Applying an Action Research Model to Assess Student Understanding of the Central Limit Theorem in Post-Calculus Probability and Statistics Courses

M. Leigh Lunsford[1], Ginger Holmes Rowell[2], Tracy Goodson-Espy[3]
Department of Mathematics and Computer Science, Longwood University[1]
Department of Mathematical Sciences, Middle Tennessee State University[2]
Department of Curriculum and Instruction, Appalachian State University[3]

Key Words: Assessment, Action Research, Classroom Research, Sampling Distributions, Central Limit Theorem

1. Introduction

Heeding the call of previous researchers (delMas, Garfield, and Chance 1999a), we used a classroom 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 the classroom research model to the first course of a two-semester post-calculus mathematical probability and statistics sequence taught at a small engineering and science oriented, Ph.D. granting university in the southeastern United States with an approximate undergraduate enrollment of 5000. Throughout this paper, when we refer to "sampling distribution(s)," we are only considering sampling distributions of the sample mean.

The focus of this paper is our examination of students' graphical understanding of the CLT and related concepts as presented at the Joint Statistics Meeting in August, 2005. For a more detailed analysis that includes students' numerical understanding, results from the second course of the two-semester post-calculus sequence, and comparison to results from previous studies, please see our paper and website (Lunsford, Rowell, Goodson-Espy 2005).

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

2. Phase 1: What is the problem? I.e. What is not working in the classroom?

While introductory statistics courses have been the focus of reform curricula and pedagogy (Rossman, 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 with traditional methods (Rossman and Chance 2002). Since most of our students (sciences, education, engineering and computer science majors) will only take the first course of the sequence, there is a need for the injection of more statistical concepts here. 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, our students may not develop a deep understanding of these important concepts. Thus we wanted to enhance, assess, and improve our teaching of sampling distributions and the CLT in that course.

3. Phase 2: Techniques to Address the Problem

In the Spring of 2004, as part of an NSF adaptation and implementation (A&I) grant (Lunsford, Goodson-Espy, and Rowell 2002), we attempted to infuse reform-based pedagogies into the first semester of our mathematical probability and statistics by incorporating activity and web-based materials (Rossman and Chance 1999; Siegrist 1997). We will refer to this course as Math 300 (Introduction to Probability). 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.

To teach sampling distributions and the CLT in the course we used a traditional text (Hogg and Tanis 2001) along with a computer simulation called *Sampling SIM* (delMas 2001), and an activity for use with the simulation. The activity, *Sampling Distributions and Introduction to the Central Limit Theorem*, 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 our paper and website (Lunsford, et. al. 2005) for a copy of the activity. After an initial in-class demonstration of *Sampling SIM*, the activity was assigned as an out-of-class group project for which the students turned in a written follow-up report.

4. Phase 3: Evidence to Collect to Determine if Implementation Effective.

Building on the work of previous researchers, we used a quantitative assessment tool graciously provided to us by Robert delMas. Please see our paper and website (Lunsford, et. al. 2005) for a copy of this assessment tool. We used the assessment tool as both a pretest and a posttest for both courses. The pretest was administered before covering sampling distributions and the central limit theorem. We did not return the pretest to the students nor did we give the students feedback regarding their pretest performance. The posttest was administered on the last day of class as an in-class quiz (each student had turned-in a report from the activity during the previous class period). A qualitative assessment tool developed by the authors was also given to the students 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.

4.1 Graphical Understanding

To assess our students' graphical understanding of sampling distributions and the CLT, we examined Question 5 of the quantitative assessment tool. This was an assessment item that was more graphical in nature and most directly related to the *Sampling SIM* program and the activity used. In this question, 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, to the right, 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) whereas Question 5e asked students to select which graph represents a distribution of sample means for 500 samples of size $n = 25$ (Answer=E).

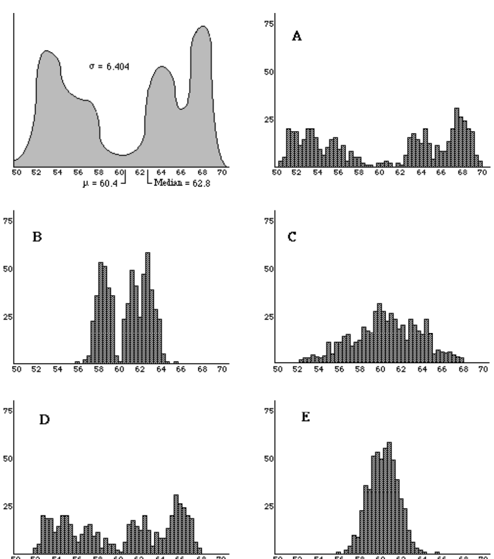
4.1.1 Correct Choice of Sampling Distribution

We first examined how well our students were able to choose the correct sampling distribution in Question 5. The percent of students that correctly identified the sampling distribution for large sample sizes ($n = 25$) on the pretest was 16.7%. The percent correct increased to 77.8% on the posttest. With only 33% correct on the posttest, 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.

4.1.2 Reasoning Pair Classifications

Following the work of delMas, et. al. (2000, 1999 a, b), we next examined reasoning pair classifications for our Math 300 students. The idea is to classify student reasoning about the shape and variability of the sampling distribution as the sample sizes increases from small to large. For a complete description of the reasoning pair classifications, please see our paper and website (Lunsford, et. al. 2005). In the first column of Table 1, on the next page, we show our Math 300 students posttest reasoning pairs for Question 5. Please see Figure 1 below for a quick reference to the graphs for this problem.

Figure 1. Population Distribution (Top Left) and Possible Sampling Distributions (A-E) for the Irregular Population. (Used with permission from Garfield, delMas and Chance 2000)



In the first row of Table 1, we see that 5 students (27.8%) gave the answer that graph C was the sampling distribution for a sample size of $n = 4$ and graph E was the sampling distribution for $n = 25$ (i.e. reasoning pair (C, E)). These students have the *correct* classification below the reasoning pair. In row two we have 2 students who answered B and E for $n = 4$ and $n = 25$, respectively, which we consider a *good* classification (here we differ with delMas, et. al. (2000) 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 because while both sets of students had the correct graph for $n = 25$, the students in the second row did choose a distribution with less variability than the population for $n = 4$ but the students in the third row did not. However, 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.

students (12 out of 18) did not choose the correct sampling distribution for $n = 4$.

4.1.3 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 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. A student is defined to demonstrate *consistent graphical reasoning* if the sampling distribution chosen is consistent with their stated expected variance *and* shape of the sampling distribution as compared to the population (even if

Table 1. Math 300 Posttest Comparison of Irregular Distribution Reasoning Pairs and *Consistent Graphical Reasoning* for $n = 4$

Posttest Reasoning Pair Irregular Population Distribution (Question 5) ($n = 4$; $n = 25$)	**Number (Percent) of Students ($N = 18$)**	**$n = 4$ Sampling Distribution Shaped More Like: (Question 5(b))**		**Variability of $n = 4$ Sampling Distribution compared to Population: (Question 5(c))**		***Consistent Graphical Reasoning***
		Answer	**Number of Students**	**Answer**	**Number of Students**	
(C, E) *Correct*	5 (27.8%)	Normal	5	Less	5	5
(B, E) *Good (Large-to-Small Normal)*	2 (11.1%)	Pop.	2	Less	2	2
(A, E) *Large-to-Small Normal*	7 (38.9%)	Pop.	7	Same	4	4
				Less*	3	0*
(A, B) *Large-to-Small Population*	1 (5.6%)	Normal*	1	More*	1	0*
(E, D) or (E, C) *Small-to-Large*	2 (11.1%)	Normal	2	Less	2	2
(C, D) *Other*	1 (5.6%)	Normal	1	Less	1	1
Totals	**18**	**Normal** **Pop.**	**9** **9**	**Less** **Same** **More**	**13** **4** **1**	**14**

There are several interesting items to note from Table 1. First, of the 6 students (33.3%) who were able to choose the correct sampling distribution for $n = 4$ most 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 top three reasoning categories. We wanted to see if the data could give us any insight as to why most of our their choice of sampling distribution is incorrect). We call this measure *consistent* because if the sampling distribution they chose was not the same (in terms of shape *and* variability) as what they say they expected, there was some inconsistency in their answer.

We saw significant improvement in our students from pretest to posttest in their demonstration of *consistent graphical reasoning*. Of our 18 students, the number who demonstrated *consistent graphical reasoning* increased from 3 to 14 (16.7 to 77.8%) for the small sample size ($n = 4$) and from 2 to 15 (11.1 to 83%) for the large sample size ($n = 25$). 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. We remark that all of the students who were *correct* were also *consistent* for $n = 4$. Please see our paper and website (Lunsford, et. al. 2005) for additional details on how we computed the number of students that demonstrated *consistent graphical reasoning*.

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 surprised at the low percent of our students displaying *consistent graphical reasoning* on the pretest. Because the pretest for this class was administered late in the semester (around the tenth week of class) we expected these students to graphically understand shape and spread and hence be consistent, even if not correct, with their choice of sampling distribution. However, upon contemplation 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.

Second, from these 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.

4.2 Comparison of Reasoning Pairs and *Consistent Graphical Reasoning*

In Table 1 above we also represent a deeper look into our data to try to determine how our students may be misunderstanding concepts about sampling distributions and the CLT. Recall the reasoning pairs in the first column 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 $n = 4$. The next two columns indicate the students' answers for the expected shape and variability of the sampling distribution when $n = 4$ compared to the irregular population from which the samples had been drawn. The last column shows the students who have been classified as showing *consistent graphical reasoning*. An asterisk has been placed by the students in that column who did not show *consistent graphical reasoning* and along the corresponding row on where their reasoning failed. Note that 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 1 we make some observations and conjectures about our students' understanding of sampling distributions and the CLT. First we observe 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 for $n = 25$. We suspect that many of our students were not recognizing how quickly the sampling distribution becomes unimodal as n increases. This is 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 are 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 think 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 σ^2/n, via mathematical expectation). So, as they did regarding the shape, these students may be thinking the variability result does not "kick in" until the sample size is greater than 30. We found it interesting that 15 (83%) of our students stated it was "true" that if the population standard deviation equals σ then the standard deviation of the sample means in a sampling distribution (using samples of size n) from that population is equal to $\sigma/\sqrt{n}$. All of the students in the top four rows of the Table 1 answered "true" this question, except one who did not answer the question. Thus we believe that our students where 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 Qualitative Results

Our students generally showed a positive response to the use of activities and simulations and believed that they contributed to their learning. In addition to observing this in the classroom, we also saw it in our 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. On 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). For more details of our qualitative results, please see our paper and website (Lunsford, et. al. 2005).

5. Phase 4: Conclusions, Conjectures, and What Should be Done Next

While in general our post-calculus probability and statistics students are more sophisticated mathematically than our algebra-based introductory-level statistics students, we should not necessarily expect them to have good graphical interpretation, comparison, and reasoning skills concerning sampling distributions even if they understand the basic theory and are 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. This could be thought of as common sense: *If you want your students to understand certain concepts, then these concepts need to be part of what you emphasize in their assignments.*

We believe the classroom research model enabled us to gain insight into our students' understanding of the CLT and related concepts. 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. For more conclusions and conjectures based on our research and information on resources for activities and assessment tools, please see our paper and website (Lunsford, et. al. 2005).

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, and the willingness of Drs. Allan Rossman, Beth Chance, California Polytechnic University, and Kyle Siegrist, University of Alabama in Huntsville, to share 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 Beth Chance, 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. For more results and details of the results presented in this paper, please see our paper and website (Lunsford, et. al. 2005).

References

delMas, R. (2002). *Sampling SIM*, version 5.4. http://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*, v.7, n.3, http://www.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," presented at the 1999 AERA Annual Meeting, http://www.gen.umn.edu /research/stat_tools/area1999/comparison.htm

delMas, R., Garfield, J., and Chance, B. (1999c), "The Role of Assessment in Research on Teaching and Learning Statistics," presented at the 1999 AERA Annual Meeting, online at http://www.gen.umn.edu/research/stat_tools/aera1999/assessment.htm

delMas, R., Garfield, J., and Chance, B. (2002), "Assessment as a Means of Instruction," presented at the 2002 Joint Mathematics Meetings, http://www.gen.umn.edu/research/stat_tools/jmm_2002/assess_instruct.html

Garfield, J., delMas, R., and Chance, B. (2000), *Tools for Teaching and Assessing Statistical Inference*, partially funded by the NSF, hosted at: http://www.gen.umn.edu/research/stat_tools/

Garfield, J. (2001). *Evaluating the Impact of Educational Reform in Statistics: A Survey of Introductory Statistics Courses*. Final report for NSF grant REC-9732404. Minneapolis, MN.

Hogg, R.V., and Tanis, E.A. (2001), *Probability and Statistical Inference, 6th Ed.*, Prentice Hall.

Lunsford, M. L., Goodson-Espy, T. J., 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 NSF Course, Curriculum, and Laboratory Improvement awards DUE #s-0126600, 0126716, 0350724, 126401, Summer 2002, http://www.mathspace.com/NSF_ProbStat/.

Lunsford, M. L., Rowell, G. H., Goodson-Espy, T. J. (2005), "Classroom Research: Assessment of Student Understanding of Sampling Distributions of Means and the Central Limit Theorem in Post-Calculus Probability and Statistics Classes," in submission and online at http://www.mathspace.com/NSF_ProbStat/.

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 NSF Course, Curriculum, and Laboratory Improvement award #DUE-9950476, June 1, 1999, http://www.rossmanchance.com/iscat/ .

Rossman, A. and Chance, B. (2002). "A Data-Oriented, Active Learning, Post-Calculus Introduction to Statistical Concepts, Methods, and Theory," ICOTS 2002, retrieved July 2005, http://www.rossmanchance.com/iscat/ICOTSpaper02.htm.

Rossman, A. and Chance, B. (2006), *Investigating Statistical Concepts, Applications, and Methods*. US: Thomson Brooks/Cole.

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

Siegrist, K. (1997), *Virtual Laboratories in Probability and Statistics*, funded by the NSF Course, Curriculum, and Laboratory Improvement award DUE-9652870, Jan. 1997 – Dec. 1999, http://www.math.uah.edu/stat/ .