TEACHING STUDENTS THE STOCHASTIC NATURE OF STATISTICAL CONCEPTS IN AN INTRODUCTORY STATISTICS COURSE[1]

MARIA MELETIOU-MAVROTHERIS
Cyprus Ministry of Education
meletiu@spidernet.com.cy

CARL LEE
Department of Mathematics,
Central Michigan University, MI 48859, USA
lee1c@cmich.edu

SUMMARY

The article argues that the persistence of student difficulties in reasoning about the stochastic, despite significant reform efforts, might be the result of the continuing impact of the formalist mathematical tradition, affecting instructional approaches and curricula and acting as a barrier to instruction that provides students with the skills necessary to recognize uncertainty and variability in the real world. It describes a study driven by the conjecture that the reform movement would have been more successful in achieving its objectives if it were to put more emphasis on helping students build sound intuitions about variation. It provides an overview of how the conjecture guiding the study was developed and linked to classroom practice, and briefly discusses the experiences and insights gained from a teaching experiment in a college level, introductory statistics classroom, which adopted a nontraditional approach to statistics instruction with variation at its core. By contrasting students' intuitions about the stochastic prior to instruction to their stochastical reasoning at the completion of the course, it illustrates the potential of the instructional approach as an alternative to more conventional instruction.

***Keywords*:** Statistics education research; Stochastic; Variation; Introductory statistics; Classroom experiment

1. INTRODUCTION

Significant reform efforts have been lately witnessed in statistics education, which have led to a movement away from statistics instruction emphasizing the abstract and the memorization of formulas and procedures. More importantly, these efforts have led to a general acknowledgment that learning occurs most effectively when students engage in authentic activities. Although many students are still being taught in traditional classrooms, there is already a large number of statistics instructors who have adopted alternative approaches to their teaching and many statistics classrooms are experiencing wide incorporation of technology. But, as Hawkins (1997) points out, for reform efforts to be successful, it is also necessary to change attitudes and expectations about statistics education. Changing long-held beliefs and attitudes towards statistics is proving to be quite difficult (Wilensky, 1993). Many people still view statistics as *"a branch of the older discipline of mathematics"* (Glencross and Binyavanga, 1997, p. 303). This affects statistics instruction and hampers the reform efforts.

The linear and hierarchical approach adopted by statistical courses and syllabuses, is testimony to the profound and continuing effect of the formalist mathematics culture on statistics education. The structure of almost every introductory statistics course is to first start with descriptive and exploratory data analysis, then move into probability, and finally go to statistical inference. It is assumed that this simplifies the process of learning by gradually leading students from more basic to more complex connections (Steinbring, 1990). However, presenting statistical content as a sequenced list of curricular topics might lead to compartmentalization of knowledge and fail to communicate to students the interconnectedness of the different statistical ideas they encounter in the course.

---

[1] *Statistics Education Research Journal, 1(2), 22-37, http:/fehps.une.edu.au/serj*
International Association for Statistical Education

Statistical methods were developed to help us filter out any "signals" in data from surrounding "noise". "Signals" are the messages, the meanings we find in *explained variation*, the patterns that we have not discounted as being transient. "Noise" is the *unexplained variation* that remains after we have "removed" all patterns. Randomness and probability theory are "human constructs" created to deal with unexplained variation. We use probability theory to model and describe phenomena in the world for which no patterns can be discerned, assuming that they had been randomly generated. Thus, "*what probability is can only be explained by randomness, and what randomness is can only be modeled by means of probability*" (Steinbring, 1990, p. 4). Stochastical knowledge is created as "*a relational form or linkage mechanism between formal, calculatory aspects on the one hand, and interpretative contexts on the other*"(Steinbring, 1990, p.5). However, the classroom culture often comes in sharp contrast with this conception of stochastical knowledge as being developed through a "self-organized" process that balances the objective aspects of a situation and the formal means employed to model and describe it. The linear, completely elaborated and hierarchical structure of knowledge presentation encourages the development of the chance concept as a concrete, totally clear and unambiguous generalization defined by methodological conventions. Steinbring (1990), who analyzed teaching episodes from several different classrooms in order to see how the concept of chance was introduced, found that in all of those episodes "chance" was first introduced through performing and discussing a chance experiment. An attempt was then made to describe the experimental outcomes using a rule or a simple stochastical model. Naturally, there was always variation observed between the theoretical predictions and the empirical data. The pattern of justification for the observed variation, regardless of its size, always was that the difference between the empirical result and the theoretical prediction was produced by "chance" (Steinbring, 1990). The difference between theory and experiment was thus neutralized, with chance degenerating into "*a substitute for justification, which serve[d] to deny the importance of the difference between theory and empirical facts in probability*" (Steinbring, 1990, p. 14).

The assumptions posed in the statistics classroom are often too simplistic. Although not necessarily denying underlying causal explanations in case of chance events, a probabilistic approach views them as impractical and adopts a 'blackbox' model (Biehler, 1994, p. 10). However, as Biehler (1994) indicates, the assumption of independence is not plausible in many real-world contexts: "*Even coin flipping can be done in a way that independence has to be rejected in favor of serial correlation, and physical theories can be developed to explain some aspects of coin flipping*" (p.10). Borovcnik and Peard (1996) warn that instruction has traditionally underrated the complexity and dangers of using pseudo-real examples that conflict with students' emotions or with their common sense.

The over-emphasis of the traditional mathematics curriculum on determinism and its "*orientation towards exact numbers"* (Biehler, 1997, p. 187) affects statistics instruction, becoming an obstacle to the adequate judgment of stochastic settings. The law of large numbers is often presented as a canon in the statistics classroom, giving students the false impression that the stabilization of the relative frequency of repeated sampling to the ideal value is *guaranteed*. Similarly, instruction leaves students with the impression that a larger random sample *guarantees* a more representative sample. There is a deterministic mindset and an over-reliance on rules and theorems, forgetting that the uncertainty and variability accompanying all finite statistical processes implies that a sample is almost never totally representative of the population from which it was selected. People have difficulties in distinguishing between the real-world problem and the statistical model. At one extreme are many people who use statistical methods for solving real-world problems in the same way they would use an artificial mathematics problem coming out of a textbook. On the other extreme, we find people who distrust statistics completely because, unlike mathematics, it deals with uncertainty. Both of these two extreme attitudes suggest inadequate understanding of statistics as a decision-support system (Biehler, 1997).

Current practices in statistics education have evolved from a background quite different from today's needs and possibilities. Hoerl, Hahn, and Doganaksoy (1997), argue that we must completely rethink the sequence of topics in order to achieve the objectives for introductory education, which for them should be to "*help students unlearn their deterministic view of the world"* (p. 152) and recognize uncertainty as a characteristic of reality. Chance and Rossman (2001) discuss four perspectives on the sequencing of topics in an introductory statistics course and lay out the merits and drawbacks of each perspective. In this article, we describe the experiences obtained from a teaching experiment that implemented an alternative path to statistics instruction with variation at its core, conjecturing that this path would lead to stronger and deeper understandings by helping students see the "big picture" of statistics (Moore, 1997). By contrasting students' intuitions about the stochastic prior to instruction to their stochastical reasoning at the completion of the course, we attempt to illustrate how the instructional approach employed in the study proved a promising alternative to more conventional instruction.

2. DESIGN OF THE STUDY

2.1. BACKGROUND

The motivation for the study was provided by the results of a previous study of twenty-two students who had just completed an introductory statistics course. The results of that study (Meletiou, Lee, and Myers, 1999), agreed with the main findings of research in the area of stochastics education. Similarly to the research literature, we found in that previous study that the students we interviewed, regardless of whether they came from a lecture-based classroom or from a course following the PACE (Projects-Activities-Cooperative Learning-Exercises) model, had poor intuitions about the stochastic and tended to think deterministically. PACE is an approach developed by Lee (1997, 2000), that attempts to provide a structured framework for integrating projects and hands-on activities conducted cooperatively in a computer-based classroom environment.

After witnessing students' poor statistical reasoning and their deterministic mindset, we concluded that student difficulties might stem from inadequate emphasis paid by instruction to the notion of variation, and to the connections among statistical ideas. This led to a decision to modify the PACE model. The course described in the next sections, which was taught by Lee, although still having a format similar in many ways to that followed in previous semesters, differed significantly in both curricular emphases and structure. A nontraditional approach to statistics instruction that had variation as its central tenet, and perceived learning as a dynamic process subject to development for a long period of time and through a variety of contexts and tools, was adopted.

2.2. VARIATION AS THE CENTRAL TENET OF STATISTICS INSTRUCTION

Statistical thinking is concerned with learning and decision-making under uncertainty. Variation is a critical source of uncertainty. It is the fact that all processes vary which creates the need for statistics. It is the need to deal with variation through measurements that provides a (numerical) basis for comparison that produces data (Snee, 1999). We use statistical tools to analyze this data and observe the pattern that exists despite (or because of) the variation. Thus, according to Snee (1999), the elements of statistical methods are variation, data, and statistical tools. Understanding of variation and using this understanding to improve the performance of processes is the core competency and it should be the focus of statistical education, research, and practice (Snee, 1999). Understanding what data is relevant and how to construct proper methods of data collection and analysis enhances successful application of this core competency (Snee, 1999).

The central element of statistical thinking is variation, and instruction should aim at providing students with the skills necessary to be able to notice and acknowledge it, to explain and deal with it. But, if variation is indeed to be "the standard about which the statistical troops are to rally" (Wild and Pfannkuch, 1999, p. 235), we have to arrive at a common conceptualization of statistics instruction in terms of variation. Wild and Pfannkuch (1999) offer the following three "variation" messages as a starting point: (1) variation is omnipresent; (2) variation can have serious practical consequences; and (3) statistics give us a means of understanding in "a variation-beset world" (see also Cobb, 1992). The subsequent messages of the statistics classroom provide information about tools and methods statistics offers us to make sense of the omnipresent variation.

Pfannkuch's (1997) epistemological triangle, views variation as the broader construct underlying statistics instruction (Figure 1). In encouraging students to develop their understanding of the concept of variation, Pfannkuch's epistemological triangle aims at the same time at promoting richer understanding of all the other main statistical ideas. The epistemological triangle indicates that for conceptualization of variation, a combination of subject and context knowledge is essential (Pfannkuch, 1997). The inter-linked arrows indicate the strong linkage that has to be created between statistical tools and the context of the problem. The assumption underlying the epistemological triangle is that the concept of variation would be subject to development over a long period of time, through a variety of tools and contexts (Pfannkuch, 1997).

Pfannkuch's model bases instruction on contexts directly connected to students' experience, it recognizes that adequate statistical reasoning requires more than understanding of different ideas in isolation. It demands "*integration between students' skills, knowledge and dispositions and ability to manage meaningful, realistic questions, problems, or situations*" (Gal and Garfield, 1997, p. 7). Content is no longer a sequenced list of curricular topics taught in isolation, but "*an interrelated repertoire of conceptual tools that can assist one in making sense of, and gaining insight and prediction over interesting phenomena*" (Confrey, 1996).

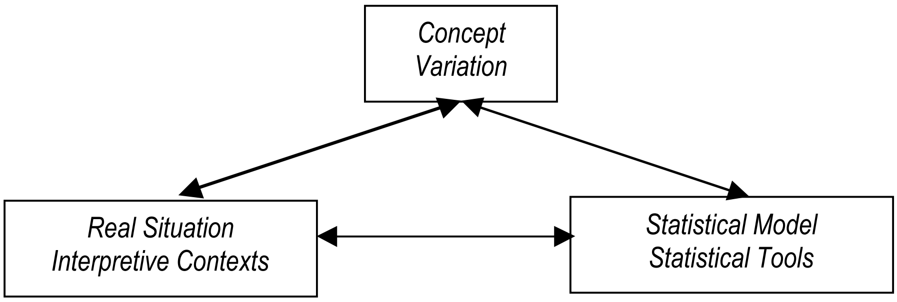

Figure 1. Pfannkuch's Epistemological Triangle

The conjecture driving our study was that the reform movement would be more successful in achieving its objectives if it were to put more emphasis on helping students build sound intuitions about variation and its relevance to statistics (Ballman, 1997). Pfannkuch's epistemological triangle, which calls for the re-structuring of statistics instruction by offering a nontraditional path with variation at its core, seemed well suited for meeting our research aspirations, and was employed in the study to guide curriculum development and instruction.

2.3. CONTEXT OF STUDY AND PARTICIPANTS

The site for the study was an introductory statistics course in a mid-size Midwestern university in the United States. One of the authors, Lee, was the instructor of the course. The study lasted over the span of five weeks. The course began on the last week of June and ended on the first week of August. Class met four times a week, for two hours each time. The number of students in the class was thirty-three (nineteen males and fourteen females). Most of the students in the class (twenty-two students) were majoring in a business-related field of study. Only few had studied mathematics at the pre-calculus level or higher.

2.4. DATA GENERATION

In examining students' learning progress and outcomes, a variety of both qualitative and quantitative data gathering techniques were employed. By assessing students' understanding prior to instruction, and then monitoring changes in their thinking throughout the course, the study attempted to develop a detailed description of the processes students go through in order to become able to intelligently deal with variability and uncertainty.

Given the number of students involved, it was impossible to observe closely every single student in the course. Therefore, we chose to study two groups of students. The primary group consisted of a subset of eight students (five males and three females) and the secondary group encompassed the whole class. Although data from both groups were used in the analysis, our focus was on investigating and describing the learning experience of students in the primary group. The selection criterion for the primary group was willingness to participate in the study.

Next, we explain the specific data that was generated. We have found it useful to describe the data generation process separately for each of three phases of the course: (a) beginning of course, (b) duration of course, and (c) end of course.

Beginning of course

In order to be able to follow the students' conceptual development process, a good understanding of their thinking prior to instruction is required. A diagnostic questionnaire with ten, mostly open-ended, questions was given to all students on the first day of class to assess their intuitive understanding of variability prior to instruction. Additionally, we conducted individual interviews of students in the primary group. The interviews, which were audio-taped, were semi-structured. In the first part of the interview, we posed students in the primary group a set of open-ended questions, which aimed at further investigating their intuitive understandings of variability. The second part of the interview was a follow-up of the questionnaire on variability completed by the whole class. We went over the questionnaire with the students in order to clarify the reasons for the different responses they had provided. Despite the open-ended nature of the diagnostic questionnaire, one-to-one communication with students in the primary group allowed a more thorough investigation of student reasoning.

Duration of course

The data gathering techniques employed during the course included: (1) direct and participant observations, (2) interviews of students, (3) video-taping of group activities, (4) pre- and post-activity assessments, (5) field notes, and (6) samples of student work. Drawing data from several different sources permitted cross-checking of data and interpretations.

Although in this article we do not present findings from the analysis of the data collected during the course, the insights drawn from this analysis did play an important role in shaping the interpretation of results from the end-of-course assessment presented in Section 5. The continuous monitoring throughout the course of the effect of instruction on student learning was constantly supplying us with valuable information on their levels of concept attainment. The assessment strategies used to support and evaluate students' conceptual development helped students further clarify their reasoning strategies.

End of course

In order to assess students' understanding at the completion of the course, a diagnostic questionnaire with fifteen open-ended questions was administered to the whole class. Additionally, we again conducted individual interviews with each of the students in the primary group. During the interviews, we went over the diagnostic questionnaires students had taken at the beginning and the end of the course, as well some assessment tasks they had completed during the course, and also reminded students of some of the responses that they had provided during the interview at the beginning of the course. We prompted students to express their agreement or disagreement with responses they had given earlier in the course, so that we could check whether their reasoning had changed in any way since then.

3. FINDINGS FROM BEGINNING-OF-COURSE ASSESSMENT

Findings from the assessment at the outset of instruction further supported the conjecture that variation is neglected, and its critical role in statistical reasoning is under-recognized. In this section, we present findings from two tasks included in the diagnostic questionnaire given to the whole class, and three tasks completed by students in the primary group during the follow-up interview, which are indicative of the tendency we observed at the beginning of the course for students to think deterministically and to have difficulties in differentiating between chance variation in the data and variation due to some form of underlying causality.

Students' performance in a question in the pre-assessment adapted from Rubin et al (1990), which examined how they balanced the ideas of sampling variability and sampling representativeness, illustrates their propensity to underestimate the actual level of random variability. In this question (see *Gummy Bears Question* in the Appendix), students were told that the Easter Bunny was distributing many packets of 6 Gummy Bears at the Easter Parade that he had made up by grabbing handfuls of Gummy Bears out of a large barrel containing two million green and one million red Gummy Bears that had been randomly mixed, and were asked to estimate the number of green Gummy Bears in a packet. Everyone gave *"4 green, 2 red"* as the estimate, but also everyone realized that *"not every student got exactly 4 green every time because there's variability".* They intuitively understood that probability is the limiting relative frequency, which only approximately holds for real data: *"That is just the mathematical way of figuring it, that number will fluctuate"; "Expected ratios are a general rule, not a formula for each individual occurrence*"; *"It is nearly impossible for the ratio to hold perfectly, unless the Easter Bunny uses his Easter magic."* However, although students did recognize that *"there will be a variation on the pattern of green bears in each bag, because of the random grabbing of the bears when they were placed in the bags*", when asked to estimate the proportion of packets with 4 greens, almost all of them underestimated the effect of sampling variability and greatly overestimated this proportion. Only two students gave estimates that came close to 33%, the actual probability of 4 greens (found by modeling the situation as a Binomial distribution). The estimates that the rest of the students gave ranged from 50%-92%. Several students wrote that they expected 66% of the packets to have 4 greens in them. Rubin et al. (1990) who gave this question to high school seniors also noticed that students answered this question by focusing on samples that mirrored the population proportion of 2G:1R. They over-relied on sample representativeness, underestimating the frequency of samples near the tails of the distribution and overestimating the frequency of the modal sample.

The tendency to underestimate the effect of sampling variability and expect small samples to match population properties was also witnessed in a question in the pre-assessment taken from Garfield and delMas (1990). The question (see *College Students' Interview Question* in the Appendix) described how a worker of a student organization went about conducting a survey at a certain college where half the students were women and half were men and the several measures he took to ensure good representation of all students. Students were told that out of the last 20 students interviewed, 13 were women and 7 were men and were asked whether they thought there would be more women or more men in the next 20 students interviewed. Only 35% of the students correctly stated that one should expect about an equal number of men and women, as who has been selected so far does not affect who will be next selected. Thirty-two percent argued that since more women than men were selected so far, they expected the opposite trend to start happening, 16% tried to find causes behind a difference that – given the small number of people interviewed this far – could be easily explained by chance variation while the rest, employing the "law of small numbers", thought that the trend of selecting more females than males should continue.

The tendency to underestimate the role of chance variation was more pronounced in real world-contexts. Although students did seem aware of the dangers involved when drawing conclusions from small samples, when asked to make their own judgments based on data, they often ignored these dangers and, exaggerating the reliability of the information provided, did not hesitate to use small samples as a basis for inferences. The responses of students in the primary group during the follow-up interview to the *Birth Defects* task (taken from Pfannkuch and Brown, 1996) are indicative. In this task (see Appendix), students were told that each year approximately seven children are born in New Zealand with a missing limb. They were shown a map of New Zealand divided into five regions, with the number of children born with this abnormality in each region during the previous year located on the map. According to the map, there were no children born with the abnormality in either the top or bottom region, while the number of children born with a missing limb in each of the other three regions were three, three, and two respectively. They were asked to comment on this, given the information that in New Zealand one-third of the population lives in the top region and one-sixth of the population in each of the other regions.

Pfannkuch and Brown (1996) found that students' understanding of variation in small samples was poor in this context. Whereas an analysis combining both probabilistic and deterministic thinking would have been more appropriate, all of the students they interviewed gave deterministic explanations, and it was only after repeated probing that some suggested the need for more data. Our findings were very similar. We observed very strong deterministic reasoning in all of the students. George, for example, *"wouldn't want to live in the middle of New Zealand*", and Julie was convinced that there must be an outside factor causing the difference: *"There is always a chance that anything can happen but, 3 and 0 in the other…there must be a reason for that.."*

Pfannkuch and Brown (1996) conjectured that students' neglect of probabilistic thinking might be the rich experience they have with similar controversial data often appearing in the media and seldom being explored from a statistical perspective. When asked what they think of the possibility of obtaining the outcome {3,3,3,4,4,5,5} (order unimportant) when rolling a fair die 7 times, no student found such an outcome surprising. They approached this problem very differently from the New Zealand one although it is analogous – obtaining 1 or 2 on the die corresponds to the top region of the map where one-third of the population lives, and obtaining a 3,4,5, or 6 corresponds to each of the other regions. Similarly, in our study, students found such a result quite likely due to the small sample size that allows extreme outcomes: *"I think nothing of the results. After a thousand throws each number will be picked around 1/6 of the total throws."*

The different way in which students approach the two problems indicates how much more prone we are to look from a stochastical perspective at standard probability tasks than problems situated in certain real-life contexts (Pfannkuch and Brown, 1996). Students are at fault in searching for causes behind data of such nature, where a lot of other factors besides chance might influence the occurrence of birth defects. Still they should realize that a sample of only 7 children has too little information in it to help us find causes since small samples often have large natural variation even when no causes are operating. They should have shown the same sensitivity to the effects of sample size they showed in the *Child Psychologist* question (see Appendix) taken from Garfield and delMas (1994). In that question, students interviewed were asked to judge the validity of conclusions drawn by a child psychologist who, after studying 5 infants and finding that 4 showed preference for the one toy, concluded that most infants would show a preference for this toy. Every single student interviewed challenged the psychologist's conclusions. Tim for example said: *"4 out of 5, I know it's good for like 4 out of 5 dentists prefer this* *kind of toothpaste, whatever on the commercials, but I would say you need at least a 100 kids...I could get my 5 sons and persuade 4 of them."* His response comes in sharp contrast to how he responded to the *Birth Defects* question:

Int.: *Just by looking at the map, do you see any connection between where one lives and how many kids are born with a missing limb*?

Tim: *Oh, yeah. They correlate because the 1/3 that lives there has 0 because probably there are more doctors and more hospitals and only 1/6 lives there, so there must be something going wrong there. So yes, there has to be a reason*.

Int.: *Do you see that the numbers are small? Do you think this is something you should take into account*?
Tim: *Why?*

4. COURSE DESCRIPTION

4.1 CURRICULUM

The design of the intervention was guided by our conjecture that if statistics curricula were to put more emphasis on helping students improve their intuitions about variation and its relevance to statistics, we would be able to witness improved comprehension of statistical concepts. At the same time, the time constraints and confines of the curriculum were also taken into account. Instruction included the set curriculum typically covered in an introductory statistics course, but was expanded in such a way as to include throughout the course activities that aimed at raising students' awareness of variation. The different topics were approached through the lens of the conjecture. Adjustments to the curriculum were also guided by the following two principles (adapted from Wild and Pfannkuch, 1999):

1. *Complementarity of theory and experience*: Statistical thinking necessitates a synthesis of statistical knowledge, context knowledge, and information in the data in order to produce implications, insights and conjectures. If the statistics classroom is to be an authentic model of the statistical culture, it should model realistic statistical investigations rather than teaching methods and procedures in a sequential manner and in isolation. The teaching of the different statistical tools should be achieved through putting students in authentic contexts where they need those tools to make sense of the situation. Students should come to value statistical tools as a means to describe and quantify the variation inherent in almost any real-world process.

2. *Balance between stochastic and deterministic reasoning*: Instruction should view as an important precursor of statistical reasoning students' intuitive tendency to come up with causal explanations for any situation they have contextual knowledge about. It should present statistical thinking as a balance between stochastic and deterministic reasoning and should stress that statistical strategies, based on probabilistic modeling, are the best way to counteract our natural tendency to view patterns even when none exists, to distinguish between real causes and ephemeral patterns that are part of our imagination.

4.2 CLASSROOM SETTING

The typical setting during a class session was such that it encouraged "statistical enculturation". The instructor's knowledge and behavior contributed towards the creation of an authentic model of the "statistical culture" (Biehler, 1999). It was a setting that modeled realistic statistical investigations and in which statistical dispositions such as appreciation of data were valued and nurtured. The instructor was trying to increase students' awareness of variation, to help them realize that it is the existence of variation, which creates the need for statistical investigations. He would keep on emphasizing that the reason we use statistical tools is to describe trends and patterns and deviations from those patterns existing in the data because of the variation inherent in every process.

No method or procedure was taught in isolation. In contrast to more typical approaches where reference to problems is made to demonstrate statistical content, reference to statistical content in this class was made (in students' mind at least) to help understand a situation, to assist a statistical investigation. The emphasis was on statistical process and along the way students got to learn different statistical methods and procedures. The hope was that by putting students in situations where they needed tools such as the standard deviation, they would realize their usefulness and not wonder why anyone would ever bother to invent them (Erickson, 2000).

Unlike conventional instruction that tends to focus on mechanical application of methods and procedures, the emphasis of this course was on recognizing applicability and interpreting results in context (Wild and Pfannkuch, 1999). In introducing binomial distributions, for example, the main goal was not to teach the formal properties of the binomial distribution but to help students recognize a binomial setting and understand why we can apply this population distribution to model a certain variable and in what ways this is useful. Students were first given a description of five different situations that could be modeled using the binomial distribution, and were asked to work with their group in order to figure out their common properties. Group work was followed by a whole class discussion during which students laid out the main properties of the binomial distribution. Introduction of the probability formula describing binomial distributions was done only after students had brought up several examples of situations in the real world that could be modeled using the binomial distribution.

In order to simplify mathematical relations and build links to students' intuitions, the course emphasized the use of analogies from students' everyday experience in contexts familiar to students. Instruction stressed the complexity of real-life situations rather than making simplistic assumptions that would conflict with students' common sense. When, for example, discussing independent events, and after students had given typical examples of independent events such as coin tossing and die rolling, the instructor asked the class whether the success of a *"free throw"* of a basketball player is independent from the success of his previous *"free throw"*. Students argued that it depends on how the player responds to pressure, on how well he did on the previous throw etc. The instructor agreed remarking that, *"in real life it's hard to say with a straight yes or no"*. He did not reject students' causal explanations although "hot hand" is an example often used by many statistics educators in their pleading for probabilistic reasoning. Tversky and Gilovich (1989) showed, using empirical data, that a binomial model well explains runs (streaks) in basketball player failures. According to this model, the chance of success in a shot is independent from the previous shot, and Tversky and Gilovich, and subsequently many teachers and researchers, concluded that people's tendency to detect patterns (hot hands) is often unwarranted. One need not look for specific causes like nervousness since chance patterns produced by a completely random process well explain the data. However, as Biehler (1994) has pointed out, even when the binomial model well explains the variation in a dataset, one should not exclude the possibility of alternative models that give better prediction and that suggest causal dependence of individual throws. The instructional approach in this class clearly made students aware of this perspective. Similarly, when talking about slot machines in a casino, the instructor noted: *"Although in theory when you put a coin and you pull it down and then you put another coin and you pull it down, although those two events should be independent, mechanically they might not be."*

4.3 EMPHASIS ON THE PROCESS OF CONJECTURING AND DISCOVERY

The idea of making conjectures ran throughout the course. Students would state what they believed might or might not be true, and then looked critically at the data to evaluate their statements. While the instructor encouraged students to make conjectures he, at the same time, also tried to help them understand that conjecturing is not enough – one has to evaluate one's predictions by looking closely at the data and making comparisons (Erickson, 2000). Evaluation of conjectures would typically begin informally by using one or more graphical displays. Students would be encouraged to describe the main features of the distribution displayed by the graph(s), always emphasizing the need to take into account not only the center, but also the spread. Students would look at the displays and try to give explanations for the patterns observed and for the departures from those patterns. Sometimes these explanations would be proposals for a possible model to summarize the dataset.

The evaluation of conjectures would then become more quantitative. An analysis using appropriate numerical summaries would be made to support or refute the conjectures originally made by students. At the start of the course, the analysis was made using simple numerical summaries. Eventually, more tools were added to students' repertoire and the mathematization of the data gradually became more formal. Even when the data agreed with their initial conjecture, the instructor would encourage students to also come up with alternative explanations and see that there can be multiple explanations for a phenomenon, in the hope that this would make them "*less likely to assume that their data 'proves' the obvious cause*" (Erickson, 2000, p.2). The issue of unrealistic or unrelated conjectures that went beyond the information provided by the data was often raised during the process of conjecturing and discovery.

4.4 EMPHASIS ON DATA PRODUCTION AND VALIDITY OF MEASUREMENTS

A special emphasis of the course was on data production issues. Unlike many other statistics courses where study design issues are discussed as a separate topic and almost never appear again, they were continuously brought up in this course. Throughout the course, instruction was stressing the fact that data are numbers collected in a particular context that are studied for a purpose (Rossman, 1996), and that the quality of the conclusions we draw depends on how the data were obtained. When, for example, students were examining graphs, the instructor would point to them that observed patterns in the data depend to a great extent on how the data were obtained and that they might be misleading if data collection had not been properly done. When discussing inferential methods, he would stress that the validity of inferences drawn is based on the assumption of probability-based data production. With regards to the inferential advantages of a larger sample size, students were given several examples of situations in which there was bias in the sample selection process and/or the measurement system in order to realize that, in such situations, increasing the sample size would probably not lead to more valid conclusions. The issue of data production has been stressed by statistics educators as a critical part of the needed reforms in introductory statistics instruction (e.g. Moore, 1997; Garfield, 1995; Hogg, 1992). It was emphasized at each possible chance during the course.

5. FINDINGS OF END-OF-COURSE ASSESSMENT

In our previous study of PACE and other statistics students (Meletiou et al., 1999), we had witnessed superficial knowledge of statistical concepts and a tendency to think deterministically and seek causes behind ephemeral patterns in the data. In contrast, students in the current study were found to have much better understanding of the relationship between chance and regularity, to reason much more effectively about the stochastic.

The results of the assessment at the end of the course were very encouraging. They indicated that students in the current study recognized that, in addition to knowing about the center of a distribution, one also needs information about its spread. All of the students acknowledged that whenever comparing measures of center one always ought to also take "the overall spread" into account. Most of them had good understanding of the meaning and purpose of the different numerical summaries they had learned in class. It was for example very impressive that, in contrast to our previous research findings where almost no student really understood what standard deviation means, the majority of students in this study had a pretty good grasp of the meaning and use of standard deviation. They explained that one calculates standard deviation to get *"information about the distribution between the scores, the distance…outside the center"*, *"to figure out the deviation, the average deviation of the scores from the mean."* They also all knew that, in addition to standard deviation, measures such as *"the range of the box"* (interquartile range), also give us information about the spread of a dataset. They understood that mean and standard deviation are not the only two measures that define the shape of the distribution. And although still having some difficulties with constructing and interpreting graphs, these students' understanding was much more sophisticated than that of students in the previous studies we had conducted.

Several tasks were given to students at the end of the course to investigate their understanding of the relationship between sampling variation and sampling representativeness. One of the tasks was one of three versions of an assessment item used by Shaughnessy, Watson, Moritz, & Reading (1999) in an exploratory study on student understanding of variation. A total of 235 primary students (grades 4 to 6) and 89 secondary students (grades 9 and 12) from the US and Australia had participated in that study. This version, which the authors called the CHOICE version, was given to a total of 105 students. Students had to choose, among five possible lists, the one that is most likely to represent the number of reds drawn by five students who each drew 10 candies out of a bowl of 100 wrapped candies that had 50 reds (see *Candies Question* in the Appendix).

In analyzing student responses, the same procedure as that of Shaughnessy et al. (1999) was followed. Responses were scaled both on the basis of their use of centers and of their use of spreads. For the "centering" scale, student responses were categorized as LOW, FIVE or HIGH. Responses for which the mean number x of reds was $4 < x < 6$, were classified as FIVE, otherwise they were classified as either LOW or HIGH. For the spread scale, the following categories were used: NARROW, REASONABLE, and WIDE. Responses in which the range was 7 or more are pretty unlikely to occur and were classified as WIDE, and so are those with ranges less than or equal to 1, which were classified as NARROW. Ranges between 2 and 7 were considered REASONABLE. According to the scale, responses can be classified as follows:

Table 1. Classification of Student Responses

Response	*Center Classification*	*Spread Classification*
A: {8,9,7,10,9}	HIGH	REASONABLE
B: {3,7,5,8,5}	FIVE	REASONABLE
C: {5,5,5,5,5}	FIVE	NARROW
D:{2,4,3,4,3}	LOW	REASONABLE
E: {3,0,9,2,8}	FIVE	WIDE

The expected number of reds drawn by each student is 5. Under an ideal situation, response C would have the highest probability of occuring. However, in answering this question, students ought to take into account not only the center of the distribution, but sampling variation also. Based on a binomial model of the problem, one standard deviation away from the expected number of reds in a sample of ten candies is 1.581. Although this question considers only five students, a fairly small sample, a reasonable spread to expect in the outcomes would be between 0 and 3, but, very unlikely to always be exactly zero. The best response is therefore B, which is centered on 5 and is also a reasonable response in terms of spread.

Table 2 compares the performance of students in the Shaughnessy et al. (1999) study, with that of students in this study.

Table 2. Results of Current Study vs. Results of Shaughnessy et al. (1999) study

Classification	*Shaughnessy et al. Study* %	*Current Study* %
Center		
Low	13	0
Five	56	100
High	27	0
Unclear	4	0
Spread		
Narrow	16	12
Reasonable	76	88
Wide	4	0
Unclear	4	0
Correct		
Five, Reasonable	35	88

Students in the current study did better in estimating both center and spread. Instruction seems to have been particularly effective in helping them take both spread and center into account. Whereas in the Shaughnessy et al. (1999) study, only 35% of the students belonged to the FIVE, REASONABLE category (i.e. chose response B), in this study the percentage of students belonging to this category was 88%. Since helping students move away from "uni-dimensional" thinking and be able to integrate center and variability into their analyses and predictions should be one of the main goals of statistics instruction, the results are encouraging. It is an important accomplishment of instruction given that in the Shaughnessy et al. (1999) study, although most students' measures of spread were reasonable, they predicted values that were either too high or too low on the centering scale. Also, in that study, the use of words explicitly referring to variation was quite rare. In contrast, students in the current study gave explanations that indicated they were integrating ideas of spread and center:

"Because 50% of the candies are red, the handfuls should be close to 5 reds each time so B. Not C because it's random, there is a margin of error".

"Because they all range around 5 per pick, as would a sample with 50% reds. The others seem too far away or impossible, like C".

"Because the average that would be expected should be 5 with some variation above and below the expected value".

"It's unlikely with .50 probability of reds that anyone got 0 or 10 or straight 5's. There are .50 reds and so we would expect to see more of those but this is a random sample and thus there should be some variability. We expect to have some below 5 and some above. B shows that".

" It's all about variance, but "central-tendency" must always be counted".

Of course, one should take into account the fact that students in the Shaughnessy et al. (1999) study were primary and high school students, whereas the present study deals with college students having completed a statistics course. Nonetheless, in that study, while a steady growth across grade levels on the "centering" criterion from 34% at Grade 4 to 83% at Grade 12 was observed, there was "an apparent oscillation on the variability criterion across grade levels." The researchers noted "a high spike occurring in our Grade 9 students, and then a drop off at Grade 12, for both the REASONABLE and the FIVE, REASONABLE categories." They speculated that the steady growth in the FIVE category is an indication of the considerable focus of school curricula on "center". A possible explanation they saw for the oscillation at Grades 9 and 12 is that Grade 9 students participating in their study were spending more time on data analysis than the higher level mathematics students, whose school work on probabilities might have interfered with their reasoning about this problem. The exposure to probability did not seem to interfere with the reasoning of the students in the current study.

Table 3. Results of end-of-course assessment compared on "College Interviewer" Question

Response	*Beginning %*	*End %*
A: The worker seems to interview more women than men. There could be several reasons for this. Perhaps women are more willing to talk about their opinions. Or, maybe the worker goes to areas of campus where there are more women than men. Either way, the worker is likely to interview more women than men out of the next 20 students.	16	6
B: Since half of the students on this campus are men and half are women, you would expect a 50/50 split between the number of men and women the worker interviewed. Since there tended to be more women than men so far, I expect the opposite trend to start happening. Out of the next 20 students the worker interviews, there will probably be more men than women so that things start to balance out.	32	24
C: Half the students on this campus are men and half are women. That means that the worker has a 50/50 chance of interviewing a man or a women. It should not matter how many men or women the worker has interviewed so far. Out of the next 20 students interviewed, about half should be men and half women.	35	64
D: So far, the trend seems to be more women to be interviewed than men. Out of the next 20 students the worker interviews, I would expect the same thing to happen. The worker will probably interview more women than men out of the next 20 students.	17	6

At the completion of the course, students were given again the same question given at the beginning which was describing how a worker of a student organization went about conducting a survey at a certain college where half the student population were women and half were men, and was asking students to predict whether there would be more women or more men in the next 20 students the worker interviews given that out of the last 20 students interviewed 13 were women and 7 were men (see *College Students' Interview Question* in the Appendix). Student performance at the end of the course compared to their performance on the same question prior to instruction, is another example of the positive effect of instruction in helping improve students' probabilistic reasoning (see Table 3).

Sixty-four percent of the students at the end of the course, compared to 35% of them in the pre-assessment, realized that due to the independence of random samples, one should still expect that, out of the next 20 students interviewed, about half should be men and half women (choice C). There were still a considerable proportion of students (24%) employing the balancing strategy and arguing that they expected the opposite trend to start happening, but in general, students' performance was much improved.

Students were, at the completion of the course, much less prompt compared to the beginning, to assume that short-term fluctuations in the data must be causal and develop causal explanations. When for example we reminded students in the primary group, during the end-of-course interview, of the *Birth Defects* question (see

Appendix), they all had a different opinion about this situation compared to the beginning of the course. They stressed that the number of children is so small that one cannot give causal explanations. Tim, for example, who had argued in the beginning of the course that the probability of giving birth to a child with a missing limb correlates with where one lives, now pointed out: *"There is not enough information to…it's not a big enough…it's only 7 people. It's not enough number of subjects to understand what's going on."* Lucas remembered that when he first saw this question he was thinking, *"there might be something wrong with the sanitation or the water, something like that."* Now though, he realized that *"this is only one year so, last year or the year before, they could have had 3 down here and 2 up there and 2 over here. You have to look at many years to see what's happening."* He added: *"That's why I liked this class. I learned to look at the big picture of things."*

The efforts of instruction to present statistical thinking as a balance between deterministic and stochastical reasoning succeeded in helping students move away from "uni-dimensional" thinking and integrate center and variation into their analyses and predictions. Although not totally letting go of their deterministic mindset, students were much more willing to interpret situations using a combination of stochastic and deterministic reasoning. The course increased significantly student awareness of sampling variation and its effects. Instruction managed to get across to them the idea that "thinking about variability is the main message of statistics" (Smith, 1999, p. 249).

6. IMPLICATIONS FOR INSTRUCTION AND RESEARCH

The expectation that students will transfer the understanding obtained through coins, dice, and games of chance to everyday contexts seems to be a naïve assumption, as previous research studies (Pfannkuch and Brown, 1996), as well as student assessment at the outset of this study, have indicated. The skills required to understand variation in random devices are very different from the skills required to understand variation in real-life contexts. Instruction needs to take into account the great variety of prior beliefs, conceptions, and interpretations that students bring to each situation.

This study investigated how a teaching pedagogy focusing on data and variation centered on students' experiences could promote understanding of the stochastic nature of statistical concepts. Unlike more typical approaches, which attempt to develop probabilistic reasoning through standard probability tasks, the model employed in the study based instruction on realistic contexts directly connected to students' experience. Although the conclusions drawn from the study focus on a single group of students for a short duration of time and thus could not be used to make general inferences about the population of students taking introductory statistics courses, assessment of student learning at the end of the course suggests that the teaching pedagogy implemented in this study did help students improve their reasoning about the stochastic and might deserve further investigation. The simultaneous focus of this model on variation and on the process of statistical investigation seems to be a promising alternative to more conventional instruction, where the linear and consecutive structure of the course comes in sharp contrast with the complex nature of stochastical knowledge.

There is a lot still to be learned regarding students' reasoning about variation. More research needs to be carried out to investigate intuitions about variation of students of different age groups and different backgrounds. Through conducting this study, we have come to realize that assessment of thinking about variation is heavily reliant upon both the types of assessment tasks employed and the context in which the tasks are situated. Students come to a situation with a wide range of skills and knowledge and offer responses that are difficult to anticipate (Cohen and Checile, 1997). We, similarly to Pfannkuch and Brown (1996), documented students' neglect of probabilistic thinking when interpreting certain real world phenomena and their tendency to come up with causal explanations for short-term fluctuations in data that could be easily explained by natural variation. Conversely, we found that in certain occasions a students' response might be erroneous due to poor knowledge not of statistical content but of the context of the situation, or due to misunderstandings about what the question is asking (Jolliffe, 1994). The wealth of information that emerged out of this study shows the advantages of using a variety of assessment tasks when investigating students' reasoning about variation. To get a more complete profile of student intuitions, future research as well as instruction on variation ought to also use a variety of assessment tasks and multiple-forms of assessment that complement each other. Additionally, the instructional materials and assessment items used by educators and researchers need to be opened up for scrutiny so that they can be gradually improved by the statistics education community.

Interested readers could, by contacting us, be provided with all the teaching materials and assessment tasks employed in the study. In addition, by keeping all the data collected in a well-organized and retrievable form, we can easily make them available to any researchers challenging the findings and seeking to reanalyze the data.

APPENDIX

Gummy Bears Question

Suppose you took your little nephew on an Easter parade. At the parade, the "Easter Bunny" handed out packets of Gummy Bears to all of the students. Each packet had 6 Gummy Bears in it. To make up the packets, the Easter Bunny took 2 million green Gummy Bears and 1 million red Gummy Bears, put them in a very big barrel and mixed them up from night until morning. Then he spent the next few hours making up the packets of six Gummy Bears. He did this by grabbing a handful of Gummy Bears and filling as many packets as he could. Then he reached into the barrel and took another handful, and so on, until all the packets were filled with 6 Gummy Bears.

a) When you get home from the parade, you open up your packet. How many green Gummy Bears do you think might be in your packet? Can you explain how you got that?
b) Do you think all the students got *n* greens, where *n* is the number of Gummy Bears you gave in part (a)? Can you explain why?
c) If you could look at the packets of 100 students, how many students do you think got n greens?
d) Remember that the Easter Bunny was starting with 2 million greens and one million reds. Did he run out of one color long before the other when he was filling the bags or did they both last until near the end? Why?

College Students' Interview Question

Circle the best answer to the following problem:

At a nearby college, half the students are women and half are men. A worker for a student organization wants to interview students on their views about recent changes in the federal government's funding of financial aid. The worker wants to get a good representation of the students, and goes to as many different areas on campus as possible. Three or four students are interviewed at each place the worker visits. Out of the last 20 students interviewed, 13 were women and 7 were men. Now, you do not know what time of day it is, to which part of campus the worker has already gone, or where the worker is going next. Out of the next 20 students the worker interviews, do you think more will be women or men?

a) The worker seems to interview more women than men. There could be several reasons for this. Perhaps women are more willing to talk about their opinions. Or, maybe the worker goes to areas of campus where there are more women than men. Either way, the worker is likely to interview more women than men out of the next 20 students.
b) Since half of the students on this campus are men and half are women, you would expect a 50/50 split between the number of men and women the worker interviewed. Since there tended to be more women than men so far, I expect the opposite trend to start happening. Out of the next 20 students the worker interviews, there will probably be more men than women so that things start to balance out.
a) Half the students on this campus are men and half are women. That means that the worker has a 50/50 chance of interviewing a man or a women. It should not matter how many men or women the worker has interviewed so far. Out of the next 20 students interviewed, about half should be men and half women.
b) So far, the trend seems to be more women to be interviewed than men. Out of the next 20 students the worker interviews, I would expect the same thing to happen. The worker will probably interview more women than men out of the next 20 students.

Birth Defects Question

Every year in New Zealand approximately seven children are born with a limb missing. Last year the children born with this abnormality were located in New Zealand as shown on the map.

Figure 1

What do you think? (In New Zealand, it is common knowledge that one-third of the population lives in the top region and one-sixth of the population in each of the other regions.)

Die Toss Question
A fair die is tossed 7 times resulting in the outcome 3,3,3,4,4,5,5 (order is unimportant). What do you think of these results?

Child Psychologist Question
A child psychologist is engaged in studying which of two toys infants will prefer to play with. Of the first five infants studied, four have shown a preference for this toy. The psychologist concludes that most infants will show a preference for this toy. Do you think the psychologist has drawn a valid conclusion?

Candies Question
A bowl has 100 wrapped hard candies in it. 20 are yellow, 50 are red, and 30 are blue. They are well mixed up in the bowl. Jenny pulls out a handful of 10 candies, counts the number of reds, and tells her teacher. The teacher writes the number of red candies on a list. Then, Jenny puts the candies back into the bowl, and mixes them all up again. Four of Jenny's classmates, Jack, Julie, Jason, and Jerry do the same thing. They each pick ten candies, count the reds, and the teacher writes down the number of reds. Then they put the candies back and mix them up again each time.

I think the teacher's list for the number of reds is most likely to be (please circle one):

a) 8,9,7,10,9
b) 3,7,5,8,5
c) 5,5,5,5,5
d) 2,4,3,4,3
e) 3,0,9,2,8

Explain your reasoning.

REFERENCES

Ballman, K. (1997). Greater emphasis on variation in an introductory statistics course. *Journal of Statistics Education, 5*(2).

Biehler, R. (1994). Probabilistic thinking, statistical reasoning, and the search for causes: Do we need a probabilistic revolution after we have taught data analysis? In J. B. Garfield (Ed.), *Research papers from the Fourth International Conference on Teaching Statistics*. Minneapolis: The International Study Group for Research on Learning Probability and Statistics.

Biehler, R. (1997). Students' difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 176-197). Voorburg, The Netherlands: International Statistical Institute.

Biehler, R. (1999). Discussion: Learning to think statistically and to cope with variation. *International Statistical Review, 67*(3), 259-262.

Borovcnik, M. & Peard, R. (1996). Probability. In A.J. Bishop (Ed.), *international handbook of mathematics education* (pp. 239-287). Netherlands: Kluwer Academic Publishers.

Chance, B. L. & Rossman, A. J. (2001). Sequencing topics in introductory statistics: A debate on what to teach when. *The American Statistician*, *55* (2), 140-144.

Cobb, G. (1992). Teaching statistics. In L.A. Steen (Ed.), *Heeding the call for change: suggestions for curricular action* [MAA Notes No. 22] (pp. 3-33). Washington, DC: The Mathematical Association of America.

Cohen, S. & Chechile, R. A. (1997). Overview of ConStatS and the ConStatS assessment. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 110-119). Voorburg, The Netherlands: International Statistical Institute.

Confrey, J. (1996, April). Strengthening elementary education through a splitting approach as preparation for reform algebra. Presented at the *annual meeting of the American Educational Research Association*, New York, NY.

Erickson, T. (2000). *Data in depth. exploring mathematics with Fathom*. Emeryville, CA: Key Curriculum Press.

Gal, I. & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal and J. B. Garfield (Eds.), *The assessment challenge in statistics education.* Burke, VA: IOS Press.

Garfield, J. & delMas, R. (1990). Exploring the stability of students' conceptions of probability. In J. B. Garfield (Ed.), *Research papers from the Third International Conference on Teaching Statistics.* University of Otago, Dunedin, New Zealand.

Garfield, J. B. & delMas, R. C. (1994). Students' informal and formal understanding of variation. In J. B. Garfield (Ed.), *Research papers from the Fourth International Conference on Teaching Statistics.* Minneapolis: The International Study Group for Research on Learning Probability and Statistics.

Garfield, J. B. (1995). How students learn statistics. *International Statistical Review, 63*(1), 25-34.

Glencross, M. J. & Binyavanga, K. W. (1997). The role of technology in statistics education: A view from a developing region. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 301-308). Voorburg, The Netherlands: International Statistical Institute.

Hawkins, A. (1997). Myth-conceptions. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. vii-viii). Voorburg, The Netherlands: International Statistical Institute.

Hoerl, R., Hahn, G. & Doganaksoy, N. (1997). Discussion: Let's stop squandering our most strategic weapon. *International Statistical Review, 65*(2), 147-153.

Hogg, R.V. (1992). Towards lean and lively courses in statistics. In F. Gordon and S. Gordon (Eds.), *Statistics in the twenty-first century* [MAA Notes No. 26] (pp. 3-13). Washington, DC: Mathematical Association of America.

Jolliffe, F. (1994). Why ask why? In J. Garfield (Ed.), *Research papers from the Fourth International Conference on Teaching Statistics*. Minneapolis: The International Study Group for Research on Learning Probability and Statistics.

Lee, C. (1997). Promoting active learning in an introductory statistics course using the PACE strategy. *Proceedings of the 6th International Symposium on Mathematics Education* (pp. 199-206). Mexico City.

Lee, C. (2000). Computer-assisted approach for teaching statistical concepts. *Journal of Computers in the Schools, 16* (1), 193-208.

Meletiou, M., Lee, C. M. & Myers, M. (1999). The role of technology in the introductory statistics classroom: Reality and potential. *Proceedings of the International Conference on Mathematics/ Science Education and Technology.* San Antonio, Texas.

Moore, D. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123-165.

Pfannkuch, M. (1997). Statistical thinking: One statistician's perspective. In J. Garfield & J. Truran (Eds.), *Research papers on stochastics education* (pp. 171-178).

Pfannkuch, M. and Brown, C. M. (1996). Building on and challenging students' intuitions about probability: Can we improve undergraduate learning? *Journal of Statistics Education, 4*(1).

Rossman, A. J. (1996). *Workshop statistics: Discovery with data.* New York: Springer-Verlag.

Rubin, A., Bruce, B. & Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. In J. B. Garfield (Ed.), *Research papers from the Third International Conference on Teaching Statistics.* University of Otago, Dunedin, New Zealand.

Shaughnessy, J. M., Watson, J., Moritz, J. & Reading, C. (1999). School mathematics students' acknowledgment of statistical variation. For the NCTM Research Precession Symposium: *There's More to Life than Centers*. Paper presented at the 77th Annual NCTM Conference, San Francisco, California.

Smith, T. M. F. (1999). Discussion. *International Statistical Review, 67*(3), 248-250.

Snee, R. D. (1999). Discussion: Development and use of statistical thinking: A new era. *International Statistical Review*, *67*(3), 255-258.

Steinbring, H. (1990). The use of chance-concept in everyday teaching - Aspects of a socially constituted epistemology of mathematical knowledge. In J. B. Garfield (Ed.), *Research papers from the Third International Conference on Teaching Statistics.* University of Otago, Dunedin, New Zealand.

Tversky, A. & Gilovich, T. (1989). The cold facts about the "hot hand" in basketball. *Chance, 2(1),* 16-21.

Wild, C. J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223-265.

Wilensky, U. (1993). *Connected mathematics - Building concrete relationships with mathematical knowledge.* Ph.D. Thesis, Massachusetts Institute of Technology.

MARIA MELETIOU-MAVROTHERIS
11 Achilleos St.
2370 Ayios Dometios
Nicosia, Cyprus