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		a. B G G B G B	c. Both
	OR	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)

Reasoning Category	Post-test: 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.