

Issues in Assessing Conceptual Understanding
in Probability and Statistics

Clifford Konold
Scientific Reasoning Research Institute
University of Massachusetts
Amherst, MA 01003

August 1993

ABSTRACT: Research has shown that adults have intuitions about probability and statistics that, in many cases, are at odds with accepted theory. The existence of these strongly-held ideas may explain, in part, why learning probability and statistics is especially problematic. One objective of introductory instruction ought to be to help students replace these informal conceptions with more normative ones. Based on this research, items are currently being developed to assess conceptual understanding before and after instruction.

KEY WORDS: Instruction; pre-post testing; student intuitions

Presented at the Annual Meeting of the American Statistical Association, San Francisco CA, August 1993. This research was supported by funding from the National Science Foundation (Grant #MDR-8954626). The opinions expressed here are the author's and not necessarily those of the Foundation.

Issues in Assessing Conceptual Understanding in Probability and Statistics

For the past 15 years, I've been part of a research group investigating student understandings of various probabilistic and statistical concepts. We have focused primarily on concepts where the intuitions of adults are at odds with accepted probability or statistical theory. These have included interpretations of probability (Konold, 1989; Konold, 1991; Konold, Pollatsek, Well, Lohmeier, & Lipson, in press), randomness (Falk & Konold, in press; Konold, Lohmeier, Pollatsek, Well, Falk, & Lipson, 1991), conditional probability (Pollatsek, Well, Konold, Hardiman, & Cobb, 1987), the law of large numbers (Well, Pollatsek, & Boyce, 1990), and conceptions of the mean (Pollatsek, Lima, & Well, 1981). The kind of data we have looked at has included simple responses on questionnaire items, complex responses to problems posed in individual interviews, and, most recently, the results of instruction in short tutoring sessions.

During the last five years I have been trying to apply our research findings to the development of software and curricula. In the process, I have found myself struggling with the question on the minds of every curriculum developer and instructor: How do we determine whether curricula are having the intended effect on students thinking? In this paper, I summarize three major findings of our research that I think have important implications for assessment of conceptual development in probability and statistics. Briefly, these are that 1) students come into our courses with some strongly-held yet basically incorrect intuitions, 2) these intuitions prove extremely difficult to alter, and 3) altering them is complicated by the fact that a student can hold multiple and often contradictory beliefs about a particular situation.

Major Findings

- 1. Students have theories or intuitions about probability and statistics prior to instruction, many of which are at odds with conventional thinking.**

In a number of studies beginning in the early 70's, Tversky and Kahneman showed that the typical adult was not a statistician in miniature, that the reasoning of the statistical novice was fundamentally different from that of the expert. They posited a

number of judgment heuristics that guide everyday reasoning under uncertainty. For example, consider the following problem from Kahneman and Tversky (1972):

All families of six children in a city were surveyed. In 72 families the exact order of births of boys and girls was G B G B B G. What is your estimate of the number of families surveyed in which the exact order of births was B G B B B B?

The correct answer is, of course, that the sequences are equally likely. Thus 72 is the best estimate of the frequency of BGBBBB within the families sampled. However, over 80% of these subjects gave estimates of fewer than 72. Kahneman and Tversky (1972) argued that this answer is based on reasoning according to a "representativeness heuristic." According to this heuristic, the likelihood of drawing a particular sample is judged by considering how *similar* the sample is to the parent population – much like judging kinship based on parent-child resemblances. There are two general ways in which BGBBBG is considered highly representative of births. First, in contrast to BGBBBB, it has a near equal number of boys and girls, better reflecting the fact that the individual outcomes are equally likely (or very nearly so). Second, in contrast to BBBGGG it appears more haphazard and hence indicative of the randomness of sex determination.

I have been following a somewhat different line of inquiry, investigating how novices *interpret* probabilities. For example, I've looked at what information college-age students derive from a probabilistic weather forecast. In an interview study (Konold, 1989), several subjects interpreted the forecast "70% chance of rain today" as the prediction "It's going to rain today." I have termed reasoning of this form the "outcome approach." Asked for a probability of some event, people reasoning according to the outcome approach do not see their goal as specifying probabilities that reflect the distribution of occurrences in a sample, but as *predicting* results of a *single* trial.

Given the desire for predictions, outcome-oriented individuals translate probability values into yes/ no decisions. A value of 50% is interpreted as total lack of

knowledge about the outcome, leaving one no justification for making a prediction. As one subject volunteered in Konold (1989):

"If [the forecaster] said 50/50 chance, I'd kind of think that was strange . . . that he didn't really know what he was talking about . . . (p. 68).

Values sufficiently above or below 50% are encoded as "yes" or "no" predictions, respectively, as illustrated in this subject's response to the question of the meaning of the number in the phrase "70% chance of rain."

Well, it tells me that it's over 50%, and so, that's the first thing I think of. And, well, I think of the half-way mark between 50% and say 100% to be like, well, 75%. And it's almost that, and I think that's a pretty good chance that there'll be rain. (Konold, 1989, p. 66).

This decoding of probability values into decisions is illustrated in Figure 1. In this illustration, all values in the white area on the right are mapped onto a "yes" decision, those in the gray area onto "don't know" and those in the white area on the left onto "no." Given this understanding of probability values, outcome-oriented individuals will assert that the forecaster has erred if it fails to rain on the day for which the 70% forecast was made.

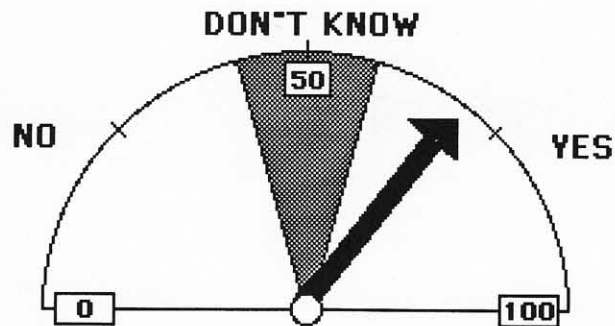


Figure 1. The outcome approach of mapping probability values onto predictions

Below is a related item that we have recently used to assess the impact of instruction on the outcome approach (In Konold & Garfield, 1993, as adapted from Falk, in press, problem 5.1.1, pg. 111).

Weather problem. The Springfield Meteorological Center wanted to determine the accuracy of their weather forecasts. They searched their records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days.

The forecast of 70% chance of rain can be considered very accurate if it rained on:

- a. 95% - 100% of those days.
- b. 85% - 94% of those days.
- c. 75% - 84% of those days.
- d. 65% - 74% of those days.
- e. 55% - 64% of those days.

We have administered this item at the beginning of a number of statistics courses and workshops. Figure 2 shows the frequency of responses to the various options by 199 students.

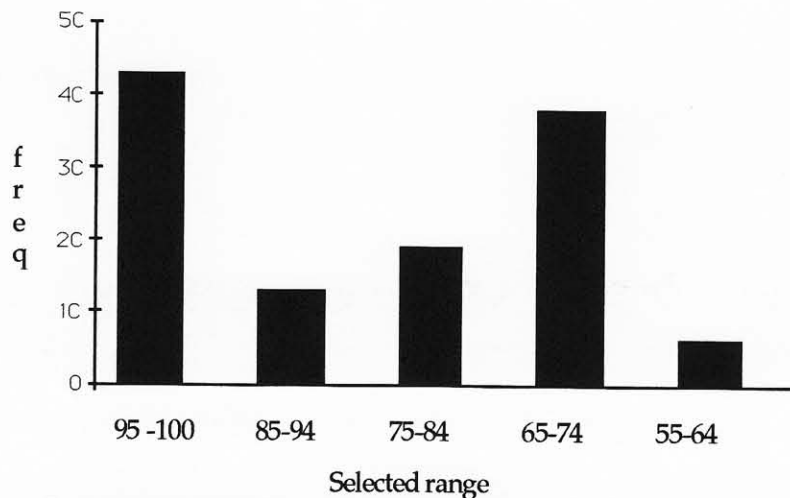


Figure 2. Frequency of pre-instruction responses of 199 students to the weather problem

Roughly 32% of the students selected the correct range 65-74, which includes within it the normative value of 70%. The majority of students (36%), in accord with the outcome approach, selected the highest range, suggesting that they expect rain to occur nearly all of the time when it has been forecast with a 70% probability. For these students, there seems to be little quantitative information conveyed in a probability

value -- a probability of 80% would seem to communicate no additional strength of belief over one of 70%.

Performance on these and similar items demonstrates that many students come into a course on probability and statistics with some strongly-held intuitions about probability and statistics that conflict with what they are about to be "taught." One approach would be to ignore these intuitions and continue teaching much as we currently do, on the belief that we can "overwrite" student beliefs with more appropriate ones. This brings me to our second finding.

2. Students' theories are difficult to alter.

I've heard teachers on the first day of class admonish students "Forget everything you know about x ." They hope for the situation depicted in Figure 3, where the information conveyed by the teacher will be accurately encoded by the students.

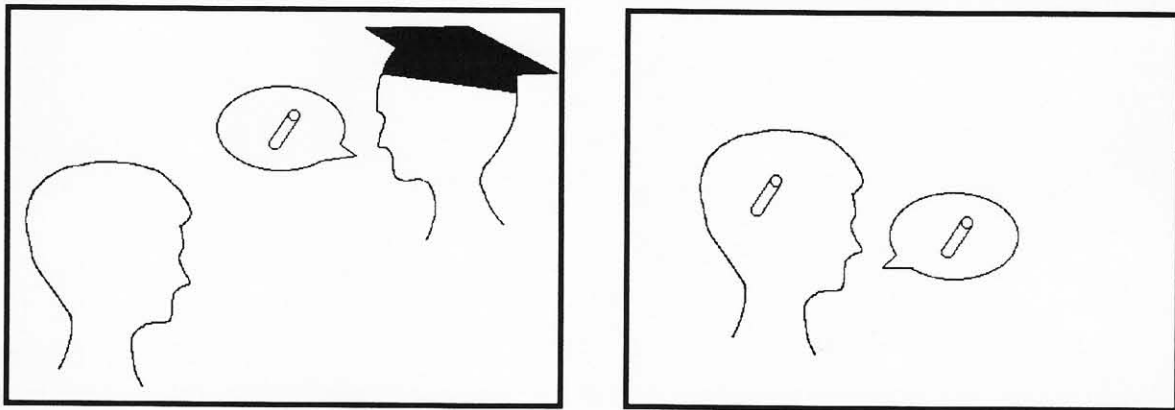


Figure 3. Representation of view of learning in which information to be learned is copied into the mind of the learner.

The "forget-everything" request implies an understanding that prior knowledge can influence what one learns, but reveals a failure to accept that there is no alternative. When presented new information, we have no other option than to relate it to what we already know -- there is no blank space in our minds within which new information can be stored so as not to "contaminate" it with existing information. Learning in the classroom involves the student weaving selected and interpreted teacher outputs into an existing fabric of knowledge. In this way, learning is both limited and, at the same time, made possible by prior knowledge. This "constructive" view of learning, as

depicted in Figure 4, explains the frequent gap between what students report and what we, as teachers, thought we clearly communicated.

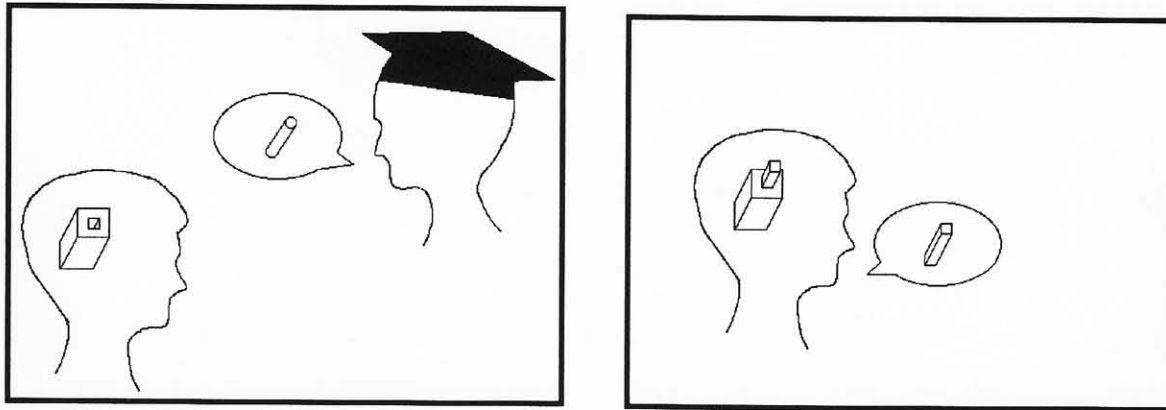


Figure 4. Representation of view of learning in which information to be learned is always interpreted by and integrated with the existing knowledge of the learner.

My research objective has been to try to determine from student outputs, as depicted in the bubble on the right, the form of their theories, as shown in the schematic of the student's head on the left. A useful model of student thinking allows one to interpret and understand the logic of students' statements which might otherwise seem bizarre.

We have a variety of data suggesting that these intuitions are persistent and, to this point, survive our best teaching efforts.

- In most of our questionnaire and interview studies of the beliefs among college students we've found little if any difference among student responses as a function of prior statistics instruction.
- We designed and tested curricula that focuses in particular on conceptual development and have achieved only small to moderate gains over this instruction on most of the concepts we assess. For example, Figure 5 compares the frequency of

various responses to the weather problem both before (black) and after (gray) instruction for the 199 students previously mentioned. The percent of these students responding correctly on this item increased only 6% over instruction.

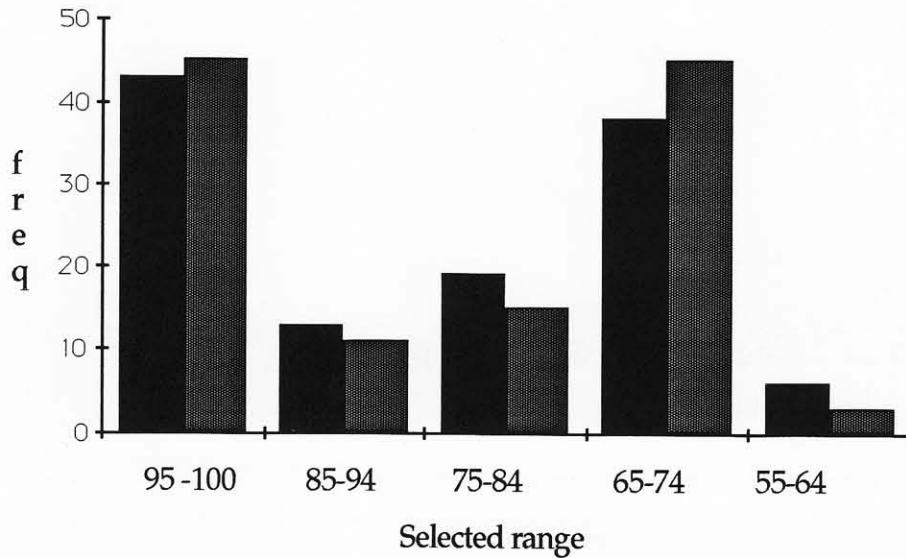


Figure 5. Frequency of pre- (black) and post- (gray) instruction responses of 199 students to the weather problem

- Our efforts to teach basic ideas related to the law of large numbers in short tutoring interviews have been only moderately successful (Konold, Well, Pollatsek, & Lohmeier, 1993).
3. A student can hold multiple and often contradictory beliefs about a particular situation.

One of the instructional approaches we've been testing involves having students articulate their theories and then put them to the test, for example, by conducting computer simulations to determine the relative likelihood of various outcomes of some event and compare these to their expectations. One of the difficulties we have encountered in using this approach is that students have multiple, and often conflicting, perspectives from which they can reason. Furthermore, they can switch among these perspectives in reasoning about a single situation. Consider the following problem:

Coin problem: Part 1. Which of the following sequences is most likely to result from flipping a fair coin 5 times?

- a. HHHTT
- b. THHHT
- c. THTTT
- d. HTHTH
- e. All four sequences are equally likely

Coin problem: Part 2. Listed below are the same sequences of Hs and Ts as listed above. Which of the sequences is least likely to result from flipping a fair coin 5 times?

- a. HHHTT
- b. THHHT
- c. THTTT
- d. HTHTH
- e. All four sequences are equally unlikely

When we first administered part 1 of this problem to undergraduates, we expected the majority of subjects would apply the representativeness heuristic and thus answer *b*. However, roughly 70% of the subjects correctly responded *e*, that the sequences were equally likely. We became suspicious of the reasoning underlying their answer after reading the accompanying written justifications, some of which were of the form "Any of the sequences could occur." Of course this is true, but this is not the same as the claim that they are all equally likely. The next time we administered this problem we included part 2, which appeared on the same page as part 1. In several administrations of this 2 part-item slightly over half of the subjects who select *e* on part 1, do not select *e* on part 2; rather, they indicate that one of the other sequence is "least likely." Clinical interviews with 20 subjects using this two-part item (Konold, et al., in press), suggest that this inconsistency results from subjects applying different perspectives to the two parts of the problem. In part one, many subjects think they are being asked, in accord with the outcome approach, to predict which sequence *will* occur. Because each sequence has such a small probability of occurring, these subjects select answer *e*. They do not, however, mean by this that the sequences have the same probability, but rather that they can't rule out the occurrence of any of them. In part 2 many of these same subjects now apply the representativeness heuristic, suggesting, for example, that the sequence THTTT is less likely because it contains an excess of T's. We speculate that the change in perspective from the outcome approach to the representativeness heuristic is cued by the fact that we normally do not make

predictions about the non-occurrence of events (e.g., which horse will not win the race). Because they cannot readily apply the prediction scheme of the outcome approach to part 2 of the problem, they revert to the representative heuristic.

The results on the coin problems also illustrate the danger of drawing conclusions about the state of a person's knowledge on the basis of performance on a single item, or, for that matter, a series of items.

Some Caveats

In stressing the difficulty of altering various student conceptions, I certainly do not mean to discourage attempts to teach probability and statistics concepts. Rather I want to suggest that we do more than we currently do to determine the effects of instruction on whatever concepts we hope our instruction fosters. Had I not been testing for conceptual development, I'm sure I'd feel the curricula I have designed was magnificent. Designing a set of assessment items has forced me to articulate my major objectives and to evaluate the effectiveness of the materials I'm designing. I hope the curriculum is improving, and I shall find out. While I have been disappointed by the lack of change on the items, I have developed more confidence in the items if for no other reason than that they are not easily taught to.

Let me emphasize that students in the classes I've assessed do reasonably well on in-class tests and quizzes which are given soon after the relevant topics are introduced and typically include items which closely resemble those they have been working on. But I do not regard most teacher-generated quizzes and tests as measures of conceptual change. In courses I teach, I avoid including on quizzes and tests items that are very similar to the ones I use pre- and post-instruction because if I did so I would undermine the usefulness of these items in measuring real rather than apparent conceptual change. I have no doubt that I could, in two weeks, get students performing at high levels on the items I've presented here if that were my sole objective. I would do this by giving them many similar problems and discussing with students the correct solution.

I also should point out that the concepts I have been researching and trying to assess are those which are particularly problematic for students. This is not a sufficient reason to regard these concepts as central objectives in a course on probability and statistics. Certainly there are many concepts and skills that students are learning which

the items I've been using do not address, and I would not recommend that the narrowly conceived items I have discussed become a general indicator of the success of some instruction. More generally, I worry about how items meant to assess achievement or readiness can begin to drive a curriculum, and too often in ways that are detrimental to conceptual change. I would prefer to learn nothing about how student's cognitions might be changing during instruction than, in the process of trying, to undermine my instructional objectives. Indeed, I think many classroom quizzes and tests do just this. In trying to assess competencies at the most rudimentary level, or to not overwhelm students with problems that require legitimate problem solving and reasoning, the teacher inadvertently leads students to believe that routine skills and memorized formulae are the important stuff.

References

- Falk, R. (in press). Understanding probability and statistics: A book of problems. AK Peters: Wellesley, MA.
- Falk, R., & Konold, C. (1994). Random means hard to digest. Focus on Learning Problems in Mathematics, 16, 2-12.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3(3), 430-453.
- Konold, C. (1989). Informal conceptions of probability. Cognition and Instruction, 6, 59-98.
- Konold, C. (1991). Understanding students' beliefs about probability. In E. v. Glasersfeld (Eds.), Radical constructivism in mathematics education (pp. 139-156). Kluwer.
- Konold, C., & Garfield, J. (1993). Statistical reasoning assessment. Part 1: Intuitive thinking. Unpublished manuscript, SRRI, University of Massachusetts, Amherst, MA.
- Konold, C., Lohmeier, J., Pollatsek, A., Well, A., Falk, R., & Lipson, A. (1991). Novice views on randomness. In R. G. Underhill (Ed.), Proceedings of the Thirteenth Annual Meeting of the International Group for the Psychology of Mathematics Education — North American Chapter, 1 (pp. 167-173). Blacksburg, VA.
- Konold, C., Pollatsek, A., Well, A. D., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. Journal for Research in Mathematics Education, 24, 392-414.

- Konold, C., Well, A. D., Pollatsek, A., & Lohmeier, J. (1993). Teaching the law of large numbers via dynamic sampling. Paper presented at the Annual Meeting of PME-NA, Pacific Grove, CA.
- Pollatsek, A., Lima, S., & Well, A. (1981). Concept or computation: Students' misconceptions of the mean. Educational Studies in Mathematics, 12, 191-204.
- Pollatsek, A., Well, A., Konold, C., Hardiman, P., & Cobb, G. (1987). Understanding conditional probabilities. Organizational Behavior and Human Decision Processes, 40, 255-269.
- Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. Journal of Organizational Behavior and Human Decision Processes, 47, 289-312.