Assessing Mathematical Problem Solving Behavior in Web-Based Environments Using Data Mining

Moshe Leiba

School of Education, Tel Aviv University, Ramat-Aviv, Tel-Aviv 69978
Israel
moshelei@post.tau.ac.il

Abstract. Over the years, mathematical problem solving research has focused on describing the process, as well as on understanding attributes affecting it, and assessing its outcomes. Most of the research in this field is qualitative, and this is understandable due to the fact that cognitive and meta-cognitive investigation involved in problems solving are complicated to be traced. Nowadays, when many problem solving environments are implemented using the web, innovative research methodologies may be applied for assessing problem solving behavior in large populations. The core of this research entails the development of a correspondence scheme between the logged traces of the students and the observed problem solving behavior. Furthermore, patterns of problem solving behavior and the factors influencing them are to be investigated in large population.

Keywords: web based learning environments, mathematical problem solving, log files, educational data mining

1 Introduction

A great emphasis is being placed on students' mathematical problem solving, and this domain is well represented in the curriculums and in the standards for school mathematics in Israel and worldwide [1]. Many problem solving environments are implemented using the Internet infrastructure which can allow innovative research methodologies to be applied for assessing problem solving behaviors. These innovative research methodologies rely on log file records, which automatically and continuously collected by Internet servers, document (almost) every action taken by three basic parameters: what was the action taken, who took it and when [2]. Following that, the core of this research entails the development of a correspondence scheme between the logged traces and the observed problem solving behavior. Furthermore, patterns of problem solving behavior and the factors influencing them are to be investigated in large population.

2 Background

2.1 Mathematical Problem Solving

Problem solving can be regarded as a situation in which an individual is responding to a problem that he or she does not know how to solve with routine or familiar procedures. Problem solving can be described as composed of three dimensions: the problem, the process and the outcome. These dimensions are detailed in the following sections.

The problem. A problem is only a problem if you don't know how to go about solving it. A problem that has no 'surprises' in store, and can be solved comfortably by routine or familiar procedures is an exercise [3, p.41]. Problems may vary in aspects such as: substance, structure, process to be carried, nature, modes of presentation or representation, and in their components and interactions among them. Jonassen [4] described differences among problems in terms of their structuredness, complexity, and abstractness:
Structuredness: Well-structured problems require the application of a finite number of concepts, rules, and principles being studied to a constrained problem situation. Ill-structured

problems, on the other hand, possess problem elements that are unknown or not known with any degree of confidence have multiple solutions, solution paths, or no solutions at all.
Complexity: The number of issues, functions, or variables involved in the problem; the degree of connectivity among these properties; the type of functional relationships among these properties; and the stability among the properties of the problem over time.
Abstractness: Problem solving activities are situated, embedded, and therefore dependent on the nature of the context or domain, because solving problems within a domain relies on domain-specific cognitive operations.

The process. Polya's [5] seminal work suggested that solving a problem involves 4 phases (or episodes): (a) understanding the problem; (b) developing a plan; (c) carrying out the plan; and (d) looking back. Hence, the problem solving process is described as linear progression from one phase to the other. Schoenfeld (1985) observed that during problem solving, students display distinct categories of behavior, also called episodes. Crucial episodes are: analyzing the problem, selecting appropriate mathematical knowledge, making a plan, carrying it out, and checking the answer with relation to the question asked.
Schoenfeld [6], [7], [8], [9], [10] contributed a framework of different factors (attributes) that affect students' abilities to solve problems. In his framework, four components comprise the major aspects of students' problem solving: (a) Resources: formal and informal knowledge about the content domain, including facts, definitions, algorithmic procedures, routine procedures, intuitive understandings of mathematics, and relevant competencies about rules of discourse; (b) Heuristics: strategies and techniques for approaching a problem; (c) Control: the ways in which students monitor their own problem solving process, use their observations of partial results to guide future problem solving actions, and decide how and when to use available resources and heuristics; and (d) Beliefs: what one believes about mathematics, mathematical tasks, and what it means to do mathematics.

The outcome. This dimension consists of the assessment of the problem solving outcome, which involves the assessment of the outcome creativity. Research on creative thinking identified three key components of a creative product: fluency, flexibility and novelty [11]. Fluency refers to the number of ideas generated in response to a prompt; flexibility refers to apparent shifts in approaches when generating responses to a prompt; and novelty – to the originality of the ideas generated in response to a prompt.

2.2 Web-based Learning Environments for Mathematical Problem Solving

Web based learning environments (WBLE) can enhance students' problem solving by providing an environment to engage in playful exploration, test ideas, receive feedback, and make their understanding public and visible [12]. Underwood et al. [13] also recommend that technological tools should support multiple solution strategies and approaches, employ multiple representations, and link between representations. Web-based tools developed for mathematical problem solving include the dynamic geometry system [14], the PACT cognitive tutors [15], and ActiveMath [16].
Healy and Hoyles [17] found that in the context of a dynamic geometry environment, the presence of particular technological resources (features) can either enable or limit certain actions, hence affecting the available heuristics a student may use during his or her problem solving. As students are solving a problem, not only do they need to implement heuristics and utilize resources, but they also need some mechanism to evaluate their progress so that they are aware of, and are critically examining, their own decision making.

2.3 Educational Data Miming (EDM)

When engaging in Mathematical problem solving in web based learning environments, students leave traces of their activity in the form of log file records, which document every action taken by three basic parameters: what was the action taken, who took it and when [2]. Discovering and extracting educational information from these log files using data mining techniques is called Educational Data Mining (EDM).

The term Data Mining or Knowledge Discovery in Databases (KDD) refers to the automatic extraction of implicit and interesting patterns from large data collections [18]. EDM is the application of data mining techniques to educational settings and is an emerging methodology in education aiming to gain insights on the learners' performance in several levels (e.g., cognitive, meta-cognitive) on large populations [19], [20].
Most research about online learners' activity on the web usually focuses on operational variables, (e.g., time patterns, pace, order of contents viewed), while higher-level cognitive variables, describing the characteristics of the learners' online learning, are less studied. This is of no surprise: traditional research methodologies fail to cope with the complex gathering of information about the online learner.
Building on the existing body of literature, this research will focus on gaining new information about the problem solving process, using an emerging methodology, Educational Data Mining. In doing so, we pursue a better understanding of the cognitive and meta-cognitive processes involved in problem solving, acquiring specific information about major problem solving attributes influencing the problem solving process in large populations.

3 Research Objectives and Expected Significance

The main purpose of the suggested research is to explore cognitive and meta-cognitive processes during problem solving activities in online environments being used in elementary schools. To this end, log files of the learning environments – automatically and continuously being collected – and Data Mining tools and techniques will be used to portray the online learner's behavior. Bridging the gap between the analyzed logged data and the problem solving processes is the core of this research and its main contribution. Three objectives have been defined for this research:

1. Developing a conceptual framework, a computational mechanism and a correspondence scheme between them for assessing mathematical problem solving behavior in Web-based environments by means of logged data.
2. Identifying different patterns of online problem solving behavior using quantitative analysis for large population (N>1,000).
3. Examining the effect of different attributes (e.g., problem type, student pre-achievement level, creativity) on problem solving behavior.

The significance of the proposed research is two-fold. On the practical level, results of this study will enable the automatic assessment of problem solving processes for large populations; this will aid both instructors and researchers to better understand students' behavior. On the theoretical level, association between problem solving attributes and phases will be empirically tested on large population; this will shed light on the nature of problem solving processes and will provide the current body of knowledge on problem solving with empirical evidences.

4 Methodology

Large scale assessment of learners' problem solving has been a challenge for both researchers and instructors for many years. The main reason is the nature of the problem solving process, which involves attributes from cognition and meta-cognition dimensions – and the notion that these dimensions are traditionally examined using qualitative tools and methods with small-scale populations.
The basic assumption of our research is that problem solving behavior is reflected in the student's behavior while interacting with an online problem solving web-based environment; hence, traces of the student's problem solving behavior might be extracted from the log files and may shed light on this behavior for large populations in ways that were not previously possible.

Research Population. Participants in research will include 10 fifth and sixth grade students divided to two equal groups for the construction of the correspondence scheme. Later on, log files from a large population (N>1,000) of students of the same grade levels will be analyzed for the large-scale assessment of problem solving behavior.

Research Field. An online learning environment in Mathematics (developed by CET - Center for Educational Technology) was chosen, including problems that differ in structuredness, complexity and domain by using a grid on which students can construct geometry objects (e.g., dots, angles, lines, polygons), measure them (length, area, angle), and transform them (move, resize, delete). In addition, student can color squares formed by the grid. This Geometry Applet has many applications within the Israeli elementary school Mathematics curriculum, and it's being used for, e.g., representing fractions, constructing polygons, finding patterns, and measurements. The applet stores the student actions in fine-grained log files.

Research Variables. Independent and dependent variables, based on the literature review, include 3 independent variables and six categories of dependent variables, as detailed in this section. The conceptual framework describing the theoretical relations between these variables is illustrated in Fig. 1.

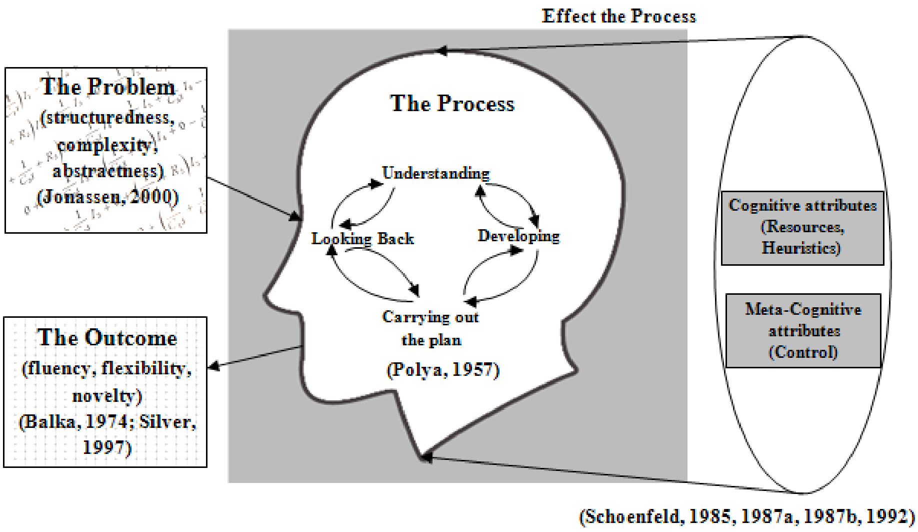

Fig. 1. The conceptual framework of the research.

Independent Variables. Three variables were defined for describing the problem:

1. Structuredness - distinguishes well-structured from ill-structured problems
2. Complexity - defined by the number of issues, functions, or variables involved in the problem. Problem difficulty is a function of problem complexity
3. Abstractness - nature of the context or domain (the battery of tasks include tasks involving fractions, geometry and patterns).

*Dependent Variables***.** Six categories of variables were defined. The first two categories (attributes related) are defined according to Schoenfeld's framework [6], [8], [10], the third category (process related) is defined according to Polya's work [5], and the fourth category (outcome related) is defined according to Belka [21] and Silver [22]; the fifth category consists of log-based variables. Following is a description of these categories:

a. Cognitive attributes – within this category, two sets of variables were defined to describe the conceptual and procedural knowledge of the students: The procedural knowledge is measured by the student's acquaintance with resources (e.g., facts, definitions, procedures, rules, intuitive understandings of mathematics), the conceptual knowledge – by the student's using of heuristics (i.e., strategies and techniques for approaching a problem).
b. Meta-cognitive attributes – this category holds variables describing the control of the student (e.g., the methods in which one monitors the problem solving process, uses

observations of partial results to guide future problem solving behavior, and decides how and when to use available resources and heuristics).

c. Process description – the way the student is progressing from one phase (of the four possible phases) to another.
d. Product assessment – this category holds 3 variables for evaluating the product/solution: *Fluency*, *flexibility*, *novelty*.
e. Log-based variables – the 4 variables to be extracted from the log files are: *Current Tool* ; *Action*; *user; and Time*

Research Instruments. Three main sets of instruments serve this research:

a. Think-aloud protocols – Think-aloud is a method for describing and analyzing thinking processes, during which the student is being asked to verbally describe thoughts and feelings - out loud and in details – simultaneously during a task operation [23].
b. Learnograms – visualization tool for presenting log-based learning variables over time [24]. Learnograms will be used in order to portray the problem solving process by means of episodes, as was done qualitatively by Schoenfeld [10].
c. Battery of tasks – several online tasks using the *Geomtery Applet* were developed, representing different types of complexity, problem space. The tasks will be planned in accordance to the research population syllabus.

Procedure. A Mixed Method [25] involving both qualitative and quantitative analysis, has been chosen for this research. Problem solving behavior will be assessed by means of qualitative research (using think-aloud protocols), as well as actual learning behavior in the online learning environment using Learnograms; patterns of problem solving processes and factors affecting them will be investigated using quantitative methods using log files.

Phase I – Constructing the Correspondence Scheme. During this phase, data from both students and log files are triangulated, aiming to reflect on students' problem solving behavior in log files. The procedure is as follows:

1. Identifying a list of behavioral variables to be extracted from the log files, based on the theoretical framework. Examples of variables include: order of actions, number of solutions, and total time for reaching the solution.
2. Characterizing a list of variables related to the cognitive and meta-cognitive attributes of problem solving, based on qualitative analysis with a sample (N=5) of students representing the discrepancy in the population background (knowledge level, achievements). To this end, observations, interviews and think-aloud protocols will be used. Examples of variables include: creativity, solution strategy.
3. Assessing the means by which attribute-related variables are reflected in the log-based behavioral variables.
4. Validating the scheme by studying a second small set of students (N=5). A second researcher will be given with a description of these students' problem solving behavior, based on their log files, and with think-aloud protocols of them. The second researcher will compare his or her understanding of the think-aloud protocols with the results from the log files analysis.

The expected result of this phase is a validated correspondence scheme for "translating" the qualitative-based outcomes to computable variables based on automatically collected data.

Phase II – Large-scale Problem Solving Characterization. Based on the correspondence scheme constructed in Phase I, the purpose of this phase is to investigate problem solving processes and attributes in large population (N>1,000) by means which were not feasible in previous research.

Distribution of variables describing the problem solving attributes and phases will be investigated, in order to better understand their expression in large population. Furthermore, relations between the variables will be examined using statistical and Data Mining methods, to acquire better understanding and more specific information about major problem solving attributes reported as influencing the problem solving process in large populations.

Phase III – Finding Factors Associated with the Problem Solving Process. A few factors will be investigated in order to associate problem variables (structuredness, complexity, and abstractness) and student variables (achievements level, creativity) with problem solving attributes (cognitive and meta-cognitive) and process. To this end, statistical and Data Mining methods will be used.

References

1. National Council of Teachers of Mathematics: Principles and standards for school mathematics. Reston, VA: Author (2000)
2. Pahl, C.: Data mining technology for the evaluation of learning content interaction. International Journal of E-Learning, 3(4), 47-55 (2004)
3. Schoenfeld, A.H.: The wild, wild, wild, wild, wild world of problem solving: A review of sorts, For the Learning of Mathematics, 3, 40–47. (1983)
4. Jonassen, D. H.: Toward a design theory of problem solving, Educational Technology Research and Development, 48(4), 63-85 (2000)
5. Polya, G.: How to Solve It; a New Aspect of Mathematical Method (2nd ed.), Doubleday, Garden City (1957)
6. Schoenfeld, A.H.: Mathematical problem solving, Orlando, FL: Academic Press (1985)
7. Schoenfeld, A.H.: Cognitive science and mathematics education: An overview, in A.H. Schoenfeld (Ed.), Cognitive Science and Mathematics Education, pp. 1-31. Lawrence Erlbaum Associates, London (1987a)
8. Schoenfeld, A.H.: What's all the fuss about metacognition? In A.H. Schoenfeld (Ed.), Cognitive Science and Mathematics Education, pp. 189–215. Hillsdale, NJ: Lawrence Erlbaum Associates (1987b)
9. Schoenfeld, A.H.: Explorations of students' mathematical beliefs and behavior, Journal for Research in Mathematics Education, 20, 338–355 (1989)
10. Schoenfeld, A.H.: Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics, In D. Grouws (Ed.), Handbook for research on mathematics teaching and learning (pp. 334–370). New York: Macmillan (1992)
11. Torrance, E. P.: The Torrance tests of creative thinking, Technical-norms manual, Bensenville, IL: Scholastic Testing Services (1974)
12. Clements, D.H.: From exercises and tasks to problems and projects: Unique contributions of computers to innovative mathematics education, Journal of Mathematical Behavior, 19, 9–47 (2000)
13. Underwood, J., Hoadley, C., Lee, H. S., Hollebrands, K. F., DiGiano, C., & Renninger, K. A.: IDEA: Identifying design principles in educational applets, Educational Technology Research and Development, 53(2), 99–112 (2005)
14. Balacheff N.G.M.: Diagnostic et raisonnement explicatif dans cabri-geometre. In B. P. Karsenty L. (Ed.), Explication et cooperation homme-machine: vers la coconstruction d'explications (1-6). CNAM (1993). (http://www.cabri.net/index-e.html)
15. Corbett, A., Koedinger, K., & Anderson, J.:. Intelligent tutoring systems, In P. P. V. Helander M. G., Landauer T. K. (Eds.), Handbook of human-computer interaction (849- 874). The Netherlands: Elsevier Science (1997)
16. Melis, E., Andrès, E., Büdenbender, J., Frishauf, A., Goguadse, G., Libbrecht, P., Pollet, M., & Ullrich, C.: ActiveMath: A web-based learning environment, International Journal of Artificial Intelligence in Education, 12(4), 385-407 (2001)
17. Healy, L., & Hoyles, C.: Software tools for geometrical problem solving: Potentials and pitfalls, International Journal of Computers for Mathematical Learning, 6(3), 235–256 (2001)
18. Klosgen, W. & Zytko, J.: Handbook of Data Mining and Knowledge Discovery, Oxford University Press, Cary, NC (2002)
19. Heiner, C., Baker, R. & Yacef, K.: Proceedings of the Educational Data Mining Workshop. ITS'2006 (2006)
20. Romero, C., & Ventura, S.: Data mining in e-learning, Southampton, UK: Wit Press.
21. Balka, D.S. (1974). Creative ability in mathematics, Arithmetic Teacher, 21, 633–636 (2006)
22. Silver, E.: Fostering creativity through instruction rich in mathematical problem solving and problem posing, Zentralblatt fuer Didaktik der Mathematik, 29(3), 75–80 (1997)
23. Newell, A., & Simon, H.: Human problem solving, Englewood Cliffs, NJ: Prentice Hall (1972)
24. Nachmias, R. & Hershkovitz, A.: Using Web mining for understanding the behavior of the online learner, Paper presented at The International Workshop on Applying Data Mining in e-Learning (ADML'07), Crete, Greece (2007)
25. Creswell, J. W.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches Thousand Oaks, CA: Sage Publications (2003)