Journal of Interactive Learning Research. VOL. 11 Number 1 2000

Learning Probability Through the Use of A Collaborative, Inquiry-Based Simulation Environment

Phil Vahey
SRI International
333 Ravenswood Ave
Menlo Park, CA 94025
philip.vahey@sri.com

Noel Enyedy & Bernard Gifford
University of California, Berkeley

Abstract

In this paper we report on the Probability Inquiry Environment (PIE), which facilitates the development of probabilistic reasoning by making available collaborative inquiry activities and student-controlled simulations. These activities guide middle school students toward a deeper understanding of probability, a domain that is becoming increasingly important in the K-12 mathematics curricula of the United States but which is notoriously difficult to learn. A study is described in which middle school students who participated in the PIE curriculum significantly outperformed students who participated in the school's traditional probability curriculum. We posit that this difference is due to the PIE curriculum fostering student collaboration as the students employ their existing intuitions as building-blocks for formal knowledge. This provides evidence that a productive learning environment should not be based solely upon the logical structure of the target domain, but should also account for both students' intuitive conceptions of a domain, as well as aspects of social interaction that shape students' experiences. We then show the importance of intuitions and social interaction by analyzing a case study in which students articulated and revised their initial understandings of probability as they interacted with PIE—collaboratively making predictions, evaluating data, and interpreting representations.

Every day, people are called upon to make decisions based on statistical and probabilistic information. Public opinion polls, advertising claims, medical risks, and weather reports are just a few of the everyday activities that draw on an understanding of probability. In addition, probability is applicable to many academic disciplines outside of mathematics, as it is routinely used in the professional activities of biologists, geneticists, and psychologists alike.

Recently, teaching and learning probability at the middle school level has been recognized as an important strand in the mathematics standards at both the national (NCTM, 1989) and state levels (California Department of Education, 1992; 1999). These standards call for students to be able to make predictions based on theoretical probabilities and empirical data, model probabilistic situations by representing all the possible outcomes for compound events and understand and appreciate the pervasive use of probability in the real world.

Designing new and effective methods of probability instruction, however, presents a difficult pedagogical challenge. It is well recognized that both lay people and professional scientists often make mistakes when reasoning probabilistically (Konold et al., 1993; Tversky & Kahnemann, 1982). This paper will discuss one instructional environment that was designed to help middle school students come to a more normative understanding of probability theory.

An Overview of the Paper

We begin by describing our instructional approach, in which we advocate student construction and appropriation of concepts by making available student-controlled interactive simulations, dynamic-representations, and contextualized learning activities. Next, we discuss students' understanding of probability theory and then go on to describe the Probability

Inquiry Environment (PIE). PIE consists of a number of carefully crafted activities designed to encourage students to articulate and refine their understandings of probability while using the tools and practices of mathematics and probability. This is followed by a description of a study in which PIE was implemented in two seventh grade public school classrooms. We then present detailed qualitative data from one pair of students as they investigate one of the computer activities, and show how the features of PIE contributed to student conversations and student understandings. We conclude with generalizations from our work, and recommendations for others creating collaborative environments for active knowledge construction.

Our Instructional Approach

There has been a recent movement to use technology-mediated inquiry (i.e., using computer software tools in helping students to make predictions about a situation, gather empirical data to test the predictions, and then compare the initial predictions to the data and drawing conclusions) as a method for teaching science (White, 1993a) and mathematics (Lampert, 1995; Richards, 1991). This movement has been particularly strong in the area of stochastics (Shaughnessy, 1992; Konold, 1991; Lajoie, et al. 1996; Hancock, 1992; NCTM, 1989; Newman, et al. 1987; Rosebery & Rubin, 1989). For guidance in creating productive activities, we build from several areas of research:

(i) existing student understandings can play a productive role in the development of expertise, and instead of viewing non-normative understandings as deficiencies that must be replaced, we should work to build on what is productive in these understandings (see Smith, et al., 1993; see also diSessa, 1988; Lampert, 1986; Minstrell, 1989; White, 1993a, 1993b),

(ii) students construct meaning and understanding while participating in a social context (Brown & Campione, 1996; Collins et al. 1989; Hall & Rubin, 1998), and student learning is facilitated when students make their existing ideas explicit in order to evaluate these ideas with respect to findings in the domain and the ideas of others (Konold, 1991; Linn, 1995).

(iii) Educators should design activities that provide intermediate tools and models that maintain fidelity with the relevant aspects of expert analysis while students participate in authentic tasks (Gordin et al. 1994; White, 1993a, 1993b),

(iv) students should engage in such activities *before* the concepts, terms, and definitions of the domain are introduced to the class. In this way students have the opportunity to actively engage in the domain, to appreciate the complexity of the domain, and to investigate their own understanding of the domain. Then, in class discussions, the concepts, terms, and definitions are not “out of the blue”, but build on students’ experiences and enrich their understanding of a situation they consider both problematic and relevant (Enyedy, Vahey & Gifford, 1998).

In designing an instructional environment based on this stance, we chose to create a progression of activity sets, where each activity set is interesting and highlights an important aspect of the domain to be studied, and each activity set extends the findings from the previous activities (cf. White 1993a, 1993b). In this environment, students are engaged in authentic activities from the beginning, they can create understandings that are consistent with some aspects of their existing reasoning, and, as they are systematically exposed to aspects of the domain, their understandings can be revised, and need not be completely overhauled. This allows students to expand the understandings that are most appropriate for specific situation, and allows students to recognize which understandings are inappropriate for specific situations. We now describe the overall structure of the Probability Inquiry Environment.

The Structure of PIE

Each of the activity sets in the PIE three week curriculum have a computer-based and hands-on collaborative inquiry activity, in which the students work in pairs, followed by a whole class discussion designed to help students understand and formalize the result of the activity. Throughout PIE students were asked to determine if a game of chance was “fair”. Evaluating the fairness of games is an activity middle school students consider authentic and legitimate, and students spontaneously invoke probabilistic reasoning when asked if certain games of chance are fair (Vahey, 1996). This finding is consistent with findings from the literature on moral development (Thorkildesen, 1995), as well as the stochastics education literature. For instance, students have been shown to spontaneously bring up ideas of fairness when asked to survey a population (Lajoie et al., 1995), or when asked to collect data that will be used to measure the relative merits of different stereo systems (Hancock et al., 1992). Investigating fairness is a context consistent with the notion of building from students’ understandings while having them engage in tasks that are meaningful to them.

During the computer-based activities students are engaged in an explicit process of inquiry. The inquiry cycle used in PIE consists of six steps: Rules, Try, Predict, Play, Conclude, and Principles. In *Rules,* the software shows the students an animated introduction to the current game. In *Try,* the students get a chance to experiment with the representations and controls of the simulation. This was done to allow the students some amount of familiarity with the environment before asking them to make predictions.

In *Predict,* we created questions that highlight aspects of the game that are particularly salient to a standard understanding of probability. These questions were designed to have students articulate cohesive explanations in which they have some commitment, consider alternate perspectives, and be sensitized to future events that may support the prediction or call that prediction into question. PIE also provides interactive representations, such as draggable bar charts, that provide students with a physical object that can be discussed, leading to more productive conversations than are otherwise possible (Enyedy et al., 1997).

In *Play,* the software simulates the playing of the game, providing several resources and capabilities to facilitate productive collaboration. Some of the most important of these resources are: an animated probability tree that highlights the current state of the point; a representation that shows how each team scores; where appropriate, a bar chart that can toggle between showing the distribution across each of the outcomes or the points for each team; and a table which shows this total numerically. Each of these is a shared representational resource that can be used by the students to jointly construct an explanation of the current situation, and each of these provides the students with intermediate tools that bring the analysis of experts into the realm of the students (Enyedy et al., 1997). Additionally, to help guide the students toward the salient aspects of the simulation, PIE provides Observation Questions that stop the simulation at pre-determined events and ask the students to reflect on their current understandings and the current status of the simulation

In *Conclude*, the students compare their predictions to the data from the simulation. Students are again asked their ideas from multiple viewpoints, but now the environment also provides data from their simulations as another resource for constructing understanding. Finally, in *Principles*, the environment scaffolds the students to jointly articulate what they can generalize from the activity. In *Principles*, students are no longer asked about the fairness of the games that were just played, but instead are asked to form generalizations that transcend the particular game.

These activities are then followed by real-world activities in which students flip coins and roll dice as they investigate probability without using the computer simulations. In the full-class discussions, each pair then relates their findings from both the computer-based and real-world activities, and discusses how to best determine if the game is fair. In these discussions students raise issues such as the relative utility of the different representations, methods they used for deciding if the game is fair, and how to formalize their knowledge in the language of mathematics (Enyedy et al., 1998).

This process is designed to scaffold students in the construction of a normative view of the domain of probability. Before discussing the specific PIE activities, let us move to a discussion of the literature on students' probabilistic reasoning: how can we use the research on people's probabilistic reasoning in conjunction with our instructional approach to provide a solid foundation for the specific activities in the Probability Inquiry Environment?

The Literature on Students' Understanding of Probability

Due to the recent interest in teaching probability in middle school and high school, there have been a number of reviews of the literature on students' probabilistic reasoning, and we refer the reader to these articles for a detailed review of the current state of the literature (Konold, 1991; Lajoie, et al., 1995; Metz, 1998; Shaughnessy, 1992). Interestingly, each of these point to the need for a more detailed understanding of young students' probabilistic reasoning. In this section we provide a telegraphic summary of some of the more influential findings on students' understandings of probability, and provide a starting point for a more detailed understanding of young students' probabilistic reasoning.

Contradictions abound in the research on probabilistic reasoning, with some research showing that people are inherently poor at reasoning probabilistically, and other research showing that people are inherently good at reasoning probabilistically. For example, in the literature on adults' probabilistic reasoning, Kahneman and Tversky's seminal work asserts that, not only are people poor at probabilistic reasoning, but this poor reasoning is innate and highly resistant to change. They state that much of our probabilistic reasoning can be described by *cognitive illusions* (Kahneman & Tversky, 1996; Tversky & Kahneman, 1982), described as *heuristics* and *biases*. Heuristics are shortcuts that people employ when making decisions under uncertainty, and although these heuristics are often productive, employing these heuristics can lead to systematic errors, which they call biases.

Directly rebutting this conclusion, others assert that people are inherently *good* at probabilistic reasoning, and people do *not* suffer from cognitive illusions. These researchers show that people are sensitive to minor changes in the wordings of questions. By varying how the situation is presented, and by performing a different analysis on the responses, they show that people can reason about complex issues in ways that are consistent with probability theory, contradicting the claim that people are simply falling prey to cognitive illusions (Cosmides & Tooby, 1996; Gigerenzer, 1996). And, others present a more moderate view, stating that people may at times suffer from cognitive illusions, but this is only one of a number of reasoning strategies that people may use (Konold et al., 1993).

Contradictions also abound in the literature on children's understanding of probability. For example, Piaget & Inhelder (1975) state that children can only understand probability once they reach the age of formal operations, because it is only at the age of formal operations that children can differentiate the necessary from the possible, and so can reason without imposing causality on all phenomena. Then, once students reach the stage of formal operations, they spontaneously begin to understand important aspects of probability theory, such as the outcome space and the indeterminacy of events. This contrasts with the view of Fischbein (1975; Fischbein et al., 1991), who disagrees on both points. Fischbein states that children *can* understand probability theory before reaching the age of formal operations. Further, once children reach the age of formal operations, they do *not* suddenly appreciate the constructs of probability theory. Instead, students construct understandings that are sometimes consistent with probability theory and sometimes inconsistent with probability theory, and it is the role of instruction to guide students as they construct their understandings.

Given this chaotic state of the literature, how can we hope to make sense of students' probabilistic reasoning in a way that can aid us in the creation of instructional environments? In the next section we will discuss an approach where people are not judged at being "good" or "poor" at probabilistic reasoning. Instead, we attempt to understand how student reasoning is consistent with and inconsistent with normative probabilistic reasoning so that we can create inquiry activities that build upon students' prior understandings.

Recent Models of Probabilistic Reasoning

We are developing a detailed model of students' probabilistic reasoning that considers their performance along several distinct, although related, dimensions (cf. Horvath & Lehrer 1998; Vahey, et al., 1997). We then move on to compare student reasoning to normative reasoning in the domain of probability, finding areas in which students reason in a manner consistent with probability theory, and finding where their reasoning differs from probability theory. This research acts as a guide in creating instructional activities that help students to build upon their existing understandings as they come to a more normative understanding of probability theory.

The dimensions in the model dynamically emerged from verbal analysis of students engaging with a prototype version of the Probability Inquiry Environment (Vahey et al., 1997). The different ideas raised by students in this pilot study corresponded to four aspects of probability theory: randomness, the outcome space, the probability distribution, and the validity of data. By *randomness* we mean the non-determinability of an event, and what predictions one can make of this non-determinable mechanism. By *outcome space* we mean the set of all possible outcomes. By *probability distribution* we mean the probabilities assigned to the different outcomes in the outcome space. By the *validity of data* we mean how data can be used in reasoning, including when and how data is considered to be useful. Although each of these aspects was just introduced as being independent from the others, these aspects are in fact related, both in formal probability theory, and sometimes in students' reasoning. When looked at in this light, we see that much of student reasoning can be used as building blocks toward normative probabilistic reasoning.

When discussing *randomness*, for example, almost all students understood that variation is to be expected from trial to trial. However, some students also believe randomness means that nothing at all can be predicted about future events, even to the point of excluding data (as will be discussed vis-a-vis data validity). It is important to note, though, that this expected lack of predictability could be viewed as a normative understanding of randomness that has been over-extended. That is, it is completely reasonable to say that one cannot guess the *next* outcome of a random process, and so students are in many ways "correct" when they apply this idea to short-run data. It is only when extending this idea to long-run data that this reasoning becomes non-normative.

This research has also shown that students invoke ideas of *outcome space* and *probability distribution* when reasoning about probabilistic situations, but it is in using the outcome space and the probability distribution that students deviate most from normative probabilistic reasoning. Although students often do apply a notion of "the things that can happen", students have great difficulty in fully enumerating an outcome space, in differentiating between different outcomes in the outcome space, and in seeing the relevance of the outcome space in certain situations. Students do, however, often assign a probability distribution over different sets of outcomes (although such a probability distribution typically does not conform to the laws of formal probability theory). By building on students' productive ideas we can help students come to a normative view of the outcome space and the probability distribution.

Finally, students have varying ideas about the *validity of data*. Although some students fell prey to the well-documented "myth of small numbers" and believed results that are based on a small sample, others did not believe the data when the data was in conflict with their expectations, even after a large number of trials. This suggests that there is an interaction between people's expectations and the validity that they are willing to attribute to data. This interaction is well known in the science learning community, where it has been shown that there are many possible ways for people to react to data that is conflict with their understandings, and indeed, the acceptance of such data is an uncommon occurrence (Chinn & Brewer, 1993; Gunstone, 1991; Strike & Posner, 1992). This interaction between expectations, results, and explanations can be used as a basis for activities that require the use of data and simulations in testing hypotheses about probabilistic phenomena.

The PIE Activities

We used this framework to determine the nature and order of the specific inquiry activities. We noted that students had many ideas about the validity of data and the importance (or lack thereof) of a large sample. Because students were going to be interacting with data throughout the curriculum, we determined that students should be introduced early in the curriculum to the law of large numbers. Additionally, students have great difficulty in enumerating the outcome space. Because the outcome space is one of the most important constructs in probability theory, we determined that this should be a main goal of instruction. In keeping with our theoretical framework, each activity is designed to highlight one of these aspects of probability in such a way that students can use their existing understandings to make sense of the activity, and each activity is designed to build from the ideas of previous activities. We will provide short descriptions of all of the activities, and only the third PIE activity, the Three Coin Game, will be described in detail. This will give the flavor of all the activities, and this description will serve as background for the in-depth case study to be presented later.

Activity Set 1: The Horse and Bunny Game

In this set of activities students become familiar with a qualitative form of the Law of Large Numbers: although probabilistic phenomena appear irregular and unpredictable in the short term, such phenomena have long term regularities. This was done in an activity set in which students predict the outcome of both a small number of flips of a coin (5) as well as a large number of flips (100).

In order to create a gaming situation that interests the students, the computer activity takes the form of a race: coins are flipped, and if they come up heads, they go to the left side of a balance-scale, and if they are tails, they go to the right side (see Figure 1). After each flip a horse moves if the scale is balanced (defined as between 40% and 60% heads), and a bunny moves if the scale is unbalanced. To help students see how the position of the scale relates to team scoring, we created a set of five “scoring zones”, where each zone is labeled with either a bunny or a horse. To highlight the difference between short term irregularities and long term regularities, the students are asked to predict who would be ahead after five flips, and who would be ahead after 100 flips. As the students run the simulation, the bunny wins more games to 5 (that is, the scale is usually unbalanced), whereas the horse

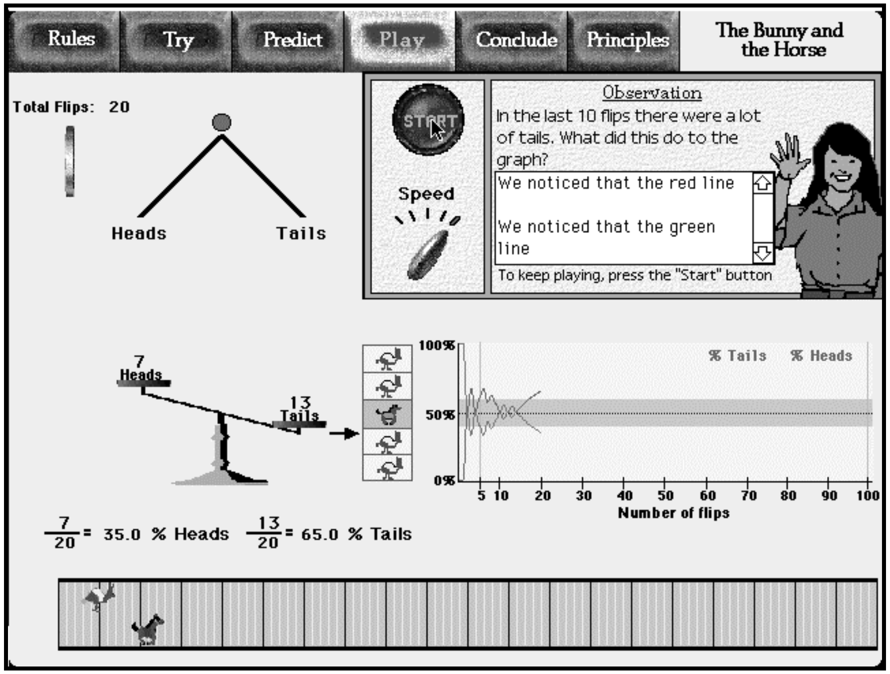

Figure 1: The Horse and Bunny Game (stopped at an Observation Question)

almost always wins the games up to 100 (that is, the scale is usually balanced). Students then decide, in pairs, on the principle that best describes the findings from this activity (for example, “when you flip a coin a few times you cannot tell what is going to happen”).

The students then participate in a hands-on activity called “50-50 Flip Off”. In this activity pairs of students flip a coin five times, each time recording if the flip was heads or tails. Then for this series of five flips, the students mark if it was “About half” heads and tails (defined by the class as 3-2). The students do this 10 times each, so each pair generates fifty coin flips. The students then calculate what percentage of those fifty coins were heads and what percentage were tails. This array of data is then reported to the entire class, with the teacher leading a discussion that compares how many of the five flip series were "about half", to what percentage of the fifty flips were heads and tails. From this the students discover that a higher percentage of the fifty flips than the five flips were "close" to fifty percent.

The class then engages in a full-class discussion where they talk about the different groups' results, and then decide on the principle that best describes the outcome of these activities. By highlighting the relative "unevenness" of the coins after five flips, as compared to the relative "evenness" of the coins near 100 flips, we help the students see the short term irregularity of a random process, as well as the long term regularity of a random process.

Activity Set 2: The Two Penny Game

The students are then introduced to the *outcome space* in an activity set that consists of different on-line and real-world activities. In the on-line PIE activity, the Two-Penny game, two teams are competing for points. The Twins score a point whenever both coins come up the same (heads-heads or tails-tails), and the Jumbles score a point whenever both coins come up differently (heads-tails or tails-heads). A simple "counting strategy" is one way to determine that this game is fair, as each team scores on two out of the four possible outcomes and all outcomes are equally likely. A probability tree that enumerates all the possible outcomes and visually presents the scoring combinations for each team is always on the screen, as is a dynamic bar chart that shows scoring either by each combination of coins, or by each team (see figure 2).

This game is designed to help students see that the number of outcomes is the determining factor in fairness. This is not necessarily obvious, as many people even adults, believe that the Jumbles will win more, because the "mixed" outcomes of HT or TH seem more "random" than HH or TT (cf. Tversky & Kahneman, 1982)).

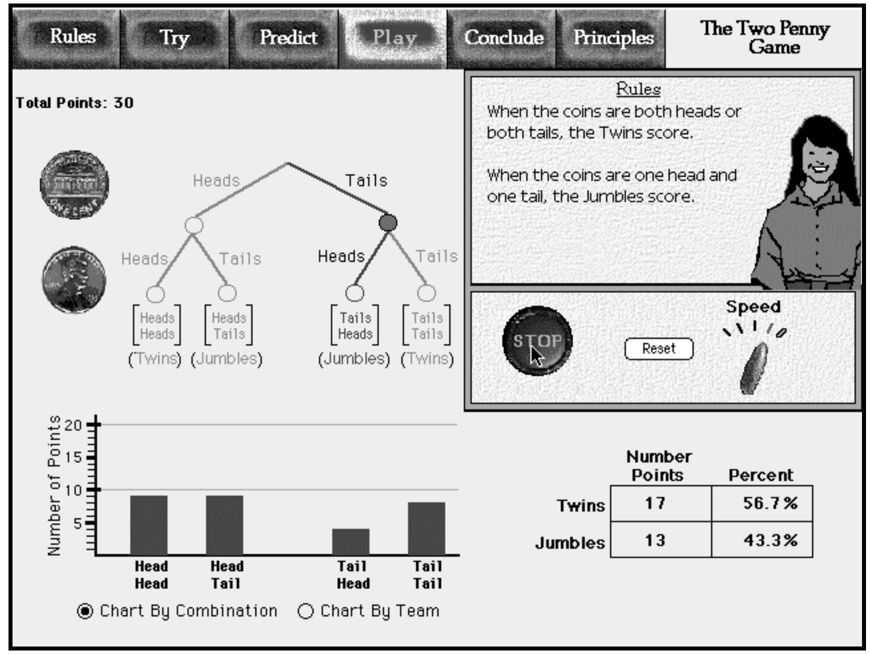

Figure 2: The Two Penny Game

The corresponding real-world activity is similar, but the scoring is different: in this game there are three teams, the Skulls (score when the coins are both heads), the Crossbones (score when the coins are both tails), and the Jumbles (score when one coins is heads and the other is tails). Together the on-line and real-word activities help students bring up a wide range of intuitions. For instance, Jumbles *do* come up more than either head-head or tail-tail in the real-world game, but Jumbles come up the same number of times as head-head and tail-tail (combined) in the on-line game. These activities lead to a class discussion that evaluates the role of the outcome space in deciding if the games are fair.

Activity Set 3: The Three Coin Game (a detailed description)

The Three Coin Game is described in detail, as this description will serve as background for the case study in a later section. Three coins are flipped: Team A scores a point on five of eight possible outcomes, and Team B scores on three of the possible outcomes. Because each outcome is equally likely, this game is unfair in favor of Team A. Each game is played to 200 points. Although this game is based upon the same principles as the two-penny game, this similarity is not necessarily obvious to students.

In the prediction questions students are first asked if this game is fair (Figure 3a). They are next asked to predict how often each outcome will occur, and show this by dragging the eight bars of a bar chart to their expected heights.

Finally, the students are asked to predict the points for each team, again by dragging a set of bar charts, and this time the bars of the chart are linked so that the total for each team always adds up to 200 (Figure 3b). It is important to note that, although a statistician might state that each of these questions are simply different wordings of the same basic question, that is not obvious to most students. It is by asking this series of questions that we help students see the relationships between their ideas, and construct a more coherent understanding.

In each of these prediction questions the students are asked to justify their answers, and as the students choose answers and manipulate bar charts, PIE generates sentence-starters for the students' justification. For example, in Figure 3b, PIE generated the text "We think Team B will score more points because".

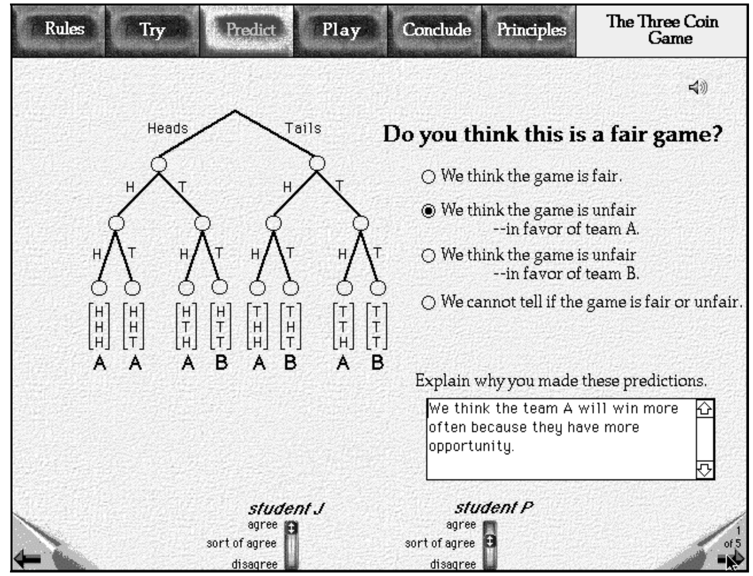

Figure 3a: Three Coin Game Prediction 1

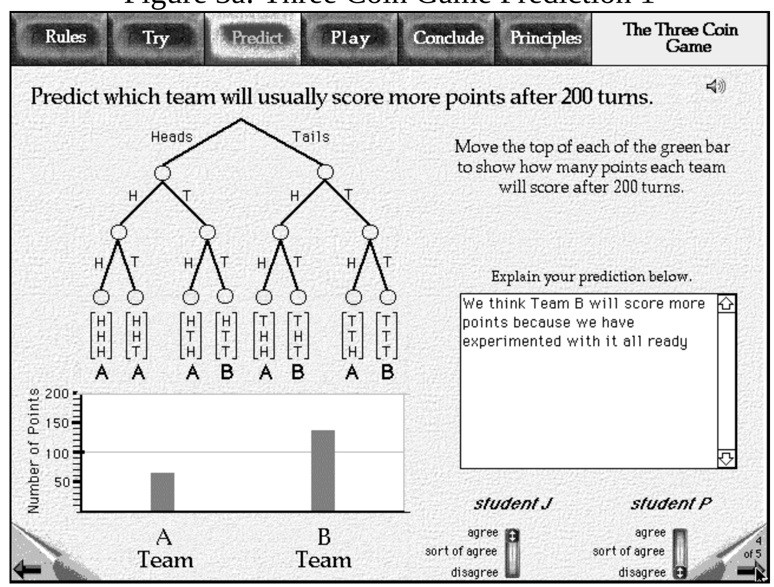

Figure 3b: Three Coin Game Prediction 3

Students then proceed to *Play* (Figure 4), where they run simulations of the three coin game. The students start the game by pressing the Start/Stop button, and control the speed via the speed control. At the four “normal” speeds, the coins in the upper left spin until each lands on heads or tails. As the coins land on heads or tails, a token animates down the probability tree, highlighting the current path, and dimming the other paths. Once all three coins are decided, the token animates down into the bar chart, and the chart updates to reflect this point. Additionally, the scoreboard at the lower right automatically updates. At any time the students can choose to view the bar chart by combinations or by teams, and at any time the students can change the speed of the game, or stop and start the game. At 20 points the game stops and “Karen” (the on-screen character) announces who is ahead, team-specific music and animation play, and a table shows the number of games each team has won. The students then start the game again, often choosing the fastest speed (“superfast”), where PIE can play 200 points in under a minute. Again, at 200 points, the game stops, Karen announces who has won, team-specific music and animation play, and a table shows the number of games each team has won. Every second game Karen introduces an “Observation Question”, asking the students if they think that the game is fair or unfair now, and providing students with a text entry space to write their current ideas. After the students play four games up to 200, Karen suggests that they go to Conclude, although they are free to go to Conclude at any time, and are free to stay in Play as long as they like.

In *Conclude* students revisit their predictions, but now are given data from the last three games as a resource for evaluation. By having students compare their predictions to results from several games they are given the opportunity to recognize patterns in the data. In *Principles* the students are then given the opportunity to create their own fair game, and are asked to provide a method for estimating how often an event will occur.

The corresponding real-world activity has the students investigate the outcomes of the role of two dice. Although these activities have very different surface features, the same principle can be applied to both of these games (for instance, one pair wrote “We can estimate how often each team will score by how many chances they have to get points.”). After a class discussion in which the class decides on the most appropriate principle for this activity, the teacher helps the students see how their principle leads directly to the formula for computing a subset of outcomes from a set of equally probable outcomes: P = # favorable outcomes/total outcomes.

Activity Set 4: The Two Spinner Game

The fourth activity introduces students to a continuous outcome space through the use of non-equal area spinners (Figure 5). This activity allows student to see the limitations of counting strategies and the formula P = # favorable outcomes/total outcomes, as Team A scores on only one out of four outcomes (and Team B scores on three out of four), yet Team A scores more points than Team B. However, by creating an area model of probability, in which one looks at each spinner as having four equally probable outcomes, of which three are orange and one is yellow, we can link this continuous outcome space to discrete outcomes (Figure 6). These activities show that situations with very different surface features can all be described by a formal use of the outcome space, which is an important part of the PIE curriculum.

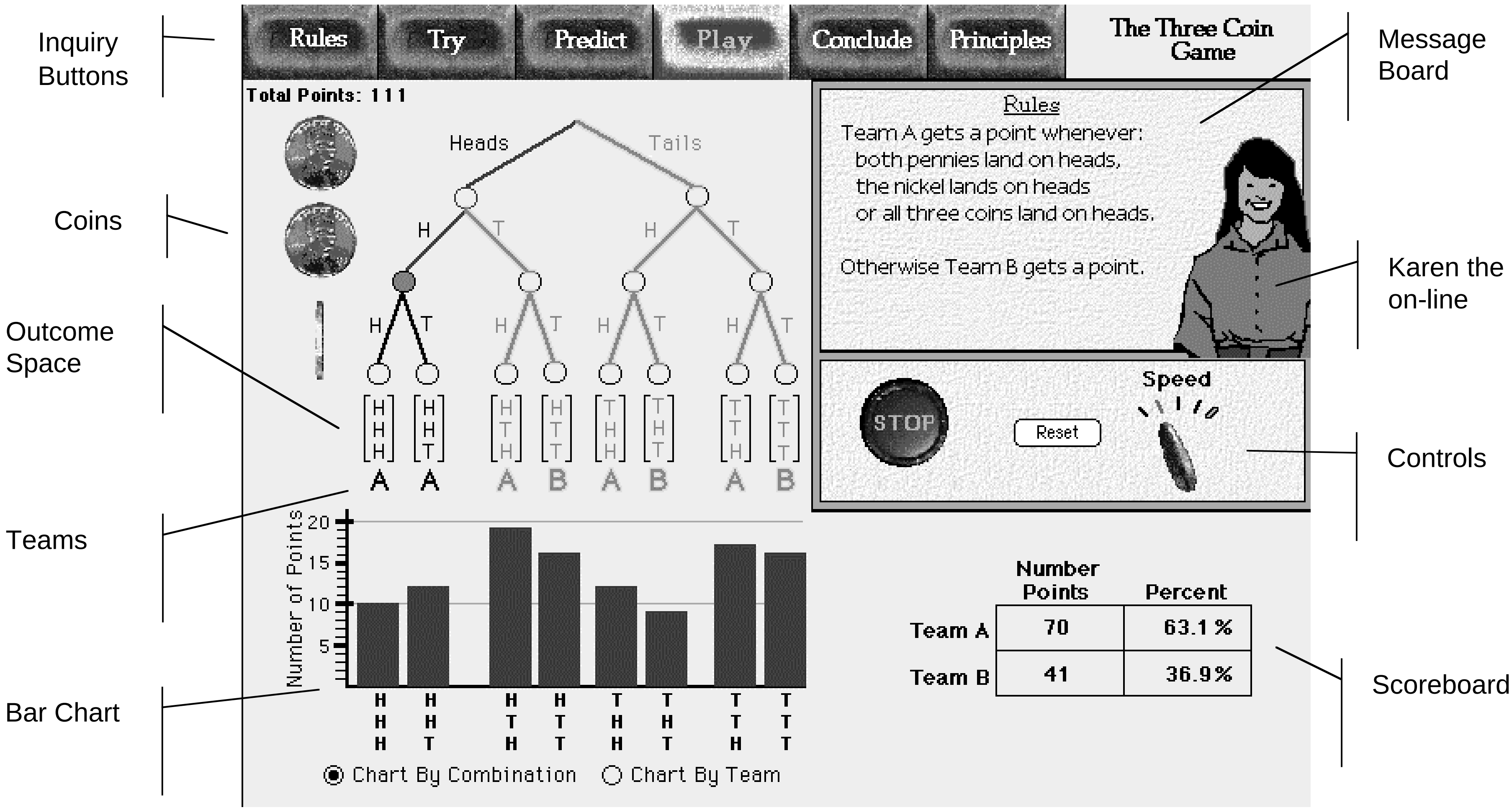

Figure 4: The Three Coin Game

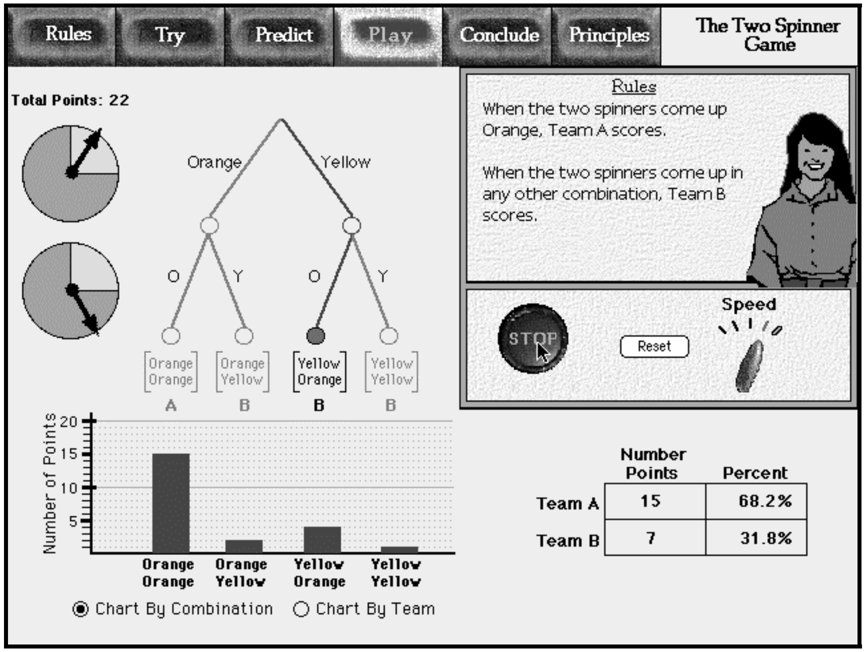

Figure 5: The Two Spinner Game

	O	O	O	Y
O	OO	OO	OO	OY
O	OO	OO	OO	OY
O	OO	OO	OO	OY
Y	YO	YO	YO	YY

Figure 6: The outcome space for PIE's Spinners

Activity Set	Game/ Activity	Rules	Fair	Target Understandings
1: on-line	Horse and Bunny	Bunny moves when scales unbalanced, Horse moves when scales balanced	-na-	aw of Large Numbers
1: off-line	50-50 Flip Off	Determine how often coin flips are "about half" heads and tails	-na-	aw of Large Numbers
2: on-line	Two Penny Game	Twins score on H-H and T-T Jumbles score on H-T and T-H	Yes	mportance of outcome pace
2: off-line	Two Penny Game	Jumbles score on H-T and T-H Skulls score on H-H Crossbones score on T-T	No	mportance of outcome pace
3: on-line	Three Coin Game	Team A scores on 8 outcomes Team B scores on 3 outcomes	No	mportance of outcome pace
4: off-line	Two Dice Game	Sum 2 dice, outcome moves one place	No	mportance of outcome pace
4: on-line	Two Spinner Game	Team A scores on 1 outcome Team B scores on 3 outcomes	No	mportance of robability distribution nd outcome space

Table 1: Summary of activities

The Study

The PIE curriculum was implemented for three weeks in two seventh grade classes (n = 45) in an urban middle school that serves a diverse group of students. During this same period a comparison group consisted of two classes taught by the same teacher as the PIE group (n = 54). The comparison group's unit covered the same topics as the PIE group during the same time in the academic year, but was taught in the traditional manner for this school. The traditional method in this case did not use a computer but instead had the students play and analyze a number of games of chance.

Students in both the PIE classes and comparison classes were given paper-and-pencil pre- and post-tests of the probability concepts addressed by the unit. Items on these tests were derived from standardized tests (National Center for Educational Statistics, 1994), suggestions from the NCTM (1989), items from the research literature on probabilistic reasoning (Tversky & Kahneman, 1982; Konold, 1991), and specific items we designed to assess the probability intuitions relevant to the instructional objectives of the probability courses. The post-test given to students in both the experimental and comparison groups consisted of a set of questions that were similar to those found on the pre-test (although the post-test questions were harder), as well as a set of questions that were new to the post-test. The score for each student was based on their performance on the multiple choice questions (1 point each) combined with their performance on the short-answer questions (1 point each). The short answers were scored correct if the student justified their answer with an appropriate mathematical construct. We used a blind scoring process so that while scoring the test the researcher did not know to which group the student belonged.

In addition, the teacher wanted to make sure that the test was “fair” to all his students (i.e. not biased toward the students in the PIE curriculum). As a result, although some questions did involve coin flipping, fairness, etc., the questions were not simple restatements of the PIE activities. That is, all questions on the pre- and post-tests involved some sort of transfer from the PIE curriculum.

Quantitative Results

A three-way analyses of variance (ANOVA) was carried out on three between-subject factors on the post-test: condition (experimental and comparison), gender (male and female), and standardized test score (split on the median for this sample). This analysis revealed a significant main effect of condition (**F** = 9.7, p < .01), a significant main effect of standardized test level (**F** = 45.7, p < .01; see Figure 7a), no main effect of gender (**F** = 1.3, p = .25; see Figure 7b), and no interactions were found. Additionally, t-tests found no significant differences between the two groups on the pre-test (t (89) = .21, p > .5), but a significant difference on the post-test (t(97)=3.4, p <.001; see Figure 8).

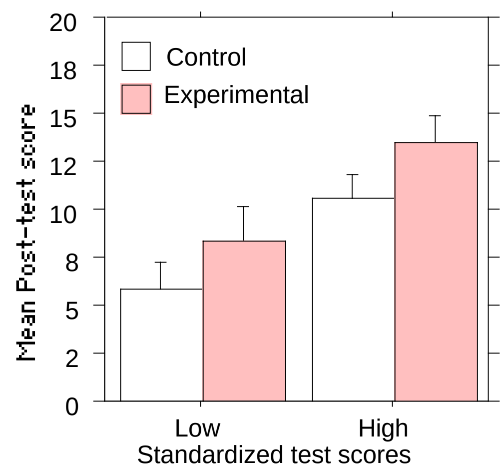

Figure 7a: Standardized tests and condition

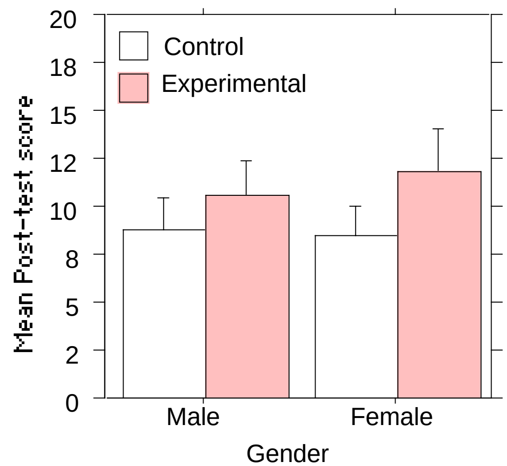

Figure 7b: Gender and Condition

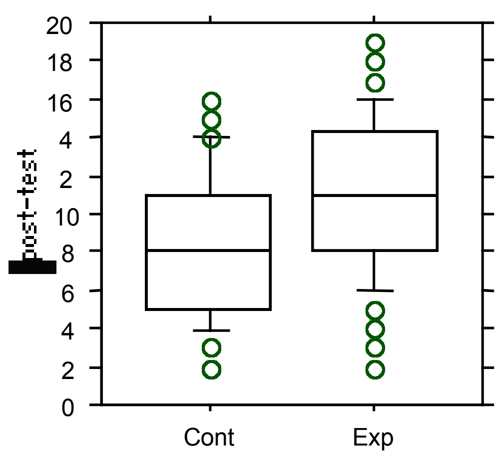

Figure 8: Post-test scores

These findings provide evidence of the effectiveness of the PIE curriculum: the PIE curriculum was similarly beneficial for students regardless of gender or scholastic achievement as measured by standardized tests, and the students in the PIE curriculum significantly outperformed the students in the comparison condition. Given these general findings, how did the PIE students compare to the comparison students on questions related to the focus of the PIE curriculum, the outcome space and the law of large numbers? To answer these questions, let us look at some of the specific post-test questions.

Test Item: Three Penny Flips

The first question that we will analyze is called the Three Penny Flips question, found in Figure 9. In this question students are asked to enumerate all the outcomes of three coin flips, calculate the chance of getting all heads, and state if any combinations of coins are more likely. Because these questions lay at the heart of the PIE curriculum, we would expect the students in the PIE condition to significantly outperform students in the comparison condition. This is the case in the first two parts of this question (for the enumeration section, $\mathbf{X}^2(1) = 11.1$, $p < .01$; for the calculation section, $\mathbf{X}^2(1) = 6.2$, $p < .02$). But, perhaps surprisingly, there was not a significant difference between groups in simply stating if any outcome was more likely ($\mathbf{X}^2(1) = 1.3$, $p = .26$). However, there was a significant difference in the number of *explanations* coded as correct ($\mathbf{X}^2(1) = 6.6$, $p = .01$). An explanation of this result requires another level of analysis, what we call “facet analysis”. We take the term “facet” from Hunt and Minstrell (1994), who state that a facet is a mental resource or knowledge element employed to explain a particular phenomenon. Each student explanation was coded for the main facet used as justification. Figure 10 shows a subset of the most common facets used on this question, and Table 2 describes the facets.

Three pennies are flipped one at a time. What are all the possible combinations (in terms of heads and tails) for the coins to land?

Draw a diagram, picture or chart that shows all the possibilities for the three coin flips.

What is the chance of getting three heads?

1/8

Are any of the combinations more likely than others? Why or why not?

No, all eight outcomes are equally likely

Figure 9: The Three Penny Flips Question (examples of acceptable answers in Italics)

This analysis of facets shows that there were important differences between the way the PIE students and the comparison students justified there answer to the question “Are any of the combinations more likely than others?” For instance, we see that the PIE students were more likely to state that all outcomes were equally likely (equally probable outcomes) and were also more likely to explicitly reference the number of outcomes when stating that the outcomes were equally likely (equiprobable-quant). Compare this to the students in the comparison condition, who were more likely to state that “anything can happen”. This was not scored as a correct answer, even though it may seem similar to “equally probable outcomes.” However, we have found, and it has been documented elsewhere (Konold et al., 1993), that students often use the simplistic reasoning of “anything can happen” to state that predictions simply cannot be made about random phenomena, because “anything can happen” (for example: Q: what is more likely, any combination of two heads and one tail, or three heads? A: You can’t tell, because coin flipping is random, and anything can happen). Finally, we see that students in the comparison condition were also more likely to state that “runs” were less likely, which is another well-documented non-normative idea often employed by students.

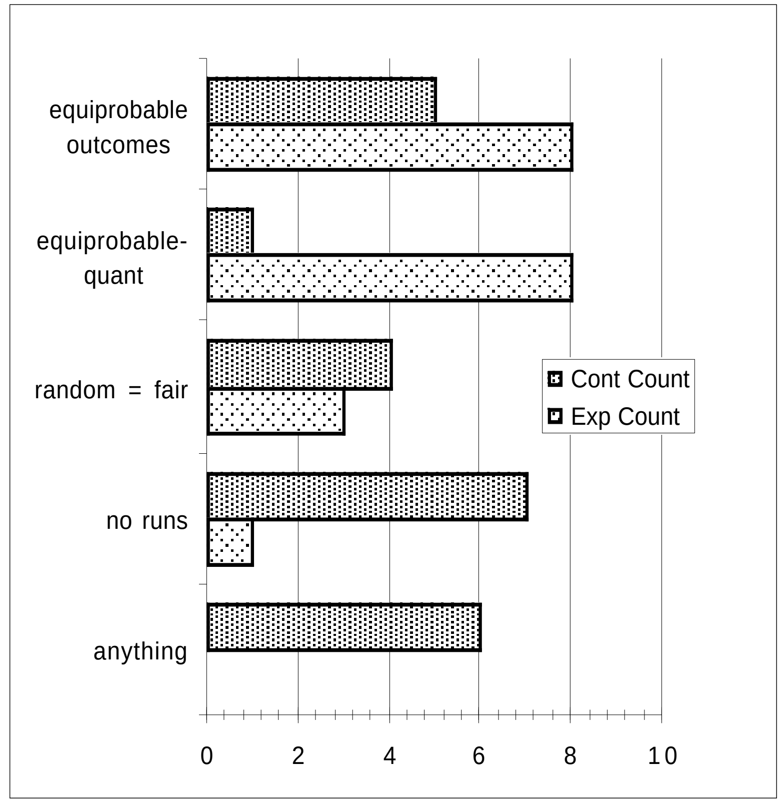

Figure 10: A subset of facets employed on the Three Penny Flips Question

This analysis shows that students in the PIE condition had more success in enumerating the outcome space, using the outcome space in a calculation, and then justifying why all possible outcomes of three coin flips were equiprobable. Let us now look at a question designed to probe students understanding of the law of large numbers.

Test Item: Fair or unfair coin

In this question students are asked to judge the fairness of a coin that is flipped 100 times, and comes up heads 47 times, and tails 53 times (see Figure 11). Because a t-test shows that we cannot reject the null hypothesis (a one-sample t-test, with heads = -1, tails = 1 gives us t (99) = .6, p = .55), we scored both that the coin is fair, and that one cannot tell if the coin is fair, as correct. Because a substantial focus of the PIE curriculum was determining the fairness of different situations, and because this is a direct application of the law of large numbers (in which students should note that a small difference after 100 flips should be expected), we expect that the students in the PIE curriculum would substantially outperform students in the comparison condition. However, as in the previous analysis, this was not the case for the forced-choice part of this question ($\mathbf{X}^2(1) = .02$, $p = .89$). A closer look at the data shows a ceiling affect for this question, with over ninety percent of students answering this question correctly. And, this ceiling affect may be explained by looking at the facets employed by students on this question. Figure 12 shows a subset of the most common facets employed on the fair or unfair coin question.

Suppose you flipped a coin 100 times and got 47 heads and 53 tails. Would you say:

(a) the coin is unfair in favor of Heads

(b) the coin is unfair in favor of Tails

(c) the coin is fair

(d) I cannot tell if the coin is fair or unfair

Why or why not?_____*When flipping a coin 100 times, such a small difference is to be expected.*

Figure 11: The Fair or Unfair Coin Question (examples of acceptable answers in italics)

An analysis of the explanations shows that the PIE students significantly outperformed the comparison students in providing normative explanations ($\mathbf{X}^2(1) = 6.2$, $p < .02$). Figure 12 shows that, although the modal answer for both groups was that the coins were “close enough” to even (“you can’t expect the coin to come out exactly 50:50. 53:47 is close enough”), a substantial amount of students in the comparison condition stated that “random = fair” (“cause when flipping a coin it is all luck”), an answer that was not coded as correct. So, as with the three penny flips question, students in the comparison condition were more likely to rely upon simplistic notions of randomness in their explanation, whereas students in the PIE condition were able to employ more normative facets of probability in their explanations.

Facet	Explanation	Example
equiprobable outcomes	state that all outcomes are equiprobable, without further justification	"yes the game is fair because each chance is equal"
equiprobable quant	state that all outcomes are equiprobable, and there are *n* outcomes	"All 8 combinations are equally likely"
Random = Fair	any random process must be fair	"it is fair because it is all chance"
No runs	outcomes such as HHH or TTT shouldn't happen much	"because 3 heads in a row is highly unlikely"
anything (can happen)	in a probabilistic situation, one can expect anything to happen	"they are all the same cause anything can happen"

Table 2: Description of facets employed on the Three Penny Flips Question

Facet	Explanation	Example
close enough	an evaluation of data in which quantities are taken to be close enough to be considered equal	"it's almost equal in the amounts, so it doesn't matter if one side has a few more."
Random = Fair	any random process must be fair	"it is fair because it is all chance"
LOLN (law of large numbers)	an explicit evaluation in which data based upon a large sample is more compelling than data based on a small sample.	"Again, the law of large numbers will even things out, but such a small sample size is hardly reliable"
can't tell	one can't say anything about the results of a random phenomena	"You can't tell what's going to happen"
anything (can happen)	in a probabilistic situation, one can expect anything to happen	"they are all the same cause anything can happen"

Table 3: Description of facets employed on the Fair or Unfair Coin Question

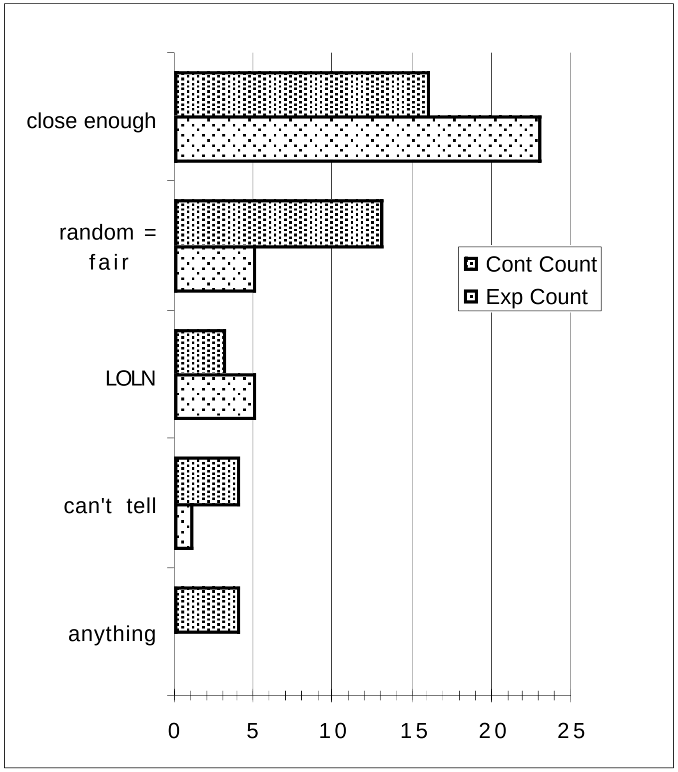

Figure 12: A subset of facets employed on the Fair or Unfair Coin Question

At this point we have seen that students in the PIE condition outperformed students in the comparison curriculum on questions relating to the outcome space and the role of data. However, we will now see that students in the PIE curriculum did not perform well on all aspects of our model of probability.

The Quiz Question and the Basketball Question

Two other questions were designed to probe students' understanding of the outcome space: the Quiz Question and the Basketball Question (Figure 13). However, students did not recognize the questions as being related to the outcome space. Instead, these questions show the importance of having students recognize that one can apply understandings of *randomness* to real world phenomena that are not based on prototypically probabilistic phenomena.

The Basketball question asks students about the chances of a free shooter making her shots, and the Quiz question asks the probability of guessing three True/False questions correctly. From the perspective of probability theory, both of these questions should be answered by invoking an outcome space isomorphic to coin flipping (where, say, heads = guessing correctly or making the shot). Given this, we expect that the students in the PIE condition would outperform students in the comparison condition. And, although we can claim that students in the PIE condition performed marginally better than the comparison students on these questions (Basketball: $\mathbf{X^2}(1) = 3.7$, $p = .054$; Quiz: $\mathbf{X^2}(1) = 6.3$, $p < .02$), this would be missing the point. In fact, only *three* students (all in the PIE group) answered the Basketball question normatively, and only *five* students (again, all in the PIE group), answered the Quiz question normatively.

When answering the Basketball question, most of the students simply relied on causal reasoning ("well it depends if the player knows how to play and shoot well. Maybe she makes one or both, you never know. She has a fifty-fifty chance") or their notions of "50%" ("still 50% because each time it is always a 50-50 chance"). When answering the quiz question students again relied on notions of 50%, often attempting to incorporate the number three (the only number in the question) into their answers ("it is a 50/50 chance that you would get 1 right so three right would be 1/3 of a chance"), or they relied on their understanding of test-taking ("think very hard and read it over").

A quiz you are taking has three "true or false" questions. You know none of the answers and guess at all three.

a. What is the probability of guessing three correct answers?

1/8

Time has run out in the big basketball game, and the score is a tie. However, one of the schools players was fouled and gets to shoot two free-throws (each worth 1 point). If either one of the shots (or both) is made the school wins. The player has an average free-throw percentage of 50%. What is the chance that she will make at least one free-throw and win the game?

3/4

Figure 13: The Quiz Question and the Basketball Question

Looking back to the PIE curriculum and our model of probabilistic reasoning, we see that perhaps this should have been expected. Because the PIE curriculum ignored issues of recognizing randomness, one of the four components of probabilistic reasoning (the curriculum used only prototypically random devices such as coin flipping and dice tossing), it seems as though the students did not learn to recognize random phenomena in "real world" situations. That is, the students did not even recognize the outcome space as being relevant in answering these questions, and the students relied more on their understanding of test-taking and basketball shooting. If we consider performance on such questions important, we must ensure that students have the opportunity to engage in activities where they can see the invariance of the outcome space across different situations, and we may also have to help students see the invariance of randomness and probabilistic reasoning across different situations.

Qualitative Results: A Case Study of J and P

In this section we take a detailed look at how one pair of students, J&P, articulated and revised their initial understandings of probability as they interacted with PIE—collaboratively making predictions, evaluating data, and interpreting representations. This qualitative analysis will give a feel for student interactions, and show how the features of PIE contributed to student conversations and student understandings.

We join J & P as they interact with *Try* of the Three Coin Game[1]. As discussed earlier, the game is unfair, as Team A scores on five of eight outcomes. In *Try*, PIE allows students to choose the outcomes of the coin flips. In this case, J is intentionally choosing the coins so that Team B scores all the points. In Excerpt 1 we see P asking J's rationale for this, and J simply states that "I want to be the man". We claim that J was using this feature of PIE to *predict* that B would win more games, and justification for this will be forthcoming.

[J is making Team B score all of the points]

P: Why you got to do B?

J: Cause I want to be the man.

Excerpt 1: J has Team B score most of the points

Jumping the first prediction question, it appears that J and P agree that the game is unfair. Excerpt 2 shows J explicitly stating that the game is unfair, P stating that Team A will win more often, and both students jointly constructing the statement that Team A has more opportunity (their response is recreated

[1]The transcription conventions used are derived from Gumperz (1982). Brackets ([ ]) are used to identify comments by the researcher, equal signs (=) are used to identify where student speech overlaps, and ellipses (...) are used to identify a pause.

in Figure 3a). This is entirely consistent with the normative view that Team A has more outcomes (or, as P and J say, "slots", or "opportunity"), but we will see that J and P have conflicting interpretations of these statements.

J: We think the game is unfair.
P: We think that team A will win more often.
J: It, I mean they =
P: =have more.
J: Have ... more =
P: =Slots.
J: Opportunity.
typed: We think the team A will win more often because they have more opportunity.

Excerpt 2: Team A has more opportunity

In the third prediction question the students are asked to predict how often each team will score by dragging bars to the appropriate heights (their response is recreated in Figure 3b). In Excerpt 3 we see J state that *Team B* will score more points. We consider the student response to this question to be a pivotal point in this activity, as this is the first time in this activity that the two students explicitly recognize their different expectations.

J: [Singing. Moves B's bar higher than A's bar]
P: Why you put B?
J: I think B is gonna win. And you see all the other times I win. Like with the bunny had more opportunity =
P: = The horse, yeah, I know that. But you shoulda put, on the first one, you said you think A will win because they have more opportunity
J: No I didn't say A would win. I think the game is unfair in A's favor, because they have more opportunity =
P: =correct, correct, correct.
J: Because we...have, how do you spell experience? [J typing]
P: E-e-x-p-e-r-e-r-i-m-e-n-t. Experiment.
J: Because we have experimented.
P: Is that how you spell it? [to T, who is walking by]
T: Perfect. So B's winning?
P: Yeah, on the Try thing. Well, J was making B win, but
J: Don't even worry about that. What you wanna do here, you gonna agree?
P: I'm gonna disagree [moves agreement bar to disagree]
Typed: We think Team B will score more points because we have experimented with it all ready.

Excerpt 3: J states that B is going to score more

This is a complex interaction, and will require some explanation. In this excerpt we notice some misunderstandings that were hidden in previous exchanges. For instance, J believes that the term "in favor of" has the opposite meaning than we intended (and, we have found that a small but substantial number of students have difficulty with this wording). We also see that J is making (what is to us) an unjustified analogy to the Horse and Bunny game when he says "And you see all the other times I win. Like with the bunny". What J is stating here is that the "scoring zones" from the Horse and Bunny Game are analogous to the outcomes in the Three Coin Game, and that even though the bunny had more zones, the horse won in the end. However, this analogy breaks down for two reasons: in the Horse and Bunny Game each scoring zone represents a large number of outcomes, not a single outcome; and each scoring zone is not equally probable, whereas each outcome in the Three Coin Game is equally probable. We also see that J states that they have "experience" with the situation (which P changes to "experiment"). This provides evidence that his seemingly flippant "I wanna be the man" comment (Excerpt 1) meant that J was constructing data in Try in an attempt to predict the final outcome of the game, and this prediction is based on his prior experiences with the Horse and Bunny Game.

Note that P immediately recognizes and points out J's apparent inconsistencies. Once J explains his reasoning, P seems to agree with J for much of this exchange, even helping to construct their typed response. However, when the teacher intervenes to ask them about their answer, P states that he does not believe their jointly constructed prediction ("Well, R was making B win, but"). He then explicitly states that he disagrees with their response ("I'm gonna disagree"). At this point the stage is set for the students to run the simulation and (potentially) rectify their disagreements, which is important in that it motivates students to compare their expectations to the results of the simulation, and then defend, justify, and modify their understandings based on their interactions.

Before continuing with this, though, let us belabor one important detail of these interactions that relates to the rationale behind the design of PIE: we see that many would consider prediction question 1 (is this game fair or unfair) and prediction question 3 (how often will each team score) closely related. However, note that question 1 was fully discussed by the two students, yet there remained a critical (and hidden) difference between their expectations. It was only in question 3, where the students interacted with a standard representation (a bar chart) in a non-standard way (that is, not by using it as a historical record of events, but by using it as a prediction tool) that these differences became visible to both participants. Let us now turn to J and P as they run the simulation.

J and P have run several simulated games, and have seen Team A win every time. We join them as A wins yet another game to 200, "Karen" (the computer agent, denoted as "K" in the transcript) announces the winner, and then asks them an observation question (Excerpt 4).

K:	At 200 turns, team A is ahead. [computer plays Team A's animation]
P:	A be whooping!
K:	do you think the game is fair or unfair now?
J:	Up yours, team A.
P:	what do you think about this game now? [paraphrasing the observation question]
J:	This game is bullshit... Well my prediction was right, this is game is unfair.
	[starts new game]
K:	At 20 turns team A is ahead.
P:	Well my prediction is right that, see mine was somewhere close to that prediction. But yours is like, umm, totally off because team A is way up there and team B is right there.
J:	You didn't say, you didn't say that team A was gonna win, you said yeah that's right =
P:	= I did, I did say team A =
J:	=but didn't you agree with me? =
P:	= I said team A
J:	P, did you or did you not agree with me?
P:	I said team A. I said team A. [repeating this in a taunting manner]
J:	Shut your ass.

Excerpt 4: P and J realize that Team A is winning

In this (somewhat heated) exchange, we see that both J and P now agree on the results: Team A wins more of the games. We also see J (possibly prompted by the observation question), attempt to salvage some of his prediction, by stating "Well my prediction was right, this is game is unfair." P, however, states that J's prediction was inaccurate, which leads to an argument about exactly what transpired in Predict. It is not until the teacher comes over (Excerpt 5) that the students discuss the precise differences in their expectations, and give backing arguments, such as P repeating his concern with J's making Team B win in Try. Next, when P states that he thought A would win because "they had more opportunity", we see J try to change the meaning of his prediction, stating "that's what I said."

P:	He put team B was going to be winning. [to teacher, who walked over to them]
T:	Did you agree or disagree?
P:	I disagree.
T:	Why do you think, why do you think, oh cause B when you experimented, B was winning?
P:	Yeah, but he was making them win. That was.
T:	Look at this, why do you think team A is winning?
P:	Cause they had more opportunity.
J:	That's what I said.
P:	They got more slots.
T:	Why didn't you go that way when you predicted?
P:	He said remember the bunny and the horse =
J:	= The horse didn't have all the opportunities, and the horse still won.
T:	Right, right, this is different. When you get to conclude, talk about what you know now.
P:	Yeah, yeah we will. [clicks conclude]

Excerpt 5: Explaining reasoning to the Teacher

These excerpts provide an example of the way that an instructional environment must account for students' desire to link their existing understandings with current situations. To make sense of the current situation J draws on an analogy to a prior game, and defends his initial prediction with statements such as “Well my prediction was right, this is game is unfair”, and, when P says “Cause they had more opportunity”, J rejoins with “That’s what I said”: although J has already determined that his prediction did not accurately describe the outcome of this situation, he is still drawing upon some aspects of his predictions in his attempt to make sense of the situation. Simply expecting students to state that they are “wrong”, and then accept what is presented as “right” cannot work. Instead, students should be presented with situations where they can discover what understandings are in agreement with normative theory and what are in conflict with normative theory, and build from areas of agreement.

Let us belabor one more detail of PIE. This entire exchange was prompted by an observation question which, by stopping the flow of the activity, provided an impetus for a conversation that allowed the two students to contrast the prior, typed predictions to the results of the simulation, and allowed P to state his disagreement with the original predictions. This question proved valuable even though the students never typed a response to the question.

Moving on from this point in the activity, we have evidence that J moved from his initial beliefs, to understanding that the outcome space is an important determinant in deciding if a game is fair, one of the instructional objectives of this activity. This evidence comes later in their investigation, during Principles, when the students were asked to explain how to make a fair game. J stated “Both teams have an even amount of chances.” However, this is an admittedly limited result, as J never does reexamine his analogy to the Horse and Bunny Game to understand why it breaks down. Let us examine this issue a little more closely.

It is important to acknowledge that, when building an environment for knowledge construction, it is impossible to determine all possibilities and contingencies beforehand. That is, any computer activity will not be able to take into account all possible student reactions. In fact, this is exactly why the PIE computer activities were designed as one part of a curriculum with a number of mutually reinforcing activities including computer-based and real-world activities, as well as full-class discussions.

However, as we were not prepared for J’s analogy, we missed a potentially valuable learning activity: an explicit comparison of the Horse and Bunny Game and the Three Coin Game. In retrospect, it seems as though this comparison would not only be potentially valuable to J, but to other students, as such a comparison would foreground the exact definition of the outcome space, and in future implementations of PIE we will consider this as a potential new activity. This finding points to the importance of looking closely at student interactions, and iteratively improving environments for knowledge construction based on the repertoire of intuitions and reasoning strategies the students employ throughout the curriculum.

Conclusions

We feel that this study has relevance not only to the creation of math and science learning environments for middle school students, but for the creation of any environment built to aid people in knowledge construction. In the PIE curriculum the students were not made to memorize terms and definitions (trials, outcome space, etc.), but instead were provided with a situation in which their intuitive ideas would lead to productive conversations about the phenomena, and probability was used not as an end in itself, but was used in the service of a task that was considered authentic to the students. It is only after participating in these activities that the students formalize their ideas using definitions and formulas, and at that point the students have a reason for doing so.

We also saw that students may employ non-standard uses of terms or ideas, many of which can be anticipated and used—such as student intuitions about "mixes happen more". Others, however, cannot always be anticipated, such as misapplying an analogy to a prior situation. In either case, we saw that collaboration brought these issues to the forefront, as collaboration required that these ideas be explained, justified, and then, occasionally, challenged. This use of collaboration, combined with presenting different perspectives, helps to ensure that student ideas are articulated and alternatives are presented, which sensitizes the students to the learning experience.

When asking students questions from slightly different perspectives, and providing them with different resources, the students were able to articulate a variety of (sometimes conflicting) ideas, and consider alternate perspectives. This is important, because when discussing probability or any other complex domain, people often answer similar questions with dramatically different reasoning (Bell, 1997; diSessa, 1988; Konold et al. 1993). When designing environments built to aid people in knowledge construction, designers must ensure that people are presented with a variety of ways in which to consider phenomena, and are provided a variety of ways in which to express these ideas.

Finally, we saw that both the environment itself, as well as human participants, can function in a manner that facilitates student reflection on the activity. This was illustrated through the use of observation questions: the current activity was paused and the computer agent, Karen, made an observation about the game. This observation prompted a student conversation in which they compared their predictions to the results of the game. The case for a human participant was made later in that same exchange, when a series of questions by the teacher resulted in the students moving away from simply discussing their predictions, to formulating ideas that help to explain the current results.

Acknowledgments

This research is funded in part by grants from the U.C. Urban Community-School Collaborative, the Berkeley chapter of Sigma Xi and the National Science Foundation's Science and Design Traineeship. We would like to thank Jesse Ragent for his help in the design and implementation of PIE in his classrooms.

References

Bell, P. (1997). "Using argument representations to make thinking visible for individuals and groups" In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Computer Support for Collaborative Learning '97* (pp. 10-19), Toronto, Canada: University of Toronto Press.

Brown, A., & Campione, J. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser, (Eds.), *Innovations in learning: New environments for learning* (pp. 289-325), Mahwah, NJ: Lawrence Erlbaum Associates.

California Department of Education. (1992). *Mathematics frameworks for California public schools: Kindergarten through grade twelve*. Sacramento, CA: California Department of Education.

California Department of Education. (1999). *Mathematics frameworks for California public schools: Kindergarten through grade twelve*. Sacramento, CA: California Department of Education.

Chinn, C. & Brewer, W. (1993). Students' responses to anomalous data: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1-49.

Collins, A., Brown, J.S. & Newman, S. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L.B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 453-494), Hillsdale, NJ: Lawrence Erlbaum Associates.

Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty. *Cognition*, 58, 1-73.

diSessa, A. (1988) Knowledge in pieces. In G. Forman, & P. Pufall (Eds.), *Constructivism in the computer age. The Jean Piaget symposium series* (pp. 49-70), Hillsdale, NJ: Lawrence Erlbaum Associates.

Enyedy, N., Vahey, P. & Gifford, B. (1998). "...It's fair because they each have two": The Development of a Mathematical Practice Across Two Social Contexts, In A. Bruckman, M. Guzdial, J. Kolodner & A. Ram (Eds.), *International Conference of the Learning Sciences 1998* (pp. 91-97). Atlanta, GA: AACE.

Enyedy, N., Vahey, P. & Gifford, B., (1997). "Active and supportive computer-mediated resources for student-to-student conversations", In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Computer Support for Collaborative Learning '97*, pp. 27-36, Toronto, Canada: University of Toronto Press.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Norwell, MA: D. Reidel.

Fischbein, E., Nello, M., & Marino, M. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics*. 22(1), 523-549.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review.* 103(3), 592-596.

Gordin, D.N., Polman, J.L., & Pea, R.D. (1994). The Climate Visualizer: sense-making through scientific visualization. *Journal of Science Education and Technology*. 3(4), 203-226.

Gumperz, J., (1982). *Discourse strategies*. New York: Cambridge University Press.

Gunstone, R.F. (1991) Constructivism and metacognition: Theoretical issues and classroom studies. In R. Duit, F. Goldberg, & H. Niedderer (Eds.), *Conference on research in physics learning: Theoretical issues and empirical studies* (pp. 129-140), University of Brennan, Germany: University of Brennan Press.

Hall, R. & Rubin, A. (1998). ...there's five little notches in here: Dilemmas in teaching and leaning the conventional structure of rate. In J. Greeno & S. Goldman (Eds.) *Thinking Practices* (pp. 189-236), Hillsdale, NJ: Lawrence Earlbaum and Associates.

Hancock, C., Kaput, J., & Goldsmith, L. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337-364.

Horvath, J., & Lehrer, R. (1998). A model based perspective on the development of children's understanding of chance and uncertainty. In S. Lajoie (Ed.), *Reflections on statistics: Learning, teaching and assessment in grades K-12* (pp. 121-148), Hillsdale, NJ: Lawrence Erlbaum Associates.

Hunt, E., & Minstrell, J. (1994). A cognitive approach to the teaching of physics. In K. McGilly (Ed.) *Classroom lessons: Integrating cognitive theory and classroom practice*, (pp. 51-74), Cambridge, MA: MIT Press.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review* 103(3), 582-591.

Konold, Clifford. (1991) Understanding students' beliefs about probability. In E. von Glaserfield (Ed.), *Radical constructivism in mathematics education*, (pp. 139-156), Boston, MA: Kluwer Academic Publishers.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, D. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics education*, 24(1), 392-414.

Lajoie, S. Jacobs, V. & Lavigne, N. (1995). Empowering students in the use of statistics. *Journal of Mathematical Behavior*, 14(1), 401-425.

Lampert, M. (1995). Managing the tensions in connecting students' inquiry with learning mathematics in school. In Perkins, Schwartz, West, & Wiske (Eds.), *Software goes to school: teaching for understanding with new technologies,* (pp. 213-232). New York, NY: Oxford University Press.

Linn, M., & Songer, N. (1991). Teaching thermodynamics to middle school students: What are appropriate cognitive demands? *The Journal of Research in Science Teaching*, 28(10), 885-918.

Metz, K. (1998). Emergent ideas of chance and probability in primary grade children. In S. Lajoie (Ed.), *Reflections on statistics: Learning, teaching and assessment in grades K-12*, (pp. 149-174), Hillsdale, NJ: Lawrence Erlbaum Associates.

Minstrell (1989). Teaching science for understanding. In Resnick & Klopfer, (Eds.), *Toward the thinking curriculum: Current cognitive research*, (pp. 129-149). Alexandria, VA: ASCD.

National Center for Educational Statistics (1994). *NAEP 1992 Trends in academic progress*. Office of Educational Research and Improvement, US Department of Education, Washington, DC: US Government Printing Office.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Newman, C., Obremski, T., & Scheaffer, R. (1987). Exploring probability. Dale Seymour Publications: Palo Alto, CA.

Piaget, J, & Inhelder, B., (1975). *The origin of the idea of chance in children*. New York, NY: WW Norton & Co

Richards, J. (1991). Mathematical discussions. In E. von Glaserfeld (Ed.), *Radical constructivism in mathematics education*, (pp. 13-51). Boston, MA: Kluwer Academic.

Rosebery, A., & Rubin, A. (1989). Reasoning under uncertainty: Developing statistical reasoning. *Journal of Mathematical Behavior, 8*, 205-219.

Smith, J.P., diSessa, A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3(2), 115-163.

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. 465-493), New York, NY: Macmillan.

Strike, K. & Posner, G. (1992). A revisionist theory of conceptual change. In R. Duschl and R. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice,* (pp. 147-176), Albany, NY: SUNY Press.

Thorkildsen, T. (1995). Conceptions of social justice. In W. Kurtines & J. Gewirtz (Eds.), *Moral development: an introduction,* (pp. 424-447). Needhan Heights, MA: Allyn & Bacon.

Tversky, A. & Kahnemann, D. (1982). Judgment under uncertainty: Heuristics and biases. In Kahnemann, Slovic, & Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*, (pp. 3-22). New York, NY: Cambridge University Press.

Vahey, P. (1996) How students' conceptions of fairness informed the design of an interactive instructional probability environment. Paper presented at the American Education Researchers Association annual conference, New York, NY.

Vahey, P, Enyedy, N., & Gifford, B. (1997). Beyond representativeness: Productive intuitions about probability. In M. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, (pp. 769-774), Hillsdale, NJ: Lawrence Erlbaum Associates.

White, B. (1993a). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction* 10 (1), 1-100.

White, B. (1993b). Intermediate causal models: A missing link for successful science education? In R. Glaser (Ed.), *Advances in Instructional Psychology*, (pp. 177-252), Hillsdale, NJ: Lawrence Erlbaum Associates.