Secondary Students' Concepts of Probability

KEYWORDS:
Teaching;
Misconceptions.

Richard W Madsen
University of Missouri, Columbia, Missouri, USA.

Summary
Students develop concepts of probability without formally studying the discipline and some of their concepts are at variance with those taught in the classroom. A survey of 200 students in five schools in Missouri was undertaken in an attempt to learn about preconceptions. The results are discussed in this article.

◆INTRODUCTION ◆

STUDENTS are sometimes presented with information in the schools as well as through the media that is expressed in probabilistic terms. They may be told that smoking increases the risk of health problems such as lung cancer, emphysema, or heart disease or that using seat belts when riding in a car reduces the risk of severe injury in the case of an accident. They are told that the probability of winning a lottery is small, but they are reminded that it can be done by seeing pictures of happy winners. They are told that unprotected sexual relations can lead to AIDS as well as a variety of sexually transmitted diseases, but that they can practice 'safe sex' if they use protection that has a high probability of being effective. While it is desirable to present this information in probabilistic terms, it is important to know how the information is actually understood by the recipients. That is, it is important to know how the students interpret statements about probability when they have not yet had formal teaching about it. Further, in teaching about probability, our teaching may be more effective if we know what the misconceptions are so that they may be addressed directly.

Green (1979 and 1983) and Truran (1985) have discussed students' understanding of probability concepts for students younger than the group that I have studied. Konold (1991) reports on beliefs of college students about probabilities. He bases much of his work on direct interviews with students who explain to him their thought processes when addressing certain probability problems. If you are interested in a more formal discussion of models of the way students perceive probability, you may wish to read Kahneman, Slovic, and Tversky (1982), Garfield and Ahlgren (1988) or Konold (1989). In these papers there are several models discussed. I will describe some of them briefly and report how the results of the survey I conducted support those models.

One model is based on 'representativeness'. When asked to identify an outcome that is 'most likely' to occur, students who relate probability to representativeness will choose an outcome that appears representative. For example, in a series of 6 coin flips, we know that all 26 sequences of H's and T's are equally likely. Given a choice among several possible sequences, students will often choose an outcome such as HHTHYI as being more likely than HHHHHH. A possible explanation is that since you would expect half heads and half tails in the long run, a sequence such as HHTHT~ with 3 H's and 3T's is more likely to be observed than a run of all 6 H's.

A second model is based on the 'outcome approach'. A student views a probability as a means of predicting the outcome on a single trial and does not view it in terms of relative frequency of occurrence. If a probability greater than 0.5 is assigned to an event, then the student using this outcome approach would say that the event "should" occur and if it does not occur, then the assigned probability is wrong. (Writers in the news media are often guilty of using this interpretation. See Madsen (1991).) For example, if a weather forecast says that there is a 70% chance of rain and if it does not rain, a student using this interpretation will say that the forecast was wrong

A third model is based on 'availability' or on experience. In trying to determine if an event is likely or unlikely (probable or improbable), the response is based on how readily an example of the event comes to mind. Konold (1991) points out that "people may determine the probability of winning a lottery by trying to recall people they know, or know of, who have won. Presumably, this is one reason that state lotteries and sweepstakes like to advertise using the names and photos of past winners."

◆SURVEY OF STUDENTS ◆

In trying to learn about student concepts of probability, I wanted to survey students from a variety of schools. This was in no way a random survey as I worked through teachers who were personal acquaintances. Specifically, I contacted some of my former students who were teaching at the post-elementary level and was able to collect data from five different schools, some in small towns and some in larger cities. The students surveyed ranged in age from 13 to 19. The students were in general math classes and the classes did not include probability or statistics as part of their content. Although I myself am sceptical of data sets based on round numbers (because they seem to be artificial rather than real), it did happen that the survey was completed by exactly 200 students from among the five schools.

The instructions to the students included the following:

"One of the response options is 'I am unsure'. Please check that option rather than making a total guess. However, if you have some idea about an answer, for example if you can eliminate one option as incorrect or unreasonable, feel free to give the answer you feel is most likely the correct one."

These instructions were given to try to make the results more reliable. Random answers based on guesses would not help us learn about the student's understanding of probability.

◆STUDENT RESPONSES ◆

There were 13 questions on this survey, but to conserve space only 11 of them will be given here. The best response is indicated in bold face type. The percent of responses are shown (in parentheses) to the right of each choice. Some comments about the most frequent responses will be made in the next section.

In question 1, nearly half the students chose response (a) or (b), each of which has 3 H's and 3 T's in the sequence. This indicates that many of them might be using the representativeness of the sequence to reflect probability. In question 2, the choice of 6 H's (c) was chosen over half the time, possibly with the idea that it is the least representative and hence least likely. It is interesting to note that 49 of the 88 who answered (e), "one is as likely as the other", to question 1 did **not** give the same response to question 2. In fact 39 of the 88 chose (c) rather than (e). This inconsistency gives an indication that some students do not understand "one is as likely as the other" to mean that each has the same probability. Also, the responses to question 3 indicate that students do not perceive the outcomes on successive flips as independent since only about a third of them chose response (c).

On question 4, I was surprised to see that only 47% correctly chose Gold as their best guess as to what face would come up. It may be that some of those who answered (d), "one is as likely as the other", do not really have in mind the equally likely concept but rather the concept that either is a possibility. Question 5 relates to the same scenario as that of question 4. One would not expect the students to apply the binomial distribution and calculate exact probabilities here. If the students think 'proportionately' they might reason that 5 of 6 faces are Gold, hence 5 of 6 outcomes would be Gold. Those who think of predicting the outcome for each individual toss might reason that Gold is most likely on each toss, hence it is reasonable to predict Gold every time, leading to all 6 coming up Gold. In fact of the 22 who answered that they would be more likely to observe 6 Gold rather than 5 Gold and a Black, 16 had correctly answered on question 4 that Gold was the most likely face.

On question 6,40% used the 'equally likely' outcome principle in spite of the fact that the faces were a different shape than the base (i.e. a non-symmetric solid) and that they did not know the height of the pyramid. On question 7, I would not expect many students to be able to use independence and complements to obtain an answer. By giving the answer that doubling the number of plays doubles the probability of winning at least once, the students have not thought about the extension of that idea to 20 or more plays.

Questions 8 and 10 are story problems in different contexts that allow for possible answers to see if students choose an option that reflects an 'outcome approach' to interpreting probability. I.e. if a stated probability is high (specifically more than 0.50) then persons using this approach believe that the proper interpretation is that it is a prediction that the event will occur. If it does not occur, then the stated probability (or something else) must be wrong. On these two problems 106 students chose one of the options which reflects an outcome approach at least once.

In addition to having difficulty in applying probability to occurrences of single trials, people often have a misunderstanding of interpretation of multiple trials. In question 9, if harmful wastes are present and if Bill the custodian does not use protective gloves even one time, then he will have been unprotected. i.e. the probability of full protection is zero. Certainly if he only wears gloves half the time, then the probability of full protection is zero. However 48.5% of the students chose to give the answer where the probability is one-half of the original value. Similarly in question 11, the assumption is that the nurses would have daily exposures with a 95% probability of being protected each day. With multiple exposures, the overall probability would be much reduced, hence the number (out of 100) who would be protected throughout the year would be considerably fewer than *95.* Only 14% chose this response. Note that if a nurse had exposure for 200 days out of the year, then the probability of protection through the entire year would be $0.95^{200} = 3.5*10^{-5}$.

In reviewing the responses of the students to these questions we find that students with no formal teaching in the area of probability do indeed have some notion of probability. Although students were given the choice of the response "I am unsure" on each question, very few made this choice. We might interpret that as meaning that the students believe that their notion of probability is correct, although their choices of responses indicates otherwise. By having some idea as to what the preconceived ideas of probability are, teachers may be able to show the students where their ideas are inconsistent with the accepted definitions of probability, thus be more effective in their teaching.

References

Garfield, J. and Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics:
Implications for research. *Journal for Research in Mathematics Education,* 19, 44-63.

Green, D.R. (1979). The Chance and Probability Concepts Project, *Teaching Statistics*, 1(3), 66-71.

Green, D.R. (1983). School Pupils' Probability Concepts, *Teaching Statistics*, 5(2), 34-42.

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

Konold, C. (1991). Understanding students' beliefs about probability. In von Glasersfeld, E. (Ed.), *Radical Constructivism in Mathematics Education,* Holland, Kluwer, 139-156.

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

Madsen, Richard (1991). Earthquakes and Snowstorms: Judging Probability Statements, *Stats,*
5, 7-10.

Truran, J. *(1985).* Children's Understanding of Symmetry, *Teaching Statistics*, 7(3), 69-74.

See next page for questionnaire

Questionnaire

1, Say you flip an ordinary quarter several times in succession, with H representing a Head coming up and T representing a Tail. The notation HT means that in two successive flips a Head occurred followed by a Tail. If you flip the quarter 6 times in succession, which of the following sequences are you most likely to observe:
(a) HTHTHT (.285) (b) HHTHTT (.16) (c) HHHHHH (.02) (d) 'ITHTHT (.04)
(e) among (a)-(d), one is as likely as the other (.44) (f) I am unsure *(.055)*

2. If you flip an ordinary quarter 6 times in succession, which of the following sequences are you least likely to observe:
(a) HTHTHT (.075*)* (b) HHTHTT (.025) (c) HHHHHH (.585) (d) TTHTHT (.03) **(e) among (a)-(d), one is as likely as the other** (.24) (f) I am unsure (.045)

3. If you did flip an ordinary quarter 6 times in succession and observed the sequence TTTTTT, then what would you expect to observe on the next flip?
(a) H (.25) (b) T (.345) **(c) among (a)-(b), one is as likely as the other** (.34) (d) I am unsure (.065)

4. An ordinary die is a cube with one face painted BLACK and the other faces painted GOLD. If this die is tossed once, what color would you guess as the one which would come up?
(a) Black (.07) **(b) Gold** (.47) (c) there is not sufficient information to be able to answer (.175) (d) among (a)-(b), one is as likely as the other (.23) (e) I am unsure *(.055)*

5. An ordinary die is a cube with one face painted BLACK and the other faces painted GOLD. If this die is tossed six times, would it be more likely that (I) you would observe Gold all six times or (II) you would observe 5 Gold and 1 Black?
(a) I (.11) **(b) II (.375***)* (c) there is not sufficient information to be able to answer (.135) (d) among (a)(b), one is as likely as the other *(.275)* (e) lam unsure *(.105)*

6. A solid is made in the shape of a small pyramid with the base being a square 2 cm on a side and each of the other four faces being an isosceles triangle. The base is numbered with a 1 and the other faces are numbered 2, 3,4, and 5. If the pyramid is tossed, then the probability that it will come to rest on face number 1 is
(a) 1/5 (.40) (b) 4/3's as likely as any other face (.175) **(c) there is not sufficient information to be able to answer (.115)** (d) I am unsure (.31).

7. In playing a game such as roulette, the probability that you will win if you bet on a certain combination of numbers is 1/20. If you bet that combination on two plays, then the probability of winning at least once is
(a) 1/20 (.155) (b) 2/20 (.505) **(c) .0975 (.045)** (d) .0025 *(.035)* (e) there is not sufficient information to be able to answer (.035) (1) I am unsure (.225).

8. In giving the weather forecast, Linda Frost stated that there was a 90% chance of getting rain the next day. Based on this forecast, you cancelled a planned picnic. However it turned out that the next day was in fact sunny. What does this information tell you about Linda's ability to forecast the weather?
(a) she is very bad as a forecaster (.11) (b) it is likely that she is very bad as a forecaster (.19) (c) she is a good forecaster but happened to be unlucky (.245) **(d) there is not sufficient information to be able to answer (.34)** (e) I am unsure (.115).

9. It is recommended that a custodian use latex gloves when handling trash that might contain harmful wastes. If the trash does indeed contain such harmful wastes, then, if gloves are always used, the probability of full protection during the course of a year is .98. Bill is a custodian who only remembers to use gloves **half the time.** If harmful wastes are present, then the probability of full protection during the course of a year is
(a) .98 (.065) (b) .96 (.045) (c)√1.98 (.055) (d) .49 (.485*)* (e) .02 (.02) (f) .01 (.00) **(g) 0 (.035)** (h) I am unsure *(.285).*

10. Surgical type masks can be used to protect a visitor from exposure to the disease of a patient, such as the disease of infectious tuberculosis. A manufacturer of these masks claims that each time a mask is used it will protect the user from exposure to disease with probability .95. Assume that one of your friends used a mask during a single visit and still contracted a disease. Based on this would you say
(a) the manufacturer's claim was too high. (.25) (b) your friend didn't really use the product or else the disease would not have been contracted (.14) **(c) there is not sufficient information to be able to be sure about the claim (.33)** (d) I am unsure (.28).

11. Assume that the manufacturer's claim (that each time a surgical mask is used it will protect the user from disease with probability .95) is in fact true. If one hundred nurses use masks faithfully when treating patients and if they treat patients daily, then after a year
(a) 95 of them will be protected from the disease (.19) (b) between 92 and 98 of them (approximately 95 ± 3) will be protected from the disease (.28) **(c) the number who are protected from the disease will tend to be considerably fewer than 95 (.14)** (d) I am unsure. (.39)