Students' Conceptions of Probability

Joan Garfield and Robert delMas - Minneapolis, Minnesota, USA

1. Introduction and background

A three-stage model was used in developing and evaluating an instructional unit on probability. The first stage is the identification of misconceptions. The second stage involves the development of an instructional approach which is based on a theory or model of learning. The design of a measure of target behaviours and the selection of an appropriate instructional model or theory feed directly into the third stage: assessment, and the results of the assessment feed back into the design.

The focus of this paper is on the first and third stages of the model, both of which depend on designing ways to identify misconceptions. In previous studies, researchers have used changes in performance on individual items to evaluate the effectiveness of instructional interventions. The instrument designed and used in this study differs from previous instruments, not in the content of the items, but in the way responses to items are analysed. Instead of considering responses to single items, pairs of items are designed so that meaningful error patterns can be identified. The identification of error patterns allows assessment that goes beyond the reporting of gain scores. Once error patterns are identified, an intervention can be evaluated according to the types of misconceptions (i.e. error patterns) that are affected.

Much of the work on misconceptions of probability has been done by psychologists (Kahneman, Slovic and Tversky, 1982). One misleading heuristic, representativeness, refers to the idea that an occurrence is probable to the extent that it seems "typical". If a sample does not resemble the characteristics of the population it came from, it seems less likely than a sample that more closely resembles the population. People using this heuristic tend to not understand sample variability and judge equally likely samples from a population to have different probabilities of occurrence.

The representativeness heuristic is also used to explain the "gambler's fallacy" or "law of averages". Many people believe that after a long run of heads, a tail is more likely to result. While the "law of large numbers" guarantees that large samples will be representative of the population from which they are drawn, another misconception, "law of small numbers", asserts that this applies to small samples as well.

2. The "Reasoning About Chance Events" instrument

This test was designed to assess the impact of Coin Toss, a tutorial software program, on students' reasoning. Coin Toss simulates tosses of a fair coin to demonstrate basic probability concepts, including variability of samples, effect of increased sample size, independence, and randomness. Items used in previous research by Shaughnessy (1981), Konold (1989b), and delMas (1989) were adapted to construct this instrument. There are a variety of item types on the test; some multiple choice, some open-ended, and some branching questions which ask students to select the best rationale for their chosen answer. Questions 1, 2, 3 and 17, and Questions 4, 7, and 8 are listed in the Appendix.

Subjects in this study were 78 first and second year college students. Students were assigned a chapter to read which provided a basic introduction to probability. During the following class, they took the "Reasoning About Chance Events" as a pre-test, then they were given instructions on how to use the Coin Toss software along with a workbook in which to record data generated by the computer simulations. One of the authors met with the students during class at the end of the week and engaged the students in discussion about their experience. Discussion ended after 30 to 45 minutes. "Reasoning About Chance Events" was completed as a post-test.

3. Analysis of results

Questions 1 and 2 were designed to identify the use of a representativeness heuristic known as the law of averages. The initial relative frequencies and the changes from pre-test to post-test presented in Table 1 are similar to those reported by Shaughnessy (1977) for a group of students who received an experimental activity-based unit on probability. Subjects were not asked to provide a reason for Question 2. Reasons were inferred from subjects' joint responses to Questions 1 and 2.

TABLE 1
Reasoning categories for Questions 1 and 2

Reasoning Category		Response pattern Question 1	Question 2
Correct Reasoning		c. Both	c. Both
Random & Even		a. B G G B G B	a. B G G B G B
Even	OR	a. B G G B G B	c. Both
		a. B G G B G B	b. B B B G G G
Random		c. Both	a. B G G B G B

Table 1 presents four response patterns for Questions 1 and 2: Correct Reasoning, Random & Event, Even, and Random. "Correct Reasoning" is displayed when a subject provides a correct answer to both questions. The "Random & Even" type of reasoning has two components. First, randomness is defined as alternation or perceived irregularity in the birth order of boys and girls. Second, the split between boys and girls should be even and expressed locally in the short-run sequence. The third type of reasoning, "Even", requires only that the number of boys and girls be equal. Finally, the "Random" type of reasoning requires only that the sequence appear random. The scheme identifies two reasoning patterns (i.e. the Random & Even, and Random patterns) in addition to patterns similar to those identified by Shaughnessy (i.e. the Correct and Even patterns).

TABLE 2

Pre-test and post-test cross-tabulation of Questions 1 and 2 (N = 75)

	Post-test				
Reasoning Category	Correct on Both	Random & Even	Even	Random	Total
Pre-test	Row %	Row %	Row %	Row %	Column %
Correct on Both	75	6	13	6	21
Random & Even	34	40	17	9	47
Even	36	9	50	5	29
Random	50	0	50	0	3
Total Row %	44	23	27	7	100

Table 2 presents a cross-tabulation of response patterns for the pre-test and post-test. The categories in Table 1 capture the pre-test and post-test response patterns for 75 (96%) of the 78 subjects. From the results it appears that 21% of the subjects do not demonstrate use of the law of average heuristic on the pre-test. About 29% of the subjects initially display an Even pattern, which is quite close to the percentage reported by Shaughnessy (1977). In addition, many subjects (46%) displayed the Random & Even response pattern. The Random response pattern occurs infrequently.

Identification of subjects' initial response patterns permits an investigation of the differential effects of the intervention. First, although the majority of subjects still display a misconception on the post-test, the overall number of subjects displaying correct reasoning doubled from pre-test to post-test. Second, fully 75% of the subjects who displayed correct reasoning on the pre-test had the same response pattern on the post-test. The instruction does not appear to have led a large number of subjects with correct response patterns to develop misconceptions. Third, a large number of subjects who had response patterns representative of misconceptions displayed correct reasoning patterns on the post-test. Thirty-four percent of those who initially had a Random & Even pattern and 36% of those with an Even pattern on the pre-test gave correct responses to both Questions 1 and 2 on the post-test. Although subjects with misconceptions tended to maintain their misconceptions, a large number did change to correct response patterns.

Questions 3 and 17 were designed to identify use of the law of small numbers (Tversky and Kahneman, 1971), and measure subjects' understanding of the relationship between sample size and variance. Question 3 is an elaboration of another question used by Shaughnessy (1977). The correct response to Question 3 is the smaller hospital. The relative frequency of correct responses on the pre-test is lower than that found by Shaughnessy. In addition, while subjects in Shaughnessy's experimental class showed a significant gain from pre-test to post-test on this item (from 27% to 67% correct responses), subjects in the present study did not. Only 66 subjects (85%) responded to Question 17 on both the pre-test and the post-test. This is apparently a function of its placement in the sequence of test items. Question 17 has the same logical form as Question 3, however, it is less complicated. It also refers to the tossing of a coin and can be easily identified as a situation which involves chance events. More correct responses were given to Question 17 than to Question 3 (41% vs 17%).

Again, much is gained by considering students' combined responses to Questions 3 and 17. The analysis of response patterns for Questions 3 and 17 is somewhat different in comparison to the analysis of Questions 1 and 2. Based on the argument that subjects are more likely to respond correctly to Question 17 than to Question 3, three response patterns were defined. In the first response pattern, a subject gives the "small" response for both items. In the second response pattern, a subject is correct on Question 17 but incorrect on Question 3. The third response pattern occurs when a subject gives incorrect responses to both items. This scheme can be considered to represent three states of misconception: no misunderstanding, a moderate misunderstanding, and a strong misunderstanding of how and when sample size affects variance.

TABLE 3

Pre-test and post-test cross-tabulation of Questions 3 and 17 (N = 60)

	Post-test			
Reasoning Category	Correct on Both	Random & Even	Random	Total
Pre-test	Row %	Row %	Row %	Column %
Correct on Both	100	0	0	10
Incorrect on 3 - Correct on 17	30	45	25	33
Incorrect on Both	8	18	75	57
Total Row %	23	27	50	100

Table 3 cross-tabulates response patterns for Questions 3 and 17 on pre-test and post-test. The three response patterns are respectively labelled Correct on Both, Incorrect on 3 - Correct on 17, and Incorrect on Both. Sixty (91%) subjects could be placed into one of the three response categories. The first observation is that very few subjects were correct on both items on the pre-test (10%), and, although the number doubled, a few were correct on both items on the post-test (23%). Second, there is considerable stability in response patterns for subjects initially categorised as Correct on Both or Incorrect on Both in response pattern for subjects who were initially identified as Incorrect & Correct appear to be more random. Forty-five percent maintained an Incorect & Correct pattern, while 30% changed to Incorrect on Both and 27% changed to Correct on Both. Therefore, the instruction does not appear to have had a strong influence on students' misconceptions about the relationship between sample size and variance.

Stability of students' conceptions: In order to explore the stability of students' conceptions of probability, responses were compared to parallel items which had different contexts. On Items 7 and 8, responses varied quite a bit between problems although the basic probability concepts were identical. More students (63%) appear to have used the frequentist strategy in responding to Item 7, which involves a guessing game, than on Item 8 (35%). The context in Item 8 of student interviews seems to lead more students to choose response "a", which represents casual reasoning (26%).

Items 5 and 6 are also parallel problems, both referring to drawing with replacement from a can with equal numbers of blue and red marbles. In each question, students were asked to identify the most likely outcome for the next draw, either Red, Blue, or both equally likely. In Item 5 the preceding set of four draws was Blue, Red, Blue, Blue, whereas in Item 6 they were Blue, Red, Red, Red. In comparing responses to Items 5 and 6, only 36 (47%) students answered both items correctly. However, ten students who were correct on Item 5 appear to have switched to a "law of averages" strategy on Item 6, by selecting response "Blue". These students seemed to expect the second series to be more likely to balance out into an equal number of blue and red marbles than the series in Item 5. Students who initially appeared to be using a "law of averages" strategy on Item 5 (25%) were most likely to also use it on Item 6 (22%), while students who chose response "Blue" and seemed to use a "law of small numbers" strategy on Item 5 (12%), were less likely to use this strategy on Item 6 (8%).

In the next phase of analysis, responses across non-parallel items were compared. Responses on Items 5 and 6 were examined for students who appeared to be consistently using a representativeness heuristic in response to Items 1 and 2. Of these 59 students, only one-fourth also appeared to be using the representativeness heuristic on Items 5 and 6, by choosing responses indicating the "law of averages" or "law of small numbers" strategies.

To determine if there was a consistent frequentist response across items, responses to six of the questions (all but Item 4, which assessed the concept of randomness) were examined. Six of the students (8%) responded correctly to all six items on the pre-test, while 12 (15%) appeared to have consistent frequentist responses on the post-test. Only one student gave a consistent "law of averages" response across the items on both pre- and post-test, and no students gave a consistent "law of small numbers" response on either pre- or post-test.

Changes from pre-test to post-test: Although the number of correct (frequentist) responses improved from pre-test to post-test, there are more interesting results to note. It is apparent that a majority of students changed their responses from pre- to post-test, some from incorrect to correct and some from correct to incorrect. Even those with a consistent pattern of response for pairs of items on the pre-test (e.g. frequentist responses to Items 7 and 8) tended to change their responses from pre-test to post-test.

Twelve percent of the students originally showed a pre-test pattern of sometimes using frequentist strategies and sometimes using law of averages strategies. These students tended to select all frequentist responses on the post-test. This pattern appeared for Items 5 and 6 as well as for Items 7 and 8.

On Item 4, a majority of students (56%) initially chose incorrect answers on the pre-test indicating that "Susan" was more likely to have made up her coin toss results. Students' explanations for this choice indicated that they believe that Susan's sequence had too many long runs and was too far from a 50-50 split of heads and tails. Of these students, almost half switched to selecting "Mary" on the post-test. These students appeared to be influenced by seeing repeated random series of coin tosses simulated on the computer. Only 7 of the 27 students who initially picked "Mary" went from correct to incorrect responses by selecting "Susan" on the post-test.

4. Discussion

There are probably many situations where a single item cannot completely capture a subject's behaviour. In this paper, we have explored a set of items where consideration of response patterns across item pairs provides more information than consideration of responses to the individual items alone. The analytical framework involved more than a score based on the count of correct responses to a family of related problems. Total correct scores can indicate whether or not change has occurred and, perhaps, provide a relative measure for the degree of change, but they generally do not tell us what has changed. The analyses across problem pairs provided not only information on the degree of change, but also permitted the mapping of conceptual changes as inferred from subject response patterns.

Three types of information were provided by consideration of response patterns across pairs of items which were not provided by separate item analysis. First, response patterns provided a measure of response consistency. Consideration of single items under- or over-estimated response frequency or produced conflicting estimates of response frequency. Second, the response patterns measured the extent to which subjects did or did not apply a misconception. Third, response patterns helped identify some of the problem features attended to by subjects. These three types of information gave a more detailed picture of subject responses than produced by an item-by-item analysis.

In contrast to the research literature documenting persistent ways of reasoning about chance events, this study suggests that students do not appear to be using one consistent strategy (e.g. the law of averages) to solve probability problems. Instead, strategies seemed to vary from item to item. One reason may be that the context of problems may obscure the relevant mathematical details. Rather than see that two problems can be solved using identical strategies, students appear to be influenced by the problem content. It did appear that on "easier" problems (less clouded by context) more students used a frequentist strategy, while on more complex problems students would use an incorrect strategy.

Students who gave more consistently frequentist responses on the post-test tended to have given partly frequentist responses on the pre-test. Although these students often seemed to use correct strategies on the pre-test, they were perhaps not yet sure enough of themselves, or were too distracted by the problem content to respond in a consistent way to all problems. The instructional unit appeared to be most helpful for these students in increasing the consistency of their frequentist response on the post-test.

A disappointing result is that the Coin Toss intervention did not appear to affect students' thinking in a systematic way. Although more students appeared to use the frequentist strategy after the unit, and the part-frequentists seemed to become more stable frequentists, many students also changed from one incorrect to another incorrect strategy or from correct to incorrect strategies.

A goal for most teachers is to design instruction to enable students to improve their probabilistic reasoning and to avoid being misled by their intuitions and biases. This study indicates that students may also need special help in learning to avoid being misled by the context of problems in order to use frequentist strategies more consistently. Students need to recognise that it is often better to use the rules and logic of probability rather than rely solely on their own intuition and assumptions which may be incorrect. It is important to determine how we can help students develop a meta-cognitive awareness of how they solve probability problems, to develop new habits of challenging their assumptions, and to give cues to themselves in solving problems. For example, they need to ask themselves: "Is it possible that I'm making an incorrect assumption?" or "Is there information that I'm not using that I should be using?".

Based on recommendations from previous research, the Coin Toss program was designed to repeatedly confront students' intuitions and assumptions. However, it was only a small prototype intervention. The program appeared to have the most impact on students' ability to solve coin-based problems, and did not transfer much to more applied problems. Although many students in this study appeared to learn the basic concepts of randomness, runs, and sample variability in the context of simulated coin tossing, they were often unable to apply this knowledge in solving problems not based on coin tossing. A similar modelling program using various types of data rather than only coins may help overcome this problem. A modelling program with these capabilities is currently being developed as part of the "Chance-Plus" project (Konold, 1989b). While previous research has examined responses to single items, speculating on student mis-conceptions leading to responses chosen, this study indicates that such an approach may be misleading and that it is difficult to predict how students will respond to probability questions based on their responses to related questions.

The results of this study suggest that there is a need for development and evaluation of alternative instructional units designed to confront students' misconceptions, so that students may learn to overcome their biases, ignore incorrect intuitions, and to instead rely more on the basic rules and logic of probability. If we can learn how to confront and change the resistant misconceptions students have and to increase the stability of their correct conceptions, we can better prepare our students to deal with uncertainty in an uncertain world.

Appendix : Reasoning about chance events - items

The first three questions have to do with babies being born at hospitals. For each of these questions, assume that about half the babies born in the world are boys, and the other half, of course, are girls. (An asterisk (*) denotes correct response.)

Question 1. Consider a community hospital in a large suburban area. Below are two possible sequences of births for six babies born at the community hospital (B stands for boy and G stands for girl). They represent the exact order in which the boys and girls might have been born. If about half the newborns are boys and half are girls, which sequence is more likely to represent the last six babies born at the hospital?

a. B G G B G B   b. B B B B G B   c.* Both sequences are just as likely

Question 2. Imagine you are at the same hospital described in Question 1 above. Six new children have been born at the hospital. Again, two possible sequences of boys and girls for the six births are presented below. Which sequence to you think is more likely to have occurred?

a. B G G B G B   b. B B B G G G   c.* Both sequences are just as likely

Question 3. Chisago Lakes is a small city which is about an hour's drive north of Minneapolis. According to statistics from the Minnesota Department of Health, the Chisago Lakes Hospital averaged about 20 births *each month* in 1984. By contrast, the University of Minnesota Hospitals in Minneapolis averaged about 60 births each month in 1984. Here's the question: Over an entire year (12 months), which hospital would you expect to have *more months* where 60% or more of the newborns are boys?

a.* Chisago Lakes Hospital (the small one)
b. University of Minnesota Hospitals (the large one)
c. Both hospitals should have about the same number of months

Question 17. Shelly is going to flip a coin 50 times and record the percentage of heads she gets. Her friend Diane is going to flip a coin 10 times and record the percentage of heads she gets. Which person is more likely to get 80% or more heads?

a. Shelly   b.* Diane   c. Both are just as likely to get 80% or more heads

Question 4. A teacher asked both Mary and Susan to toss a coin 50 times and record every time whether the coin landed heads or tails. For each head that came up they recorded an 'H', and for each tail they recorded a 'T'. Below are the two sets of results.

Mary: T H T H H T T H H T T H T H T H T T H T T H H T H
T T T H H T H H T H H T T H T H H H T T H T T H T

Susan: H T T H H H T H H H H T H T T H H H T T H T T H H
H T T H T T T H H H T H H H H H H T H T H T H T H

Now, one of the women tossed her coin and recorded the actual results. The other woman cheated and made up her results. Which woman cheated, Mary or Susan? Circle the best answer.

a.* Mary cheated   b. Susan cheated

Now, check the circle(s) next to the reason(s) listed below which come close to your reasons for who you think cheated.

MARY

- O Mary's sequence is too regular - it alternates back and forth between heads and tails too much.
- O Mary's sequence doesn't have enough heads - it's too far from a 50/50 split.
- O Mary's sequence comes too close to a 50/50 split between heads and tails.
- O Mary has too many short-runs.

SUSAN

- O Susan's sequence is too irregular - there should be more alternations back and forth between heads and tails.
- O Susan's sequence doesn't have enough tails - it's too far from a 50/50 split.
- O Susan's sequence comes too close to a 50/50 split between heads and tails.
- O Susan has too many long-runs.

O I just made a guess - I really can't say for sure.

Question 7. A father and his son play a game once every day. The father holds a nickel in one hand, places his hands behind his back, and then switches the nickel back and forth between his hands several times. His son then guesses which hand holds the nickel. If the son guesses correctly, he gets to place the nickel in his bank. If the son guesses incorrectly, he wins nothing and must wait until the next day to play again. Over the last 16 days, the son guessed correctly on 5 days and incorrectly on 11 days. What will the son do more often, guess correctly or incorrectly, the next 16 times he plays this game? Circle the best answer.

a. If the son is just guessing, you would expect a 50/50 split between right and wrong guesses. So far the son was right on only 6 out of 16 guesses. I think the opposite trend is bound to start happening soon, so that the son is more likely to be right the next 16 times he plays with his father.

b.* The son should have a 50% chance of guessing the correct hand. Just because he was wrong more often than not out of 16 times does not change the chances. The son is just as likely to guess wrong as to guess right the next 16 times he plays. So, I think he would guess about half right and half wrong.

c. The son just does not seem to be good at guessing which hand is holding the coin. The son's trend seems to be to guess the wrong hand. I think he will tend to do the same thing the next 16 times and guess the wrong hand more often.

d. It appears that the father has an advantage in this game. The son may not be old enough to pick up on his father's strategy for switching the coin from hand to hand, or the father knows his son well and knows how to fool him. I think the son is more likely to be wrong the next 16 times he plays with his father.

Question 8. At a nearby college, half the students are women and half are men. A worker for a student organisation 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 parts of the 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? Circle the best answer.

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 the 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 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.

c.* Half the students on this campus are men and half are women. That means the worker has a 50/50 chance of interviewing a man or a woman. 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.

d. So far, the trend seems to be for 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.

References

delMas, R (1988) *The Effect of Activity-Based Instruction and Directed Evaluation of Predictions on Misconceptions of Probability.* Unpublished doctoral thesis, University of Minnesota.

Kahneman, D, Slovic, P and Tversky, A (1982) *Judgement Under Uncertainty : Heuristics and Biases.* Cambridge University Press, Cambridge.

Konold, C (1989a) Informal conceptions of probability. *Cognition and Instruction* **6(1)**, 1-40.

Konold, C (1989b) *A Computer-Based Curriculum for Probability and Statistics.* University of Massachusetts, Amherst.

Shaughnessy, J M (1977) Misconceptions of probability : an experiment with a small group, activity-based, model-building approach to introductory probability at the college level. *Educational Studies in Mathematics* **8**, 295-316.

Shaughnessy, J M (1981) Misconceptions of probability : from systematic errors to systematic experiments and decisions. In: A Shulte (ed) *Teaching Statistics and Probability, 1981 Yearbook, National Council of Teachers of Mathematics*, 90-100.

Shaughnessy, J (in press) Research in probability and statistics : reflections and directions. In: D Grouws (ed) *Handbook of Research on Mathematics Teaching and Learning.*

Tversky, A and Kahneman, D (1971) Belief in the law of small numbers. *Psychology Bulletin* **2**, 105-110.