agenda. Through the teacher’s feedback, the students start to review their actions and create new understandings for what they do.

In this study, the teacher appropriated students’ utterances with several objectives: to reinforce the legitimacy of an interpretation as the right ‘kind’ in spite of not being fully correct, to simply refocus

attention on the question, to redirect their attention, to encourage certain initiatives, and implicitly to discourage others (by not referring to certain remarks). The students appropriate from the teacher a reinterpretation of the meaning of what they do. For example, they appropriate from her answers to their inquiries (e.g., what phrasing an hypothesis or interesting phenomena may mean), from her unexpected reactions to their request for explanation (e.g., “*You suggest that there are very short names and very long ones.*”), and from inferring purpose from the teacher’s answers to their questions (e.g., “*About the length of surnames. OK?*”). Appropriation by the teacher (to support learning) or by the students (to change the sense they make of what they do) seems to be a central mechanism of enculturation: entering and picking up the points of view of a community or culture (Schoenfeld, 1992; Resnick, 1988). In this process, the teacher is considered as an ‘enculturator’. As shown in this study, this mechanism is especially salient when students learn the dispositions that accompany using the subject matter (data analysis) rather than its skills and procedures.

4.4. LIMITATIONS OF THE STUDY

The two students described in this study were considered by their teacher to be both able and verbal. Their choice was aimed to enable the collection and analysis of focused and remarkably detailed data in order to draw, in very fine strokes, the “picture” of their emerging statistical reasoning about variability. Even when a phenomenon seems important and the data interpretation was validated and agreed upon, the question of the idiosyncrasy of the identified phenomenon may remain open. Therefore, in other studies, the data and interpretations from students in the same class or from other classes assist in checking for generalizability of the phenomena (cf., Ben-Zvi, 2002).

In presenting the students with tasks based on comparing two groups of equal size, some complications are avoided. This is both an advantage and disadvantage for the overall aims of this study. Research shows that the group comparison problem is one that students do not initially know how to approach and the challenge may remain even after extended periods of instruction (e.g., Bakker & Gravemeijer, 2004). Avoiding some of the complexity of proportional reasoning, the key for handling groups of different size, simplifies the task and may help researchers focus on and expose students' reasoning about variability. In this study, students were “pushed” to consider other complex statistical issues, such as integrating measures of variation and center and comparing measures *within* each group and *between* groups. However, it should be acknowledged that the study of students’ statistical reasoning about variability in comparing groups is not complete without incorporating tasks of comparing unequal data sets.

5. IMPLICATIONS

The idiosyncratic aspects of this study restrict the provision of broad recommendations. However, several conclusions that are tied to specifics of this study and its results, in the context of results from similar studies, can be drawn. The learning processes described in this paper took place in a carefully designed environment. This environment included: a curriculum built on the basis of expert views of EDA as a sequence of semi-structured, yet open, leading questions within the context of extended meaningful problem situations (Ben-Zvi & Arcavi, 1998), timely and non-directive interventions by the teacher as representative of the discipline in the classroom (cf., Voigt, 1995), and computerized tools that enable students to handle complex actions (change of representations, scaling, deletions, restructuring of tables, etc.) without having to engage in too much technical work, leaving time and energy for conceptual discussions (cf., Ben-Zvi, 2000).

In learning environments of this kind, from the very beginning students encounter, develop, and work with ideas, concepts, cognitive tools and dispositions related to the culture of EDA, such as making hypotheses, summarizing data, recognizing trends and variability, identifying interesting phenomena, comparing distributions and handling numerical, tabular and graphical data representations. Skills, procedures and strategies, such as creating and interpreting graphs and tables or calculating statistical measures, are learned as integrated in the context and at the service of the main ideas of EDA.

It can be expected that beginning students will have difficulties, of the type described, when confronting the problem situations of the curriculum. However, it is proposed that what *A* and *D* experienced should be an integral and inevitable component of a meaningful learning process if it is to have lasting effects. If students were to work in environments such as the above, teachers are likely to encounter the following learning phenomena:

- * Students' prior knowledge would and should be engaged in interesting and surprising ways, possibly hindering progress in some instances but making the basis for construction of new knowledge in others,
- * many questions that would either make little sense to the students, or, alternatively, will be re-interpreted and answered in different ways than intended, and
- * students' work that would inevitably be based on partial understandings, which will grow and evolve.

This study suggests that in order to help students gradually build a sense of the meaning of the data and statistical task with which they engage, multiple factors can and should be planned. These include appropriate teacher guidance, peer work and interactions, and more importantly, ongoing cycles of experiences with realistic problem situations.

Given that it is difficult to tease out the effects of what students learned or could or couldn't do from the enculturation processes and support of the teacher, further study is recommended that focus more attention on the role of teachers and what they should do, or learn to do, in order to promote statistical reasoning about variability. Much of students' progress in the current study is influenced by their interactions with the teacher that helped them adopt the statistical perspective but did not instruct them in exactly what to do or how to reason. The role of the teacher which is considered as an 'enculturator' deserves further exploration.

It is generally recommended that students be provided with multiple opportunities to engage with data in group-comparison tasks. The role of comparing unequal-size groups in promoting reasoning about variability, which was not studied here, should be further explored. The students in this study have gained from reading and interpreting multiple types of conventional data representations. The role of student-invented data representations and new graphical tools available through educational software and Internet has to be investigated to better expose the many ways variability is noticed, measured, and modeled by students. It is hoped that the complexity involved in group-comparison tasks can push students to think about the meaning of what they do and how they reason in statistics, develop relevant actions and interpretations, and be more critical of their actions and interpretations.

REFERENCES

- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Ben-Zvi, D. (2000). Toward understanding of the role of technological tools in statistical learning. *Mathematical Thinking and Learning*, 2(1&2), 127–155.
- Ben-Zvi, D. (2002). Seventh grade students' sense making of data and data representations. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics, Cape Town, South Africa*. [CD-ROM] Voorburg, The Netherlands: International Statistical Institute.
- Ben-Zvi, D. (2004). Reasoning about Data Analysis. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 121–146). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Ben-Zvi, D., & Arcavi, A. (1998). Toward a characterization and understanding of students' learning in an interactive statistics environment. In L. Pereira-Mendoza (Ed.), *Proceedings of the Fifth International Conference on Teaching Statistics* (Vol. 2, pp. 647–653). Voorburg, The Netherlands: International Statistical Institute.

- 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. (1997a). *Statistical investigations with spreadsheets—Student's workbook* (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.), *Research on the role of technology in teaching and learning statistics* (pp. 45–55). Voorburg, The Netherlands: International Statistical Institute.
- Ben-Zvi, D., & Ozruso, G. (2001). *Statistical investigations with spreadsheets—Teacher's guide* (In Hebrew). Rehovot, Israel: Weizmann Institute of Science.
- 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* (pp. 68–100). Berlin: Springer-Verlag.
- Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review*, 65(2), 167–189.
- Biehler, R. (2001, August). Developing and assessing students' reasoning in comparing statistical distributions in computer supported statistics courses. Paper presented at the Second International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-2), Armidale, Australia.
- Bright, G. W. & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 63–88). Mahwah, NJ: Lawrence Erlbaum.
- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1(1), 5–43.
- Gal, I., & Garfield, J. B. (Eds.). (1997). *The assessment challenge in statistics education*. Amsterdam, Netherlands: IOS Press.
- Gal, I., Rothschild, K., & Wagner, D. A. (1990, April). Statistical concepts and statistical reasoning in school children: Convergence or divergence? Paper presented at the annual meeting of the American Educational Research Association, Boston.
- Garfield, J. (1995). How students learn statistics. *International Statistical Review* 63(1), 25–34.
- Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337–364.
- 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*, (pp. 9–24). Haifa, Israel: The Technion.
- Hershkowitz, R., Dreyfus, T., Schwarz, B., Ben-Zvi, D., Friedlander, A., Hadas, N., Resnick, T., & Tabach, M. (2002). Mathematics curriculum development for computerized environments: A designer-researcher-teacher-learner activity. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 657–694). London: Erlbaum.
- Hunt, D. N. (1995). Teaching statistical concepts using spreadsheets. In the *Proceedings of the 1995 Conference of the Association of Statistics Lecturers in Universities*. Nottingham, UK: The Teaching Statistics Trust. [Online: <http://www.mis.coventry.ac.uk/~nhunt/aslu.htm>]
- Konold, C (2002). Teaching concepts rather than conventions. *New England Journal of Mathematics*, 34(2), 69–81.
- Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin & D. E. Schifter (Eds.), *A research companion to principles and standards for school mathematics*, (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.
- Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics*, (pp. 151–167). Voorburg, The Netherlands: International Statistical Institute.

- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, 27, 29–63.
- Magidson, S. (1992, April). From the laboratory to the classroom: A technology-intensive curriculum for functions and graphs. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Makar, K., & Confrey, J. (2004). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 353–374). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- 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.
- Meletiou, M. (2002). Conceptions of variation: A literature review. *Statistics Education Research Journal*, 1(1), 46–52.
- Meletiou, M., & Lee, C. (2002). Student understanding of histograms: A stumbling stone to the development of intuitions about variation. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town, South Africa*. [CDROM] Voorburg, The Netherlands: International Statistical Institute.
- Mokros, J., & Russell, S. J. (1995). Children's Concepts of Average and Representativeness. *Journal for Research in Mathematics Education*, 26(1), 20–39.
- Moschkovich, J. D. (1989). Constructing a problem space through appropriation: A case study of guided computer exploration of linear functions. An unpublished manuscript available from the author.
- Moschkovich, J. D., Schoenfeld, A. H., & Arcavi, A. A. (1993). Aspects of understanding: On multiple perspectives and representations of linear relations, and connections among them. In T. Romberg, E. Fennema & T. Carpenter (Eds.), *Integrating Research on the Graphical Representation of Function*, (pp. 69–100). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 17–46). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Reading, C., & Shaughnessy, M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 201–226). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Resnick, L. (1988). Treating mathematics as an ill-structured discipline. In R. Charles & E. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 32–60). Reston, VA: National Council of Teachers of Mathematics.
- Resnick, T., & Tabach, M. (1999). *Touring the land of Oz - algebra with computers for Grade Seven* (in Hebrew). Rehovot, Israel: Weizmann Institute of Science.
- 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* (pp. 334–370). New York: Macmillan.
- Schoenfeld, A. H. (1994). Some notes on the enterprise (research in collegiate mathematics education, that is). *Conference Board of the Mathematical Sciences Issues in Mathematics Education*, 4, 1–19.
- Shaughnessy, J. M., & Ciancetta, M. (2002). Students' understanding of variability in a probability environment. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town, South Africa*. [CDROM] Voorburg, The Netherlands: International Statistical Institute.
- 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* (Vol. I, pp. 205–237). Dordrecht, The Netherlands: Kluwer Academic Publishers.

- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*, (Edited by M. Cole, V. John-Steiner, S. Scribner, & E. Souberman). Cambridge, MA: Harvard University Press.
- Voigt, J. (1995). Thematic patterns of interaction and sociomathematical norms. In P. Cobb & H. Bauersfeld (Eds.), *Emergence of mathematical meaning: Interaction in classroom cultures*, (pp. 163–201). Hillsdale, NJ: Erlbaum.
- Watson, J. M. (2001). Longitudinal development of inferential reasoning by school students. *Educational Studies in Mathematics*, 47, 337–372.
- Watson, J. M., & Kelly, B. A. (2002). Can grade 3 students learn about variation? In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town, South Africa*. [CDROM] Voorburg, The Netherlands: International Statistical Institute.
- Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, 34(1), 1–29.
- Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37(2), 145–168.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–265.
- Yackel, E., & Cobb, P. (1996). Socio-mathematical norms, argumentation and autonomy in mathematics. *Journal for Research in Mathematics Education*, 27(4), 458–477.

SOFTWARE

- Excel*, Microsoft Corporation, <http://www.microsoft.com/office/excel/>.
- Fathom* (Fathom Dynamic Statistics Software), B. Finzer, Key Curriculum Press, 1150 65th Street, Emeryville, CA 94608, USA. <http://www.keypress.com/fathom/>.
- Mini-Tools*, Peabody College, Vanderbilt University, principal investigator: P. Cobb, http://peabody.vanderbilt.edu/depts/tandl/mtd/Proj6_CMT/6MiniTools.html.
- Tinkerplots*, the Statistics Education Research Group at the University of Massachusetts, Amherst, principal investigator: C. Konold, <http://www.umass.edu/srri/serg/projects/tp/tpmain.html>.

DANI BEN-ZVI
 Faculty of Education
 University of Haifa
 Mount Carmel
 Haifa 31905
 Israel