Classroom Research: Assessment of Student Understanding of Sampling Distributions of Means and the Central Limit Theorem in Post-Calculus Probability and Statistics Classes

M. Leigh Lunsford
Longwood University

Ginger Holmes Rowell
Middle Tennessee State University

Tracy Goodson-Espy
Appalachian State University

Abstract

We applied a classroom research model to investigate student understanding of sampling distributions of sample means and the Central Limit Theorem in post-calculus introductory probability and statistics courses. Using a quantitative assessment tool developed by previous researchers and a qualitative assessment tool developed by the authors, we embarked on data exploration of our students’ responses on these assessments. We observed various trends regarding their understanding of the concepts including results that were consistent with research completed previously (by other authors) for algebra-based introductory level statistics students. We also used the information obtained from our data exploration and our experiences in the classroom to examine and conjecture about possible reasons for our results.

1. Introduction

Heeding the call of previous researchers (delMas, Garfield, and Chance 1999a), we used a classroom (or action) research model to investigate students’ understanding of concepts related to sampling distributions of sample means and the Central Limit Theorem (CLT). It was our goal to build on previous researchers’ work when implementing our teaching methods and assessing our students (delMas, et al. 1999a, 1999b, 1999c, 2002). We applied a classroom research model to a traditional two-semester post-calculus probability and mathematical statistics sequence taught at a small engineering and science oriented, Ph.D. granting university in the southeast United States with an enrollment of approximately 5,000 undergraduate students. We used assessment data collected on our students to investigate their learning of sampling distributions of means and the CLT, to find areas of student misunderstanding, and ultimately, to improve our teaching of these concepts. In our post-calculus probability and statistics classes, we used teaching methods and assessment tools that were similar to those used by previous authors (delMas, et al. 1999a, 1999b, 1999c, 2002) in introductory level statistics courses. We were interested to see if we would observe in our advanced classes the same types of students’ misunderstandings of sampling distributions of means and the CLT found in introductory level statistics courses (delMas, et al. 1999a, 1999b, 1999c, 2002; Lee and Meletiou-Mavrotheris 2003). Throughout this paper, when we refer to “sampling distribution(s),” we are only considering sampling distributions of the sample mean.

Though our class sizes were relatively small, which is not uncommon in post-calculus probability and statistics courses at many colleges and universities, we hope our results will be none the less interesting to other instructors of these courses. In summary, we concluded that we could not necessarily expect our post-calculus introductory-level probability and statistics students to have good graphical interpretation and reasoning skills concerning sampling distributions even if they appeared to understand the basic theory and were able to perform computations using the theory. We also found that the ability to recall facts about sampling distributions did not imply an ability to apply those facts to solve problems, and thus, our students needed to practice with concepts in order to develop proficiency. In addition, we believed that some of our students confused the limiting result about the shape of the sampling distribution (i.e. as n increases the shape becomes approximately normal, via the CLT) with the fixed (i.e. non-limiting) result about the magnitude of the variance of the sampling distribution, regardless of its shape (i.e. for random samples of size n the variance of the sampling distribution is \( \sigma^2 / n \), via mathematical expectation, where \( \sigma^2 \) is the population variance). Lastly, we discovered that the use of computer simulation, only for demonstration purposes, was not sufficient for developing deep graphical understanding of concepts associated with sampling distributions and the CLT. Thus, in our future classes, we will combine these simulations with well-designed activities or assignments. This paper provides the details of our methods and results along with some changes we intend to implement which will, hopefully, improve the teaching of sampling distributions and the CLT in our post-calculus probability and mathematical statistics course.

2. Background

While there are several philosophical and practical approaches to action research, also called classroom research model, we followed the model applied by delMas, et al. (1999a) to gain “insight into problems, their definitions, and their sources” (p.1). Such an approach to action research is a developmental one in which teachers examine their instructional practices in their own classrooms (Noffke and Stevenson 1995; Feldman and MInstrell 2000). The purpose of
action research, according to Feldman and Minstrell, is two-fold. It is aimed at improving individual “teaching practice in order to improve students’ learning” and improving one’s understanding of the specific educational situations in which one teaches in order to contribute to the knowledge base of teaching and learning (p.432). This form of action research does not claim to develop explanations for all related instructional cases; rather it intends to interpret what was observed in the classroom of the teacher-researcher. In order to demonstrate that the interpretation of the classroom data is valid, one must use triangulation, consideration of alternative perspectives, and testing through practice. In the present case, we achieved triangulation through recording instructors’ field notes, student interviews and surveys, and collection of student reports and tests. Triangulation insures that a variety of data has been collected; however, to extract value from these data they must be analyzed from different perspectives. In our study, the field notes consisted of personal instructional notes concerning particular lessons, notes written within the text, notes written concerning particular activities and overall student reactions to lessons and activities. One of the main purposes of the field notes in this context is to help the instructor record, and thus remember specifically, what was done in a particular lesson. Then later, when one examines assessment data concerning the lesson or data concerning students’ impressions of the effectiveness of the lesson, one may analyze those data in the context of what the instructor believes happened. This also allows the instructor to remember in subsequent semesters, what instructional strategies were used, judge the effectiveness of the strategy, and make decisions about whether to repeat the lesson or modify it. Readers interested in further arguments concerning the validity and reliability of action research designs may consult Ball (2000) and Tobin (2000). Another excellent resource concerning action research may be found in Parsons and Brown (2002).

For our classroom research, we initially analyzed the data by considering our instructional goals during the lessons under analysis. Following our initial analyses, we examined the students’ responses and looked for possible explanations for their responses beyond our instructional goals. Student surveys and interviews were also collected. As these survey and interview results were not shared with the class instructor until after class grades were posted, students were encouraged to be open and honest in their reactions to class activities. Ultimately, our goal was to use our results to inform and improve our teaching of this material. However, we hope that others who teach probability and statistics courses will find our results interesting and useful.

3. Methodology

We will describe our work via the phases of Action Research Model as applied to statistics education in the model presented by delMas, et al. (1999a).

3.1 What is the problem? What is not working in the classroom?

While introductory statistics courses have been the focus of reform curricula and pedagogy (Rossmann, Chance, and Ballman 1999; Garfield 2001), the typical two-semester mathematical probability and statistics sequence has not received the same degree of attention and is generally taught using more traditional methods (Rossmann and Chance 2002). By “typical” we refer to the sequence that consists of mostly probability topics in the first semester, culminating with sampling distributions and the CLT, and mostly statistical concepts in the second semester. Since many students (especially those in the sciences, education, engineering and computer science) will only take the first course of the sequence, there is a need for the injection of more statistical concepts into this course. This especially applies to sampling distributions and the CLT. Due to the generally late, short, and fast coverage of sampling distributions and the CLT in the first semester of our mathematical probability and statistics sequence, students, particularly those who only complete that course, may not develop a deep understanding of these important concepts. We wanted to enhance, assess, and improve our teaching of sampling distributions and the CLT in that course. In addition to incorporating new teaching methods in our classes, we also wanted to measure our students’ understanding of the concepts. We wanted our teaching methods, and future changes to our methods, to be informed by the analysis of our students’ assessment data.

3.2 Techniques to Address the Problems

In the Spring of 2004, as part of a National Science Foundation adaptation and implementation (A&I) grant (Lunsford, Goodson-Espy, and Rowell 2002), we attempted to infuse reform-based pedagogies into the first semester of the mathematical probability and statistics sequence at a public university in the southeast United States by incorporating activity and web-based materials (Rossmann, Chance and Ballman 1999; Siegrist 1997). One of our many goals with this grant was to assess how well our students were understanding sampling distributions and the CLT. We also taught the second course of the sequence during the same semester and thus decided to use some of the same assessment instruments in this course although it was not part of our A&I work. We will refer to these two courses at our institution as Math 300 (Introduction to Probability) and Math 400 (Introduction to Mathematical Statistics).

Although they are both post-calculus mathematics courses, Math 300 is primarily a service class for computer science and engineering majors. In the Spring of 2004, seventeen of the thirty-five students originally enrolled in the course were computer science majors. Only three of the students were mathematics majors with the remainder of the students coming from engineering (seven students) and other disciplines (eight students) including the other sciences, business, and graduate school. The second course of the sequence, Math 400, consisted mostly of mathematics majors. Five of the eight students originally enrolled in the Math 400 course had taken Math 300 the previous semester (Fall 2003) while the remaining students had taken Math 300 even earlier. Math 400 is a required course for mathematics majors seeking a state high school teaching certificate but is an elective for other mathematics majors. Roughly half of the students in the class were seeking licensure. The Spring 2004 Math 400 course contained both typical weak and strong students and thus was not an unusual class in terms of student ability.

In teaching sampling distributions and the CLT in the Math 300 course, we used a traditional text (Hogg and Tanis 2001) along with a computer simulation called Sampling SIM (delMas 2002), and a corresponding activity, Sampling Distributions and Introduction to the Central Limit Theorem, which was slightly modified from one provided by Rossman, et al. (1999) and is an earlier version of an activity by Garfield, delMas, and Chance (2000). Please see Appendix A for a copy of the activity. After an initial in-class demonstration of Sampling SIM, we assigned the activity as an out-of-class group project for which the students turned in a written report. This typed report was a technical document in which students were required to clearly state the problems clearly, explain their approaches, and submit their solutions and conclusions. In order to incorporate more activities into our Math 300 course (as part of our A&I grant), we omitted most of the chapter on Multivariate Distributions (Hogg and Tanis 2001, chap. 5). The coverage of sampling distributions and the CLT occurred near the end of the semester; however, we spent more time on these topics than when we had previously taught the course. This included exposing our students to an application of the CLT by covering the sections of the textbook dealing with confidence intervals and sample sizes for proportions (Hogg and Tanis 2001, sections 7.2, 7.5, and 7.6). These sections were not typically covered in our Math 300 course.

We taught the Math 400 course principally with a lecture method using the same text as Math 300. As with most second courses in the typical post-calculus probability and statistics sequence, the topics covered included an overview of probability topics (including the CLT) and coverage of standard statistical topics (Hogg and Tanis 2001, chapters 7 and 8). We used simulations, including Sampling SIM and simulations from the Virtual Laboratories in Probability and Statistics (VLPS) (Siegrist 1997), exclusively for in-class demonstrations. While we did use the statistical software package, Minihab, for computations, in contrast to the Math 300 course, we did not use any activities in-class or out-of-class nor did we incorporate simulations into any homework assignments. The same professor taught both courses.

3.3 Evidence Collected to Determine Effectiveness of the Implementation.
We endeavored to measure how well students understood sampling distributions and the CLT before and after coverage of the topic in Math 300, and upon entering and completing the Math 400 course. Building on the work of previous researchers, we used a quantitative assessment tool (Appendix B) provided to us by delMas. In addition to containing the same (or similar) graphically-oriented questions used in delMas, et al. (2002), the assessment tool also contained questions of a fact-recollection nature, as well as some straightforward computational questions. We used the assessment tool as both a pretest and a posttest for both courses. In the Math 300 course, we administered the pretest before covering sampling distributions and the CLT. However, unlike delMas, et al. (2002), we did not return the pretest to the students nor did we give the students feedback regarding their pretest performance. We administered the posttest on the last day of class as an in-class quiz. The students had turned in their reports from the activity the previous class period. For the Math 400 course, we gave the pretest at the beginning of the semester and the posttest at the end of the semester. As with the Math 300 students, we did not return the pretests to the students nor were the students given any feedback on their pretest performance. We also gave a qualitative assessment tool (Appendix D), developed by the authors, to the students in both classes at the beginning of the semester and at the end of the semester. This tool measured students’ attitudes and beliefs about several aspects of the course including their use of technology and their understanding of concepts.

From the quantitative assessment tool we collected information about student learning of the sampling distribution and CLT topics via pretest and posttest scores. In addition to seeing how our results compared to previous studies (delMas, et al. 1999a, 1999b, 1999c, 2002), we explored the data to try to describe our students’ understanding and reasoning. We also wanted to determine if there were any noticeable differences in the assessment results of our second semester students versus our first semester students. Thus we examined student responses on the quantitative assessment tool to extract trends and examine reasoning skills that might explain our student responses. Details of these results are given in the next section of this paper.

4. Results

Below we give the details of our results from the quantitative and qualitative assessment tools.

4.1 Overall Scores and Comparisons

The quantitative assessment tool consisted of 27 questions, each of which was assigned 1 point for correct or zero points for incorrect (no partial credit was given). Table 1 below shows a summary of the overall scores for each class.

| Total Number of Questions = 27 | Average Number of Correct Answers (Percent Correct) | Standard Deviation in Number (Percent) |
|--------------------------------|-----------------------------------------------------|----------------------------------------|
| Math 300 Pretest (N = 18)     | 9.9 (36.6%)                                         | s = 2.20 (11.8%)                       |
| Math 300 Posttest (N = 18)    | 18.5 (68.5%)                                        | s = 4.26 (15.8%)                       |
| Math 300 Paired Difference Post - Pre (N = 18) | 8.6 (31.8%)                                        | s = 5.45 (20.2%)                       |
| Math 400 Pretest (N = 7)      | 12.2 (45.0%)                                        | s = 3.81 (14.1%)                       |
| Math 400 Posttest (N = 7)     | 18.9 (69.9%)                                        | s = 2.97 (11.0%)                       |
| Math 400 Paired Difference Post - Pre (N = 7) | 6.7 (24.9%)                                        | s = 2.69 (20.0%)                       |

As expected, we did see significant improvement in student performance from pretest to posttest in both classes. Even with our small sample sizes the paired differences of posttest minus pretest scores were significantly greater than zero for the Math 300 class (t(17) = 6.7, p = 0.0001) and for the Math 400 class (t(6) = 6.6, p = 0.0003). However, while we did see improvement, we were disappointed with the low percentage of correct responses for these mathematically inclined students on the posttest. We were also somewhat surprised to see that the average percent correct on the posttest was not very different between the Math 300 and Math 400 students. This could be a function of several factors including the types of students enrolled in the two courses and when the tests were administered. For the Math 300 students the pretest and posttest were given within a three week period of study of sampling distributions and the CLT. As previously mentioned, the Math 400 students were tested at the beginning and end of the semester, during which time a wide variety of topics were taught. Thus the Math 300 students’ scores may be artificially high due to their recent coverage of the material. It is also interesting to observe that the average score on the Math 400 pretest (given at the beginning of the semester) shows low retention of CLT concepts from the Math 300 course. Lastly, we note that the median increase in the number of questions answered correctly from pretest to posttest as measured by the paired differences was 10.5 out of 27 (38.89%) for our Math 300 and 6 (22.22%) for our Math 400 students.

Because the Math 400 students had an extra semester of statistics, we were curious if there were any noticeable differences between their posttest performance and the posttest performance of the Math 300 students. Given the teaching and assessment methods used in each class, it seemed logical to divide the assessment tool questions into two broad categories: those that were more graphical and those that were more fact recollection and computational. Questions 4 and 5 were the assessment items that were more graphical in nature and most directly related to the Sampling SIM program and the corresponding activity used in Math 300. They were also very similar or the same as the assessment items used when developing and assessing the Sampling SIM program to improve students’ statistical reasoning (delMas, et al. 1999a, b, 2002). In Question 5, the students were given the graph of an irregular population distribution and five possible choices for the histogram of the sampling distribution (assuming 500 samples each of a specified size n). Figure 1 below shows the population distribution and sampling distribution choices (Choices A through E). Question 5a asked students to select which graph represents a distribution of sample means for 500 samples of size n = 4 (Answer = C). Question 5e asked students to select which graph represents a distribution of sample means for 500 samples of size n = 25 (Answer = E). Question 4 was similar but involved a skewed distribution (Appendix B). Questions 4 and 5, each with 7 items, contained 14 of the possible 27 questions on the assessment tool. The 13 remaining questions consisted mostly of items that asked the students to recall facts about sampling distributions of means and/or to apply these facts to perform routine probability computations.

We also wanted to determine if there were any noticeable differences in the assessment results of our second semester students versus our first semester students. Thus we examined student responses on the quantitative assessment tool to extract trends and examine reasoning skills that might explain our student responses. Details of these results are given in the next section of this paper.
Figure 1. Population Distribution (Top Left) and Possible Sampling Distributions (A-E) for the Irregular Population Distribution (please see the Assessment Tool Question 5 in Appendix B).

Table 2 below shows how each class performed on the posttest with the 27 assessment questions divided into the following two categories:

- Graphical: Questions 4 and 5 (7 points each for a maximum score of 14).
- Fact/Computational: The remaining questions on the assessment (a maximum of 13 points).

We found that our Math 400 students actually did quite well on the Fact/Computational type questions answering an average of 11.3 (86.8%) correct with a very small standard deviation of 0.76 (2.81%). This was much better than our Math 300 students who averaged 8.9 (68.9%) correct with a standard deviation of 2.75 (10.19%). We believe that our results in Table 2 make sense in terms of what was covered and emphasized in each of these classes. Because the Math 400 students had more experience throughout the semester applying results about sampling distributions to solve problems, it should not be surprising that they performed well on questions that asked them to recall and apply those results. However, even though we demonstrated the Sampling SIM software and other graphical simulations in-class, our Math 400 students did not seem to be able to extend their knowledge to a more graphical realm.

|                              | Math 300 Post (N = 18) | Math 400 Post (N = 7) |
|------------------------------|------------------------|-----------------------|
| Graphical Questions 4 and 5 Average Number (Percent) Correct with Standard Deviation. (Max Number Correct = 14) | 9.6 (68.3%) s = 2.66 (9.9%) | 7.6 (54.1%) s = 2.93 (10.9%) |
| Fact/Computational Average Number (Percent) Correct with Standard Deviation. (Max Number Correct = 13) | 8.9 (68.9%) s = 2.75 (10.2%) | 11.3 (86.8%) s = 0.76 (2.8%) |
| Average Total Number (Percent) Correct with Standard Deviation. (Max Number Correct = 27) | 18.5 (68.5%) s = 4.26 (15.8%) | 18.9 (69.9%) s = 2.97 (11.0%) |

4.2 Graphical Reasoning, Consistency, and Comparisons

To assess our students’ graphical understanding of sampling distributions and the CLT, we further examined Questions 4 and 5. This enabled us to compare our students’ performance to results from previous studies as well as to examine different types of reasoning skills displayed by our students.

4.2.1. Correct Choice of Sampling Distribution

We first examined how well our students were able to choose the correct sampling distribution in Questions 4 and 5 (parts 4a, 4e, 5a, and 5e of the quantitative assessment tool in Appendix B). By requiring the students to choose the best histogram of the sampling distribution, given a graph of the population and a sample size, these questions were more graphically oriented and thus more similar to the Sampling SIM activity given in the Math 300 class. Question 4 pertained to a skewed right population distribution and Question 5 dealt with an irregular population distribution. Figure 2 below shows
the percent of students in each class that chose the correct sampling distribution by population shape (skewed or irregular) and by sample size (small or large). First, we note that like previous studies (delMas, et al. 1999a, b, 2002), our students had a more difficult time correctly choosing the graph of the sampling distribution when the sample size was small \( n = 4 \) than when the sample size was large, especially in the case of the skewed distribution. For the Math 300 students, the percent of students that correctly identified the sampling distribution for large \( n \) for the skewed and irregular distributions on the pretest was 16.7% and 5.6%, respectively. The percent correct increased to 55.6% and 77.8%, for the skewed and irregular distributions respectively, on the posttest. However, the increase was not as dramatic for small \( n \). For readers interested in comparing pretest and posttest correct responses on a pair wise basis, we have included tables in Appendix C, Section C.4 for this purpose. In contrast, we see little or no improvement in the Math 400 students and consider the posttest percent correct to be disturbingly small (Figure 2). We note that this may be due to sampling variation and the very small number of students assessed in Math 400; however, it is still cause for concern, especially considering these were upper-level mathematics majors.

![Correct Identification of Sampling Distribution](image)

Figure 2. Correct Identification of Sampling Distributions for Math 300 and Math 400 Students.

4.2.2 Reasoning Pair Classifications

Following the work of delMas, et al. (1999a, b, 2000), we next examined reasoning pair classifications as determined by the pretests and posttests for the Math 300 students. The goal was to classify student reasoning about the shape and variability of the sampling distribution as the sample size increased from small to large. For a complete description of the reasoning pair classifications, please see Appendix C, Section C.1. In Table 3 below we show our Math 300 students posttest reasoning pairs for the irregular distribution (Question 5). Please see Figure 1 above for a quick reference to the graphs for this problem. The first column of Table 3 gives the reasoning category as described by delMas, et al. (1999a, b). The second and third columns give the number and percent, respectively, of students whose answer was categorized in this reasoning category. The last column is the reasoning pair given by the students. For example, in the first row we see that the 5 students (27.8%) in the “correct” reasoning category gave the answer that graph C was the sampling distribution for \( n = 4 \) and graph E was the sampling distribution for \( n = 25 \) (i.e. reasoning pair (C, E)). In the second row we had 2 students who answered B and E for \( n = 4 \) and \( n = 25 \), respectively, which we considered a good classification (here we differed with delMas, et al. (1999a) who classified this response as large-to-small normal). In the third row we had 7 students with a reasoning pair of (A, E). Observe that the reasoning pair in the second row is better than the pair in the third row. This is because the students in the second row chose a distribution with less variability than the population for \( n = 4 \), while the students in the third row did not. Note that all of the students in these two rows chose a sampling distribution for \( n = 4 \) with a shape more like that of the population. Thus the reasoning pairs in the first column are essentially ranked from the best (i.e. correct) to worst answers. Lastly, we note that the results in Table 3 were an improvement from the pretest results. Please see Appendix C for more details.

![Correct Identification of Sampling Distribution](image)

Table 3. Math 300 Students Posttest Reasoning Pairs for the Irregular Distribution (Question 5)

| Reasoning Category (delMas, et al. (1999a)) | Number of Students \((N = 18)\) | Percent of Students \((N = 18)\) | Posttest Reasoning Pair Irregular Distribution \((n = 4, n = 25)\) |
|--------------------------------------------|---------------------------------|---------------------------------|---------------------------------------------------------------|
| Correct                                    | 5                               | 27.8%                           | (C, E)                                                        |
| Good (Large-to-Small Normal)               | 2                               | 11.1%                           | (B, E)                                                        |
| Large-to-Small Normal                      | 7                               | 38.9%                           | (A, E)                                                        |
There are several interesting items to note from Table 3. First, of the 6 students (33.3%) who were able to choose the correct sampling distribution for \( n = 4 \), five were also able to choose the correct sampling distribution for \( n = 25 \). Also, all of our 14 students (77.8%) who chose the correct sampling distribution for \( n = 25 \) appeared in the better (top three) reasoning categories. Clearly our students were having a difficult time with choosing the sampling distribution for the small sample size \( (n = 4) \). The most common answer for the sampling distribution for \( n = 4 \) was graph A (please see Figure 1 above). In Appendix C we show the distribution of our Math 300 students’ answer pairs among the reasoning categories for both the pretest and posttest for both the irregular (Question 5) and the skewed (Question 4) population distributions. Our results are consistent with previous studies (delMas, et al. 2002) in: 1) showing improvement in students’ reasoning from pretest to posttest; 2) illustrating students’ difficulties in interpreting the skewed distribution; and 3) demonstrating students’ struggles with finding the correct sampling distribution for \( n = 4 \). Please see Appendix C for more details.

Of interest to us was the large number of our students (12 out of 18) who did not choose the correct sampling distribution for \( n = 4 \). Before we could conjecture about why this was happening, we needed to determine if our students were getting incorrect answers because of graphical misconceptions about shape and/or variability (such as confusing “variability” with “frequency”) or because of some misunderstanding of sampling distributions. Student graphical misconceptions of shape and variability have been studied and documented at the introductory statistics level by several authors (Lee and Meletiou-Mavrotheris 2003, Pfannkuch and Brown 1996).

### 4.2.3 Consistent Graphical Reasoning

To determine if our students were getting incorrect answers because of graphical misconceptions about variability and/or shape or because of some misunderstanding of sampling distributions, we defined and computed a measure of consistent graphical reasoning using the question for the irregular population (Question 5). This essentially measured how well our students could distinguish between the shape and spread of the sampling distribution they chose and the shape and spread of the population. In addition to asking the students to choose a sampling distribution, the assessment tool asked them to compare the variance and shape of the sampling distribution they chose to the population variance and shape. Below we have restated (via renumbering) the questions for Question 5, the irregular distribution, for \( n = 4 \) (see also Figure 1 above and the Assessment Tool in Appendix B):

**Question 5:**

(a) Which graph represents a distribution of sample means for 500 samples of size 4?

(circle one) \( A \) \( B \) \( C \) \( D \) \( E \)

Answer each of the following questions regarding the sampling distribution you chose for Question 5(a):

(b) What do you expect for the shape of the sampling distribution? (check only one)

_____ Shaped more like a NORMAL DISTRIBUTION.

_____ Shaped more like the POPULATION.

(c) Circle the word that comes closest to completing the following sentence:

less more

\( I \) expect the sampling distribution to have the same VARIABILITY than/as the POPULATION.

Because the shape and variability of the sampling distribution graphs for the irregular population were very clear to compare to the population distribution, we computed our consistent graphical reasoning measure based on this population only. A student was defined to demonstrate consistent graphical reasoning if the sampling distribution chosen was consistent with their stated expected variance and shape of the sampling distribution as compared to the population (even if their choice of sampling distribution was incorrect). We called this measure consistent because if the sampling distribution they chose was not the same (in terms of shape and variability) as what they said they expected, then there was some inconsistency in their answer. Please see Appendix C for details on how we computed the number of students that demonstrated consistent graphical reasoning.

In Table 4 below is a comparison of the correct sampling distribution chosen versus consistent graphical reasoning from pretest to posttest for the Math 300 students. We saw significant improvement in the Math 300 students from pretest to posttest for both their selection of the correct sampling distribution and their demonstration of consistent graphical reasoning. Also, while only 33.3% (6 students) correctly identified the sampling distribution for the irregular population with \( n = 4 \) on the posttest, 77.8% (14 students) were consistent in their actual choice for the sampling distribution and their stated expected shape and variance of the sampling distribution as compared to the population. Furthermore, all of the students who were correct were also consistent for \( n = 4 \).

| Large-to-Small Population | 1 | 5.6% | (A, B) |
|---------------------------|---|------|--------|
| Small-to-Large            | 2 | 11.1%| (E, D) or (E, C) |
| Other                     | 1 | 5.6% | (C, D) |

| Sample Size \( n = 4 \) | Sample Size \( n = 25 \) |
|--------------------------|--------------------------|
| Determining the Sampling Distribution form the Irregular Population for Math 300 Students | Pre-Test | Post-Test | Pre-Test | Post-Test |
| % Selecting the Correct Sampling Distribution | 5.6% | 33.3% | 5.6% | 77.8% |
In contrast, for the seven Math 400 students, we saw little or no improvement in consistent graphical reasoning from the beginning of the semester to the end of the semester with at most only three students demonstrating consistent graphical reasoning (for \( n = 25 \) on the posttest) and only one student demonstrating consistent graphical reasoning for the remaining items (\( n = 4 \) pre and post, and \( n = 25 \) pre). While this may be due to the small sample size, these results are depressingly consistent with the low percentage of correct identification of the sampling distribution for the Math 400 students as was previously shown in Figure 2.

We find the consistent graphical reasoning results interesting for several reasons. First, via our NSF grant, we were using activities and graphical devices such as applets throughout the semester in our Math 300 class. Thus we were very surprised at the low percent of our students displaying consistent graphical reasoning on the pretest. Recall that the pretest for this class was administered late in the semester (around the tenth week of class). Thus we expected these students to graphically understand shape and spread and hence be consistent, even if not correct, with their choice of sampling distribution given their stated expected shape and spread of the sampling distribution as compared to the population. However, upon further examination we realized that the Sampling Distributions activity was the only assignment we gave that actually had the students investigating shape and spread in a graphical setting, albeit in the context of learning about sampling distributions and CLT. In the Math 400 class we also demonstrated graphical concepts using simulations and applets (including a teacher demonstration of Sampling SIM when we reviewed the CLT). However, we assigned no student activities during class time or outside of class, and the homework assignments were essentially of a theoretical or computational nature. We conjectured that by assigning the Math 300 students to work through the Sampling Distribution activity using the Sampling SIM applet (instead of the teacher only demonstrating Sampling SIM in class, as was done in Math 400), we enabled the Math 300 students to develop better graphical comparison skills than our Math 400 students.

Based on our results we do not believe that on the posttest the majority of our Math 300 students were having major difficulties with consistent graphical reasoning (such as confusing frequency with variance). Rather it appears that our students had some misunderstandings about sampling distributions. Thus we decided to further examine the reasoning pair rankings in light of our consistent graphical reasoning measure.

### 4.2.4 Comparison of Reasoning Pairs and Consistent Graphical Reasoning

In Table 5 below we looked deeper into our data to try to determine how our students may be misunderstanding concepts about sampling distributions and the CLT. The table shows the detail of how our students’ responses to Question 5 (with \( n = 4 \)) were used to categorize their answers in terms of consistent graphical reasoning. The first two columns of the table are contained in the last three columns of Table 3 above. Recall the reasoning pairs are essentially in order from the best to worst answers. Also recall the first component in the answer pair is the distribution the student chose for the sampling distribution for \( n = 4 \). In the next two columns are the students’ answers when they were asked what they expected for the shape and variability, respectively, of the sampling distribution when \( n = 4 \) compared to the irregular population from which the samples had been drawn. The last column shows the number of students who were classified as showing consistent graphical reasoning. This table allows us to see where the students’ graphical reasoning was incorrect. For the students that were not counted as demonstrating consistent graphical reasoning in the last column, an asterisk has been placed by their answers to Questions 5(b) and 5(c) to show where their reasoning failed. For example, three of the students who had the answer pair (A, E) did not display consistent graphical reasoning because they said they expected the sampling distribution to have less variability than the population (which is correct!) but they chose a sampling distribution (A) that did not have this property.

Using Table 5, we made some observations and conjectures about our students’ understanding of sampling distributions and the CLT. First we observed that of our 9 (50%) students who said they expected the sampling distribution to have a shape more like the population, all had chosen a sampling distribution with this property and were thus consistent in terms of shape. Also, all of these students chose the correct sampling distribution.

| Posttest Reasoning Pair | Number (Percent) of Students (\( N = 18 \)) | \( n = 4 \) Sampling Distribution | Variability of \( n = 4 \) Sampling Distribution | Number with Consistent Graphical Reasoning |
|-------------------------|---------------------------------------------|-----------------------------------|-----------------------------------------------|---------------------------------------------|
| (C, E)                  | 5 (27.8%) Normal                            | 5                                | 5                                            | 5                                           |
| (B, E)                  | 2 (11.1%) Pop.                             | 2                                | 2                                            | 2                                           |
| (A, E)                  | 7 (38.9%) Pop.                             | 7                                | Same                                         | 4                                           |
|                        |                                            |                                  | Less*                                         | 3                                           |
|                        |                                            |                                  |                                               | 0                                           |
| (A, B)                  | 1 (5.6%) Normal*                          | 1                                | More*                                         | 1                                           |
| (E, D) or (E, C)        | 2 (11.1%) Normal                           | 2                                | Less                                         | 2                                           |
| (C, D)                  | 1 (5.6%) Normal                            | 1                                | Less                                         | 1                                           |
| Totals                  | 18                                          | 9                                | 14                                           |                                             |

Table 5. Math 300 Posttest Comparison of Irregular Distribution Reasoning Pairs and Consistent Graphical Reasoning for \( n = 4 \)
distribution for \( n = 4 \). We suspected that many of our students were not recognizing how quickly the sampling distribution becomes unimodal as \( n \) increases. This was not surprising since students are used to thinking of the CLT as a limiting result that doesn’t really take effect until the “magic” sample size of 30. Next we observe that the majority of our students (13 out of 18) correctly stated that they would expect the sampling distribution to have less variability than the population. For the two students who chose E for the sampling distribution, it may have been because they were not able to graphically estimate the magnitude of the standard deviation. For the three students who answered “less” but who chose A for the sampling distribution, we were not sure if they did so because they were either not able to estimate the magnitude of the standard deviation or they may have been confusing variability with frequency (due to the difference in heights of the histogram bars versus the height of the population distribution). Lastly, we thought the four students who answered “the same” (and were thus consistent in their choice of A for the sampling distribution) may be confusing the limiting result about the shape of the sampling distribution (i.e. as \( n \) increases the shape becomes approximately normal, via the CLT) with the fixed (i.e. non-limiting) result about the variance of the sampling distribution, regardless of its shape (i.e. the variance of the sampling distribution is \( \sigma^2 \), via mathematical expectation). As with the shape, these students may be thinking the variability result does not “kick in” until the sample size is greater than 30. Because of the nature of classroom research, we note that these subgroups consist of very small numbers of students and there could be other explanations for what we observed. We believe it would be very interesting to extend this initial research to see if our observations and conjectures above held for other classroom researchers or larger studies. Furthermore, we would want to follow up with personal interviews of students who used inconsistent graphical reasoning skills to gain additional insights into where the disconnect in learning occurs.

These observations led us to consider whether our students really knew and understood the basic result about the variance of the sampling distribution of the sample mean. Question 9d of the assessment tool (Appendix B) gave a quick statistic on this concept: 15 (83%) of our students stated it was “true” that if the population standard deviation equals \( \sigma \) then the standard deviation of the sample means in a sampling distribution (using samples of size \( n \)) from that population is equal to \( \frac{\sigma}{\sqrt{n}} \). All of the students in the top four rows of the table answered “true” to question 9d (except one who did not answer the question). Thus, we believed that our students were able to validate this fact when it was presented to them yet they did not understand it well enough to extend their knowledge to the graphical realm.

### 4.3 Computational Skills versus Theoretical Knowledge

Before looking at the assessment tool results, we believed our Math 300 students would perform well on the computational and theoretical questions requiring the use of the standard deviation of the sampling distribution. This was because these were the items that were most emphasized in the text and homework problems from the text. While we were happy to see an 83% correct response rate to Question 9d (a true/false question about the standard deviation of the sampling distribution), we were dismayed that just 39% and 44% of our Math 300 students were able to correctly compute a probability for a sample mean (sampling from a normal distribution) by using the standard deviation of the sampling distribution and the Empirical Rule (Questions 2 and 3, respectively, of the assessment tool). While this was up from the 11% and 20% correct response rates on the pretest, we wanted to see a larger percent correct for our mathematically inclined students. Again, before we conjectured about why our students had such a low correct response rate to Questions 2 and 3, we had to first determine if they were missing these questions because of a lack of understanding of the Empirical Rule. In particular, did our students know that approximately 68% of the data is located within one standard deviation of the mean for normally distributed data? Based on the data in Table 6 below, we found that our Math 300 students were correctly applying the Empirical Rule. Of the five possible multiple choice answers to Question 2 they only selected two: Answer c which uses the population standard deviation with the Empirical Rule or answer d which uses the sampling distribution standard deviation with the Empirical Rule. However, even when students have correctly identified the theoretical result in Question 9d, many still used the incorrect (population) standard deviation in their computation in Question 2. The results for Question 3 were similar.

| Answer to Question 9d | Answer to Question 2 | TOTAL |
|-----------------------|----------------------|-------|
| (std. dev. of Sampling Distribution is \( \frac{\sigma}{\sqrt{n}} \)) | a, b, or c (uses standard error) | d (uses pop. std. dev.) |
| True correct answer | 0 | 8 | 7 | 15 |
| False | 0 | 3 | 0 | 3 |
| Total | 0 | 11 | 7 | 18 |

We looked at other comparisons of theoretical knowledge of the standard deviation of the sampling distribution versus computational ability using that knowledge. Questions 6 and 9d of the assessment tool were essentially fact recollection questions concerning the magnitude of the standard deviation of the sampling distribution (\( \frac{\sigma}{\sqrt{n}} \)) given the sample size \( n \) versus the magnitude of the standard deviation of the population (\( \sigma \)). Questions 2 and 3 required students to use their knowledge of sampling distribution variability to perform a probability computation using the sample mean. In general, we found our Math 300 students performed well on their recall of theoretical knowledge of the standard deviation of the sampling distribution (11 of 18 answered both 6 and 9d correctly while 17 answered at least one correctly) but were not able to apply that theoretical knowledge (9 out of 18 missed both Questions 2 and 3). We were not sure why our Math 300 students were not correctly applying their theoretical knowledge but we conjectured that it could be because they have not fully realized the concept that “averaging reduces variability” and also did not recognize the sample mean as a random variable with a different standard deviation than the population from which the sampling was done. We also recall from Section 4.2.4 above that our Math 300 students were having a difficult time graphically understanding this concept for small \( n \). We believed this may be due to the coverage late in the semesters of these concepts and thus a lack of practice with the concepts. On the other hand, when we examined these same questions on the posttest of our Math 400 students, we saw that the
majority understood the theoretical results and were able to apply them (6 of 7 answered both 6 and 9d correctly, 5 out of 7 answered both Questions 2 and 3 correctly, and 4 out of 7 answered all four questions correctly). This is also seen by their high score with low standard deviation on the Fact/Computation component of the post assessment tool (see Table 2 above). We should also note here that from the pretest results at the beginning of the semester, we saw poor retention by the Math 400 students of concepts as only 1 of 7 students answered both 6 and 9d correctly while 6 students answered at least one of these correctly, but 4 students missed both Questions 2 and 3 with only 1 student getting all four questions correct. However, by the end of the semester, via repeated application of sampling distribution concepts, we saw that the Math 400 students seem to demonstrate more accurate knowledge and application of the theory.

4.4 Qualitative Results

Students enrolled in the Math 300 class completed two surveys, each given as a take home assignment at the beginning (pre-survey) and end (post-survey) of the semester, and an end-of-course e-mail interview administered by an external evaluator. All students were invited to complete all surveys and participate in the interviews. Twenty-two students participated in both the pre-survey and post-survey and eleven students completed the more open ended and in-depth end-of-course e-mail interviews. As there was some overlap between the post-survey questions and the end-of-course e-mail interviews, some of the students elected not to complete both. Because the end-of-course e-mail interviews were conducted with an external evaluator and the students were informed that the course instructor would only receive a summary of the interviews following the release of their course grades, we hoped that students would feel free to share their honest opinions. The surveys and end-of-course interview questions are included in Appendix D. Schau’s (2003) research has demonstrated relationships between students’ attitudes towards statistics and their learning of statistics. While Schau’s instrument (1999) is an excellent one for evaluating students’ attitudes towards statistics overall, we needed an instrument that would allow us to assess the value of particular activities and instructional strategies associated with our adaptation and implementation of specific curricular materials (Rossman and Chance 1999; Siegrist 1997).

The survey instruments provided information about what class topics the students found to be difficult, which class activities the students valued as contributing positively to their learning experience, and their self perceived understanding of class topics. In the surveys we asked students to rank their current understanding of 65 probability and statistics concepts, such as the Central Limit Theorem, using a Likert scale of 1 to 5 where 1 indicated low knowledge and 5 indicated high knowledge. In Appendix C in Table C.4 we show student responses to topics of relevance for this paper. In general, for topics covered in the Math 300 class, we saw noteworthy increases in student self-assessment of their knowledge of these topics. Topics for which students had an average increase of more than 2 points on the scale (maximum possible increase is 4) included their understanding of: how to find probabilities associated with continuous random variables; mathematical expectation and its properties; the normal distribution; and the Central Limit Theorem. Topics for which the average rating for their understanding was at least 4 points on the post-survey included: how to compute a sample mean versus how to compute the distribution (population) mean; how to compute a sample variance versus how to compute the distribution (population) variance; and the normal probability distribution.

The post-survey also included a section where they were asked a series of open-ended questions concerning their reactions to methods and technologies used to teach the class. The responses to these questions were analyzed for patterns in responses. From the post-survey, students mentioned the CLT activity (13 out of 21, 62%) and group activities and group/individual reports in general (17 out of 21, 81%) as “contributing positively” to their learning. They also believed that the technology used in the class (Minitab and computer simulation activities) helped them learn the material (13/21 or 62%); that methods used in presenting the activities and/or class demonstrations stimulated their interest in the material (13/21); and that the class stimulated their problem solving skills (17/21). As these results indicate, students’ responses to our use of class activities and simulations were generally positive.

The end-of-course interviews with an external evaluator included 18 questions concerning the methods used in the course. Students were asked for their reactions to specific instructional strategies, such as the use of simulations, technology, Minitab, specific activities, as well as balance of lecture versus active student participation in the class. Many respondents to these interviews (7 out of 11) reported that they had the most difficulty understanding topics associated with distributions such as understanding the differences between continuous and discrete distributions; understanding the cumulative distribution function; and understanding some of the standard distributions such as the binomial and exponential distributions. Several students mentioned random variables, including independent random variables, functions involving sums of independent random variables, and the CLT as among the most difficult topics for them to understand.

5. Conclusions, Conjectures, and What Should be Done Next

Below we give some conclusions and conjectures based on our work. Following that we give some ideas, including possible resources, for how we plan to address these issues in our future teaching of these courses. We hope these suggestions will be of use to other instructors of upper level probability and statistics courses.

5.1 Conclusions and Conjectures

1. In our experience, while our post-calculus probability and statistics students were more sophisticated mathematically than our algebra-based introductory-level statistics students, we saw that we could not expect them to have good graphical interpretation, comparison, and reasoning skills concerning sampling distributions even if they understood the basic theory and were able to perform computations using the theory. Just demonstrating graphical concepts in class via computer simulation was not sufficient for our students to develop these skills. We believed our students needed to have a directed (either through an activity or an exercise) hands-on experience (either in-class or out-of-class) with simulations that emphasize graphical representations of distributions (such as Sampling SIM or many of the simulations in the VLP S). This could be thought of as common sense: If one expects one’s students to understand certain concepts, then these concepts must be part of what is emphasized in their assignments.

This was shown by both our Math 300 students in their low pretest consistent graphical reasoning abilities and our Math 400 students in both their pretest and posttest consistent graphical reasoning abilities, and especially on their posttest scores when grouped by category (Table 2). These results were in line with observations by previous authors that computer simulation in itself does not stimulate increased understanding of concepts and thus “instructional materials and activities (need to) be designed around computer simulations that emphasize the interplay among verbal, pictorial and conceptual representation.” (delMas, et al. 1999a paragraph 3.11. Also see delMas, et al. 1999a, paragraphs 2.6-8 and 3.10-11 and Mills 2002 for an overview of the literature).

2. Knowledge of facts about the center, shape, and spread of the sampling distribution of the sample mean did not necessarily result in our students having the ability to apply the knowledge to solve computational problems. We believed our students needed experiences in applying these results in order to develop proficiency.
While our Math 300 students were certainly aware of the basic results regarding sampling distributions, they were not as proficient as our Math 400 students in applying the results to solve computational problems (see Table 2 and Section 4.3). This made sense because the Math 400 students had much more exposure to problem solving using these statistical concepts while the Math 300 students were introduced to the concepts at the end of the semester and thus had only a cursory exposure to problem solving with the concepts.

3. We suspected that many of our students did not fully understand the CLT in that they did not recognize how quickly the sampling distribution becomes unimodal as \( n \) increases.

We saw this when we did a close examination of our Math 300 students reasoning pairs and consistent graphical reasoning. This was not surprising since students were used to thinking of the CLT as a limiting result that doesn’t really take effect until the “magic” sample size of 30.

4. Our students were having a difficult time with variability both graphically and computationally. We believed that some of our students were not able to graphically estimate the standard deviation of a normal distribution or may have been confusing variability with frequency. Also, our students may have been confusing the limiting shape result for sampling distributions (via the CLT) versus the fixed variability result. We believed that our students had not fully realized the “averaging reduces variation” concept. We also believed our students may not fully understand that for a fixed sample size, the sample mean was a random variable and thus had a distribution with a shape, center, and spread. Again, we thought that our students needed to apply these results to develop proficiency.

We saw this particularly with our Math 300 students regarding their graphical reasoning pairs and consistency for small sample sizes and particularly in their application of the Empirical Rule in Questions 2 and 3. However, via more practice with these concepts, our Math 400 students seemed to grasp these concepts better (in the non graphical realm) in terms of knowing and applying the theory.

5. Concepts such as probability distributions and especially sampling distributions and the CLT were difficult to understand, even for post calculus students.

Difficulty in understanding these concepts has been well documented for students in introductory level statistics classes. We must “not underestimate the difficulty students have in understanding basic concepts of probability” and not “overestimate how well (our) students understand basic concepts.” (Utts, Sommer, Acredolo, Maher and Matthews 2003, section 5.2). Although the above quotes referred to algebra-based introductory level statistics, we believed that they are also applicable to post-calculus probability and statistics students. Certainly in our experience and via our classroom research we have found this to be the case. Given that a large number of our students reported difficulty with distributions (7 out of 11 responses on the end-of-course interview), it was not surprising that they also had difficulty understanding sampling distributions and the CLT.

6. Our Math 300 students generally showed a positive response to the use of activities and simulations and believed that they contributed to their learning. We saw this not only in the qualitative results above but also observed this in the classroom.

5.2 What Should Be Done Next

By using the assessment tools we were able to better evaluate what our students were learning and where our students were having difficulties with the concepts. To complete the action research cycle, based on what we have learned about our students’ understanding of these concepts via our assessments, we now offer some ideas, including resources, of what we plan to try in our future classes to improve and assess our teaching.

- To encourage better graphical reasoning skills, we will use activities and exercises early in the Math 300 semester that have students working graphically and computationally with some of the standard distributions they are learning about (especially skewed distributions such as the Poisson and Chi-Square distributions). These activities and exercises should lead students to explore and estimate parameters such as the distribution shape, mean and spread in a problem solving context. Resources for activities and exercises may include Rossman and Chance (2006), Nolan and Speed (2000), and Terrell (1999).

- To encourage better understanding of random variables and their distributions we will increase our use of simulation as a tool in activities and exercises throughout the semester. This will be done on several fronts:
  - Exploring random behavior by generating sample data from known distributions.
  - Exploring the characteristics (such as shape, center, and spread) of distributions of random variables before and after covering them theoretically.
  - Using techniques such as Monte Carlo simulation to approximate solutions to problems before solving them with the theory.

We note that a large percent of our students reported having difficulty with understanding distributions. We hope that by having a distribution-centered approach students will have a clearer understanding of random variables and their distributions before they learn about sampling distributions. In addition to using the FLPS (Siegrist 1997) for in-class demonstrations and assignments, we also think it is important for students to create and run their own simulations via Minitab, Excel, etc. (Rossman and Chance 2006).

- To avoid confusion of key concepts, we need to make sure activities and class presentations are clear and precise regarding the results about sampling distributions. In particular, we would like to improve the activity we used for Sampling Distributions (Appendix A) to make a clearer distinction between the CLT and the expected value and variance of the sampling distribution of the sample mean. We may also add a computational component to the activity or create a new activity or exercise to emphasize these differences.

- To continue using classroom research to improve our teaching, we will find, develop, and use assessment tools to obtain quantitative information regarding our students’ understanding of concepts. Ideally we will use assessment items over the course of several semesters to compare student performance as we try new teaching methods. An excellent source for assessment items is the ARTIST (Assessment Resource Tools for Improving Statistical Thinking) website (Garfield, delMas, and Chance 2002).

We found the action research model useful for assessing our students’ understanding of sampling distributions and the CLT. While we had some preliminary ideas about our students understanding of these topics, it was enlightening to formally explore their understanding through the assessment data. Because of our research and the work of others (i.e. Rossman and Chance 2002), we have also questioned the purpose of our Math 300/400 sequence. What skills do we expect our students to have upon completion of the entire sequence or upon completion of only Math 300? How different are the answers to these questions if we have an “applied” versus a “theoretical” approach to the course? How well does the Math 300/400 sequence prepare our majors to teach high-school statistics, apply statistics in the real world, and/or go to graduate school? What will constitute “best practices” in teaching for achieving our (new) goals in the Math 300/400 sequence? We believe this is an exciting time to be teaching probability and statistics at all levels. We are excited by
the work of our colleagues who are addressing the above questions in their research and development of teaching materials and methods. We also encourage instructors to engage in classroom research to assess how well their teaching is meeting their goals.

Appendix A: Sampling Distributions and Introduction to the Central Limit Theorem Activity

Download Appendix A as a Word document.

(This activity was provided by and slightly modified with permission from Rossman, et al. (1999) and is an earlier version of an activity by Garfield, delMas, and Chance (2000) which can be found on-line at www.tc.umn.edu/~delma001/stat_tools/).

Concepts: Random samples from populations, parameters, statistics, sampling distributions, empirical sampling distributions, the Central Limit Theorem.

Prerequisites: The student should be familiar with random variables, distributions (probability and empirical probability), expected value, and statistics such as the sample mean and sample variance.

Recap: In our last class we used simulation (the software package Sampling Sim) to examine the sampling distribution for the sample mean statistic $\bar{x}$. First we saw that the sample mean $\bar{x}$ is a random variable. Our investigation of the empirical probability distribution of $\bar{x}$ by taking many samples of the same size, $n$, from the same population resulted in the following observations about the sampling distribution of $\bar{x}$:

- **Population Parameters**: mean = $\mu$, standard deviation = $\sigma$;
- **Sample Statistics**: mean = $\bar{x}$, standard deviation = $s$;
- **Observations about the sampling distribution of $\bar{x}$**:
  - shape: Bell shaped (i.e. normal shaped) distribution for “large enough” sample sizes, $n$.
  - center: Distribution of $\bar{x}$ centered at the population mean $\mu$.
  - spread: Spread of $\bar{x}$ depends on sample size, $n$. Spread decreases as $n$ increases (actually spread is $\sigma/\sqrt{n}$).

Our simulation results to compute the sampling distributions for the sample mean statistic $\bar{x}$ illustrated the **Central Limit Theorem**. This theorem says the following about the sampling distribution of the sample mean $\bar{x}$:

- The mean of the sampling distribution of $\bar{x}$ equals the population mean $\mu$, regardless of the sample size or the population distribution, i.e. $E[\bar{x}] = \mu$.
- The standard deviation of the sampling distribution of $\bar{x}$ equals the population standard deviation $\sigma$ divided by the square root of the sample size, regardless of the population distribution, i.e. $Var[\bar{x}] = \frac{\sigma}{\sqrt{n}}$.
- As the sample size gets larger, the shape of the sampling distribution of $\bar{x}$ approaches a normal distribution (i.e. it is approximately normal for “large” sample sizes), regardless of the population distribution, and it IS normal for ANY sample size when the population distribution is normal.

In this take-home activity, you will run the Sampling Sim program to investigate the sampling distribution for $\bar{x}$ and thus see the Central Limit Theorem in action. The first part of this activity takes you through how to use the Sampling Sim program and reviews some basic concepts along the way (parts (a)-(i)). Once you know how to use the simulation and understand what information it is providing, please hand in parts (j)-(q).

To download Sampling Sim, go to the website: www.gen.umn.edu/research/stat_tools/ then click on the Software button and download the proper compressed file for your machine (zip for windows machines). I will also have this software installed on the computer labs in Madison Hall and the MLC.

**Scenario: Professor Lectures Overtime**

Let $X =$ amount of time a professor lectures after class should have ended. Suppose these times follow a Normal distribution with mean $\mu = 5$ min and standard dev $\sigma = 1.804$ min.

a. Draw a rough sketch (and label) this distribution.
b. Is $\mu$ a parameter or a statistic?
c. Suppose you record these times for 5 days $x_1, x_2, ..., x_5$, and calculate the sample mean $\bar{x}$. Is $\bar{x}$ a parameter or a statistic?

to investigate the sampling distribution of these $\bar{x}$ values, we will take many samples from this population and calculate the $\bar{x}$ value for each sample. Open the program Sampling SIM by double clicking on its icon.
Click the DISTRIBUTION button and select “Normal” from the list. You should see a sketch similar to what you drew in (a).

From the Window menu, select “Samples.”

Click Draw Samples and one observation from the population is selected at random (Note: the program may be very slow the first couple of times you click this button). This is one realization of the random variable \( X \).

d. How long did the professor run over this time?

e. Click Draw Samples again, did you observe the same time?

f. Change the value in the Sample Size box from 1 to 5 and click Draw Samples. How does this distribution compare (roughly) to the population distribution?

g. Click Draw Samples again. Did the distribution of your 5 sample values change?

h. Change the sample size from 5 to 25 and click Draw Samples. Describe how this distribution differs from the ones in (f) and (g). How does the shape, center, and spread of this distribution compare to that of the population (roughly)? (The mean of this distribution is represented by \( \bar{X} \), the standard deviation of this distribution is represented by \( s \). Compare these values to \( \mu_A \) and \( \sigma_A \).)

i. Click Draw Samples again. Did you get the same distribution? The same \( \bar{X} \) and \( s \) values?

The main point here is that results vary from sample to sample. In particular, statistics such as \( \bar{X} \) and \( s \) change from sample to sample. You will now look at the distribution of these statistics.

From the Windows menu, select Sampling Distribution. Move this window to the right so you can see all three windows at once. You should see one green dot in this window (it will be small and on the x-axis). This is the \( \bar{X} \) value from the sample you generated in (i). In the Sampling Distribution window, click on “New Series” so it reads “Add More.” Click the Draw Samples button. A new sample appears in the Sample Window and a second green dot appears in the Sampling Distribution window for this new sample mean. Click the Draw Samples button until you have 10 sample means displayed in the Sampling Distribution window. Note: You can click the F button in the Samples Window to speed up the animation. Record the values displayed in the “Mean of Sample Means” box and in the “Standard Dev. of Sample Means” box. These values are empirical. Compare these to the theoretical values predicted by the Central Limit Theorem.

Mean of Sample Means ________________

Standard Dev. of Sample Means ______________

Be very clear you understand what these numbers represent. If not, ask your instructor!

j. In the Population window, click on NORMAL to change the population to one of the following: Bimodal, Skew-, Skew+, Trimodal, U-Shaped, Uniform. Note, this changes the population mean \( \mu_A \) and standard deviation \( \sigma_A \) as well. Please indicate which population distribution you are using and also the population mean and standard deviation.

k. Now change Sample Size to 1 and number of samples to 500 (you definitely want to make sure you have the F button pressed in your Samples window to speed up the animation!). Click the Draw Samples button. Record the following information:

- Describe the shape, center (the mean of the sample means), and spread (the standard deviation of the sample means) of the Sampling Distribution of the \( \bar{X} \) values. In particular, how do the shape, center, and spread compare to the population distribution? You can click the purple population outline (upper left corner of Sampling Distribution window) for easier visual comparison.
- Now click the blue normal outline in the Sampling Distribution window. Which outline (population or normal) appears to be a better description of the sampling distribution of the sample mean \( \bar{X} \) values?

l. Change the sample size to 5 (keep number of samples at 500) and click the Draw Samples button. Give the information asked for in part (k).

m. Change the sample size to 25, click the Draw Samples button, and answer the same questions. Give the information asked for in part (k).

n. Change the sample size to 50, click the Draw Samples button, and answer the same questions. Give the information asked for in part (k).

o. Complete the table below. Are the theoretical values predicted by the Central Limit Theorem (CLT) close to the empirical values you got when you ran the simulations above?

| Sample Size \( n \) | Population Mean | Empirical Mean of Sample Means | Theoretical Mean of Sample Means (via the CLT) | Population Standard Deviation | Empirical Standard Deviation of Sample Means | Theoretical Standard Deviation of Sample Means (via the CLT) |
|---------------------|-----------------|-------------------------------|--------------------------------------------|-------------------------------|-------------------------------------------|-----------------------------------------------------------|
p. Repeat parts (j)-(p) for another non-normal population. Clearly indicate which population you use!

q. Briefly summarize your results in terms of the Central Limit Theorem.

Appendix B: Sampling Distribution Assessment Tool

Download Appendix B as a Word document.

(This document is the same content as the “Sampling Distribution Posttest” from Garfield, J., delMas, R., and Chance, B. (2000), tools for Teaching and Assessing Statistical Inference, hosted at www.tc.umn.edu/~delma001/stat_tools/ and used with permission.)

1. In a geology course, students were asked to determine the weight of rock samples. One instructor asked her students to weigh a rock several times on the same scale. This rock is known to weigh exactly 1000 grams. However, the scale is not completely accurate and sometimes it is off in either direction by 25 grams or less. After a lot of practice, one student weighed the rock 20 times, then computed and recorded the average of the 20 weighings. After a lot of practice, a second student weighed the rock 5 times, then computed and recorded the average of the 5 weighings.

How would you expect the average weight recorded by the first and second student to compare? (circle one)

a. The student who weighed the rock 20 times would have a more accurate average.
b. The student who weighed the rock 5 times would have a more accurate average.
c. Both averages would be equally accurate.
d. It is impossible to predict which average would be more accurate.

2. Weight is a measure that tends to be normally distributed. Suppose the mean weight of all women at a large university is 135 pounds, with a standard deviation of 12 pounds. If you were to randomly sample 9 women at the university, there would be a 68% chance that the sample mean weight would be between: (circle one)

a. 119 and 151 pounds.  b. 125 and 145 pounds.
c. 123 and 147 pounds.  d. 131 and 139 pounds.
e. 133 and 137 pounds.

3. If you took a random sample of 36 university women from the population described in question 2 above, there would be a 68% chance that the sample mean weight would be between: (circle one)

a. 119 and 151 pounds.  b. 125 and 145 pounds.
c. 123 and 147 pounds.  d. 131 and 139 pounds.
e. 133 and 137 pounds.

4. The distribution for a population of test scores is displayed below on the left. Each of the other five graphs labeled A to E represents possible distributions of sample means for random samples drawn from the population.

Population Distribution

4a) Which graph represents a distribution of sample means for 500 samples of size 4? (circle one)  A  B  C  D  E

4b) How confident are you that you chose the correct graph? (circle one of the values below).
Answer each of the following questions regarding the sampling distribution you chose for question 4a.

4c) What do you expect for the shape of the sampling distribution? (check only one)
- Shaped more like a NORMAL DISTRIBUTION.
- Shaped more like the POPULATION.

Circle the word between the two vertical lines that comes closest to completing the following sentence.

4d) I expect the sampling distribution to have | less | the same | VARIABILITY than /as the POPULATION.
| more |

4e) Which graph do you think represents a distribution of sample means for 500 samples of size 16?
(circle one) A B C D E

4f) How confident are you that you chose the correct graph? (circle one of the values below).
- 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95% 100%

Answer each of the following questions regarding the sampling distribution you chose for question 4e.

4g) What do you expect for the shape of the sampling distribution? (check only one)
- Shaped more like a NORMAL DISTRIBUTION.
- Shaped more like the POPULATION.

Circle the word between the two vertical lines that comes closest to completing each of the following sentences.

4h) I expect the sampling distribution to have | less | VARIABILITY than /as the POPULATION.
| more |

4i) I expect the sampling distribution | the same | VARIABILITY than / as the sampling distribution I chose for question 4e to have | more | I chose for question 4a.

5. The distribution for a third population of test scores is displayed below on the left. Each of the other five graphs labeled A to E represent possible distributions of sample means for random samples drawn from the population.

Population Distribution

5a) Which graph represents a distribution of sample means for 500 samples of size 4?
(circle one) A B C D E

5b) How confident are you that you chose the correct graph? (circle one of the values below).
- 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95% 100%

Answer each of the following questions regarding the sampling distribution you chose for question 5a.

5c) What do you expect for the shape of the sampling distribution? (check only one)
Shaped more like a NORMAL DISTRIBUTION.
Shaped more like the POPULATION.

Circle the word between the two vertical lines that comes closest to completing the following sentence.

5d) I expect the sampling distribution to have | less | | more |
      | the same | VARIABILITY than / as the POPULATION.

5e) Which graph do you think represents a distribution of sample means for 500 samples of size 25?
(circle one)     A     B     C     D     E

5f) How confident are you that you chose the correct graph? (circle one of the values below).

20% 25% 30% 35% 40% 45% 50% 60% 70% 75% 80% 85% 90% 95% 100%

Answer each of the following questions regarding the sampling distribution you chose for question 5e.

5g) What do you expect for the shape of the sampling distribution? (check only one)

Shaped more like a NORMAL DISTRIBUTION.
Shaped more like the POPULATION.

Circle the word between the two vertical lines that comes closest to completing each of the following sentences.

5h) I expect the sampling distribution to have | the same | VARIABILITY than / as the POPULATION.
      | more |

5i) I expect the sampling distribution to have the same VARIABILITY than / as the distribution I chose for question 5a

6. The weights of packages of a certain type of cookie follow a normal distribution with mean of 16.2 oz. and standard deviation of 0.5 oz.

Simple random samples of 16 packages each will be taken from this population. The sampling distribution of sample average weight (\( \bar{X} \)) will have:

(CIRCLE ONE)

a. a standard deviation greater than 0.5
b. a standard deviation equal to 0.5
c. a standard deviation less than 0.5
d. It's impossible to predict the value of the standard deviation.

7. The length of a certain species of frog follows a normal distribution. The mean length in the population of frogs is 7.4 centimeters with a population standard deviation of .66 centimeters.

Simple random samples of 9 frogs each will be taken from this population. The sampling distribution of sample average lengths (the average \( \bar{X} \)) will have a mean that is:

(CIRCLE ONE)

a. less than 7.4
b. equal to 7.4
c. more than 7.4
d. It's impossible to predict the value of the mean.

8. Scores on a particular college entrance exam are NOT normally distributed. The distribution of test scores is very skewed toward lower values with a mean of 20 and a standard deviation of 3.5.

A research team plans to take simple random samples of 50 students from different high schools across the United States. The sampling distribution of average test scores (the average \( \bar{X} \)) will have a shape that is:

(CIRCLE ONE)

a. very skewed toward lower values.
b. skewed toward lower values, but not as much as the population.
c. shaped very much like a normal distribution.
d. It's impossible to predict the shape of the sampling distribution.

9. Consider any possible population of values and all of the samples of a specific size (n) that can be taken from that population. Below are four statements about the sampling distribution of sample means. For each statement, indicate whether it is TRUE or FALSE.

a. If the population mean equals \( \mu \), the average of the sample means in a sampling distribution will also equal \( \mu \).
   TRUE    FALSE
b. As we increase the sample size of each sample, the distribution of sample means becomes more like the population.
   TRUE    FALSE
c. As we increase the sample size of each sample, the distribution of sample means becomes more like a normal distribution.
   TRUE    FALSE
d. If the population standard deviation equals \( \sigma \), the standard deviation of the sample means in a sampling distribution is equal to \( \frac{\sigma}{\sqrt{n}} \).
   TRUE    FALSE

The distribution for a population of measurements is presented below. Suppose that ten values are going to be sampled from this population and the sample mean calculated. Some possible values for this sample mean are 1, 6, 8, and 10.
10. Which of the four possible sample mean values is MOST likely to be calculated? (circle only one)
   a. 1  
b. 6  
c. 8  
d. 10  

11. Which of the four possible sample mean values is LEAST likely to be calculated? (circle only one)
   a. 1  
b. 6  
c. 8  
d. 10  

12. Looking at the graph above, what would you guess to be the value of \( \mu \), the population mean?
   a. 0.0  
   e. 2.0  
   i. 4.0  
   m. 6.0  
   q. 8.0  
   u. 10.0  
   b. 0.5  
   f. 2.5  
   j. 4.5  
   n. 6.5  
   r. 8.5  
   v. 10.5  
   c. 1.0  
   g. 3.0  
   k. 5.0  
   o. 7.0  
   s. 9.0  
   w. 11.0  
   d. 1.5  
   h. 3.5  
   l. 5.5  
   p. 7.5  
   t. 9.5  
   x. 11.5  

*This tool was provided to the authors by Robert delMas.

Appendix C: Data Analysis Details

Download Appendix C as a Word document.

C.1. Classification of Reasoning Categories for Questions 4 and 5 of the Sampling Distribution Assessment

Using the Assessment Tool (see Appendix B) provided by Garfield, delMas, and Chance (2000), we classified students reasoning skills according to their answer pairs on Questions 4 and 5. The answer pairs for Question 4 consisted of their answers to 4a and 4e and similarly for Question 5. These are the same classifications as delMas, et al. (2002) which are a refinement of the “reasoning pairs” defined in delMas, et al. (1999a) by essentially separating the “good” responses into “good” and “L-S Normal.” Below is a description of the classifications. Tables C.3 and Table C.4 below show how all possible answer pairs are classified for Questions 4 and 5, respectively, and Figure C.1 below shows graphical examples of the different types of reasoning for Question 5 of the Assessment Tool.

C.1.1 Description of Classifications

- **Correct:** The correct graph is chosen for the sampling distribution for the small sample size and also for the large sample size.

- **Good:** Students select answer pairs in which the standard deviation of the sampling distributions chosen decreases from large to small as the sample size increases; the sampling distribution chosen for the large sample size has a normal shape but the sampling distribution chosen for the small sample size has a shape more like the population. However, the sampling distribution chosen for the small sample size must have a standard deviation less than the population distribution for an answer pair to be classified in this category.

- **Large to Small Normal:** As in the Good category, students choose answer pairs in which the standard deviation of the sampling distributions chosen decreases from large to small as the sample size increases; the sampling distribution chosen for the large sample size has a normal shape but the sampling distribution chosen for the small sample size has a shape more like the population. However, the distribution chosen for the small sample size may now have a standard deviation the same as the population.

- **Large to Small Population:** Students choose answer pairs in which the standard deviation decreases from large to small as the sample size increases, but the shape of the sampling distributions resemble the population distribution for both large and small sample sizes.

- **Small to Large:** Students choose answer pairs in which the standard deviation of the sampling distributions increases as the sample size increases.

- **Same:** Students choose answer pairs in which the standard deviation of the sampling distributions are the same as the sample size increases.

- **Other:** Any answer pair not in one of the above categories.

C.1.2 Classification Tables and Example Figure

Figure C.1 below gives graphical examples for category classification for Question 5.
Figure C.1. Examples of pairs of graphs for each type of reasoning for Question 5 (Figure 9 of delMas, et. al. (2002), used with permission).

Table C.1 and Table C.2 below show how the answer pairs for Questions 4 and 5 were classified into reasoning pairs. The rows are the possible answers for the sampling distribution chosen for the smaller sample size and the columns are the sampling distribution chosen for the larger sample size. In parentheses by each answer, we state the characteristics of the sampling distribution, in terms of its shape and standard deviation as compared to the population.

| Answer Pair Classification (4a, 4e) | 4a Answer (n = 4) | 4e Answer (n = 16) |
|-----------------------------------|------------------|-------------------|
| A (Population/Same)               | Same             | L-S Pop           |
| B (Population/Smaller)            | S-L              | Good              |
| C (Normal/Smallest)              | S-L              | Same              |
| D (Approx Normal/Smaller)         | S-L              | Correct           |
| E (Approx Pop/Same)               | L-S Pop          | L-S Normal        |

Table C.1. Question 4 Answer Pairs with Category Classification

| Answer Pair Classification (4a, 4e) | 4a Answer (n = 4) | 4e Answer (n = 16) |
|-----------------------------------|------------------|-------------------|
| A (Population/Same)               | Same             | L-S Pop           |
| B (Population/Smaller)            | S-L              | Good              |
| C (Normal/Smallest)              | S-L              | Same              |
| D (Approx Normal/Smaller)         | S-L              | Correct           |
| E (Approx Pop/Same)               | L-S Pop          | L-S Normal        |

Table C.2. Question 5 Answer Pairs with Category Classification
Below in Figure C.2 are graphs showing the percent of Math 300 students in each reasoning category (both pretest and posttest results) for both the skewed and the irregular populations (Questions 4 and 5, respectively). The figure also shows the distribution of the twenty-five possible answer pairs for each question (assuming equally likely answer pairs). Note that none of the twenty-five possible reasoning pairs for the skewed distribution were classified as other. The percent of students who were classified as having correct or good reasoning increased from 16.7% to 33.3% for the skewed population and from 11.1% to 27.8% for the irregular population. If we examine the students who were classified as correct, good, or large to small normal, then we see an increase from 27.8% to 61.1% for the skewed distribution and 11.1% to 77.8% for the irregular distribution. We note that our students seemed to have a more difficult time with the skewed distribution than the irregular distribution. In addition, we see a large decrease in the percent of students that show incorrect reasoning (small-to-large, same, or other) from pretest to posttest (55.6% to 22.2% for the skewed population and 77.8% to 16.7% for the irregular population). These trends and difficulties are consistent with previous results (delMas, et al., 2002).

Of interest to us was the large percent of our students that were classified as having “large-to-small normal” reasoning for the irregular population. Since 78% of our students chose the correct sampling distribution for the large sample size for the irregular distribution (see Section 4.2.1 above), it appears that our students were not demonstrating correct or good reasoning because of their choice of the sampling distribution for the small sample size.

Below in Figure C.2 are graphs showing the percent of Math 300 students in each reasoning category (both pretest and posttest results) for both the skewed and the irregular populations (Questions 4 and 5, respectively). The figure also shows the distribution of the twenty-five possible answer pairs for each question (assuming equally likely answer pairs). Note that none of the twenty-five possible reasoning pairs for the skewed distribution were classified as other. The percent of students who were classified as having correct or good reasoning increased from 16.7% to 33.3% for the skewed population and from 11.1% to 27.8% for the irregular population. If we examine the students who were classified as correct, good, or large to small normal, then we see an increase from 27.8% to 61.1% for the skewed distribution and 11.1% to 77.8% for the irregular distribution. We note that our students seemed to have a more difficult time with the skewed distribution than the irregular distribution. In addition, we see a large decrease in the percent of students that show incorrect reasoning (small-to-large, same, or other) from pretest to posttest (55.6% to 22.2% for the skewed population and 77.8% to 16.7% for the irregular population). These trends and difficulties are consistent with previous results (delMas, et al., 2002).

Of interest to us was the large percent of our students that were classified as having “large-to-small normal” reasoning for the irregular population. Since 78% of our students chose the correct sampling distribution for the large sample size for the irregular distribution (see Section 4.2.1 above), it appears that our students were not demonstrating correct or good reasoning because of their choice of the sampling distribution for the small sample size.

| Answer Pair Classification (5a, 5e) | 5a Answer (n = 4) | B (Population/Smallest) | C (Approx Normal/Smaller) | D (Population/Smaller) | E (Normal/Smallest) |
|-----------------------------------|-------------------|-------------------------|--------------------------|------------------------|---------------------|
| A (Population/Equal)              | Same              | L-S Pop                 | L-S Normal               | L-S Pop                | L-S Normal          |
| B (Population/Smallest)           |                   | Same                    | S-L                      |                        |                     |
| C (Approx Normal/Smaller)         | S-L               |                         | Same                     | Other                  | Correct             |
| D (Population/Smaller)            | S-L               | L-S Pop                 | Good                     | Same                   | Good                |
| E (Normal/Smallest)              |                   | Other                   | S-L                      | S-L                    | Same                |

C.2. Classification of Consistent Graphical Reasoning in Choice of Sampling Distribution for Question 5.

Below in Table C.3 and Table C.4 we show the determination of “consistent graphical reasoning” for Question 5 of the Assessment Tool (Appendix B). In the rows we have the stated variance (as compared to the population) and the stated shape of the sampling distribution. In the columns are the possible choices for the sampling distribution. For example, in Table C.3 below we see the data for the Math 300 pre- and post-tests for Question 5a (determining the
sampling distribution for $n = 4$). The values in the cells represent the number of students. Pre-test data is in parenthesis below post-test data. Thus we can see that on the post-test, 6 of the 18 students chose the correct distribution. Cells with an asterisk indicate that the student displayed “consistent graphical reasoning” even if their answer was not correct. Thus we see that for the post-test, of the 4 students that stated the variance of the sampling distribution would be the same as the population and the shape of the sampling distribution would also be the same as the population, all chose the sampling distribution that was consistent with their stated variance and shape (although it is not correct). Thus we see that 14 of the 18 students displayed “consistent graphical reasoning” on the post-test although only 6 of the 18 actually had the correct answer.

| Stated Variance of Sampling Distribution | Stated Shape of Sampling Distribution | 5a - Sampling Distribution Chosen for Irregular Population $n = 4$ (Variance of Sampling Distribution Compared to Population) | Totals |
|------------------------------------------|--------------------------------------|-------------------------------------------------|--------|
| Less than Population Normal              | (1)                                  | 6*                                              | 8 (3)  |
| Population                               | 3                                    | 2* (1)*                                         | 5 (1)  |
| Same as Population Normal                | (1)                                  | *                                               | 0 (4)  |
| Population                               | 4*                                   | (1)                                             | 4 (1)  |
| More than Population Normal              | (1)                                  | 6                                               | 2 (8)  |
| Population                               | 8                                    | 2 (5)                                           | 18 (18)|
| Totals                                   | 26                                   | 12                                              | 18 (18)|

Cells with an asterisk indicate that the student displayed “consistent graphical reasoning” even if their answer was not correct.

| Stated Variance of Sampling Distribution | Stated Shape of Sampling Distribution | 5e - Sampling Distribution Chosen for Irregular Population $n = 25$ (Variance of Sampling Distribution Compared to Population) | Totals |
|------------------------------------------|--------------------------------------|-------------------------------------------------|--------|
| Less than Population Normal              | (1)                                  | 1*                                              | 17 (2) |
| Population                               | (1)                                  | *                                               | 0 (1)  |
| Same as Population Normal                | (2)                                  | *                                               | 0 (3)  |
| Population                               | *                                    | *                                               | 0 (2)  |
| More than Population Normal              | (2)                                  | (1)                                             | 0 (4)  |
| Population                               | (1)                                  | 1 (1)                                           | 1 (6)  |
| Totals                                   | 0                                    | 1 (2)                                           | 14 (1) |

Cells with an asterisk indicate that the student displayed “consistent graphical reasoning” even if their answer was not correct.
C.3 Details of Qualitative Results

Table C.5 below shows the results of two qualitative surveys given to our Math 300 students. The pre-survey was administered at the beginning of the Spring 2004 semester and the post-survey was administered at the end of the semester. A Likert scale was used to assess student self-perception of their understanding of class topics. Students were asked to rate their knowledge of a given topic with a 1 indicating low knowledge and a 5 indicating high knowledge. In the table we show the topics, the average knowledge level reported on the post-survey by 22 students, the average difference between the post-survey and the pre-survey, and the standard deviation of the pair wise differences. Note that we are only showing topics relevant to this paper and that we actually assessed our students on more than 60 topics, some of which were not included in the course.

| Topic (Sample Size of 22 Students)                                                                 | Post Avg | Avg. Diff. (Post - Pre) | Paired Std. Dev |
|---------------------------------------------------------------------------------------------------|----------|-------------------------|-----------------|
| My understanding of how to compute a sample mean versus how to compute the distribution (population) mean. | 4.3      | 1.7                     | 1.2             |
| My understanding of how to compute a sample variance versus how to compute the distribution (population) variance. | 4.0      | 1.6                     | 1.6             |
| My understanding of empirical probability distributions versus probability distributions.            | 3.9      | 1.8                     | 1.3             |
| My understanding of how to find probabilities associated with discrete random variables.           | 3.7      | 1.9                     | 1.2             |
| My understanding of how to find probabilities associated with continuous random variables.          | 3.8      | 2.0                     | 1.2             |
| My understanding of Bernoulli trials and their associated distributions.                            | 3.4      | 1.6                     | 1.6             |
| My understanding of the Normal Probability Distribution.                                           | 4.1      | 2.1                     | 1.3             |
| My understanding of the difference between cumulative probability functions and probability mass/density functions. | 3.0      | 1.5                     | 1.6             |
| My understanding of Mathematical Expectation and its properties.                                   | 3.5      | 2.1                     | 1.2             |
| My understanding of independent random variables.                                                  | 3.9      | 1.8                     | 1.1             |
| My overall understanding of random variables.                                                      | 3.6      | 1.4                     | 1.2             |
| My understanding of the Central Limit Theorem.                                                     | 3.8      | 2.2                     | 1.5             |
| My understanding of Point Estimation.                                                              | 3.1      | 1.3                     | 1.6             |
| My understanding of Confidence Intervals for Proportions.                                           | 3.2      | 1.4                     | 1.4             |
| My overall understanding of probability.                                                            | 3.4      | 1.2                     | 1.0             |
| My overall understanding of statistics.                                                             | 3.0      | 1.0                     | 1.2             |

C.4 Paired Results for Correct Identification of Sampling Distribution

Below are four two-way tables which show the paired results for our Math 300 students for correct sampling distribution identification for Problems 4a, 4e, 5a, and 5e. The rows give the number of students who correctly or incorrectly identified the sampling distribution on the pretest and the columns give the number of students who correctly or incorrectly identified the sampling distribution on the posttest.

### Table C.6 Identification of Correct Sampling Distribution for Question 4a – Skewed Distribution with Small Sample Size

| Question 4a       | Posttest | Totals |
|-------------------|----------|--------|
|                  | Correct ID | Incorrect ID |     |
| Pretest           |           |           |     |
| Correct ID        | 0         | 0        | 0    |
| Incorrect ID      | 3         | 15       | 18   |
| Totals            | 3         | 15       | 18   |

### Table C.7 Identification of Correct Sampling Distribution for Question 4e – Skewed Distribution with Large Sample Size
Table C.8 Identification of Correct Sampling Distribution for Question 5a –Irregular Distribution with Small Sample Size

|       | Pretest | Posttest | Totals |
|-------|---------|----------|--------|
| Correct ID | 1       | 2        | 3      |
| Incorrect ID | 9      | 6        | 15     |
| Totals | 10      | 8        | 18     |

Table C.9 Identification of Correct Sampling Distribution for Question 5e –Irregular Distribution with Large Sample Size

|       | Pretest | Posttest | Totals |
|-------|---------|----------|--------|
| Correct ID | 0       | 1        | 1      |
| Incorrect ID | 6      | 11       | 17     |
| Totals | 6       | 12       | 18     |

Appendix D: Surveys and Interview Questions

Download Appendix D as a Word document.

Date: ____________________________ Current Class/Semester: ____________________________

Please list all probability and/or statistics courses that you have taken (other than this course) or write none if you have not had any previous courses of this type. If you have had some other exposure to probability and/or statistics please indicate that below.

Course Number & Name When & Where Taken Grade

Other Exposure to Probability and/or Statistics:

Please rate your knowledge of the given topic with 1 indicating low knowledge and 5 indicating high knowledge. It is important that you answer as honestly as possible. Some of these topics you may have not have been exposed to very much or at all.

1. My understanding of randomness.
   
   1 2 3 4 5

2. My understanding of how to list a sample space.
   
   1 2 3 4 5
3. My understanding of how to compute a sample mean versus how to compute the distribution (population) mean.

4. My understanding of how to compute a sample variance versus how to compute the distribution (population) variance.

5. My understanding of why the factor $n-1$, where $n$ is the sample size, appears in the denominator of the sample variance formula.

6. My understanding of basic set notation.

7. My understanding of how to create, read, and use Venn diagrams.

8. My understanding of empirical probability distributions versus probability distributions.

9. My understanding of how to calculate the probability of an event.

10. My understanding of combinations and permutations.

11. My understanding of how to use counting techniques in probability problems.

12. My understanding of hypergeometric probability problems.

13. My understanding of conditional probability.

14. My understanding of the independence of events.

15. My understanding of the Multiplicative Law of Probability.

16. My understanding of the Additive Law of Probability.

17. My understanding of the Law of Total Probability.

18. My understanding of Bayes’ Theorem.

19. My understanding of the definition of a Discrete Random Variable.

20. My understanding of how to find the expected value of a discrete random variable.

21. My understanding of how to find probabilities associated with discrete random variables.

22. My understanding of how to find probabilities associated with continuous random variables.
23. My understanding of the Binomial Probability Distribution.

24. My understanding of the Geometric Probability Distribution.

25. My understanding of the Poisson Probability Distribution.

26. My understanding of Bernoulli Trials and their associated distributions.

27. My understanding of the Poisson Process and its associated distributions.

28. My understanding of the definition of a probability distribution of a continuous random variable.

29. My understanding of the Normal Probability Distribution.

30. My understanding of the Exponential Probability Distribution.

31. My understanding of the Chi-Square Probability Distribution.

32. My understanding of the difference between cumulative probability functions and probability mass/density functions.

33. My understanding of Mixed Distributions.

34. My understanding of Mathematical Expectation and its properties.

35. My understanding of Multivariable Probability Distributions.

36. My understanding of Moment Generating Functions and their properties.

37. My understanding of independent random variables.

38. My overall understanding of random variables.

39. My understanding of the Central Limit Theorem.

40. My understanding of Point Estimation.

41. My understanding of Confidence Intervals for Means.
42. My understanding of Confidence Intervals for Proportions.

43. My understanding of calculating 1-Sample Confidence Intervals.

44. My understanding of calculating 2-Sample Confidence Intervals.

45. My understanding of interpreting 1-Sample Confidence Intervals.

46. My understanding of interpreting 2-Sample Confidence Intervals.

47. My understanding of determining which type of confidence interval is needed.

48. My understanding of how to determine the sample size for a given confidence level and a given margin of error.

49. My understanding of 1-Sample Hypothesis Testing for Means.

50. My understanding of 1-Sample Hypothesis Testing for Proportions.

51. My understanding of 2-Sample Hypothesis Testing for Means.

52. My understanding of 2-Sample Hypothesis Testing for Proportions.

53. My understanding of Paired Difference Hypothesis Testing.

54. My understanding of Rejection Regions.

55. My understanding of Test Statistics.

56. My understanding of Type I error.

57. My understanding of alpha.

58. My understanding of Statistical Significance (p-values).

59. My understanding of Correlation.

60. My understanding of Regression.

61. My understanding of non-parametric or distribution-free methods.
62. My understanding of Multiple Regression.

63. My understanding of correlation versus causation.

64. My overall understanding of probability.

65. My overall understanding of statistics.

Survey Open-Ended Questions

1. Which of the class activities contributed positively to your learning of the material? Please refer to the class website to recall the activities used in class.

2. Did the class lectures help you learn the material?

3. Did the technology used in the class help you learn the material (Minitab, computer simulation activities)? Did it present you with any additional difficulties that you had to resolve? How did you resolve any difficulties?

4. Did the class stimulate your problem-solving skills? How or how not?

5. Did the methods used in presenting the activities and/or class demonstrations stimulate your interest in the material? Why or why not?

6. Did the methods used in presenting the activities and/or class demonstrations encourage your participation? How?

7. What do you think is the greatest strength of the methods used in this class?

8. What do you think is the greatest weakness of the teaching methods used in this class?

9. Do you have any suggestions for the improvement of the teaching methods used in this class?

Interview Questions for Math 300

In this E-mail Interview, I will ask some questions pertaining to the teaching strategies, materials, technology, and assignments used in this class. This is being done as part of the evaluation process for the National Science Foundation project in probability and statistics that Dr. Lunsford is co-directing. I am serving as the project evaluator and therefore I collect information pertaining to this course. Your responses will not be shared with Dr. Lunsford until after final grades have been posted. After grades have been posted, I will share a summary of your responses with Dr. Lunsford, however, individual students will not be identified. As Dr. Lunsford has shared with you, participating in this interview will result in bonus points, which will be added to your quiz scores. To aid you in answering the questions, I have included a table to refresh your memory concerning the topics covered in the class this term.

Thank you for providing us with your opinions. E-mail me if you have any questions.

Dr. Tracy Goodson-Espy

Table D-1: Major Mathematical Topics Covered

| Topic                                                      |
|------------------------------------------------------------|
| Sample Spaces, Outcomes, and Events                        |
| Relative Frequency Histograms                              |
| Basic Descriptive Statistics (Mean, Sample Mean, Variance, Sample Variance, etc.) |
| The Probability Function and its Basic Properties          |
| Methods of Enumeration (Combinatorics)                     |
| Conditional Probability and Independent Events             |
| Bayes’ Theorem                                             |
| Basic Concepts for Discrete and Continuous Random Variables |
| Mathematical Expectation                                   |
| Moment-Generating Functions                                |
Bernoulli Trials and the Binomial Distribution

The Poisson Distribution

The Uniform Distribution (Discrete and Continuous)

The Exponential Distribution

The Normal Distribution

Distributions of Two Random Variables

Independent Random Variables

Distributions of Sums of Independent Random Variables

The Central Limit Theorem

Confidence Intervals for Means

Confidence Intervals for Proportions

Sample Size

1. How would you compare this class to other mathematics classes that you have taken with regard to how much you have learned?

2. Did the lectures in this class differ from those in other mathematics courses?

3. Which mathematics topics covered in this course were easiest for you?

4. Which mathematics topics covered in this course did you find to be the most difficult?

5. What is your opinion of the text and other written materials used in this class?

6. Did you visit and use the Virtual Laboratories in Probability and Statistics website described in the course syllabus? How often? Was it helpful to you and can you describe how you used it?

7. Which aspects of this website did you find to be most helpful?

8. Were there parts of the website that you did not find helpful?

9. Did you visit any of the other websites on the Cool Probability and Statistics links included on Dr. Lunsford’s website?
   - Statistical Inference on the TI-82/83/86/89/92+
     - Cool Java applets
   - The Gallup Organization Website
   - The Pew Research Center for the People and the Press
   - Society of Actuaries
     - Were any of these websites helpful to you in learning probability and statistics?

10. Did you visit and use any of the sample quizzes and tests included on Dr. Lunsford’s website for this class? If so, which ones did you use?

11. If you did not visit the websites or use the sample quizzes and tests, why didn’t you?

12. Did you use Minitab to help with your homework? If yes, how often?

13. Did you use Excel to help with your homework? If yes, how often?

14. Did you use a calculator for this course? If so, do you think it helped you learn the material?

15. Do you think the course would be improved by having assignments/projects that require the use of Minitab and/or Excel? Why or why not?

16. What did you like about this class? What occurred that helped you to learn?

17. What did you not like about this class?

18. What suggestions would you make about how the course is taught in the future?

Acknowledgements
In developing the foundations of this classroom-based research, we appreciate the direct support of the National Science Foundation Division of Undergraduate Education (NSF DUE) matching grants (DUE-126401, 0126600, 0126716, and 0350724) and our institutions. The views and opinions expressed here are those of the authors and do not necessarily reflect the views and opinions of the National Science Foundation. We also sincerely appreciate the willingness of Drs. Allan Rossman, Beth Chance, California Polytechnic University, and Dr. Kyle Siegrist, University of Alabama in Huntsville, to share the educational materials they have developed through NSF-supported projects (DUE-9950476 and 9652870). Drs. Rossman and Chance were especially generous with their time and expertise in training us in appropriate pedagogy for using their educational materials. Furthermore, we are grateful for the willingness of Drs. Robert delMas and Joan Garfield, University of Minnesota, and Dr. Beth Chance, California Polytechnic University, to share their assessment instrument, learning activities, and research developed in part with support of NSF DUE-9752523 grant, “Tools for Teaching and Assessing Statistical Inference,” in which the software Sampling SIM was developed and its usefulness was tested. Some of the results for the Math 300 class found in this paper were also presented by the authors at the 2005 Joint Statistical Meetings at the presentation, “Applying an Action Research Model to Assess Student Understanding of the Central Limit Theorem in Post-Calculus Probability and Statistics Courses,” and are published in the corresponding proceedings. Finally, we would like to thank the three anonymous referees for their many valuable comments and suggestions which greatly improved the presentation of our results.

References

Ball, D. L. (2000), “Working on the Inside: Using One’s Own Practice as a Site for Studying Teaching and Learning,” in Handbook of Research Design in Mathematics and Science Education, eds. Kelly, A. E., and Lesh, R. A., pp. 365 - 401, Lawrence Erlbaum Publishers, Mahwah, NJ.
delMas, R. (2002), Sampling SIM, version 5.4 [On line]. www.gen.umn.edu/research/stat_tools/
delMas, R., Garfield, J., and Chance, B. (1999a), “A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students’ Statistical Reasoning,” Journal of Statistics Education [Online], 7(3). ww2.amstat.org/publications/jse/secure/v7n3/delmas.cfm
delMas, R., Garfield, J., and Chance, B. (1999b), “Exploring the Role of Computer Simulations in Developing Understanding of Sampling Distributions,” paper presented at the 1999 American Educational Research Association Annual Meeting, [Online]. www.tc.umn.edu/~delma001/stat_tools/
delMas, R., Garfield, J., and Chance, B. (1999c), “The Role of Assessment in Research on Teaching and Learning Statistics,” paper presented at the 1999 American Educational Research Association Annual Meeting, [Online]. www.tc.umn.edu/~delma001/stat_tools/
delMas, R., Garfield, J., and Chance, B. (2002), “Assessment as a Means of Instruction,” paper presented at the 2002 Joint Mathematics Meetings, [Online]. www.tc.umn.edu/~delma001/stat_tools/
Feldman, A. and Minstrell, J. (2000), “Action Research as a Research Methodology for the Study and Teaching of Science,” in Handbook of Research Design in Mathematics and Science Education, eds. Kelly, A. E., and Lesh, R. A., pp. 429 - 456, Lawrence Erlbaum Publishers, Mahwah, NJ.
Garfield, J. (2001), Evaluating the Impact of Educational Reform in Statistics: A Survey of Introductory Statistics Courses. Final report for National Science Foundation grant REC-9732404. Minneapolis, MN.
Garfield, J., delMas, R., and Chance, B. (2000), Tools for Teaching and Assessing Statistical Inference, [Online]. www.gen.umn.edu/research/stat_tools/
Garfield, J., delMas, R., and Chance, B. (2002), ARTIST: Assessment Resource Tools for Improving Statistical Thinking, [Online]. www.gen.umn.edu/artist/
Hogg, R.V., and Tanis, E.A. (2001), Probability and Statistical Inference, 6th Edition, New Jersey: Prentice Hall.
Lee, C., and Meletiou-Mavrotheris, M., (2003), “Some Difficulties of Learning Histograms in Introductory Statistics,” in 2003 Proceedings of the American Statistical Association, Statistics Education Section [CD-ROM], pp. 2326 - 2333. Alexandria, VA: American Statistical Association.
Lunsford, M. L., Goodson-Espy, T. J., and Rowell, G. H. (2002), Collaborative Research: Adaptation and Implementation of Activity and Web-Based Materials into Post-Calculus Introductory Probability and Statistics Courses, funded by the Course, Curriculum, and Laboratory Improvement program of the National Science Foundation, awards DUE-0126600; DUE-0126716, DUE-0350724, DUE-126401, [Online]. www.mathspace.com/NSF_ProbStat/
Meletiou-Mavrotheris, M. and Lee, C. (2002), “Teaching Students the Stochastic Nature of Statistical Concepts in an Introductory Statistics Course,” Statistics Education Research Journal, 1(2), 22-37.
Mills, J. D. (2002), “Using Computer Simulation Methods to Teach Statistics: A Review of the Literature,” Journal of Statistics Education [Online] 10(1).
Noll, S., and Stevenson, R. (eds.) (1995), Educational Action Research, New York: Teachers College Press.
Nolan, D. and Speed, T. (2000), Stat Labs: Mathematical Statistics Through Application, New York: Springer-Verlag.
Parsons, R. and Brown, K. (2002), California: Wadsworth/Thomson Learning.
Pfannkuch, M. and Brown, C. (1996), “Building On and Challenging Students’ Intuitions about Probability: Can We Improve Undergraduate Learning?” Journal of Statistics Education [Online], 4(1). ww2.amstat.org/publications/jse/v4n1/pfannkuch.html
Rossman, A. and Chance, B. (2002), “A Data-Oriented, Active Learning, Post-Calculus Introduction To Statistical Concepts, Methods, And Theory,” paper presented at International Congress On Teaching Statistics 2002, retrieved July 9, 2005, at www.rossmanchance.com//iscat/ICOTSpaper02.htm
Rossman, A. and Chance, B. (2006), Investigating Statistical Concepts, Applications, and Methods, New York: Thomson Brooks/Cole.
Rossman, A., Chance, B., and Ballman, K. (1999), *A Data-Oriented, Active Learning, Post-Calculus Introduction to Statistical Concepts, Methods, and Theory (SCMT)*, funded by the Course, Curriculum, and Laboratory Improvement program of the National Science Foundation, award #DUE-9950476, [Online]. www.rossmanchance.com/iscat/

Rowell, G., Lunsford L., and Goodson-Espy, T. (2003), “An Application of the Action Research Model,” in 2003 Proceedings of the American Statistical Association, Statistics Education Section [CD-ROM], 3568-3571, Alexandria, VA: American Statistical Association.

Schau, C. (1999), Survey of Attitudes Towards Statistics, [Online]. www.unm.edu/~cschau/viewsats.htm

Schau, C. (2003, August), Students’ Attitudes: The "Other" Important Outcome in Statistics Education, Joint Statistical Meetings, San Francisco, CA.

Siegrist, K. (1997), Virtual Laboratories in Probability and Statistics, funded by the Course, Curriculum, and Laboratory Improvement program of the National Science Foundation, award DUE-9652870, [Online]. www.math.uah.edu/stat

Terrell, G. (1999), Mathematical Statistics: A Unified Introduction, New York: Springer-Verlag.

Tobin, K. (2000), “Interpretive Research in Science Education,” in Handbook of Research Design in Mathematics and Science Education, eds. Kelly, A. E., and Lesh, R. A., pp. 487-512, Lawrence Erlbaum Publishers, Mahwah, NJ.

Utts, J., Sommer, B., Acredolo, C., Maher, M., and Matthews, H. (2003), “A Study Comparing Traditional and Hybrid Internet-Based Instruction in Introductory Statistics Classes,” Journal of Statistics Education [Online] 11(3). www2.amstat.org/publications/jse/v11n3/utts.html

M. Leigh Lunsford  
Department of Mathematics and Computer Science  
Longwood University  
Farmville, VA 23909  
U.S.A.  
lunsford@longwood.edu

Ginger Holmes Rowell  
Department of Mathematical Sciences  
Middle Tennessee State University  
Murfreesboro, TN 37132  
U.S.A.  
rowell@mtsu.edu

Tracy Goodson-Espy  
Department of Curriculum and Instruction  
Appalachian State University  
Boone, NC 28608  
U.S.A.  
goodsonespy@appstate.edu