Missed opportunities in research on the teaching and learning of data and chance

J. Michael Shaughnessy
Department of Mathematical Sciences
Portland State University
Portland, OR USA

This paper and the accompanying keynote talk for the 20th annual MERGA conference are dedicated to the memory of my friend and colleague Georg Schrage, August 9, 1940 - February 11, 1997.

Reflecting on a body of research work can sometimes lead to the recognition of areas of opportunity for research that have gone largely unnoticed. In this paper we consider three such opportunities in the area of research on the teaching and learning of probability and statistics: i) Following up on students' initial thinking to watch for future transitions; ii) Investigating students' thinking on variability; and iii) Posing research questions that begin with what students *can* do rather than pointing out what they *cannot* do. Situations from research tasks, past and future, are used as starting points the discussion.

Introduction

It is a great honor that you have invited me to give this address at the 20th MERGA conference. I hope very much that my thoughts and words today might plant some seeds for future research in the area of data and chance, seeds that hopefully will enable us to improve the teaching and learning of probability and statistics at various levels of education.

The area of probability and statistics within mathematics education, or stochastics as some of our European colleagues refer to it, has fascinated me ever since I first had the opportunity to teach probability to beginning college students. There are two things that have fascinated me most in my twenty odd year journey with students in stochastic reasoning. One is finding out what students think or believe about chance and data. How do they reason in situations that involve chance? The second involves documenting student growth or change as they are interacting with probability or data tasks. In reflecting on the current state of affairs of our research efforts in probability and statistics, I find that we have made a good deal of progress in the first arena--uncovering students' conceptions and beliefs about chance and data. However, we have made very little progress in the second arena, the documentation of student growth and change as they interact with chance and data tasks or curriculum materials. I believe this is a missed opportunity in our research.

In preparing for this talk, and reflecting on some of the recent literature in research on probability and statistics, there were several other themes that emerged as opportunities that we researchers in the area of chance and data may be missing . These themes became nagging ones and kept returning to me as I pondered the current research situation in our field. In particular, three themes strike me as missed, golden opportunities, and so I would like to bring them to your attention. One we have already mentioned above i) following student's thinking as they interact with materials. I was also struck by ii) the great absence of research on students' thinking about variability--where is it anyway? and, iii) the possibility of starting to do research based on what students *can* do rather than on documenting what they *cannot* do. And so, as fate would have it, I have chosen to try to plant some seeds today in this ground: a closer look at several of these "missed opportunities."

Some missed opportunities

For me it is helpful to think about these missed opportunities by starting from the context of several tasks, or situations, that have been used in research in probability and statistics.

Situation 1--Following up on Students' Thinking

In teaching probability and statistics we must not only deal with helping students to develop an understanding of and to apply some difficult mathematical and statistical ideas, but we must also deal with the psychological issues involving chance and data that can be deeply rooted in students' experiences or in their beliefs about chance phenomena. Sometimes the mathematical issues and the psychological issues are quite confounded, which makes our job all the more difficult. For example, consider the situation of drawing a sample of people from a large city in the U.S. to give their opinions on whether motorcyclists should be required to wear helmets for protective head gear. If we do a proper job of drawing our sample so that it is "representative" of the community, so that it doesn't have obvious bias, and so that it is as random as possible, we'd expect the opinion results from our sample of people to mirror the whole population, and thus to give us some realistic notion of the community's pulse on the wearing of helmets. This concept of a representative sample is a very powerful, and necessary one in the area of statistical decision making. On the other hand, consider the following version of an oft used task in research on probability and statistics.

> A fair coin is flipped 5 times in succession. Which do you feel is more likely to occur for the five flips?
>
> A) HTTHT B) HHHHH C) they have the same chance of happening.

Here are some responses to this task which came from a recent convenience sample of middle school mathematics teachers before they had a class in probability and statistics.

"I would go with A) only because it more closely approximates the ratio of 50-50, but in such a small *sample* anything is possible."

"I would say both are equally likely on any particular instance, although the long term results would gravitate to a result more like A).

" To get either sequence you'd have to multiply the probability of each being a Head or Tail (1/2) times each (of the) other toss(es), so (1/2)^5 for both."

"...(After circling C--then circles A) I changed my mind again. The chances are 50/50 no matter what, and sequence A is more likely."

" A) because I think it would be more likely to have a series of two of the same than to have five of the same."

"B) because A) is too *specific."*

"C). It is a 50-50 chance for either an H or a T, so either sequence."

The responses from these teachers are very similar to the types of responses that have been found among middle school, secondary, and college students, and which have been well documented (Fischbein, 1996; Konold, 1983; Konold et. al., 1993; Shaughnessy, 1977, 1981). They exhibit a great variety of beliefs, conceptions, and interpretations of the problem. The notion of a "representative" sample that is so helpful in the helmet survey can cause problems when applied in this context. There is no "sample" in the above question, there are just several outcomes from five flips. And yet some of my subjects felt that the "representative" notion should apply here, too, when they said "A) more closely approximates the 50-50 ratio" or "(we are)...more likely to have a series of two of the same, than to have five of the same."

Several of the responders even tried to superimpose the idea of sample on the problem. "In a small *sample* anything is possible" and "the *long term* results gravitate toward A)," as if there were an ongoing sample. Still other responses indicate what Cliff Konold has described as the *outcome approach* (Konold, 1989; Serrano, 1996). "I would say that both are equally likely in a particular instance..." and "it is a 50-50 chance for and H or a T so could be either (outcome)." Students may be just trying to predict what the next sequence would be if we tossed five coins, rather than considering these two sequences as only two among many possibilities. In fact, the question is framed in a way that might actually lead a subject to focus on single outcomes, rather than a range of possible outcomes. One of the teachers did use some knowledge about probability and did some calculations that indicate recognition of a whole range of outcomes, but was an exception.

So, here we have an example of a well documented misconception about the likelihood that certain binomial sequences will occur in a probability experiment. Now what? What are our next steps? It seems in research on probability and statistics, we have developed a tradition of putting tasks to students and documenting the types of answers and responses they give, classifying various beliefs or conceptions about chance, randomness. However, I think that all too often we stop at that point, describe what we have found, and lament the state of affairs and the difficulty students have with chance concepts. For us as mathematics educators, the variety of responses to the question above can be a good starting point for research, not just an ending point in itself. Here we have a golden opportunity to place students into an environment to investigate this type of problem over time, and to document any growth and changes in their thinking. It seems that we don't often use the fruits of our research efforts to spin off further investigations. In this we may be missing some golden opportunities to ask questions that can show what our students are capable of, rather than simply documenting what they are incapable of.

Suppose we took the point of view that our students' conceptions are *"transitional conceptions,* rather than *mis*-conceptions, that their thinking is always under construction. Where could we go from here with our questioning, what types of tasks might we pose for students, to help them begin transitioning in their thinking about binomial situations? This may be a slightly different way for some of us to think about research. The nature of evidence in our research on learning need not be static, it can be dynamic. We have an opportunity to do research on the *process* of learning.

Here is an example (Figure 1) of a visual environment for binomial experiments that I have used as a point of departure for investigating student's growth in thinking. It is an activity taken from the *Math and the Mind's Eye* series (Shaughnessy & Arcidiacono, 1993). The *Math and the Mind's Eye* series is the parent project that has led to the development of a middle school mathematics curriculum, based on visual thinking, that is currently underway in Portland, Oregon (Foreman and Bennett, 1995, 1996).

In this checkerboard environment, we place a marker on the start square. As we flip a coin, an H indicates a diagonal move on the checkerboard one square up and left, and a T a move up and right, so that we stay on the shaded squares. I generally ask students to play this game several times, flipping a coin five times and moving the marker so that they can see the path of the marker up the board as they record the string of five H's and T's. In this way, sequences of tosses can be visualized on the checkerboard as "paths" up the board. For example, in Figure 2 the path HTTHT is drawn in on the board, as is the path HHHHH.

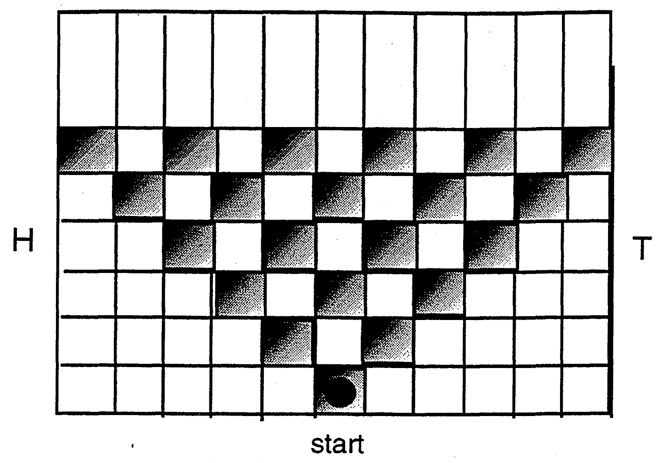


Figure 1

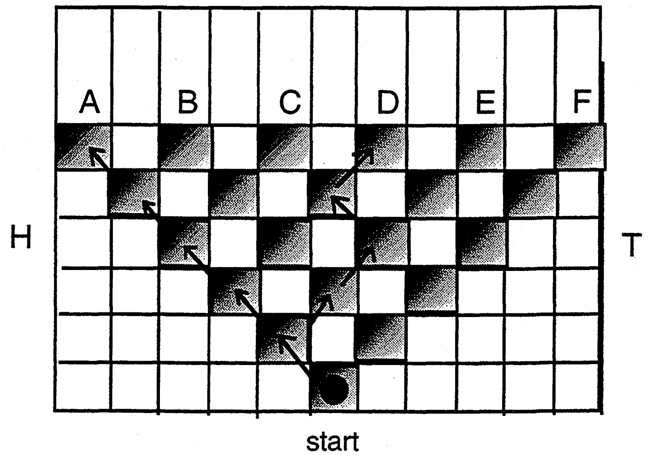


Figure 2

Here are examples of questions that I have used to investigate the development of transitional conceptions in students' thinking about binomial experiments.

1) How many paths can you find that correspond to HHHHH? How many correspond to HTTHT?

2) Altogether, how many different paths are possible on the checkerboard for five tosses?

3) How many different paths lead to each of the possible six finishing areas A - F?

As students work through this visual environment, some things inevitably come up as they discuss and analyze the game. Students' have:

a) At first predicted equal #'s for the finishing places , then, changed their mind because they see that there are more paths that lead to some finishing places than to others as they are playing it.

b) Made connections between the paths and the sequences of H's and T's. Students can draw in all possible paths (there are 32 of them) which corresponds to listing all the possible sequences of 5 H's and T's.

c) Built their own theoretical model for the chance of a given outcome, like HTTHT, by counting paths. In doing this, they often mention that their model does depend on each path having an equal chance of occurring, but this seems reasonable to assume when using a fair coin. The branching odds are 50-50 at each juncture.

d) Made connections between the number of paths and the numbers in Pascal's triangle (Figure 3).

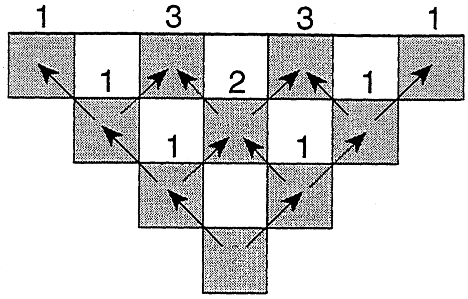


Figure 3

This particular visual model is quite powerful as a teaching tool as well as for use in a research environment, because it can be adopted for unequally likely branching, and for any path length. In other words, the two parameters in the binomial distribution, the number of trials "n", and the probability of success on a given trial "p", can be represented by the path length and the branching probabilities respectively. Students can *see* how HTTHT and HHHHH fit into the entire spectrum of possible outcomes. As researchers we can observe and document their thinking in transition. Growth and changes over time in students' thinking can also be documented using an approach like that of Watson, Collis & Moritz (1995). They investigated students' responses over grades 3, 5, 7, & 9 on various chance and data tasks and noted differences across ages. However, there are also opportunities for us to trace individual students or groups of students over time as they interact with problems and tasks. We do take advantage of these types of opportunities very often in our research, although recently some case studies of younger students engaged in probability tasks has been conducted by a group at Illinois State University (Jones et. al., 1997).

Situation 2--Where is the research on variability?

The two big concepts in the teaching and learning of data are central tendency and variability. Although there have been investigations into students' concepts and beliefs about "averages"(Russell and Mokros, 1996; Pollatsek et. al. 1981), there does not seem to be a similar tradition of research into students' ideas about variability or spread. One reason for the lack of research on variability, I believe, is that our research often mirrors the emphasis in our curricular materials, at least this seems to be the case in the area of chance and data. While there are a variety of models for centers (means, medians, modes) that have been developed and tried with students, there is not a corresponding rich array of models for spread or variability. For example, in research on averages, Russell and Mokros (1996) mention three different models of the mean that they have used to investigate elementary students' thinking about the mean--a fair divisions model, a center of balance model, and a dynamic distribution model in which students work backward from the mean to recreate various possibilities for the original distribution of the data. Russell and Mokros point out that each of these models has some limitations when used to introduce the concept of mean to children. Still another model that we have used for the mean (Figures 4 and 5) involves forming a stack of cubes to represent each data point, and then leveling off the stacks by making trades to find the mean (Foreman and Bennett, 1995).

This model works equally well when asking students to find means from distributions, or when asking them to reconstruct distributions from means--the converse problem that Russell and Mokros investigated. The cube columns can be represented on centimeter grid paper, and cubes can be "sliced" to represent trades that involve rational numbers. We are rich in models for centers in our curriculum materials and they have carried over into our research tasks, but I find us lacking in similar creative models or approaches to concepts of variability and spread.

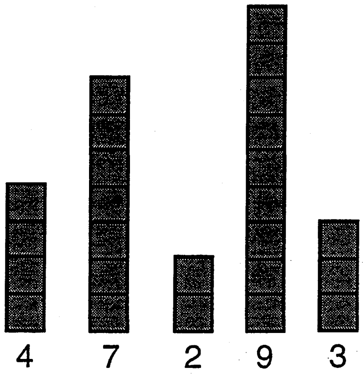


Figure 4

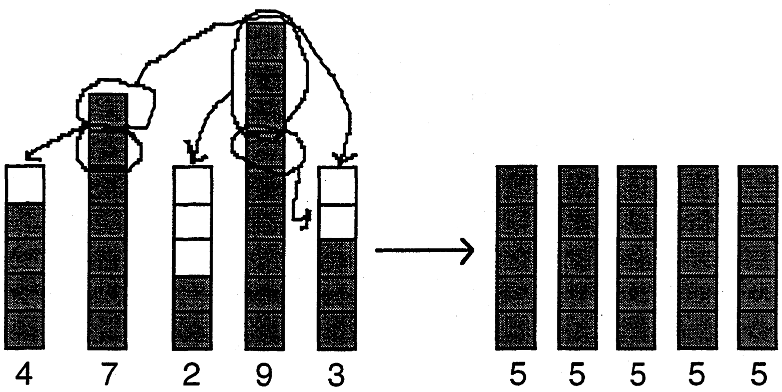


Figure 5

Batanero et. al. (1994) make this same point about the lack of research on variability, and they mention an article by Loosen et. al. (1985) which suggests that textbooks in statistics seem to put more emphasis on looking at heterogeneity in data (centers) rather than on variability in data (spreads). (Note: there is some preliminary work being undertaken by Watson & Moritz (1997) in which they solicit students' thoughts about variability by asking them to compare and analyze various hypothetical distributions of student test scores).

I have my own suspicions as to why variability has not received the attention that measures of central tendency have in our research. One reason is that statisticians have traditionally been very enamored with the standard deviation as *the* measure of spread or variability, and teachers and curriculum developers often avoided dealing with spread because they felt they couldn't do so without introducing standard deviation. Standard deviation is not only computationally messy, but difficult to motivate why it is a good choice for measuring spread, especially with beginning students. Another reason for lack of attention to variability is that centers or averages are often used to predict what *will* happen in the future, or to compare two different groups-- not always used correctly in this regard of course but nevertheless used. The incorporation of spreads or variation into these predicting or comparing processes only confounds people's ability to make clean predictions or comparisons. You see, where there is variability or spread, there must also be some waffling in our predictions or comparisons, and predictors do not like to waffle (unless of course they are politicians). Furthermore, this whole concept of variability is outside of many people's comfort zone, and may even outside their zone of belief. Jim Landwehr (1989) nicely summarized people's over reliance on means and their beliefs about variability when he noted that people

* believe that any difference in means between two groups is significant

* have unwarranted confidence in small samples

* have insufficient respect for small differences in large samples

* mistakenly believe there is no variability in the 'real world.'

Let us again look at an example, this time we consider some research evidence about peoples' intuitive conceptions of variability. Here is a version of another task that has often been used. This particular version was reported by Schrage (1983).

What would you think is more likely to occur?

A) 7 of 10 babies born are males?

B) 70 of 100 babies born are males?

C) These two have the same chance of occurring?

Here are some typical responses to this task, taken once again from the sample of middle school mathematics teachers:

C) Because the chance of getting boys or girls is the same in both situations, so both choice A) and choice B) have the same chance of occurring

C) The ratio of 7 out of 10 is the same as 70 out of 100, so it makes no difference, they are the same.

B) With 100 babies you will have a lot more chances of having males.

A) With less babies it seems more likely to be further away from 50-50.

A) The more babies you have, the closer you should come to a 50-50 split.

As in our first example we can find a variety of beliefs and interpretations of this problem. Some of the interpretations (1 and 2) focus on predicting from averages or ratios--using a 50:50 ratio of boys to girls or from the sample proportion which is .7 for both choices A and B. On the other hand, several of the responses indicate that there is an awareness that smaller samples can deviate from expectations, or conversely, that larger ones should be closer to expectations (4 & 5). Schrage (1983) reports that although 60% of his sample of 153 pre service mathematics teachers chose response A, his results are misleading because in his follow up questions most of them gave incorrect reasons for this correct response, indicating that they were totally ignoring the effect of sample size in the problem.

One difficulty I see with this item, and ones similar to it, is that it actually focuses the attention of the responders on centers, while dealing with the concept of spread. The problem is (implicitly at least) asking for the likelihood of a particular outcome (7 out of 10 boys vs. 70 out of 100 boys) rather than asking what the likely range of outcomes for having 10 children, or 100 children might be. If we want to know what our students *can* do with variability, I think we have to ask a different type of question here. I think we need to pose questions on variability that can be answered in a sampling context in which the concept of an "interval of likely values" arises, rather than in a context that forces people to compare the point value probabilities of two particular events. Here is an example of a task on variability that I have used with students.

Imagine you have a huge jar of M&M's with many different colors in it. We know that the manufacturer of M&M's puts in 40% browns. If you reached in and pulled samples of 20 M&M's at a time, what do you think would be the likely range for the number of browns you found in your samples?

Responses I have received to this question vary from 7-9 browns up to 5-12 browns. But no one yet has said "you will get 8 browns every time." So, the idea of a range of likely values in a sampling situation where the population proportion is known (40% in this case) is very accessible to students, what Fischbein might call a primary intuition (Fischbein, 1987). This can be a point of departure for further research on variability. What will students say if we change this question to samples of 100 M&M's? Then what will they say is the likely range for the browns? Nowadays it is very easy to actually "draw samples of 100 M&M's" by introducing a simulation of the M&M's experiment using some sort of sample-resample software. I find the little software package Prob-Sim (Konold & Miller, 1994) very useful for generating samples of this kind. The results from using Prob-Sim to draw 100 samples of 10 "babies" and 100 samples of 100 babies are presented in Figures 6a and 6b.

Key: Column totals indicate the number of times that n (horizontal axis) boys in 10 babies occurred.

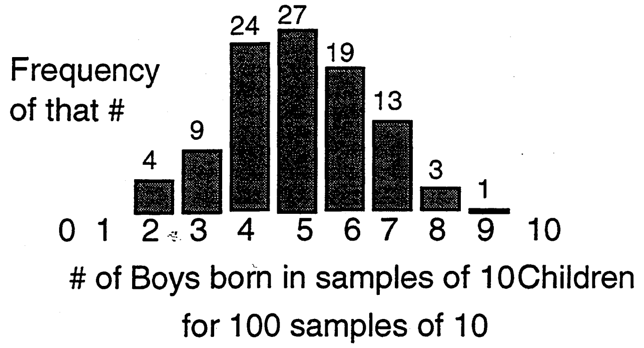


Figure 6a

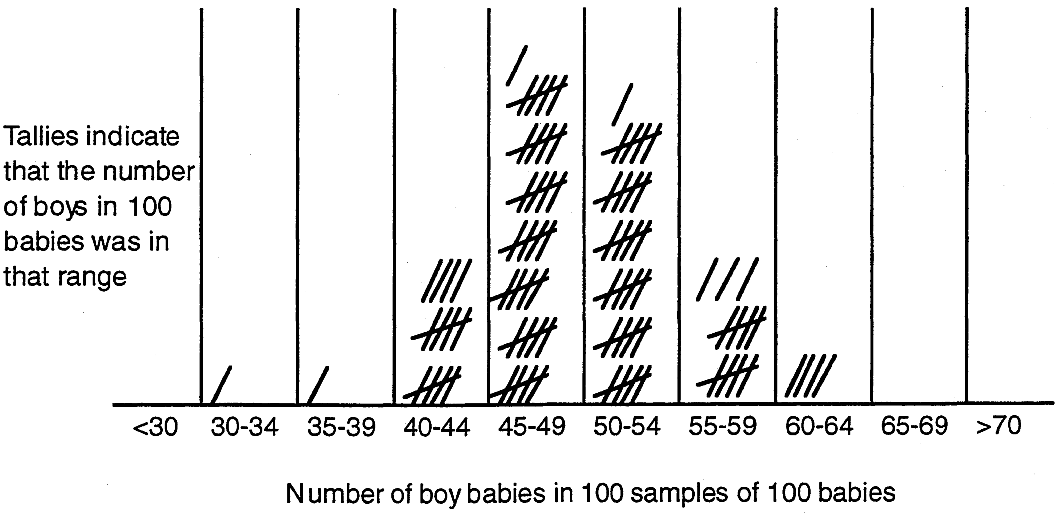


Figure 6b

Notice that there were 13 instances of 7 males born in 10 babies, however the outcome 70 (or even more) males in 100 babies did not ever occur in any of our samples of 100 babies.

Based on such resampling results students can begin to gain intuition on what the likely spread of outcomes is for a sample from a population. They can begin to gain some sense of what is likely and what is unlikely to occur by considering the entire distribution of outcomes. They can begin to develop some sense of the affect of sample size on the spread of likely outcomes. Neither a formal quantification of "spread" nor formal definitions for variability need to be used in order for us to begin to investigate students' conceptions of variability. Furthermore, if we begin our research investigations on variability based on something that students *can* do (like estimating confidence intervals for the M&M's problem), rather than something they *cannot* do (like comparing binomial probabilities in the baby problem) we may find that research on variability is an untapped well in research on data and chance. (By the way, the baby problem and the M&M problem can also be modeled in the visual checkerboard environment, since they are binomial problems. However, building a 20 X 20 or a 100 x 100 checkerboard is a little inconvenient. On the other hand, if students have previously worked in the checkerboard environment, they can visualize it in this situation).

I'd like to follow up more on the notion of starting our investigations into students' thinking on data and chance by looking at what our students *can* do rather than concentrating on what they stumble on.

Situation 3--Unveiling what our students **can** *do*

At the moment there seems to be bit of a boom in research into the teaching and learning of probability and statistics. At least one gets that impression when browsing the proceedings of a number of recent international meetings on research in mathematics and statistics education (Garfield, 1994; Puig and Gutiérrez, 1996; Phillips, 1996; Batanero, 1996). The proceedings of these meetings depict an ever growing body of evidence that deals with students conceptions and beliefs about data and chance. (In fact, there is so much research reported now in the Newsletter for the International Group for Research in Probability and Statistics— begun by Joan Garfield at the University of Minnesota and currently edited by Carmen Batanero at The University of Granada— that it is becoming very difficult to keep abreast of research developments around the world).

Although most of this research is being conducted by mathematics and statistics educators, a good deal of the theoretical underpinnings for the research into peoples' conceptions of data and chance is based on a methodology from cognitive psychology, particularly from the area of the psychology of decision making (Kahneman & Tversky 1972, 1973; Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980). There have been several past reviews of the literature which point to a large body of evidence in the area of students' conceptions and beliefs about probability and statistics, and to the psychological roots of this particular line of research (Hawkins & Kapadia, 1984; Garfield & Ahlgren, 1988; Shaughnessy, 1992; Borovcnik & Peard, 1996; Shaughnessy, Garfield, & Greer, 1996). Recently a group of psychologists at the University of Chicago has been challenging some of the claims of the earlier psychological research, and finding better "success" with students when questions are posed in terms of frequencies (Gigerenzer, 1994, 1996; Gigerenzer & Hoffrage, 1995). However, the basic methodology has remained the same: pose cognitive tasks to students, record successes and failures, but do little, if any, follow-up on their thinking, and rarely any following of the development of students' thinking over an extended period of time.

In adopting the research tasks and methodology of our brethren in psychology, we have been able to document that our students have certain beliefs and conceptions about probability and statistics that can lead them to incorrect decisions or erroneous analyses of problems. We have been able to characterize some of these erroneous conceptions. We have been able to do an excellent job of documenting what our students *cannot* do in the areas of chance and data. We have not done a very good job documenting what our students *can* do. We, like our compatriots in psychology, tend to ask research questions that will almost certainly expose the pitfalls in our students' thinking. We do not often put students in situations that will give us an opportunity to document potential growth or possible change in their thinking.

I suggest we adopt a somewhat different methodology, that we put students into open ended situations, that we ask them to try and make sense out of data, to generate their own questions about data, to design their own graphical representations of data, to in fact even create their own "measures" on data. This is a somewhat different paradigm for research on the teaching and learning of data and chance than we have taken in the past, and may even have some aspects similar to the Soviet teaching experiments.

I would like to share a couple of tasks that I have used with some middle school students in the past several years which give them an opportunity to explore data. My job is to watch what they do, and to listen, and occasionally to ask a question. The group of middle school students I worked with is a combined grade 6-7 class in a school located on the outskirts of Portland, Oregon. The class is an experimental class that has been working through the *Visual Mathematics* curriculum as it is being developed (Foreman & Bennett, 1995, 1996).

The students are each given the entire weather page for one day taken from the local newspaper. (In many cities in the U.S., there is a special multicolored page for the weather, complete with many types data sets, graphs and charts, and containing both a national and a regional weather map). The students actually have a set of about 25-30 consecutive days of this weather map among them, with each student having one day's weather summary. The maps are passed out so that students have consecutive days in their small groups of 4-5 students. The task begins with two writing stems, "We notice that...." and "We wonder about....", and asks students to make lists under each stem. After a while, the groups are asked to share and discuss what they have written in their lists. Here are some of the responses I obtained from this group of students.

We notice that:
* the weather page tells weather from all over the world
* a hurricane died down on the day we investigated
* pictures and maps explains things better (than words)
* they predicted rain 3/5 of the time on the 5 day forecast on our groups pages
* the weather pages show information from the past for Portland

We wonder:
*why Medford has the highest air pollution
*what the zeros indicate in the white column
* if the east coast always has more severe weather
* where the maps of Alaska and Hawaii are
* how it can be that the high temperature is in the morning

Of course, not all the "notices" or "wonders" that the students come up with are worthy of an effort at exploration. But several of the "wonders" that the students came up with proved quite intriguing and were investigated by groups of students in great detail. Two of these were subsequently written into the materials in the program as possibilities for teachers to use (Foreman & Bennett, 1996, p. 265).

1) Is the coldest time of day more likely to occur before or after sunrise?

2) How well do the weather folk do at predicting Hi and Lo temperatures five days in advance?

To collect data on the first of these questions, students needed to extract information for the time of day for the lowest temperature from the time-temp graphs given for each day on the weather pages (Figure 7). They also needed to check on the time of sunrise, which they decided was around 7:00 A.M. for the days (in October) they were investigating.

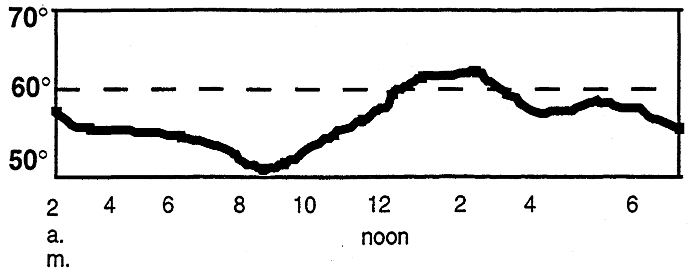


Figure 7

Some of the students represented the information for the time of day for the low temperature in a type of line plot (Figure 8).

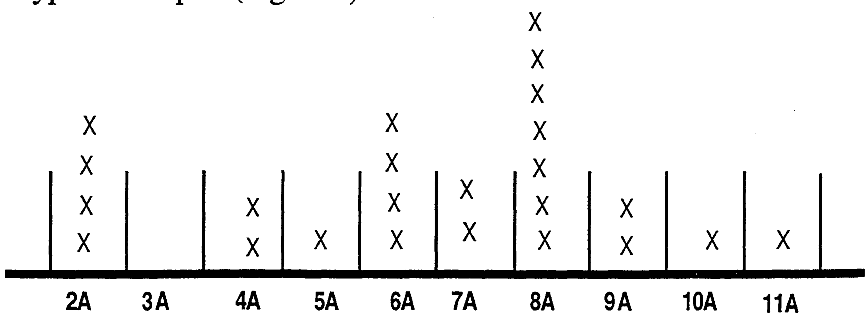


Figure 8. Time of day for the low temperature

These students were surprised that the time of the low temperature for the day occurred *after* sunrise nearly as frequently as it did *before* sunrise. Quite a discussion ensued as to why there was such variability for the coldest time of day.

The second of these questions above led students to gathering and comparing the real and the predicted temperature data. On each day, predictions are given in the weather pages for the next five consecutive days in graphics that include predicted Hi's and Lo's, and predicted weather conditions (e.g. rain, clouds, sun, etc.). The students in this sample are quite visually oriented, since the mathematics curriculum they have been working on emphasizes visual models and visual representations of ideas. So, it was no surprise that they represented the information for this problem in visual displays. Here are some examples of how students represented the information on predicted temperatures vs. actual temperatures (Figures 9, 10).

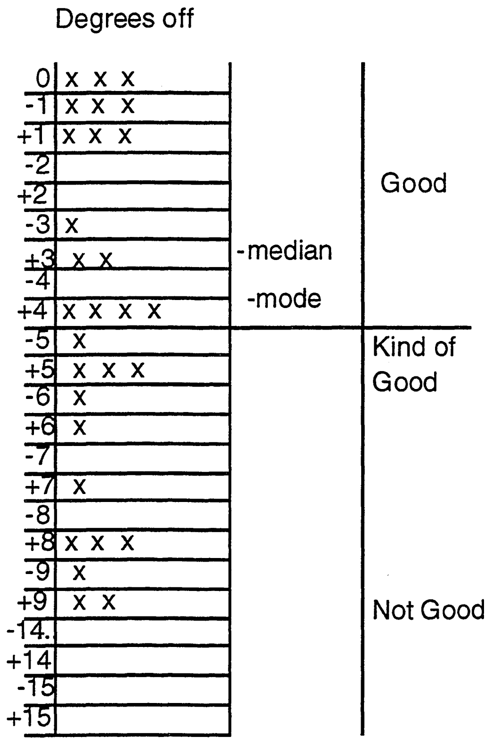
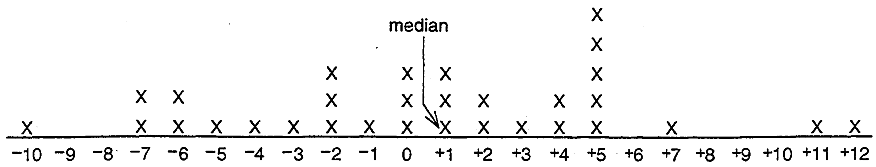


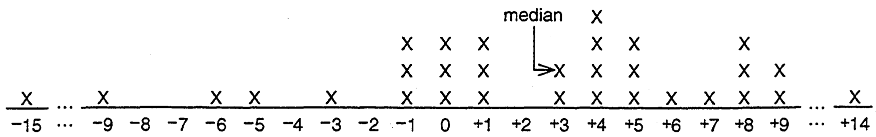
Figure 9

Degrees Off—Lows



They predicted too low 12/31 of the time and too high 16/31 of the time. They were right on 3/31 of the time. 23/31 of the time they were off by 5° or less.

Degrees Off—Highs



More than ½ the time they were high by 3 or more degrees. Most of the time they predicted high. The data clusters between −1 and +5 off. We think 14 and 15 degrees off are outliers. 20/31 of the time they were off by 5° or less.

Figure 10

After the students had generated their own representations of the data they were shown how to represent data in a scatter plot. Although the students had some experience plotting points on co-ordinate graphs, this was their first experience representing data in a scatter plot. While they were making this scatter plot, several student-generated graphical measures were created by different groups to "measure" what they meant by a "good prediction" for the five day temperature predictions. First, several students announced that the predictions were exactly correct only when the points landed on the line of equal temperatures--their own creation. This line became known as the "J. Alex" line, after one of its creators (Figure 11).

Student Plot with "J. Alex" line

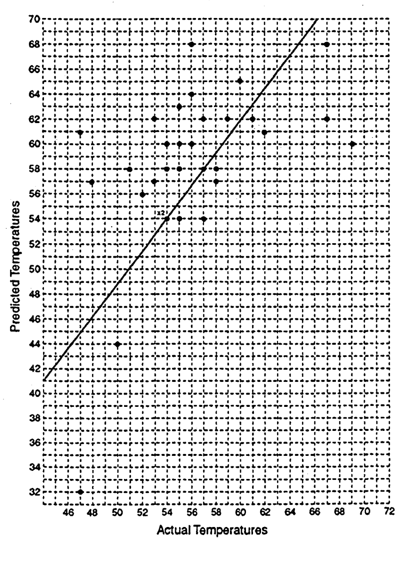


Figure 11

Second, several other students created confidence band lines (they didn't call them this, of course) around the J. Alex line and decided that for this data, a "good prediction" meant that the points had to fall within these ±5 degree bands. These lines became known as the "Morgan" lines, naturally after one of their creators (Figure 12).

Student Plot with "Morgan" lines

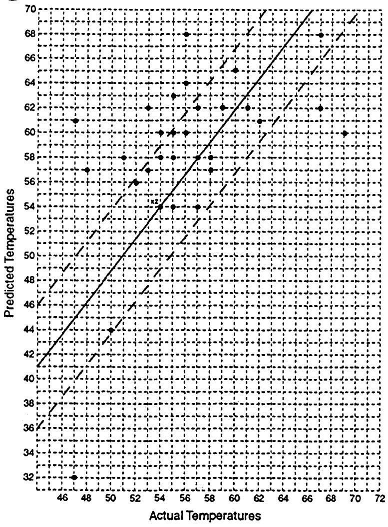


Figure 12

What was fascinating about this entire process were the many things that these students *could* do, on their own, while investigating this weather data. Given some time, the resources of their fellow students, and an interesting (to them) set of data, they were quite capable of organizing, analyzing, and representing the data their own way. Furthermore, they came up with several of their own user-constructed measures for making decisions about this data. If we, as researchers, give students an opportunity to show us what they *can* do, rather than what they cannot do, we may discover some encouraging conceptions and abilities that normally go unnoticed.

Situation 4--What next?-- Investigate students' conceptual development in variability, over time, starting with what they* can *do.

I'd like to conclude with a situation that I am beginning to work on. In pursuing this situation, I have been trying to listen to my own suggestions about some of these missed opportunities in research that we have discussed: 1) follow-up on student's thinking and reasoning by documenting growth and change over time; 2) begin to investigate student's conceptions of variability; and 3) try some research approaches that uncover what our students *can* do in problem solving in chance and data, rather than merely documenting what they are unable to do.

The data for this situation involve time lapses between blasts of one of our U.S. national treasures, the geyser Old Faithful located in Yellowstone National Park. The data were taken from the book *A Handbook of Small Data Sets* (Hand et. al., 1994), which contains data for the time intervals between blasts of Old Faithful for 15 consecutive days. The plan is to put students into a context in which variability runs rampart, in several forms, but in which there is still the possibility of detecting patterns. Students are presented a table of the data taken from one day's blasts of Old Faithful (Table 1). Then they are asked, "What do you predict for the next day's data set, and why?"

Table 1

Time intervals in minutes between Old Faithful's blasts

First Day Data

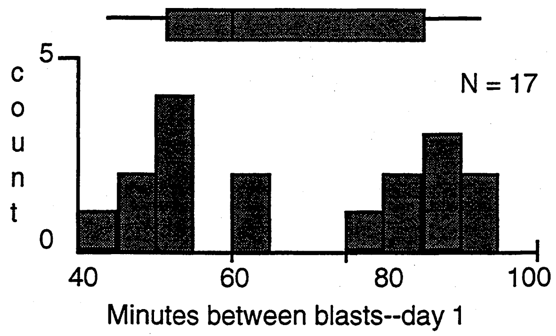
50	87	48	93	54	86	53	78	52	83	60
87	49	80	60	92	43					

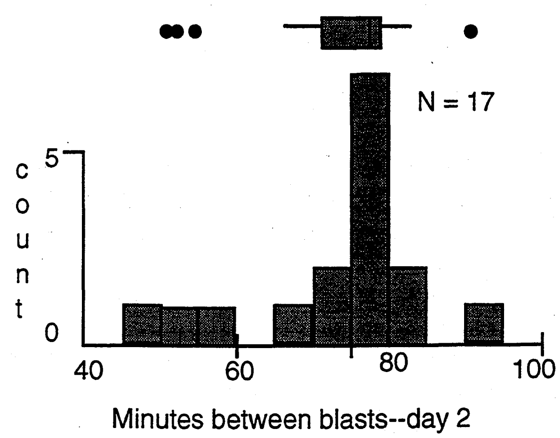
First Day Ordered Data --

43 48 49 50 52 53 54 60 60 78 80 83 86 87 87 92 93

After they make their predictions for the next day's data, students are given data for three consecutive days of Old Faithful blasts, and asked to investigate whether there are differences in the data for these three days, and how they decided. The idea is to create an environment in which students get more and more information, one in which they become detectives while exploring the ultimate question "Just how faithful is Old Faithful, anyway?"

Some visual representations from blasts on three consecutive days for Old Faithful are given in Figure 13. This data set has the potential to highlight the concept of variability, and perhaps to help counter Landwehr's concern that people believe there is "no variability in the real world." First, the time intervals between blasts given in Table 1, indicate that Old Faithful appears to alternate between long time - short time throughout the day. Thus, there is variation within a day. Second, although the three ranges for the time intervals are comparable for all three days (comparative boxplots in Figure 13), the inter-quartile ranges are very different for the three days. They tend to cluster around 70-80 minutes for day 2, for example, but are spread from the mid 50's to the mid 80's for day 1, and from the mid 60's to the mid 80's for day 2. So, there is also variation across days.





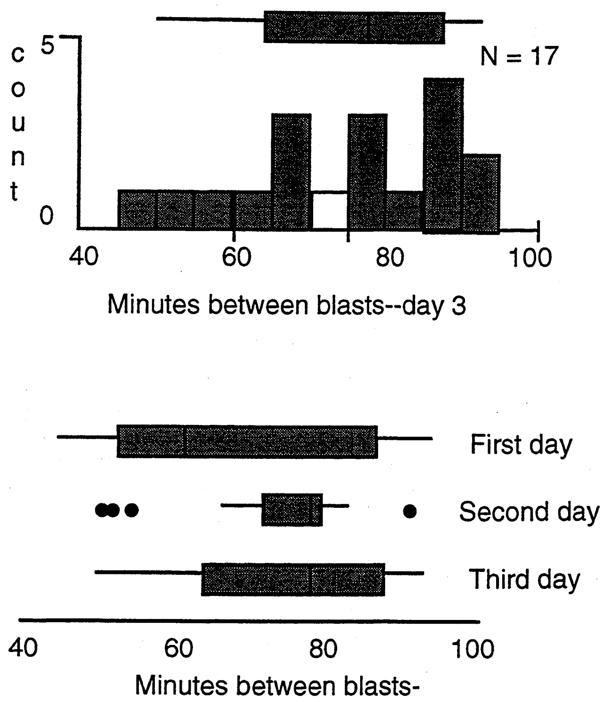


Figure 13

Clearly centers alone are not adequate to describe the Old Faithful's data set from one day to the next day, as there is tremendous variability in the range of times for the blasts, even though there are identical medians for days 2 and 3.

Since middle school students love to play detective with complex data sets, they will likely be interested in exploring questions like: Does this long-short alternation hold up over many days? Do differences between days, like days 1 & 2, continue to occur or is this an unusual phenomenon? They will also come up with questions of their own, and with ways of representing the data that might not occur to us. I believe the Old Faithful environment is an example that will prove to be a fruitful context for investigating students' development of the concept of variability.

In conclusion I would like to recommend that we investigate students thinking about data and chance starting with data sets and tasks that elicit conceptions of variability and difference, rather than tasks that focus on centers and sameness. I would also suggest that a more fruitful avenue for our research might be to begin with rather complex *multivariate* data sets, like the weather pages, rather than simple univariate data sets that can seduce us and our students into looking too much just at averages. And, as we go about our research, let's remember to follow-up on our students' thinking, and to include tasks that show us what they *are* capable of doing.

References

Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R.., & Holmes, P. (1984). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics Education in Science and Technology, 25*(4), 527-547.

Batanero, C. (Ed.). (1996). Research on the role of technology in teaching and learning statistics. IASE Roundtable Conference Papers, University of Granada, Spain.

Borovcnik, M. & Peard, R. (1996). Probability. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International Handbook of Mathematics Education,* Dordrecht: Kluwer Academic Publishers.

Fischbein, E. F.. (1987). *Intuition in Science and Mathematics.* Dordrecht: D. Reidel Publishing

Fischbein, E. F.., & Schnarch, D. S.. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education,* 28(1), 96-105.

Foreman, L.C., & Bennett, A. B. (1995).*Visual Mathematics.* Salem, OR: The Math Learning Center.

Foreman, L. C., & Bennett, A. B. (1996). *Visual Mathematics.* Salem, OR: The Math Learning Center.

Garfield, J. B. & Ahlgren, A. (1988). Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research. *Journal for Research in Mathematics Education,* 19(1), 44 - 63.

Garfield, J. B. (1994). *Research Papers from the Fourth International Conference on Teaching Statistics.* Minneapolis: The University of Minnesota.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice-versa). In G. Wright and P. Ayton (Eds.), *Subjective Probability,* New York: John Wiley & Sons.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A rebuttal to Kahneman and Tversky. *Psychological Review, 103* (3).

Gigerenzer, G. & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102* (4), 684-704.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J., & Ostrowski, E. (1994). *A Handbook of Small Data Sets.* London: Chapman & Hall.

Hawkins, A. & Kapadia, R.(1984). Children's Conceptions of Probability: A Psychological and Pedagogical Review. Educational *Studies in Mathematics,* 15, 349 - 377.

Jones, G. A., Thorton, C.A., Langrall, C. W., & Mogill, T. A. (1997). *Using students' probabilistic thinking in instruction.* Paper presented at the Research Presessions at the 75th Annual Meeting of the National Council of Teachers of Mathematics, Minneapolis, Minnesota.

Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Kahneman, D. & Tversky, A. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-232.

Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Konold, C. (1983). Conceptions about probability: Reality between a rock and a hard place. (Doctoral dissertation, University of Massachusetts, 1983). *Dissertation Abstracts International, 43, 4179B.*

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6,* 59-98.

Konold, C. (1991). Understanding students' beliefs about probability. In E. von Glasersfeld (Ed.) *Constructivism in Mathematics Education* (pp. 139-156). Dordrecht: Kluwer Academic Publishers.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J. & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24* (5), 392 - 414.

Konold, C. & Miller, C. (1994). *Data Scope*. Santa Barbara: Intellimation.

Landwehr, J. M. (1989). A Reaction to '*Alternative Conceptions of Probability: Implications for Research, Teaching, and Curriculum*'. Symposium Conducted at the 11th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA). New Brunswick, NJ: Rutgers University Press.

Loosen, F., Lioen, M., & Lacante, M. (1985). The standard deviation: some drawbacks of an intuitive approach. *Teaching Statistics, 7*(1), 29-39.

Nisbett, R. E. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Phillips. B. (Ed.). (In Press). Research on Teaching and Learning Probability and Statistics. Papers presented at ICME 8, the Eighth International Congress on Mathematics Education. The University of Sevilla, Spain.

Pollatsek, A., Lima, S., & Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics, 12*, 191-204.

Puig, L. & Gutiérrez, A. (1996). Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education, PME. Dept. de Didáctica de la Matemática, Universitat de Valéncia, Spain.

Russell, S. J. & Mokros, J. (1996). What do children understand about average? *Teaching Children Mathematics, 2* (6), 360-364.

Schrage, G. (1983). (Mis)-Interpretation of stochastic models. In R. Scholz (Ed.), *Decision Making Under Uncertainty* (pp. 351 - 361). Amsterdam: North-Holland.

Serrano, L. (1996). Significados personales e institucionales de objetos matemáticos ligados a la aproximación frecuencial de la enseñanza de la probabilidad. Ph.D. University of Granada.

Shaughnessy, J. M. (1977). Misconceptions of probability: An experiment with a small-group, activity-based, model building approach to introductory probability at the college level. *Educational Studies in Mathematics, 8,* 285-316.

Shaughnessy, J. M. (1981). Misconceptions of probability: From systematic errors to systematic experiments and decisions. In A. Shulte (Ed.), *Teaching Statistics and Probability* (Yearbook of the National Council of Teachers of Mathematics, (pp. 90-100). Reston, VA: NCTM.

Shaughnessy, J. M. (1992). Research in Probability and Statistics: Reflections and Directions. In D. Grouws (ed.), *Handbook of Research on Mathematics Teaching and Learning* .(pp. 465-494). New York: Macmillan.

Shaughnessy, J. M. & Arcidiacono, M. J. (1993).*Visual Encounters with Chance. Math and the Mind's Eye*, Unit VIII. Salem: OR: The Math Learning Center.

Shaughnessy, J. M., Garfield, J. B., & Greer, B. (1996). Data Handling. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International Handbook of Mathematics Education,* Dordrecht: Kluwer Academic Publishers.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1995). The development of concepts associated with sampling in grades 3, 5, 7, and 9. Paper presented at the Annual Conference of the Australian Association for Research in Education, Hobart.

Watson, J. M., & Moritz, J. B. (1997). Personal communication.