TECHNOLOGY, STATISTICS, AND SUBTLETIES OF MEASUREMENT: BRIDGING THE GAP BETWEEN SCIENCE AND MATHEMATICS

Tim Erickson
Epistemological Engineering
California, USA

In an ideal world, science students would act as scientists do: investigating their own questions, designing experiments, and so forth. This paper reports on curriculum development and field testing that takes a step in this open-ended direction. To do this, we have focused on integrating more data analysis into science activities; this also gives students a chance to use more mathematics, in an understandable context. This mathematics includes work with functions and variation. A closer look at plausible activities shows us that principles of measurement connect these elements; furthermore, a broad view of measurement reconnects us to our original goal: to expose students more directly to the nature of science.

INTRODUCTION

Our project integrates mathematics into the science curriculum through increased attention to data analysis. Data analysis is a natural bridge between science and mathematics; mathematics students get a welcome context in which to practice their skills, and science students get new and powerful tools to help them understand content. Yet data analysis is complicated; fortunately, new advances in technology can help students use mathematical tools they might otherwise be unable to use. We can also ask why more mathematics has not been integrated into science already. The stereotypical answer—which is largely true—is that the students "can't do the math". While we can glibly say that technology will help them, a more interesting question to ask is, "just what math is it that we want students to use in the science classroom?"

The curriculum development we have done on this project to date has, as of December 2001, consisted of developing prototype laboratory activities and doing extensive field tests. If we look at the math in these labs, we easily see *measurement* when students get their raw data. But to analyze the data (which occurs partly before measurement) students need to use many mathematical ideas; chief among these seem to be *functions* and *variation*. Interestingly, we can also look at function and variation, as we use them in data analysis, as sophisticated tools for measurement.

So let us talk about measurement. Of all the areas covered by standards documents from the United States (e.g., NCTM 2000), measurement is the weakest. In the early grades, it focuses on skills such as learning to use a ruler; later on, it extends itself only as far as units analysis and formulas for volume. Measurement error gets a brief nod; the main idea is not to carry too many decimal places. In the data analysis and probability strand, there is some discussion of residuals, but we do not discuss their causes. And while students are to "understand the distinction between a statistic and a parameter" (p. 325) the idea that you can find a parameter based on data is limited to finding a proportion based on a poll. In general, science standards documents are not much better. With the few notable exceptions (for example, the *Benchmarks* documents: AAAS 1990, 1993, 2001), they are weighed down by the massive amounts of content they seem obligated to cover (even NRC 1995, while a good document, suffers from this content-heaviness).

So let us first take a broad view of the mathematics we do when we do science. Then we will look at a sample curriculum activity field-tested in a high school, and discuss the mathematics in it. This activity highlights functions and variation, and we will be able to see how closely these are related to measurement in this context. Finally, we will comment on the importance of measurement to our subsequent curriculum development, to data analysis, and to the study of the nature of science.

THE MATH WE DO WHEN WE TALK ABOUT SCIENCE

Here is a list from Erickson (2001)—roughly ordered by complexity — of ways scientists use mathematics. I do not claim that it is comprehensive, but it is a useful starting point for discussion:

1. We use mathematics—or at least numbers and categories—to record data.

2. We use mathematical ideas, e.g., the concept of a variable, to organize our thinking and plan just what data we will collect.
3. We use mathematics (some would say computer science) to organize, clean, and transform our data, to get it ready to analyze. We may have to write a formula to tell when an event happens; or we might need to use relational formulas to look up data in another table; or we might have to clean up or transform our data—for example, to look at a change rather than an absolute amount. We often forget that this "data preparation" is an important, often time-consuming part of data analysis (Gould 2001).
4. We use mathematics to make indirect measurements. This goes beyond trigonometry for inaccessible distances. For example, knowing a symbolic relationship among variables, we first solve for the one we want to determine; then we measure the others and calculate.
5. We use mathematical functions to model phenomena. To be sure, we can calculate using formulas, e.g., to find the force of gravity on the surface of the moon. But real understanding comes when we think of the phenomenon functionally, e.g., to see how that force decreases as we get farther away. Recursively-defined functions and many types of simulations can be models as well.
6. We use mathematics to describe departures from a model. Models typically oversimplify; often we draw the most interesting conclusions from the way data depart.
7. We use mathematics to cope with variability in data. This goes beyond insisting that every measurement have a "plus-or-minus." For example, we might calculate standard errors, or repeatedly simulate an inexact measurement process.
8. We use mathematics to decide if we can reject a hypothesis based on quantitative results. Often this requires only an informal comparison of data to model, but sometimes we need statistical inference in order to assess whether an observed effect could be due to chance.
9. Mathematics—logic and statistics—also helps implement our experimental designs (e.g., to control variables both informally and using multiple regression or analysis of variance).
10. We use multiple mathematical representations—graphs, of course, but also tables, charts, and formulas—to help us understand the story the data tell and to communicate with others. The more representations we have at our disposal, the more flexible we are at exploratory data analysis and communication.

AN EXAMPLE ACTIVITY: GALILEO'S RAMPS

The Galileo's Ramps activity is a common laboratory in USA high-school Physics classes, often in year 11 or 12. Students roll ball bearings down metal tracks, using stopwatches to time how long it takes the ball to roll different distances. Measuring the time introduces the most inaccuracy. Without computer technology, teachers instruct students to take many measurements at each distance (the class we visited had traditionally done nine neasurements at each of six distances). Then students average the measurements to get a single timing for each distance, which they then plot by hand in a scatter plot of distance versus time. The point is for students: first to realize that the points fall on a parabola of the form $d = \frac{1}{2} at^2$ (a is an acceleration); second, to figure out the acceleration. An extension asks them to see how the acceleration depends on the angle of the ramp.

Problems with this arrangement included:

*   Averaging the points and plotting them by hand was extremely laborious.
*   The points looked at first glance as if they might easily be fit by a line.
*   It was not obvious that students should try a parabola, even once they decided the data were not linear.
*   It was not obvious how students were to determine a best fit parabola, even once they decided the data were quadratic.

The first change we made to the activity was to eliminate the averaging. Since students could make graphs easily, we wondered what we could do if they plotted all of the data instead of

just the means. Figure 1 shows sample student data; the students used the software package *Fathom* (KCP Technologies, 2000). We also show what the "averaged" graph would look like.
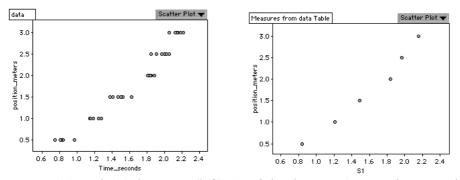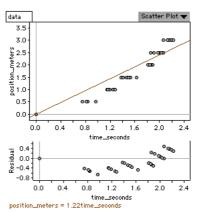

Figure 1. Sample Student Data (left). At Right, the Data Averaged over Each Position.

With just this graph, we can ask interesting questions, such as why the clusters of points are horizontal. This helps less-experienced students firm up their understanding of what the graph is telling them. But it also brings up important questions such as whether time should be on the vertical axis. Some groups decided to do the lab that way; with Fathom it is a matter of a moment to switch axes, so we could leave it up to them. Now we asked the students to try to fit a line to the data. Fathom has a *Movable Line* feature; students place a line on the graph. As they move the line with the mouse, the equation for the line updates continuously. They can even make a residual plot from that line, which also updates.

While this line is plausible at fitting the data on the graph, a student always mentions that it really should go though the origin, as it should take no time to go no distance. (And if a student had not pointed it out, we could have easily asked about it.) So we can either ask Fathom to *Lock Intercept at Zero* or add another point at the origin—which appears immediately on our graph. Figure 2 shows both.
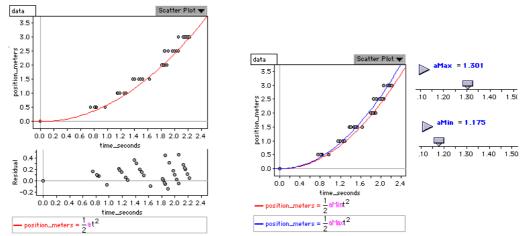

$$position\_meters = 1.22 time\_seconds$$

Figure 2. The same data with a movable line passing though (0, 0) and a residual plot. As you move the line, the equation changes at the bottom of the graph, and the residual plot updates.

Here we pointed out that the "bowed" pattern in the residuals shows that a line is not the best model for our data. This is not obvious to all students, but playing with the line, and seeing the residuals update, helps them understand the residual plot and what it means. So students tried curves. Even though they had just been studying uniform acceleration, students tried a wide variety of curves: parabolas, exponentials, even the tangent function. It is easy to plot a curve in Fathom if you can express it symbolically, and easy to get the residuals from that model. But what values should you use for the parameters the curve requires?

Fathom does not provide "movable" versions of all elementary functions as it does for the line. Neither does it automatically make least-squares fits to curves. While movable parabolas and exponentials might be a good idea for a future version, we think the lack of nonlinear least-squares fits is an advantage for learning, helping students avoid what we often see elsewhere: a

scattershot approach where they try every possible function and choose the one with the "best" fit, as measured by r^2 or its equivalent, irrespective of visible patterns in the data or (especially) any reason to expect that functional form. Instead, in a statistics class, we might well ask students to transform the data and make a linear fit. Or, as we do here, to use another of the dynamic features of the software: sliders.

A slider is a variable parameter that, when you slide it, changes a global value. This can be many things; here, we use it for a coefficient in a function. We asked students to find the range of values of the acceleration that they thought could be correct given the data. Figure 3 shows, at right, a graph with two plausible curves—and the sliders that determine these accelerations. Asking students for a *range* of parameters helps them appreciate that the variation in the data implies an uncertainty in the parameter they derive—and even gives them a clear visual representation of that uncertainty.

Figure 3. At left, the data with a parabola and its residuals; at right, the data with two parabolas plotted corresponding to the maximum and minimum "plausible" acceleration.

Note that we can also discuss what makes a plausible fit. Here, the class felt that the curves should go though, or at least be close to, all six clusters of points. Note the cluster at a distance of 2 meters—the tightest of the clusters. Since the points are so close together, they force the parameters close as well (this is a good statistical lesson should they address it later). But notice: if they had just used the means, that point would have appeared as an outlier, and they probably would have ignored it. That is, in this case, what we would see as the best, most influential data would have been thrown away.

Let us return to the issue of the form of the function. One group, for example, persisted in using the tangent function because it fit the data so well; at least as well as the parabolas when we looked at residual plots. But we could then display the data and the function on the class projector and discuss whether it was reasonable to use the tangent. Students had many cogent objections to the tangent, besides the fact that they finally remembered that uniform acceleration was quadratic. Two of the most interesting were, first, that the tangent implied what looked like an infinite speed when you reached the asymptote; and second, that the tangent required more parameters—and if you could do just as well with that one-parameter parabola, why not use it? That is, the student had come up with Occam's Razor independently. Then we could ask the critical question: what could you do to tell whether the tangent or the parabola was a better model? They all knew: take more data. Then we could ask: take more data, *where*? that they knew as well: farther out, where the tangent group's model blew up.

ABOUT THE MATHEMATICS

As suggested at the outset, this approach to data analysis lets students use more of their math-class knowledge of functions than the old, plot-one-parabola-by hand version; the ease of plotting any function, and the dynamic sliders gives students a rich experience with different elementary functions and their parameters. Then, plotting all the data gives students more

experience with and intuitive sense of variation. Asking for a plausible range of parameters connects the ideas of functions and variation.

What does this have to do with measurement? First, we can ask students directly which variable contributed more to measurement error; and they can see that immediately in the graph. More important, though, is what we're measuring. We're measuring time and distance directly. But indirectly, we're measuring *acceleration*, and that's really the point of the whole exercise. To do that, we could measure a single distance/time point and plug the results into a formula, but to really understand whether we're doing the right thing, we need to understand the *function* that relates these three quantities. And we need to assess the whole function—not just a point—by comparing that model to data (for example, with the residual plots). Finally, we can assess the uncertainty in the acceleration by looking at how variation in the data affects our estimate of that parameter.

This is deep stuff, and, like Occam's razor, points us into the nature of science—a topic often given short shrift in USA secondary science programs, because of the content we feel we need to "cover." It may be that with activities a little more open-ended than this one, students can take even more responsibility for deciding what to measure and how to measure it well; that is, they can come closer to becoming experimentalists. In one activity, described in detail in Erickson (2001), students receive a (fictitious) scientific "paper" expounding a theory that they then support or refute through an experiment of their own design.

CONCLUSIONS

Thus data analysis is at the heart of what makes science work, but students can't even get to it if data analysis is too cumbersome. Instead, we suggest that we can use statistical/data analysis packages to make it much more accessible and interesting—and consequently motivating. In our field test, students used more math, and took less time, largely because the technology took care of graphing and the laborious calculations that graphing at its most useful (e.g., plotting functions over data, making residual plots) requires.

Another, subtler benefit of technology was that it helps students escape what we might call the "tyranny of the center": by plotting *all* the data, and determining ranges of parameter values instead of a single best fit, students have a better chance to appreciate the importance of variation—both in the measurements themselves and in what that variation implies for results. Eventually these students may use more formal statistical baggage, for example, to determine a confidence interval for the acceleration. But at this stage, that's neither necessary nor particularly desired.

ACKNOWLEDGEMENT

REFERENCES

American Association for the Advancement of Science, Project 2061 (AAAS) (1990). *Science for All Americans*. New York: Oxford University Press.

AAAS. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

AAAS. (2001). *Atlas of science literacy*. New York: Oxford University Press.

Erickson, T. (June 2001). *Final Report for SBIR Phase I Project*. Submitted to the US National Science Foundation.

Gould, R. (2001). *Introduction to data analysis*. Presentation at the Tinkerplots annual meeting, Madison, WI.

KCP Technologies (2000). *Fathom™ dynamic statistics software*. Emeryville, CA: Key Curriculum Press.

NCTM (2000). *Principles and standards of school mathematics*. Reston, VA: NCTM.

National Committee on Science Education Standards and Assessment, National Research Council (NRC) (1995). *National science education standards*. Washington, CD: National Academy Press.