Developing Fourth-Grade Students' Statistical Reasoning about Distribution with *TinkerPlots* Software

By: Orly Segal Ben-Moshe

Supervised by: Dr. Dani Ben-Zvi

THESIS SUBMITED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE MASTER DEGREE

University of Haifa

Faculty of Education

Department of Education

February, 2007

Developing Fourth-Grade Students' Statistical Reasoning about Distribution with *TinkerPlots* Software

Orly Segal Ben-Moshe

Abstract

The increasing involvement of statistics in all areas of our lives has increased awareness of the need for universal statistical education. Many educational systems throughout the world have begun to include the teaching of statistics from an early age in their educational programs. Central to these programs is the development of statistical reasoning and literacy, which is considered today to be a universal desideratum for students and citizens alike.

This study deals with the development of statistical reasoning about a key concept in statistics – statistical distribution – among fourth-grade students. It deals with the empirical frequency distribution of a variable: the way in which collected data are distributed over the possible range of values of the variable.

Although understanding of the concept of distribution and its relation to other fundamental concepts is vital to the development of statistical reasoning and literacy (Garfield & Ben-Zvi, in press), previous studies have shown that it is difficult for students of various ages to take an overall view of distribution: to see the 'big picture' that emerges from a group of data. These studies show that when students assess, describe, and draw conclusions from a distribution, they concentrate on discrete values, ignore the rest of the distribution, and do not construct the 'big picture' which emerges from the group of data. Even high-school students who have learnt statistical concepts such as central measures of distribution in the framework of a yearly program of statistical studies completely ignored them in assignments requiring comparison between distributions. Instead, they made partial and local comparisons between distributions, using discrete data only (Ben-Zvi, 2004; Konold & Higgins, 2003). In the light of these findings, and in view of the crucial importance of the concept of distribution in the world of statistics, it is recommended that informal teaching of the concept of distribution should be begun at an early age, through investigation of frequency distributions, as a basis

for gradual progress towards advanced and complex understanding of the formal aspects of the concept (Cobb & McClain, 2004). This study is a contribution to knowledge in this area: it examines the possibilities of fostering informal reasoning about distribution in elementary schools.

In the spirit of the above-mentioned recommendations, the hypothesis of our study is that it is possible to improve and develop the statistical reasoning of young schoolchildren, and, in particular, informal reasoning about the concept of distribution, through a series of investigations of varied sets of data, in a rich educational environment, with the assistance of technological aids. The use of technology in processes of statistical learning and teaching can increase users' ability to work with data in complex ways, and develop conceptual reasoning, simultaneously with encounters with new difficulties and challenges (Artigue, 2002). The present study makes use of the modern educational computer program *TinkerPlots* (Konold, 2005), which deals with the teaching of data analysis to children from fourth to eighth grades. It is designed to meet the needs and abilities of beginners in the field.

In this study we examine the development of informal reasoning concerning the concept of distribution among fourth-grade students (aged 9-10) who have had no formal learning experience of data analysis, in a learning environment in which the use of *TinkerPlots* is combined with a series of data research-based investigation activities specially developed for this study. Its aims were to examine the following subjects in the framework of the learning processes which took place in the above mentioned learning environment:

a. The nature of the reasoning of fourth-grade students in relation to the concept of distribution, before the study of statistics at school.

b. Processes of the development of informal reasoning about the concept of distribution.

c. The contribution of the technological tool *TinkerPlots* to the development of reasoning about the concept of distribution.

d. The contribution of other factors in the learning environment to the development of reasoning about the concept of distribution.

The experiment took about one month (eight 90-minute lessons), in a school in the north of Israel, and encompassed three fourth-grade classes who took part in a study program which we designed, in the framework of mathematics lessons. In the light of the innovative and

exploratory nature of the research, throughout the experiment we paid special attention to three pairs of students who were chosen by the mathematics teacher as possessing a high capacity for perseverance and cooperation with the investigators throughout the research project. We also observed the other students, though less closely, in order to validate the conclusions of the analysis of the work of the three pairs under investigation. The investigator acted as an involved observer, the students' work was documented by a video camera, and an exact transcription of the text of the discussions was made.

We based this study on the definition of reasoning as a type of communication activity (also called 'discourse'), not necessarily verbal, and of learning as a process of improvement of the learner's ability to participate in a particular type of discourse, which is expressed in a change in his/her communication activity (Ben-Yehuda, Lavy, Linchevski & Sfard, 2005). Accordingly, we analyzed the communication activity of the subjects in the learning environment in order to gain knowledge of their methods of reasoning and their development.

In our analysis of the data we used a qualitative interpretative approach like that advocated by Shkedi (2003), which is based on the 'grounded theory' (Glaser & Strauss, 1967). For the analysis we also made use of the theoretical model of Konold, Higgins, Russell & Khalil (2003), who suggest four perspectives that students use in working with data.

Our findings show that before they have received formal instruction about data analysis in their school, fourth-grade students view statistical distribution locally; they concentrate on discrete values and their frequency, and on 'outstanding' values - especially the most frequent value, which they conceive of as the 'winning' value of the distribution. They look on the distribution as a 'competition between values', which is judged according to the magnitude of the frequencies; as a result, they relate to the values of a quantitative variable as nominal variables, and ascribe no significance to the order of the appearance of the values in the range, or to the distances between them.

Moreover, we discovered that reasoning about the concept of distribution developed in the course of the research project. When the students analyzed the distribution, they began to relate to another dimension – the range of the data; in other words, they began to consider differences and distances between different values in the distribution, and the dispersion of values over the range. With the help of this additional dimension of reasoning about

distribution, the students began to identify and use other characteristics of the distribution such as center, dispersion and tendency, and to establish connections between these concepts.

Work with the technological tool *TinkerPlots* enabled and encouraged the students to divide the distribution into bins of values, first into two bins, and, later on, into three (dynamic division of data into bins by means of dragging). Initially the students examined the discrete values in each bin according to their frequency, without relating to the bin as a single entity. Later, they began to consider the cumulative frequencies of the values in the different bins, to compare them, and, accordingly, to describe the general trend of the distribution. Thus, our study shows that there exists an intermediate stage between local reasoning and global reasoning, which are defined in the research literature as categories of reasoning about distribution. At the intermediate stage the students divide the distribution into bins with the aid of a cut-point, but examine the discrete values in each bin by the degree of their frequency. To this intermediate stage we have given the name 'bin-local view'. The technological tool *TinkerPlots* contributes to the transition from observation of discrete values to 'bin-local view' thanks to its simplicity, and the variety of possibilities of dynamic/visual division of groups of data which it affords. On the other hand, we discovered that in certain situations, the flexibility and openness of the software led the students to mistaken statistical conclusions. Moreover, we found that they had difficulty in dealing with distributions which dealt with unfamiliar subjects (such as the distribution of the height of the waves in Haifa harbor). We also found that distributions with certain characteristics (such as a distribution containing a number of modes whose frequency was similar to that of the other values in the distribution), and certain operations (for instance, reasoning about a typical value, or an attempt to find unusual data in the distribution) possessed the potential to improve the learning of statistical concepts.

The findings of our study broaden and deepen previously existing knowledge about the statistical reasoning of young students, and the ways in which it develops. In particular, they prove that even young students, in a suitable learning environment comprising informal data analysis, are capable of developing and internalizing fundamental statistical concepts. These findings constitute important information about the development of learning environments for the teaching of statistics at elementary school level, use of technological tools, and the choice and refinement of questions for future research.

In the light of our findings, we recommend the inception of programs of data analysis in the fourth grade. This can afford the students the possibility of the informal analysis of various sets of data, in subjects close to their day to day experience, combined with the use of suitable technological tools such as *TinkerPlots*. On the other hand, this recommendation raises the problem of the difficulties and challenges facing the educational system in providing such an environment; and this emphasizes the need for further research about the creation of suitable activities for the encouragement of statistical reasoning, training of teachers of mathematics in statistics, the function of the teacher in the learning process in the classroom, and the organization of instruction in a rich technological environment.