

UNDERGRADUATE STUDENTS' MODELS OF CURVE FITTING

Shweta Gupta
Kean University

The Models and Modeling Perspectives (MMP) has evolved out of research that began 26 years ago. MMP research uses Model Eliciting Activities (MEAs) to elicit students' models of mathematical concepts. In this study MMP were used as conceptual framework to investigate the nature of undergraduate students' models of curve fitting. Participants of this study were prospective mathematics teachers enrolled in an undergraduate mathematics problem solving course. Videotapes of the MEA session, class observation notes, and anecdotes from class discussions served as the sources of data for this study. Iterative videotape analyses as described in Lesh and Lehrer (2003) were used to analyze the videotapes of the participants working on the MEA. Results of this study discuss the nature of students' models of the concept of curve fitting and add to the introductory undergraduate statistics education research by investigating the learning of the topic curve fitting.

Keywords: [Problem solving, Curve fitting, Modeling, Models, Statistical reasoning]

Introduction

Research on the teaching and learning of curve fitting is meager when compared to other statistical topics, such as the measures of central tendency or variability. Curve fitting, as a topic, is important, not only in general mathematics, but also in specific areas that correspond with the subject, including engineering and science. Situations in which curve fitting is required frequently arise in every day life in which we are given a set of data about which we would like to make a prediction. Only few studies discussed *how to learn* and *how to teach* this topic. But even those studies simply discussed what the students understand and do not understand about curve fitting. There was no literature that discussed, in depth, how students learn this topic. Therefore, this study helps to fill this void in the curve fitting literature, and consequently the undergraduate statistics education research literature.

In order to conduct research on the way in which someone learns mathematics, the learning must actually occur on the part of the participants. Extensive research exists on students' mathematics learning through clinical interviews (Hunting, Davis, & Pearn, 1996; Hunting & Doig, 1992) and teaching experiments (Cobb, 2000; Cobb & Steffe, 1983; Simon, 1995; Steffe & Thompson, 2000). In addition, literature exists in regard to students' learning through problem solving (Clement, 2000). Learning via problem solving was used as a method for this study as approaching mathematics through problem solving can create a context which simulates real world and, therefore, justifies the mathematics rather than treating it as a means to an end.

The research that has been successful in investigating students' learning via problem solving is the models and modeling perspectives (MMP) developed by Lesh and his colleagues over the past 26 years (Lesh & Doerr, 2003; Lesh, Hamilton, & Kaput, 2007; Lesh, Hoover, & Kelly; 1993; Lesh & Lamon, 1992; Lesh, Landau, & Hamilton, 1983). MMP uses the notion of students' models to study learning of a particular mathematics topic. MMP researchers have changed the notion of problem solving as *solving* difficult mathematical problems to *modeling* complex mathematical activities. Detailed discussion on MMP is provided below.

Models and Modeling Perspectives

Models are defined as “conceptual systems that are expressed using external systems and that are used to construct, describe, or explain behaviors of other systems” (Lesh & Doerr, 2003, p. 10). MMP is the name given to the theoretical perspectives that have evolved from the research first utilized more than 26 years ago by Lesh and his colleagues (Lesh & Doerr, 2003; Lesh, Hamilton, & Kaput, 2007; Lesh, Hoover, & Kelly; 1993; Lesh & Lamon, 1992; Lesh, Landau, & Hamilton, 1983). As this study will focus on studying the nature of students’ models of curve fitting through a methodology that uses MEAs (Model Eliciting Activities), MMP will serve as a useful conceptual framework as it brings together two important, but separate research traditions: problem solving and conceptual development, in mathematics education research.

From MMP, problem solving and conceptual development in mathematics can be seen as co-developing as modeling can be seen as local conceptual development. Local conceptual development refers to the, “development of powerful constructs in artificially rich mathematical learning [problem solving] environments” (Harel & Lesh, 2003, p. 360). Lesh and Harel (2003) state that:

[W]hen problem solvers go through an iterative sequence of testing and revising cycles to develop productive models (or ways of thinking) about a given problem solving situation and when the conceptual systems that are needed are similar to those that underlie important constructs in the school mathematics curriculum, then these modeling cycles often appear to be local or situated versions of the general stages of development that developmental psychologists and mathematics educators have observed over time periods of several years for the relevant mathematics constructs. (p. 157)

In other words, during an MEA session, students go through several modeling cycles that lead to students’ conceptual development.

Objectives of the study

The research goal of this study was to investigate the nature of undergraduate students’ models of curve fitting.

Methods

MMP research uses MEAs, which are designed specifically for research purposes (Lesh & Doerr, 2003; Lesh, Hoover, Hole, Kelly, & Post, 2000). MEAs are simulations of real world situations which are often used in research for their *model-eliciting* properties. The MEAs are designed using six principles (Lesh et al., 2000). Model-eliciting property refers to the way in which MEAs are designed to encourage students to clearly express, not only their final models, but, also, the numerous models that they create, revise and reject along the way. MEAs are different from typical mathematical modeling activities in that they not only require students to clearly express their final models, but, also, elicit the students’ intermediary models.

A typical MEA session involves three distinct phases, summarized in Figure 1 below.

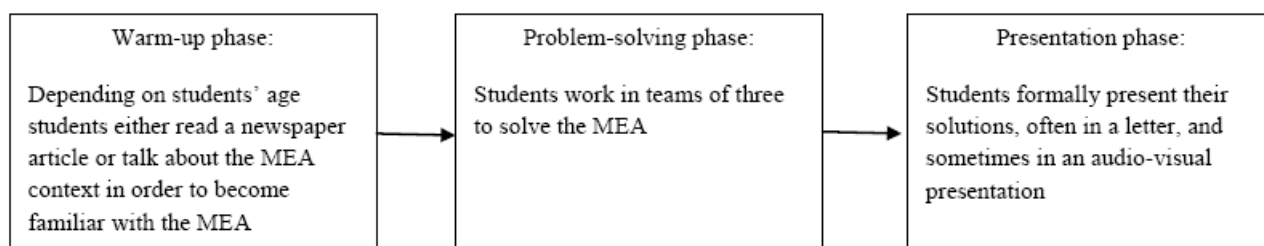


Figure 1. A typical MEA session

This study used the Time on Drill and Test Scores problem (TDTS) which is an MEA and has been used, tested and revised with different populations of students, both undergraduate (at all different levels of their program) and graduate. The core statistical ideas of this MEA are centered on the notion of fitting a line or curve to make a prediction about the situation in the MEA. The students in this study had no specific formal exposure to or instruction in these ideas prior to this MEA. Rather, this MEA was designed so that the students could readily engage in meaningful ways with the problem situation and create, use and modify the quantities in ways that would be meaningful to them and could be shared, generalized and reused in new situations. An excerpt from the TDTS problem appears in Table 1. In the TDTS problem, the problem solvers are supposed to provide the school administrators with a prediction of the test scores of students in several schools based on the time that the schools spend on the drill that teaches the information on the test. Problem solvers are given data on the time spent on the drill by 26 different schools and their respective average student test-scores. The TDTS problem was designed to elicit the notion of curve fitting for the purpose of making a prediction about the test scores.

Table 1. *Excerpts from the TDTS Problem*

Time Spent on the Drill as a Predictor of Performance on a Standardized Test		
School	Time Spent on the Drill (in minutes)	Test Scores
A	17.49	49.55
B	28.02	48.76
C	23.62	64.69
D	36.6	57.74
E	39.92	55.55

Table reports the amount of time spent on the drill each week. Then, it shows each school's average score on the end of the year test. The school administrators want to know whether the time spend on the drill is helpful in improving the test scores. For example, the administrators want to predict the end of the year test score for a school that spends 120 minutes per week on the drills. The school administrators need your help in making this prediction. In addition to making a prediction, they need you to write a letter describing how you made this prediction. They will use your process to make predictions about the test scores given any time on drill. Thus, you need to make sure that your process for making predictions will work for any time-on-drill.

Settings

The study took place in March of 2007 in an undergraduate mathematics teacher education classroom at a large mid-western university. The student demographics for a typical undergraduate mathematics teacher education classroom at this University are female and 90% white.

Procedural Details and Data Sources

Table 2. *Procedural details*

Class session #	Task	Data collected	Number of students
1(~30 minutes)	Explaining the study and its purpose to the students	N/A	14
2(~75 minutes)	TDTS problem	Artifacts produced by the students, observation notes, anecdotes shared by AIs, audiotapes and videotapes of the session.	9
3(~75 minutes)	Class Presentation	Artifacts produced by the students, observation notes, audiotapes and videotapes of the session.	12

The data sources from the class included

1. audiotapes and videotapes of all of the classroom sessions,
2. the students' worksheets and final reports detailing the development of their models, detailed field notes, and anecdotes.

Data Analysis

Iterative videotape analysis (Lesh & Lehrer, 2000) was used to analyze the videotapes. Lesh and Lehrer described multiple windows through which to view a given video. For example, each video of students' work has theoretical and physical aspects. These aspects can be analyzed in a variety of ways including analysis of isolated sessions, analysis of one group across several sessions or analysis of similar sessions across several groups, where a session implies a MEA session. This study focused only on students' models, therefore, only the mathematical perspective of the theoretical aspect was focused and field notes and the transcripts of the audiotapes and videotapes were analyzed. As only one MEA was used and focused on only one particular group in this study, only an isolated session was analyzed, see Figure 2.

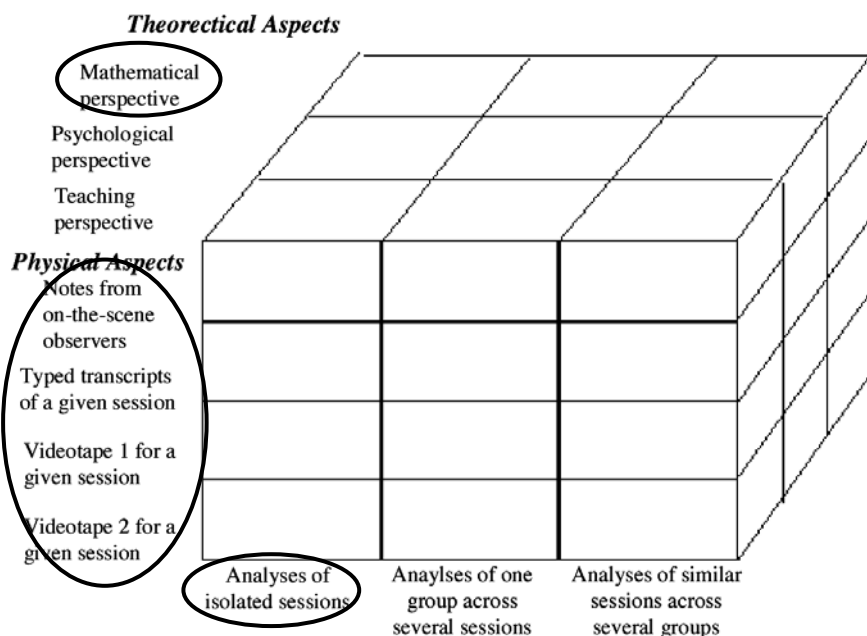


Figure 2. Multiple windows for viewing given sessions (the ovals represent the aspects that I focused on). (Lesh & Lehrer, 2000, p. 679)

Results

The results of the fine grained analysis of the cycles of conceptual development of curve fitting as displayed by the participants are given. In tracing the nature of the models of one group of 3 students, namely, Adam, Beth and Cathy, multiple cycles of increasing coordination and stability of conceptual systems that were observed in the students' responses are reported. Each cycle represents a shift in the students' thinking, providing powerful forms of information about the nature of students' models. In the present case, the cycles ranged from applying standard procedures without detailed analyses, to thinking about chunks of data, to sophisticated models of curve fitting that predict the situation. The analyses below shows that as students' work progressed on the TDTS problem, their conceptual systems evolved from being uncoordinated and unstable to becoming increasingly coordinated and

stable. As the students' conceptual systems evolved along this line of coordination and stability, they were developing the concept of curve fitting.

First Cycle: Finding Data Summary.

When the students looked at the TD and TS scatterplot, they came to the conclusion that “no relationship” existed between the TD and TS and that “no pattern” was apparent in the scatterplot. As Adam's group had a limited, but useful background, in statistics, it was natural for them to resort to the usual methods of finding mean and standard deviation, even though these numbers were not helpful in the context of the problem.

The group spent a short amount of time finding the standard deviation using the summary table (students can do a summary table in *Fathom*TM to find standard deviation for a data set) provided by *Fathom*TM. Then, they discussed their results and concluded that finding the standard deviation did not help them solve the problem. That is, standard deviation did not help them making the desired prediction about TS.

The students then decided to consider both sets of data, TD and TS, in an attempt to find a relationship between them because the students were unsure whether finding individual values such as mean and standard deviation for each TD and TS would help.

In this modeling cycle Adam's group *described* the scatterplot as having almost no correlation and found certain statistical values like standard deviation and mean. When they *translated* back to the original problem of making a prediction and tried to *verify* the usefulness of their results, they realized that finding such values did not make any sense in the context of the problem. Eventually, they started a new modeling cycle as described below.

Second Cycle: Applying Standard Procedures.

After making a scatterplot, the group spent time looking at the graph. As Caty had previously taken a traditional statistics course, in which she was taught about lines and curves of best fit, she suggested a line as a best fit for the data. It is interesting to note that while Caty thought that the shape of the scatterplot was not linear, she still suggested line as a best fit, as seen in their exchange below. One of the possible explanations could be that in her introductory statistics course she almost always used lines as a best fit for any given scatterplot without thinking about the underlying assumptions that a line of best fit makes (e.g., a strong linear relationship between the variables).

The group plotted a movable line, least squares line and median-median line, which are all built-in functions in *Fathom*TM, but were not convinced that the lines represented the best fit in regard to making the desired prediction. Therefore, they rejected the notion of linearity and moved on to discussing which curve would make the most sense. During the second cycle Adam's group *described* the relationship between TD and TS as linear and hence plotted the best-fit lines available in *Fathom*TM. They even plotted a movable line and *manipulated* it to make it fit to the scatterplot. When they *translated* back to the original problem and tried to *verify* their results they concluded that a linear relationship between TD and TS did not make any sense for the TDTS problem. So they shifted to another interpretation of the problem, that is, the third modeling cycle.

Third Cycle: Thinking About Correlation.

This group now started thinking about several functions (linear, polynomial etc...) in this cycle. They began discussing the different graphs in the third cycle in order to figure out which shape would best describe the scatterplot after rejecting the notion of linearity. After

introducing the sliders¹, which are built in to *Fathom*TM, Cathy and Beth attempted to move the sliders to fit the curve to the data, when Adam *translated* his focus back to the scatterplot he was not convinced that there was enough correlation between the TD and TS in order to make an appropriate prediction. Therefore, upon his suggestion, they returned to the original problem and started a new modeling cycle.

Fourth Cycle: Focusing on Small Chunks of Data.

The group again looked at the scatterplot and began focusing on individual data points and small groups of data points after Adam pointed out that “not enough correlation [existed] to make a prediction.” They then realized that some points on the scatterplot were “throwing off” any pattern in the plot and began concentrating on the individual points and whether each point was throwing off a potential pattern in the scatterplot.

When no obvious pattern was apparent in the scatterplot, the group decided to concentrate on the small groups of data points in order to see if a pattern existed. Then, they decided to make another scatterplot with the TDs between 20 and 30 minutes. This scatterplot did not help them because no pattern was obvious in the new scatterplot. In fact, this plot was less organized than the original scatterplot. They made the table and plotted the graph, however, the new scatterplot did not help them finding a pattern. After trying out small groups of data points, Adam’s group was convinced that they needed to look at all of the data points in order to find any patterns.

Fifth Cycle: Making a Prediction Using a Model.

The group then began arguing about patterns in the scatterplot of the whole data. Adam suggested that there was “no correlation” in the data and no obvious pattern, while Caty suggested that unless they found a pattern, they could not make the desired prediction. This argument started a discussion about the correlation between the TD and TS. Each of the students had a different idea about the correlation. Adam stated that there was “no correlation” at all, Beth stated that there may be some “coincidental correlation” and Caty stated that there was “some correlation” that may help in finding a pattern and making a prediction. As the problem asked them to make a prediction, they had to agree with Caty in order to proceed with the problem.

After the group decided that they *have to* look for a pattern in order to make a prediction, they started investigating the scatterplot more closely for a pattern. Caty suggested that they should look for “curvy lines” because the TSs increased “gradually” with the TDs. They came up with two curves on the scatterplot because, according to them, no single curve described the data the best. They discussed “somehow” combining the two curves in order to come up with a single curve to make a more accurate prediction, but did not have enough mathematical tools or skills to do that.

In their final solution, they used these curves to make their prediction. They also used the sliders to shrink and stretch the curves to fit the scatterplot. This modeling cycle culminated in their final model. During this cycle all the members of Adam’s group had different *descriptions* of the scatterplot. While *translating* back and forth from the original problem of making a prediction to the scatterplot they came up with a single idea of plotting two curves. Finally, when they *verified* their result with the original problem it made sense to them that the final prediction would lie between the two curves.

¹ *Fathom*TM sliders are used to vary the values of coefficients in the functions. You can interact with the sliders and note the changes in the functions.

From the analysis of the data, it is clear that the idea of curve fitting not only evolved, but changed significantly in the students' thinking throughout the TDTS problem session. The students encountered ideas about correlation; how to shrink, move and stretch curves; how to decide the "fit" and how to combine formulas. This simultaneous awareness of different concepts caused their idea of curve fitting to evolve.

Discussion

Results of this study provide insights into the nature of students' developing models of curve fitting. The data supports the claim that models evolve from being uncoordinated, unstable and undifferentiated to being increasingly coordinated, stable and differentiated as their work on the TDTS problem progressed.

Significance of the Study

This study contributes to several areas of the mathematics education research, including statistics education research and mathematical modeling, and problem solving in undergraduate mathematics. The significance of the products of this study can be assessed in the following ways:

1. This study adds to the existing statistics education research by investigating students' learning of a statistical topic, curve fitting, which has not been the subject of much research until now.
2. This study also introduces the use of a conceptual framework, MMP, which can be used to investigate the learning of other statistical topic areas. It offers an in-depth analysis of how students learn a particular topic via solving a real world problem.

Finding the nature of students' models also lays the groundwork for activities that could enhance students' understanding of the topic under investigation and, ultimately, improve instruction at college- and school-level mathematics and statistics.

References

- Harel, G., & Lesh, R. (2003). Local conceptual development of proof schemes in a cooperative learning setting. In R. A. Lesh & H. Doerr (Eds.), *Beyond constructivism: A models and modeling perspective on mathematics teaching, learning, and problem solving* (pp. 359-382). Mahwah, NJ: Erlbaum.
- Lesh, R., & Doerr, H. (2003). Foundations of a models & modeling perspective on mathematics teaching and learning. In R. A. Lesh & H. Doerr (Eds.), *Beyond constructivism: A models and modeling perspective on mathematics teaching, learning, and problem solving* (pp. 3-34). Mahwah, NJ: Erlbaum.
- Lesh R., Hamilton E., & Kaput J. (Eds.) (2007). *Foundations for the future in mathematics education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lesh, R., & Harel, G. (2003). Problem solving, modeling, and local conceptual development. *International Journal of Mathematics Thinking and Learning*, 5, 157-189.
- Lesh, R., Hoover, M., Hole, B., Kelly, A., & Post, T. (2000). Principles for developing thought revealing activities for students and teachers. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 591-645). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Lesh, R., & Lamon, S. J. (1992). *Assessment of authentic performance in school mathematics*. Washington, D.C.: American Association for the Advancement of Science.

- Lesh, R., Landau, M., & Hamilton, E. (1983). Conceptual models in applied mathematical problem solving. In R. Lesh, *The acquisition of mathematical concepts and processes*. New York, NY: New York Academic Press.
- Lester, F. K. (2005). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. *Zentralblatt für Didaktik der Mathematik*, 37(6), 457-467.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.