Children As Explorers:
Exploratory Data Analysis
By Junior High School Students
In A Computer Assisted Environment

By
Dani Ben-Zvi

Department of Science Teaching

PH.D. THESIS

Submitted to
The Feinberg Graduate School
The Weizmann Institute of Science

June 10th, 2001

Abstract

The overall goal of this dissertation is to advance the understanding of learning and teaching Exploratory Data Analysis (EDA) in a carefully designed computer assisted learning environment for the junior high school level. This thesis, presented as a collection of papers, documents and analyzes three dimensions of students' activity, their evolving conceptions, their interactions, and their adoption of points of view. The theoretical framework informing the analysis of students' learning assumes that competence in a new and complex domain, such as statistics, involves more than particular set of skills, strategies, or knowledge. Following Resnick (1988), the framework includes viewing competence of a domain as adopting the habits, language and dispositions of interpretation and sense making (enculturation).

Firstly, I analyze at a very fine level of detail the ways in which a pair of students began to make sense of data and data representations, as well as the process of adopting and exercising the habits and points of view that are common among EDA experts. The focus is on the ways they started to develop global views of data and their representations on the basis of their previous knowledge and different kinds of local observations. I examine how knowledge was gradually constructed through complex cognitive and socio-cognitive processes, which included their interactions with each other, the teacher, the materials and the computerized tool. I analyze the ways in which the same 'pieces' of students' prior knowledge which seemed to hinder progress, ultimately became the support for the construction of new meanings. Of special interest were the teacher's interventions, which though

short and not necessarily directive had catalytic effects, can be characterized in general as interesting instances of appropriation.

Secondly, through the analysis of students' 'research projects', I suggest an initial framework of student reasoning in the domain of EDA with an emphasis on handling data representations. I then use written assessments to characterize student sense making of statistics after the end of their experience with the curriculum. Finally, I analyze the interrelationships between curriculum design and research in order to characterize the nature of the instructional activities, including the role and impact of computerized tools, which have the potential to promote meaningful learning of EDA.

The research methodology is mainly qualitative in nature with some quantitative aspects. The subjects for this study were seventh grade students (13-year-old) of mixed ability from the experimental classrooms that used the curriculum. The analysis is based on:

1) focused and detailed data on one pair of students, which were videotaped at almost all stages of their learning statistics;
2) classroom data that was gathered in three experimental classes; and
3) summative assessment data that consist of students' 'research projects', written assessments, students' evaluations, and teacher's comments.

These data were used to characterize important phenomena related to the following questions:

1) how students choose, interpret, design, transform and use data representations?;
2) what are the contributions of student interactions with their peers and their teacher to their understanding of data representations?; and
3) how students adopt the habits and points of view that are common among EDA experts, in particular the experts' point of view on local-global approaches to data interpretation and their representations?

The dissertation shows how meaningful learning of EDA took place through complex socio-cognitive processes of enculturation, the processes of teacher-student (and, possibly, student-student) appropriation, the students' exposure to carefully designed learning arenas and relevant computerized tools, and the students' long-term involvement in constructing a 'research project.'

CONTENTS

- **References**
- **Appendices**
  - Appendix I Sequence and types of data collection activities
  - Appendix II A and D's statistical explorations sequence
  - Appendix III The Statistics Curriculum student textbook

LIST OF PUBLISHED AND NON-PUBLISHED RESEARCH PAPERS

This thesis includes synopses and full manuscripts of the following five published (or in press) research papers, and one non-published research paper.

Paper I

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

Paper II

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

Paper III

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*, 45-55. Voorburg, The Netherlands: International Statistical Institute.

Paper IV

Ben-Zvi, D. (unpublished paper, thesis version). *Seventh Grade Students' Sense Making of Data and Data Representations at the End of the SC.*

Paper V

Ben-Zvi, D., & Arcavi, A. (1998). Towards 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 of Statistics **II**, 647-653. Voorburg, The Netherlands: International Statistical Institute.

Paper VI

Page 156

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

ACKNOWLEDGEMENTS

I would like to first thank my dissertation supervisors, Abraham Arcavi and Alex Friedlander, for their ongoing encouragement and guidance. I would especially like to thank them for many hours of viewing videotapes, discussing conjectures and interpretations, and contributing their complex understanding of learning and teaching processes. I thank them for reading early drafts, struggling through telegraphic text, giving me feedback on half-baked ideas, and having to deal with convoluted prose. They encouraged me to consider statistical content; cognitive, socio-cognitive and later socio-cultural approaches to learning and research; and various aspects of the innovative educational environment. I also thank my dissertation examining committee, Lee Segel and Michal Yerushalmy. My meetings with them were important and stimulated new ideas and directions.

I am deeply indebted to the teachers and students who welcomed me into their classrooms. I especially thank A and D who agreed to be part of the case study including being videotaped, and repeatedly being 'nudged'. I owe many of the insights in this dissertation to the open and candid discussions of these students and their peers. Their curiosity, perseverance, and creativity were inspiring. I hope that the insights in this thesis will make these and other students' classroom experiences in statistics more meaningful and interesting. I would especially like to thank the three skillful and dedicated teachers, Michal Tabach, Hannah Stein, and Gila Ozruso, who always welcomed me with patience and provided every possible support and advice.

My conversations with colleagues were essential in the writing of the thesis. Conversations with Maxim Bruckheimer, Tommy Dreyfus, Ruhama Even, Nurit Hadas, Rina Hershkowitz, Anna Sfard, Tsippi Resnick, Baruch Schwarz, and others in the Weizmann Science Teaching Department, helped me to put many of the ideas in the thesis into words, and to attempt to explore new directions. During my work on the thesis, I was very fortunate to meet many colleagues and friends from all over the world .

My discussions with them were also important and supportive while I was conducting the research and writing the papers. I mention here just a few, who read and reviewed earlier versions, and provided support in many helpful ways: Janet Ainely, Carmen Batanero, Bob delMas, Iddo Gal, Joan Garfield, Brian Greer, Jonathan Moritz, Maxine Pfannkuch, Chris Reading, Jane Watson, and many others. I especially would like to thank Joan Garfield. Her dedication to statistics education, research, and curiosity were inspirational and influenced my work on this thesis.

The Weizmann Science Teaching Department provided a supportive environment where I could present 'pieces' of the thesis as 'work in progress'

and watch the videotape in great company providing helpful feedback. I would also like to thank the department administrative staff doing their outmost efforts to help in any service needed. This dissertation would not have been possible without various forms of institutional support provided by the Feinberg Graduate School and the Science Teaching Department of the Weizmann Institute of Science. Kibbutz Be'eri and its members were also supportive along the way.

And lastly, I would like to thank my family, my wife and friend, Hava, my children, Michal, Dagan, Nir, and Noa, and my parents, who understood that this thesis was an important project for me, and suffered my frequent and extended 'disappearances' for so many years. They have been my main source of energy, always thinking I could do this, even when I was most discouraged and anxious.

I am thankful to the opportunity given to me in writing this thesis. The development and growth that came with it were often painful, but ultimately rewarding.

A NOTE TO THE READER

This thesis developed through extended planning, trying ideas, testing and implementing educational materials, gathering data, writing and presenting, collecting feedback, discussing and more discussing, and so forth. Through this process, my conceptions and understandings of the issues related to children learning statistics have evolved and changed. Thus, the introductory overview of the study (Chapter 1) mostly reflects my intentions and views of what is researchable, interesting, and original at the 'start point'. However, the research evolved to show that the issues selected as my research questions were richer and wider, and thus the scope of my initial questions got enlarged and immersed in 'larger issues', without leaving aside the initial questions. I could have arranged that introduction to completely fit the 'end point', but I preferred the readers to get a feeling of how the study evolved. I hope this structure preserves this sense of growth, and provide you, the reader, with a perspective on those changes that this thesis and myself have gone through during this scholarly endeavor.

"I learnt a lot from my teachers,
and even more from my colleagues,
but from my students - I learnt the most".
(Talmud)

INTRODUCTION

Statistics is becoming ever more pervasive, and we live in a society which is ever more dependent on information. Major political, social, economic and scientific decisions are made on the basis of data. Statistics is a discipline which provides means for dealing with data. Statistical reports affecting virtually all aspects of our lives appear regularly in all the news media. Therefore, the importance of statistics literacy is becoming a major goal of the school curriculum, regardless of the professional future of the student (Gal, 2000). This thesis joins the increasing efforts and interest in the study of statistics education.

The presentation of this thesis is the final part of a 'direct Ph.D. program' at the Feinberg Graduate School, of The Weizmann Institute of Science. I began with studies leading to the M.Sc. Degree, focusing primarily on developing junior high school statistics curriculum. I reviewed relevant statistics curricula, software (Ben-Zvi, 1997a) and research literature, established a set of curriculum design principles, and wrote and tested instructional materials to be used in an innovative educational environment.

With the change of status from the M.Sc. to the Ph.D. study, I continued the cycles of writing - field-testing - research - improvements (for an account of the curriculum development, see Hershkowitz, Dreyfus, Ben-Zvi, Friedlander, Hadas, Resnick, Tabach, and Schwarz, in press). The first field experiments culminated in the production of an improved version of the student textbook (Ben-Zvi & Friedlander, 1997a, see Appendix III), the development and implementation of teachers' courses, the raising of numerous research issues involving many variables, as well as a complex pedagogical situation, and the narrowing down to the main themes of investigation (see Chapter 1).

This Ph.D. thesis is a collection of papers presenting in a coherent description how junior high school students begin to learn Exploratory Data Analysis (EDA, in the sense of Shaughnessy, Garfield and Greer, 1996) in a carefully designed, computer assisted learning environment. The thesis consists of five published (or in press) research papers, in which I was the only author or the major contributor. In addition, I include a non-published study, presented in a published paper format. Each of the six research papers is introduced by a brief synopsis[1] and is linked to the research questions and the thesis 'story' to form a logical continuum (Chapters 2, 3 and 4). At the end, I review and discuss the study, and suggest five major conclusions, theoretical implications, and practical implications for education (Chapter 5).

The first two papers analyze at a very fine level of detail the ways in which two students (A and D) began to make sense of data and data representations, as

[1] The synopsis of each paper is presented below before the full manuscript.

well as the process of adopting and exercising the habits and points of view that are common among EDA experts (Chapter 2). The third paper analyze typical statistical thinking modes of junior high school students, that is, an initial taxonomy of novices' reasoning in the domain of EDA, with an emphasis on handling data representations (Paper III ,in Chapter 3). In the fourth paper, I complement the former analysis by assessing students' understanding of statistics after the end of the Statistics Curriculum (SC) based on their responses to a 'real' data-based assessment task (Paper IV, in Chapter 3). The fifth paper analyzes the mutual relationships between the design of the SC and research on learning, in order to characterize the nature of the instructional activities, which promoted meaningful learning of statistics (Paper V, in Chapter 4; see also, Hershkowitz et al., in press). Finally, I study the impact of computerized tools in enhancing learning and understanding statistics (Paper VI, in Chapter 4). Thus, this thesis includes five chapters:

1. Overview of the study: aims and research questions, methods, a review of literature on teaching and learning EDA, and the SC learning environment.
2. A and D's case, concentrating on the construction of meanings and global views of data and data representations by one pair of students.
3. Students' statistical reasoning and their understanding of data at the end of the SC.
4. The impact of curriculum design and technology in enhancing the learning of EDA.
5. Discussion of conclusions, theoretical and educational implications.