ASSESSING THE EFFECTS OF A COMPUTER MICROWORLD ON STATISTICAL REASONING

Robert C. delMas, Joan Garfield, University of Minnesota, USA
Beth L. Chance, University of the Pacific, USA

The Sampling Distributions program and ancillary instructional materials were developed to guide student exploration and discovery. The program provides graphical, visual feedback which allows students to construct their own understanding of sampling distribution behavior. Diagnostic, graphics-based test items were developed to capture students' conceptual understanding before and after use of the program. An activity which asked students to test their predictions and confront their misconceptions was found to be more effective than one based on guided discovery. Our findings demonstrate that while software can provide the means for a rich classroom experience, computer simulations alone do not guarantee conceptual change.

INTRODUCTION

We believe that too often students develop a shallow and isolated understanding of important foundational concepts in a traditional statistics class, and do not develop the deep understanding needed to integrate these concepts and apply them in their reasoning. As teachers who have taught introductory statistics courses for several years, we were particularly disappointed in our students' continual inability to explain or apply their understanding of sampling distributions. We find this lack of understanding particularly troublesome as we see the concept of sampling distributions as crucial to the understanding of statistical inference. Perhaps one reason students experience such difficulty with sampling distributions is that this concept requires the integration of several other ideas in statistics, such as the ideas of a distribution, center, spread, a sample, and chance events. We began to develop our own methods to teach sampling distributions that would facilitate integration of these ideas and enhance understanding of the basic processes involved.

The research reported in this paper was conducted in collaboration among the three authors at their respective institutions. While these three settings do not represent every possible configuration of the college introductory statistics course, we believe they provide enough diversity to allow generalizations. The first setting was an introductory statistics course offered through the Mathematics Department at University of the Pacific, a private university in northern California. The second setting was an introductory statistics course offered through the College of Education at the University of Minnesota.

The third setting was an introductory statistics course offered through the General College, a developmental education college at the University of Minnesota. The General College is an open admissions college that admits nontraditional, underserved, underprepared students who would not normally be admitted to the University. Across the three settings, the courses enroll a range of students from freshmen to seniors majoring in the arts, social sciences, humanities, business, and engineering. Most students enroll in these introductory statistics courses to fulfill a General Education or mathematical reasoning requirement.

Several years ago, one of us (delMas) decided to develop software for the Macintosh computer to illustrate the creation of sampling distributions. While delMas has served as the software's programmer, the other two authors have been instrumental in its development through their use of the program in their respective educational settings. The design of the software is based on recommendations from the literature (Nickerson, 1995; Snir, Smith, and Grosslight, 1995) for conceptually enhanced simulations. The *Sampling Distributions* program and instructional materials were developed to facilitate guided exploration and discovery by allowing students to change the shape of a theoretical population or the size of the samples drawn, and then run a simulation by drawing a large number of random samples. In one window simply labeled "Population" students create a distribution shape by clicking on one of many preset buttons (e.g., normal, skewed, bimodal, uniform, U-shaped). Students can also use up and down arrows to "push" the distribution into any shape they wish. Students can then switch to a window labeled "Sampling Distributions" to draw random samples of a specified size from the population.

INITIAL ACTIVITY

The initial form of the activity was based on ideas from the learning and cognition literature (Holland, Holyoak, Nisbett, and Thagard, 1987; Perkins, Schwartz, West, and Wiske, 1995). Students were instructed to create a normal distribution in the Population window. They were then instructed to switch to the Sampling Distributions window where they changed the sample size in increments from $n = 5$ to $n = 100$, each time drawing 500 random samples. The students recorded the sampling distribution statistics that resulted for each sample size, described the shape, spread, and center of the sampling distributions, and related these observations to the parameters and shape of the population. After completing the last run for $n = 100$, several questions were presented to the students to help them understand the effects of sample size on shape, center, and spread of sampling distributions: What is the relationship between sample size and the spread of the sampling distributions? At what sample sizes do each of the sampling distribution statistics begin to stabilize (not change significantly as the sample size is increased)? Did the sampling distribution statistics provide good, accurate estimates of the population parameters? Overall, did the sampling distribution statistics behave in accordance with the Central Limit Theorem? After running simulations for a population with a normal distribution, students were instructed to repeat the activity for a skewed population and for a population with “an unusual shape.” The activity and questions were intended to direct student attention toward the different pieces of information that are related to the Central Limit Theorem and prompt them to test out their assumptions about the behavior of sampling distributions.

THE ASSESSMENT METHOD

We decided to design graphics-based measurement items to assess the effects of the program and activity on students’ conceptual understanding of sampling distributions. Two of the authors had already developed graphics-based items for testing students’ understanding of statistical power (Garfield and delMas, 1994). Using some of the ideas from this earlier work, a set of five problem situations were developed. Each problem situation consisted of a graph that depicted the distribution for a population and five additional graphs that represented possible distributions of sample means for samples drawn at random from that particular population. Two versions of the test were developed. Students received one version of the test as a pretest prior to the activity and the other version as a posttest after completing the activity.

The five population distributions represented a normal distribution, a positively or negatively skewed distribution, a symmetric bimodal or trimodal distribution, and two irregular distributions. Each problem consisted of two parts. Part A asked the student to select the graph that represented a distribution of sample means for 500 random samples each of a relatively small sample size ($n = 1$ or $n = 4$). Part B asked students a similar question, only the sample size was larger ($n = 9$, $n = 16$, or $n = 25$). The five histograms of possible sampling distributions were designed so that students could display several types of good and erroneous reasoning. For both parts of the question, students selected among statements to indicate the reasons for their choices. The statements were derived from statements made by students during the pilot phase of the test development. Students could also write in a reason that was not represented by the statements.

The initial activity was administered to 79 students who were registered in two different sections of introductory statistics at the private university and 22 students who took introductory statistics at the college of education. Eighty-nine students who gave responses to all pretest and posttest items are used for the analyses.

NEW ACTIVITY

Results from the initial activity demonstrated a significant, positive change from pretest to posttest. There were, however, a significant number of students who did not appear to understand the basic implications of the Central Limit Theorem. We came to recognize that good software and clear directions do not ensure understanding. We believed that the software program incorporated all the needed features to promote conceptual change, so we focused on the nature of the activity.

We looked to the ideas of Posner, Strike, Hewson, and Gertzog (1982) who had developed a model of conceptual change that has been applied in science classroom learning. The model proposes that students who have misconceptions or misunderstandings need to experience an anomaly, or contradictory evidence, before they will change their current conceptions. Ross and Anderson (1982) suggest that effective discrediting experiences are those that 1) require subjects to act upon their beliefs and 2) increase the dissonance between their expectations and observed outcomes.

We sought a way to have students make their own predictions and then test them out using the *Sampling Distributions* program. Our solution was to have the pretest became part of the activity. Students created the population presented for a problem on the pretest, then drew 500 samples at each of the sample sizes stated respectively in parts A and B of the problem. They then identified the graph that looked most like the sampling distribution produced by the program. Students then answered the following three questions about the sampling distribution chosen for part A:

1. How does the shape of the graph you chose compare to the shape of the population?
2. How does the shape of the graph you chose compare to the shape of a normal distribution?
3. How does the spread of the graph you chose compare to the spread of the population?

In addition to these same three questions, students were also asked to respond to the following two questions regarding the sampling distribution for part B:

1. How does the shape of the graph you just chose for problem B compare to the shape of the graph you chose above for problem A?
2. How does the spread of the graph you just chose for problem B compare to the spread of the graph you chose above for problem A?

The new activity placed more emphasis on comparisons of shape and spread than on the recording of parameters and statistics, and required students to make a direct comparison of their pretest "predictions" with the sampling distributions produced by the program.

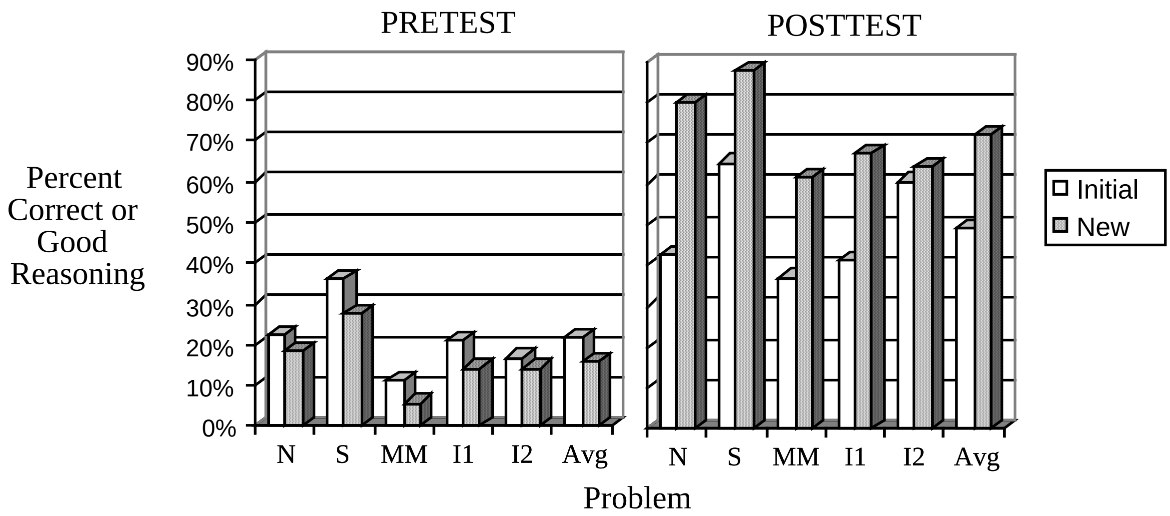
In the spring and fall of 1997, a total of 149 students used the *Sampling Distributions* microworld with the new activity. Thirty-two of the students were enrolled at the private University, 94 took an introductory statistics course through the developmental education college, and 13 took their course through the college of education. Of the 149 students, 141 who gave responses to all items on the pretest and posttest were used in the analyses.

RESULTS

We found evidence of different types of reasoning based on students' choices of sampling distributions. In the first type, "correct reasoning", students chose the correct pair of histograms for a problem. In the second type , "good reasoning", students chose a histogram for the larger sample size that was shaped like a normal distribution and that had less variability than the histogram chosen for the smaller sample size. The histogram chosen for a sample size of 4, however, resembled the population more than a normal distribution. A third type, "larger to smaller reasoning", was represented by students who chose a histogram with less variability for the larger sample size, but both histograms had the shape of the population. A fourth type, "smaller to larger reasoning", was exhibited when students chose a histogram of larger variability for the larger sample size. Comments by some of these latter students indicated that they expected the sampling distribution to look more like the population as the sample size increased. While these four categories covered about 80% to 90% of the responses for each problem, there were additional, less frequent, responses (e.g., choosing the same histogram for both sample sizes).

Figure 1 presents a graphic comparison of results from the initial and new activity. The percentage of students who gave either a correct or good response to each item on the pretest and posttest are displayed. The two groups of students were very similar on the pretest. Posttest percentages for the new activity students, however, were noticably higher than those of the initial activity students. On the average, the spring students went from correct or good reasoning on an average of 16% of the pretest items to an average of 72% of the posttest items compared to a change from 22% to 49% for the initial activity students.

Figure 1. Comparison of results from the initial (N = 89) and new activity (N = 141).



N = normal, S = skewed, MM = multi-mode, I1 = first irregular, I2 = second irregular, Avg = mean percent correct or good reasoning across all five problems.

A multivariate analysis of variances (MANOVA) was conducted for scores based on the Correct or Good criterion. The MANOVA consisted of two within and one between factors: item by test (pretest vs. posttest) by group (initial vs. new activity). A significant three-way interaction [$F(4) = 4.25$, $p = .002$] was found. Separate t-tests produced no statistically significant differences between the initial and new activity on the pretest items. With the exception of problem 5 (the second irregular population), differences between the two groups on the first four posttest items were statistically significant, and all differences were significant at $\alpha = .001$. Consistent with the t-test results, the MANOVA produced a statistically significant interaction between group and test [$F(1) = 32.84$, $p < .001$]. With the possible exception of the fifth problem, the overall results suggest that the new activity group outperformed the initial activity group on the posttest with respect to choosing either a correct or good response for each item.

We have found that a straightforward presentation of the knowledge does not necessarily lead to a sound conceptual understanding of core concepts. We now strongly believe, as is supported in the research literature in other fields, that learning is enhanced by having students become aware of and confront their misconceptions. Students learn better when activities are structured to help them evaluate the difference between their own beliefs about chance events and actual empirical results (e.g., delMas and Bart, 1989). We think the way to do this is with carefully guided experiences that use the types of computer microworlds we've developed and are exploring.

REFERENCES

Garfield, J. and delMas, R. (1994). Students' informal and formal understanding of statistical power. Paper presented at the Fourth Conference of the International Council of Teachers of Statistics (ICOTS4), Marrakech, Morocco, July 1994.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., and Thagard, P. R. (1987). *Induction: Processes of inference, learning, and discovery*. Cambridge, Massachusetts: MIT Press.

Nickerson, R. S. (1995). Can technology help teach for understanding? In D. N. Perkins, J. L. Schwartz, M. M. West, and M. S. Wiske (Eds). *Software goes to school: Teaching for understanding with new technologies*. New York: Oxford University Press.

Perkins, D. N., Schwartz, J. L., West, M. M., and Wiske, M. S. (1995). *Software goes to school: Teaching for understanding with new technologies*. NY: Oxford University Press.

Posner, G. J., Strike, K. A., Hewson, P. W., and Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66(2)*, 211-227.

Ross, L., and Anderson, C. (1982). Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic and A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Snir, J., Smith, C., and Grosslight, L. (1995). Conceptually enhanced simulations: A computer tool for science teaching. In D. N. Perkins, J. L. Schwartz, M. M. West, and M. S. Wiske (Eds). *Software goes to school: Teaching for understanding with new technologies*. New York: Oxford University Press.